Paper Code: 20MAT21C5 **Semester-I**

MATHEMATICAL STATISTICS

M.Sc. Mathematics

DIRECTORATE OF DISTANCE EDUCATION

MAHARSHI DAYANAND UNIVERSITY, ROHTAK

(A State University established under Haryana Act No. XXV of 1975)

NAAC 'A+' Grade Accredited University

Author

Dr. Poonam Redhu
Assistant Professor, Department of Mathematics
Maharshi Dayanand University, Rohtak

Maharshi Dayanand University
ROHTAK – 124 001

Paper Code: 20MAT21C5

Mathematical Statistics

M. Marks = 100
Term End Examination = 80
Assignment = 20
Time = 3 Hours

Course Outcomes

Students would be able to:

CO1 Understand the mathematical basis of probability and its applications in various fields

CO2 Use and apply the concepts of probability mass/density functions for the problems involving single/bivariate random variables.

CO3 Have competence in practically applying the discrete and continuous probability distributions along with their properties.

CO4 Decide as to which test of significance is to be applied for any given large sample problem

Section - I

Probability: Definition and various approaches of probability, Addition theorem, Boole inequality, Conditional probability and multiplication theorem, Independent events, Mutual and pairwise independence of events, Bayes theorem and its applications.

Section - II

Random variable and probability functions: Definition and properties of random variables, Discrete and continuous random variables, Probability mass and density functions, Distribution function. Concepts of bivariate random variable: joint, marginal and conditional distributions. Mathematical expectation: Definition and its properties. Variance, Covariance, Moment generating function- Definitions and their properties.

Section - III

Discrete distributions: Uniform, Bernoulli, Binomial, Poisson and Geometric distributions with their properties. Continuous distributions: Uniform, Exponential and Normal distributions with their properties.

Section - IV

Testing of hypothesis: Parameter and statistic, Sampling distribution and standard error of estimate, Null and alternative hypotheses, Simple and composite hypotheses, Critical region, Level of significance, One tailed and two tailed tests, Two types of errors. Tests of significance: Large sample tests for single mean, Single proportion, Difference between two means and two proportions.

Note: The question paper of each course will consist of five Sections. Each of the sections I to IV will contain two questions and the students shall be asked to attempt one question from each. Section-V shall be compulsory and will contain eight short answer type questions without any internal choice covering the entire syllabus.

Books Recommended:

1.V. Hogg and T. Craig, Introduction to Mathematical Statistics, 7th addition, Pearson Education Limited-2014

2. A.M. Mood, F.A. Graybill, and D.C. Boes, Introduction to the Theory of Statistics, Mc Graw Hill Book Company.

3. J.E. Freund, Mathematical Statistics, Prentice Hall of India.

4. M. Speigel, Probability and Statistics, Schaum Outline Series.

5. S.C. Gupta and V.K. Kapoor, Fundamentals of Mathematical Statistics, S. Chand Pub., New Delhi

Contents

CHAPTER-1

THEORY OF PROBABILITY AND BAYE'S THEOREM

Structure

1.1 Introduction

In real life, there is no certainty about many problems that what will happen in the future but decisions still have to be taken. Therefore, decision processes must be able to deal with the problems of uncertainty. Uncertainty creates risk and this risk must be analyzed. Both qualitative and quantitative techniques exist for modeling uncertainty. In many situations, large amounts of numerical data are available which requires statistical techniques for analysis. If an experiment is repeated under essentially homogeneous and similar conditions, we generally come across two types of situations.

(1) The result or what is usually known as the 'outcome' is unique or certain.

(2) The result is not unique but may be one of the several possible outcomes.

The phenomena covered by (1) are known as 'deterministic' or 'predictable' phenomena in which the result is predicted with certainty.

For Example:

(a) The velocity 'v' of a particle after time't' is given by v=u+at where 'u' is initial velocity and 'a' is the acceleration. The equation uniquely determines v if the right-hand side quantities are known.

(b) Ohm's Law, i.e. C=E/R , where 'C' is the flow of current, 'E' the potential difference between the two ends of the conductor and 'R' the resistance, uniquely determines the value C as soon as E and R are given.

Similarly,

The phenomena covered by (2) are known as 'unpredictable' or 'probabilistic' phenomena in which the result is predicated with uncertainty.

For Example:

(c) In a random toss of a uniform coin, we are not sure of getting the head or tail.

(d) A manufacturer cannot ascertain the future demand of his product with certainty.

Remark: In **'**Theory of probability' there are three possible states of expectation-'certainty', 'impossible', and 'uncertainty'. The probability theory describes certainty by 1, impossibility by 0 and the various grades of uncertainties by coefficients ranging between 0 and 1.

From above discussion, it is clear that we need to cope with the uncertainty which leads to the study and use of probability theory. Galileo (1564-1642), an Italian Mathematician, was the first to attempt at a qualitative measure of probability when he was dealing with the gambling. But, the first foundation of mathematical theory of probability proposed by two French mathematicians B. Pascal (1623-62) and P. Fermat (1601-65) while solving the problems proposed by French gambler. Afterward, important contributions were made by various researchers including Huyghens (1629-95), Jacob Bernoulli (1654-1705), Laplace (1749-1827), Abraham De-Moivre (1667-1754), Markov (1856-1922), and Thomas Bayes.

1.2 Basic Concepts

1. Random Experiment

An experiment is said to be a random experiment if it is conducted repeatedly under essentially homogeneous conditions, the result is not unique but may be anyone of the various possible outcomes**.** In other words, an experiment whose outcomes cannot be predicted in advance is called a random experiment. For instance, if a fair coin is tossed three times, it is possible to enumerate all the possible eight sequences of head (H) and tail (T). But it is not possible to predict which sequence will occur at any time.

2. Sample Space

The set of all possible outcomes of a random experiment is known as the sample space and is denoted by S. Each conceivable outcome of a random experiment under consideration is called a sample point. The totality of all conceivable sample points is called a sample space. For example: Sample space of a trial conducted by tossing of two coins is {HH, HT, TH, TT}. In the above experiment, it is simple to note that anyone sequence of H and/or T is a sample point whereas all the possible four sample points constitute the sample space.

3. Trial and Events

Any particular performance of a random experiment is called a trial and outcome or combinations of outcomes are termed as Events or Cases. Any subset of the sample space is an event. In other words, the set of sample points that satisfy certain requirement(s) is called an event. For example, if a coin is tossed repeatedly, the result is not unique. The tossing of coin is a random experiment and getting a head or tail is an event.

For example: In an experiment which consists of the throw of a six-faced dice and observing the number of points that appear, the possible outcomes are 1,2,3,4,5,6.

In the same experiment, the possible events could also be stated as'odd number of points', 'Even no. of points', 'Getting a point greater than 4'and so on.

4. Exhaustive Events

It is defined as total number of all possible outcomes of any trial. In other words, if all the possible outcomes of an experiment are taken into consideration, then such events are called exhaustive events e.g. when a coin is tossed three times there are eight exhaustive events, when two dice are thrown then exhaustive events are 36 and drawing two cards from a pack of cards, the exhaustive number of cases is $^{52}C_2$.

5. Favourable Events

The number of cases favourable to an event in a trial is the number of outcomes which entail the happening of the event.

For Examples:
(i)In throwing two dice, the number of cases favourable to getting the sum 5 is (1,4),(4,1),(2,3), (3,2).

(ii)In drawing a card from a pack of cards the number of cases favourable to drawing of an ace is 4, for drawing a spade is 13, and for drawing a red card is 26.

6. Mutually Exclusive Events

Two or more events are said to be mutually exclusive if the happening of one of them prevents or precludes the happening of all others in the same experiment. Two events E_1 and E_2 are said to be mutually exclusive when they cannot happen simultaneously in a single trial. In other words, if there is no sample point in E_1 which is common to the sample point in E_2 i.e., $E_1 \cap E_2 = \phi$, the events E_1 and E_2 are said to be mutually exclusive. In throwing a die, the events 1, 2, 3, 4, 5 and 6 are mutually exclusive because all the six events cannot happen simultaneously in a single trial. If it shows 3, then the event of getting 3 precludes the event of getting 1, 2, 4, 5, and 6 at the same time.

7. Equally Likely Events

The events are said to be equally likely if the chance of happening of each event is equal or same. In other words, cases are said to be equally likely when one does not occur more often than the others e.g.

if a die is rolled, any face is as likely to come up as any other face. Hence, the six outcomes 1, 2, 3, 4, 5 or 6 appearing up are equally likely.

8. Independent Events

Events are said to be independent of each other if happening of any one of them is not affected by and does not affect the happening of any one of others. In other words, two or more events are said to be independent if the happening (or non-happening) of anyone does not depend on the happening or non-happening of any other, otherwise, they are said to be dependent. For example

a)In tossing an unbiased coin, the event of getting a head in the first toss is independent of getting a head in the second, third and subsequent throws.

b) If we draw a card from a pack of well-shuffled cards and replace it before drawing the second draw, then the second draw is independent of the first draw, but if the first card is not replaced then the second draw is dependent on the first draw.

9. Complementary Events

The complement of an event A, means non-occurrence of an event A and is denoted by $\bar{A}$ or A^c. $A^c/\bar{A}$ contains those points of the sample space which do not belong to A. In tossing a coin, occurrence of Head (H) and Tail (T) are complementary events. In tossing of a die, occurrence of an even number (2, 4, 6) and odd number (1, 3, 5) are complementary events.

10. Simple (Elementary) Events

An event contains only a single sample point is called an elementary event or simple event e.g. in tossing a die, getting a number 5 is called a simple event.

11. Compound Events

When two or more events occur in connection with each other, their simultaneous occurrence is called a compound event. The joint occurrence of two or more events is called a compound event. An event is termed compound if it represents two or more simple events e.g. if a bag contains 4 white and 3 black balls. If we are required to find a chance in which 3 balls drawn are all white is a simple event. However, if we are required to find out the chance of drawing 3 white and then 2 black balls, we are dealing with a compound event because it is made up of two events.

1.3 Mathematical Probability

Definition of Probability: The chance of happening of an event when expressed quantitatively is called probability. The probability is defined in the following three different ways:

1. Classical, Mathematical or a Priori.
2. Empirical, Relative or Statistical
3. Axiomatic

1.3.1 Mathematical (or Classical or Prior) Probability

This is the oldest and simplest definition of probability. This definition is based on the assumption that the outcomes or results of an experiment are equally likely and mutually exclusive. According to James

Bernoulli who was the first man to obtain a quantitative measure of uncertainty. If a random experiment results in *N* exhaustive, mutually exclusive and equally likely cases out of which m are favourable to the happening of an event *A*, then probability of occurrence of *A*, usually denoted by *P(A)* is given by

$$P(A) = \frac{\text{number of favourable cases}}{\text{number of Exhaustive cases}} = \frac{m}{N}$$

Example 1: Two identical symmetric dice are thrown. Find the probability of obtaining a total score of 8.

Solution: The total number of possible outcomes is $6\times 6=36$. There are 5 sample points (2, 6), (3, 5) (4, 4), (5, 3), (6, 2), which are favourable to the event A of getting a total score of 8. Hence, the required probability is 5/36.

Properties:

1. The number of cases favourable to the complimentary event $\overline{A}$, i.e. non-happening of event *A* are *(N-m*) and by definition of probability of non-occurrence of *A* is given by:

$$P(\overline{A}) = \frac{\text{favourable number of cases } \overline{\text{A}}}{\text{Exhaustive number of cases}} = \frac{N-m}{N} = 1 - \frac{m}{N} = 1 - P(A)$$

$$P(A) + P(\overline{A}) = 1$$

2. Since *m* and *N* are non-negative integers, $P(A)\geq 0$. Further, since the favourable number of cases to *A* are always less than total number of cases *N*, i.e. $m \leq N$, we have $P(A) \leq 1$. Hence, the probability of any event is a number lying between 0 and 1 i.e., $0\leq P(A) \leq 1$. If $P(A)=0$ then this event is said to be impossible event. If $P(A) = 1$, then *A* is called a certain event.

The above definition of probability is widely used, but it cannot be applied under the following situations:

i. If it is not possible to enumerate all the possible outcomes for an experiment.

ii. If the sample points (outcomes) are not mutually independent.

iii. If the total number of outcomes is infinite.

iv. If each and every outcome is not equally likely.

It is clear that the above drawbacks of a classical approach restrict its use in practical problems. Yet this is still widely used for problems concerning the tossing of coin(s), throwing of die, game of cards and selection of balls of different colours from the bag etc.

The probability by classical approach cannot be discovered in the cases where situations like an electric bulb will fuse before it is used for 100 hours, a patient will die if operated for an ailment, a student will fail in a particular examination, a rail compartment in which you are traveling will catch fire, or a fan will fall on you while sitting under fan etc., under such circumstances another definition can be used.

1.3.2 Statistical (Empirical) Probability

If an experiment is performed repeatedly under essentially homogeneous and 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, assuming that the limit is finite and unique. Let an event A occurs m times in N repetitions of a random experiment. Then the ratio m/N gives the relative frequency of the event A. When N becomes sufficiently large, it is called the probability of A.

$$P(A) = \lim_{N \to \infty} \frac{m}{N}$$

The above definition of probability involves a concept which has long term consequences. This approach was initiated by Von Mises. Moreover, N is not equal to infinity. Thus, in this case, the probability is the limit of relative frequency. Whether such a limit always exists, is not definite.

Limitations of Empirical probability:

1. If an experiment is repeated a large number of times, the experiment conditions may not remain identical and homogeneous.

2. The limit in equation (1) may not attain a unique value, however, large N may be.

The two definitions of probability are apparently different. In the prior definition, it is the relative frequency of favourable cases to the total number of cases. In the relative frequency approach, the probability is obtained objectively by repetitive empirical observations, hence it is known as empirical probability. The empirical definition provides validity to the classical theory of probability.

1.3.3 Axiomatic Approach of Probability

The modern theory of probability is based on the axiomatic approach introduced by the Russian Mathematician A.N. Kolmogorov in 1930. The axiomatic definition of probability includes both the classical and empirical definition of probability and at the same time is free from their drawbacks. It is based on certain properties or postulates, commonly known as axiom. It is defined as given a sample space of a random experiment, the probability of the occurrence of any event A is defined as asset function $P(A)$ satisfying the following axioms:

(a) $P(A)$ is defined, is real and non-negative i.e. $P(A) > 0$.

(b) The probability of entire sample space is one i.e. $P(S) = 1$.

(c) If $A_1, A_2, \ldots, A_n$ are mutually exclusive events, then the probability of the occurrence of either A_1 or A_2, ...or A_n denoted by $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1)+P(A_2)+\ldots+P(A_n)$.

The above axioms are known as axioms of positiveness, certainty and unity respectively.

Probability in this approach is defined as, let S be the sample space of a random experiment with a large number of sample points N i.e. $n(S) = N$ and the number of occurrences (sample points) favourable to the event A be denoted by $n(A)$. Then the probability of an event A is equal to

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

1.4 Calculation of Probability of an Event

The probability of an event can be calculated by the following methods:

Method I: Find the total number of exhaustive cases (N). Thereafter, obtain the number of favourable cases to the event say m. Divide the number of favourable cases by the total number of equally likely cases. This will give the probability of an event. The following example will illustrate this.

Example 2. Two dice are tossed. Find the probability that the sum of dots on the faces that turn up is 11.

Solution: When two dice are tossed total number of possible outcomes =36

The number of outcomes to get a sum of 11 are (6, 5) and (5, 6) i.e. the number of cases favourable to this event is equal to 2. Hence, probability of getting a sum of 11 when two dice are thrown = $\frac{2}{36}$.

Example 3: What is the chance that a leap year selected randomly will contain 53 Sunday?

Solution: In a leap year (which consists of 366 days), there are52 complete weeks and 2 days over.The possible combinations for their two 'over' days:

1. Sunday and Monday
2. Monday and Tuesday
3. Tuesday and Wednesday
4. Wednesday and Thursday
5. Thursday and Friday
6. Friday and Saturday
7. Saturday and Sunday.

In order that a leap year selected at random should contain 53 Sunday, one of the two 'over' days must be Sunday. Since out of the above 7 possibilities, 2, viz., (1) and (7) are favourable to this event.

Required probability = 2/7.

Example 4: Among the digits 1, 2,3,4,5 at first one is chosen and then a second selection is made among the remaining four digits, find the probability that an odd digit will be selected

1. First time
2. Second time
3. Bothtimes

Solution: Total number of cases = 5×4 = 20

1. Now there are 12 cases in which the first digit drawn is odd, viz., (1,2), (1,3), (1,4), (1,5), (3,1), (3,2), (3,4), (3,5), (5,1), (5,2), (5,3) and (5,4).

The probability that the first digit drawn is odd = 12/20 = 3/5.

2. Also there are 12 cases in which the second digit drawn isodd, viz., (2,1), (3,1),(4,1), (5,1), (1,3), (2,3), (4,3), (5,3), (1,5), (2,5), (3,5) and (4,5)

∴The probability that the second digit drawn is odd =12/20 =3/5.

3. There are six cases in which both the digits drawn are odd, viz.,(1,3) ,(1,5),(3,1),(3,5),(5,1) and (5,3).

∴The probability that both the digits drawn are odd = 6/20 = 3/10.

Example 5: From 25 tickets, marked with first 25 numerical, one is drawn at random. Find chance that

1. It is multiple of 5 or 7. and
2. It is multiple of 3 or 7.

Solution: 1.Numbers (out of the first 25 numerals) which are multiples of 5 are 5, 10, 15, 20 and 25, i.e.,5 in all and the numbers which are multiples of 7 are 7, 14, 21, i.e., 3 in all.

Hence the required number of favourable cases is 5+3 = 8.

∴Required probability = 8/25.

2. Numbers which are multiples of 3 are 3,6,9,12,15,15,18,21,

i.e., 8 in all; and the numbers, which are multiples of 7 are 7, 14, 21, i.e., 3 in all. Since the number 21 is common in both the cases, the required number of distinct favourable cases are 8+3-1 = 10.

∴ Required probability = 10/25 =2/5.

Method II: The Fundamental Principle or the Fundamental Rule of Counting:

If one operation can be performed in m different ways and another operation can be performed in n different ways, then the two operations when associated together can be performed in m x n ways.

Method III: Use of Permutation and Combination in Theory of Probability:

Permutation:

The word permutation in simple language means arrangement. A permutation denoted by P is an arrangement of a set of objects in a definite order. Precisely, the number of ways of an arrangement of n distinct objects at r places is

$$^{n}p_{r} = \frac{n!}{(n-r)!}$$

Combination:

The concept of combination is very useful in understanding the theory of probability. It is not always possible that the number of cases favourable to the happening of an event is easily determined. In such cases, the concept of combination is used. The different selections that can be made out of a given set of things taking some or all of them at a time are called combinations. The number of ways of selection of r objects from a set of n objects is denoted by ${}^{n}c_{r}$, which is

$${}^{n}c_{r} = \frac{n!}{r!(n-r)!}$$

Example 6: (a) Four cards are drawn at random from a pack of 52 cards. Find the probability that

1. They are a king, a queen, a jack and an ace.
2. Two are kings and two are queen.
3. Two are black and two are red.
4. There are two hearts and two cards of diamonds.

(b)In shuffling a pack of cards, four are accidentally dropped,find the chance that the missing cards should be one from each suit.

Solution: (a) Four cards can be drawn from a well-shuffled packof 52 cards in ${}^{52}c_4$ ways, which gives the exhaustive number of cases.

1. 1 king can be drawn out of the 4 kings in ${}^{4}c_1$ ways. Similarly, 1 queen, 1 jack and 1 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 numbers of cases are ${}^{4}c_1 \times {}^{4}c_1 \times {}^{4}c_1 \times {}^{4}c_1$

Hence the required probability is

$$\frac{{}^{4}c_1 \times {}^{4}c_1 \times {}^{4}c_1 \times {}^{4}c_1}{{}^{52}c_4} = \frac{256}{{}^{52}C_4}$$

2. Required probability = $\dfrac{{}^{4}C_2 \times {}^{4}C_2}{{}^{52}C_4}$

3. Since there are 26 black cards (of spade and clubs) and 26 red cards (of diamond and hearts) in a pack of cards, the required probability $= \dfrac{{}^{26}C_2 \times {}^{26}C_2}{{}^{52}C_4}$

4. Required probability $= \dfrac{{}^{13}C_2 \times {}^{13}C_2}{{}^{52}C_4}$

(b)There are ${}^{52}C_4$ possible ways in which four cards can slip while shuffling a pack of cards. The favourable number of cases in which the four ${}^{39}C_4$ cards can be one from each suit is: ${}^{13}c_1 \times {}^{13}c_1 \times {}^{13}c_1 \times {}^{13}c_1$.

The required probability $= \dfrac{{}^{13}c_1 \times {}^{13}c_1 \times {}^{13}c_1 \times {}^{13}c_1}{{}^{52}c_4} = \dfrac{2197}{20825}$.

Example 7: What is the probability of getting 9 cards of the same suit in one hand at a game of bridge?

Solution: Since one hand in a bridge game consists of 13 cards, the exhaustive number of cases is ${}^{52}c_{13}$. The number of ways in which 9 cards of a suit can come out of 13 cards of the suits $= {}^{13}c_9$. The number of ways in which balance 13-9 = 4 cards can come in one hand out of a balance of 39 cards of the other suit is ${}^{39}c_4$.

Since there are four different suits and 9 cards of any suit can come, by the principle of counting, the total number of favourable cases of getting 9 cards of suit $= \dfrac{{}^{13}c_9 \times {}^{39}c_4 \times 4}{{}^{52}c_{13}}$.

Example 8: If the letters of the word STATISTICS are arranged randomly then find the probability that all the three T's are together.

Solution: Let E be the event that the selected word contains 3 [TTT] T'stogether. There are 10 letters in the word STATISTICS. If we consider three T's as a single letter, then we have 8 letters.

i.e. 1 ' [TTT] '; 3 'S'; 1 'A'; 2 'I';and1 'C'.

Number of possible arrangements with three T's coming together $= \dfrac{8!}{2!3!}$

And the number of exhaustive cases =Total number of permutations of 10 letters in the wordSTATISTICS $= \dfrac{10!}{2!3!3!}$ ($\because$ out of 10 letters, 3 are T's , 2 are I's and 3 are S's)

$$P(A) = \frac{\frac{8!}{2!3!}}{\frac{10!}{2!3!3!}} = \frac{1}{15}.$$

Problems:

1. Three unbiased coins are tossed simultaneously. Find the probability of getting
   (i) at least two heads
   (ii) at most two heads
   (iii) All heads
   (iv) Exactly one head
   (v) Exactly one tail.
2. A single letter selected at random from the word 'STATISTICS'. What is the probability that it is a vowel?
3. A committee of 4 people is to be appointed from 3 officers of the production department, 4 officers of the purchase department, 2 officers of the sales department and 1 chartered accountant. Find the probability of forming the committee in the following manner:
   (a) There must be one from each category.
   (b) It should have at least one from the purchase department.
   (c) The chartered accountant must be in the committee.
4. In a random arrangement of the letters of the word 'COMMERCE'.Find the probability that all the vowels come together.
5. (a) If the letters of the word 'REGULATIONS' be arranged at random, what is the chance that there will be exactly 4 letters between R and E?
   (b) What is the probability that four S's come consecutively in the word 'MISSISSIPPI'?

6. Compare the chances of throwing 4 with one die, 8 with two dice and 12 with three dice.

7. A and B throw three dice; if A throws 14, find B's chance throwing a higher number.

8. In a family, there are two children. Write the sample space and find the probability that
   (a) The elder child is a girl
   (b) Younger child is a girl
   (c) Both are girls
   (d) Both are of opposite sex.

Concept of Odds in Favour of and Against the Happening of an Event

Let n be the number of exhaustive cases in a random experiment which are mutually exclusive and equally likely as well. Let m out of these n cases are favourable to happening of an event A (say). Then, the numbers of cases against A are n-m.

Then odds in favour of event A are $m:n-m$ (i.e. m ratio $n-m$) and odds against A are $n-m:m$.

Example 9: If odds in favour of event A are 3:4, what is the probability of happening A?

Solution: As odds in favour of A are 3:4,

$\therefore$ m = 3 and n-m = 4 implies that n =7.

i.e. P(A)=m/n = 3/7.

Example 10: Find the probability of event A if

1. Odds in favour of event A are 4:3
2. Odds against event A are 5:8

Solution:

1. We know that if odds in favour of A are m:n then

$$P(A)=\frac{m}{m+n}, \Rightarrow P(A)=\frac{4}{4+3}=\frac{4}{7}$$

2. Here n-m=5 and m = 8, therefore n = 5+8 = 13

Now, as we know that if odds against the happening of event A are $n-m:m$ then

$$P(A)=\frac{m}{n}=\frac{8}{13}.$$

Problem:

1. If $P(A) = \frac{3}{5}$ then find

(a) Odds in favour of A;

(b) Odds against the happening of event A.

1.5 Some Theorems on Probability

In the last section, we have studied the probability of an event in a random experiment as well as axiomatic approach and observed that probability as a function of outcomes of an experiment. By now you know that the probability $P(A)$ of an event A associated with a discrete sample space is the sum of the probabilities assigned to the sample points in A as discussed in axiomatic approach of probability. Moreover, in practical problems, writing down the elements of S and counting the number of cases favourable to a given event often become very tedious. However, in such situations the computation of probabilities can be facilitated to a great extent by fundamental theorem of addition. In this chapter, we will learn Addition Theorem of Probability to find probability of occurrence for simultaneous trials under two conditions when events are mutually exclusive and when they are not mutually exclusive.

List of Symbols

$A \cup B$: An event which represents the happening of at least one of the events *A* and *B* i.e. either *A* occurs or *B* occurs or both *A* and *B* occur. This is also denoted as *A* or *B*

$A \cap B$: An event which represents the simultaneous happening of both *A* and *B* i.e. *A*and *B*.

$\overline{A}$: *A* does not happen.

$\overline{A} \cap \overline{B}$: Neither *A* nor *B* happens i.e. none of *A* and *B* happens.

$\overline{A} \cap B$: *A* does not happen but *B* happens.

$(A \cap \overline{B}) \cup (\overline{A} \cap B)$ Exactly one of the two events *A* and *B* happens.

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

Proof: Impossible event contains no sample point and hence the certain event *S* and the impossible event ϕ are mutually exclusive.

$$\therefore S \cup \phi = S \Rightarrow P(S \cup \phi) = P(S)$$

Hence using axiom 2 of probability;

i.e., axiom of additivity, we get

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

Hence proved.

Theorem 2: Probability of the complementary event $\overline{A}$ of *A* is given by

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

Proof: Since A and $\overline{A}$ are mutually disjoint events,

so that $A \cup \overline{A} = S \Rightarrow P(A \cup \overline{A}) = P(S)$

Hence, we have

$$P(A) + P(\overline{A}) = P(S) = 1$$
$$\Rightarrow P(A) = 1 - P(\overline{A}).$$

Theorem 3: For any two events *A* and *B*, we have

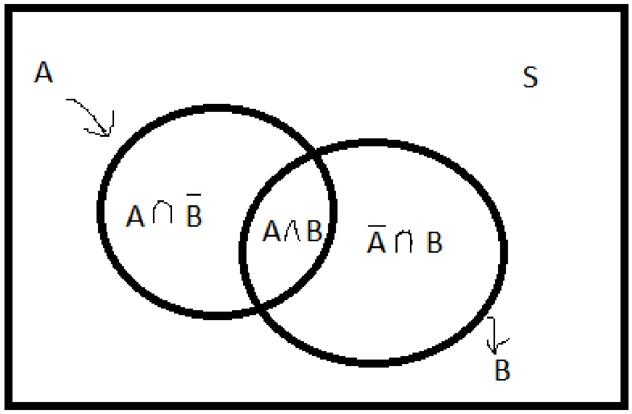
1. $P(\overline{A} \cap B) = P(B) - P(A \cap B)$
2. $P(A \cap \overline{B}) = P(A) - P(A \cap B)$

Proof: 1. From the Venn diagram, we get $B = (A \cap B) \cup (\overline{A} \cap B)$,

where $\overline{A} \cap B$ and $A \cap B$ are disjoint events.

Hence by axiom (3), we get

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



$\Rightarrow P(\overline{A} \cap B) = P(B) - P(A \cap B)$.

2. Similarly, we have $A = (A \cap B) \cup (A \cap \overline{B})$.

where $(A \cap B)$ and $A \cap \overline{B}$ are disjoint events.

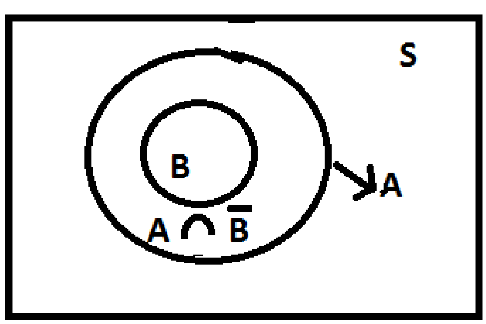
Hence by axiom (3), we have

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

$\Rightarrow P(A \cap \overline{B}) = P(A) - P(A \cap B)$.

Theorem 4: If $B \subset A$, Then

1. $P(A \cap \overline{B}) = P(A) - P(B)$
2. $P(B) \leq P(A)$



Proof:

1. When $B \subset A$, B and $A \cap \overline{B}$ are mutually exclusive events so that

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

$\Rightarrow P(A) = P[BC(A \cap \overline{B})] = P(B) + P(A \cap \overline{B})$ (By axiom (3))

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

2. $P(A \cap \overline{B}) \geq 0 \Rightarrow P(A)-P(B) \geq 0 \Rightarrow P(B) \leq P(A)$

Hence $B \subset A$

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

Hence proved.

Example 11: *A, B* and *C* are three mutually exclusive and exhaustive events associated with a random experiment. Find *P(A),* given that

$P(B) = \frac{3}{4}P(A)$ and $P(C) = \frac{1}{3}P(B)$.

Solution: It is given that *A; B* and *C* are three mutually exclusive and exhaustive events.

$\therefore A \cup B \cup C = S$

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

$\Rightarrow P(A) + P(B) + P(C) = 1$

$\Rightarrow P(A) + \frac{3}{4}P(A) + \frac{1}{3}P(B) = 1$

$\Rightarrow P(A) + \frac{3}{4}P(A) + \frac{1}{3}\left(\frac{3}{4}P(A)\right) = 1$

$\Rightarrow 2P(A) = 1$

$\Rightarrow P(A) = \frac{1}{2}$.

Problem: If two dice are thrown, what is the probability that sum is

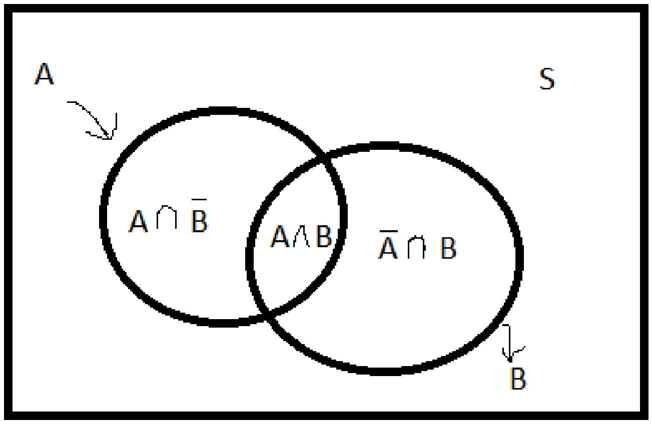
(a) Greater than 9 and

(b) Neither 10 nor 12.

Addition Theorem of Probability

Theorem 5: If *A* and *B* are any events and are not disjoint then

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



Proof: From the Venn diagram, we have $A \cup B = A \cup (\overline{A} \cap B)$

where A and $\overline{A} \cap B$ are mutually disjoint.

$\therefore P(A \cup B) = P[A \cup (\overline{A} \cap B)]$

$= P(A) + P(\overline{A} \cap B)$ (By axiom (3))

$= P(A) + P(B) - P(A \cap B)$ (By theorem (3))

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

Cor.1: For three non-mutually exclusive events *A, B* and *C*, we have

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

Proof: $P(A \cup B \cup C) = P[A \cup (B \cup C)]$

$= P(A) + P(B \cup C) - P[A \cap (B \cup C)]$ (By theorem (5))

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

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

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

Hence proved.

Cor.2: If *A* and *B* are two mutually exclusive events, then the probability of occurrence of either *A* or *B* is the sum of the individual probabilities of *A* and *B*. Symbolically

$$P(A \cup B) = P(A \text{ or } B) = P(A) + P(B).$$

Proof : Let *N* be the total number exhaustive and equally likely cases of an experiment. Let m_1 and m_2 be the number of cases favourable to the happening of events *A* and *B* respectively. Then

$$P(A) = \frac{n(A)}{n(S)} = \frac{m_1}{N}$$

and

$$P(B)=\frac{n(B)}{n(S)}=\frac{m_2}{N}.$$

Since the events A and B are mutually exclusive, the total number of events favorable to either A or B i.e. $n(A\cup B)=m_1+m_2$ then

$$P(A\cup B)=\frac{n(A\cup B)}{n(S)}=\frac{n(A)+n(B)}{N}=\frac{m_1}{N}+\frac{m_2}{N}=P(A)+P(B).$$

Theorem 6: For n events $A_1, A_2,...,A_n$, we have

$$P(\bigcup_{i=1}^{n} A_i)=\sum_{i=1}^{n} P(A_i)-\sum_{1\le i<j\le n}\sum P(A_i\cap A_j)+...+(-1)^{n-1}P(A_1\cap A_2\cap...\cap A_n) \quad ...(*)$$

Proof: For two events A_1 and A_2, we have

$P(A_1\cup A_2)=P(A_1)+P(A_2)-P(A_1\cap A_2)$ …(i)

Hence (*) is true for n=2

Let us suppose that (*) is true for n=r (say)

$$P(\bigcup_{i=1}^{r} A_i)=\sum_{i=1}^{r} P(A_i)-\sum_{1\le i<j\le r}\sum P(A_i\cap A_j)+...+(-1)^{r-1}P(A_1\cap A_2\cap...\cap A_r) \quad ...(ii)$$

Now, $P(\bigcup_{i=1}^{r+1} A_i)=P\{(\bigcup_{i=1}^{r} A_i)\cup A_{r+1}\}$

$$=P(\bigcup_{i=1}^{r} A_i)+P(A_{r+1})-P\{(\bigcup_{i=1}^{r} A_i)\cap A_{r+1}\}$$

$$=P(\bigcup_{i=1}^{r} A_i)+P(A_{r+1})-P\{(\bigcup_{i=1}^{r} (A_i\cap A_{r+1})\}$$

(By distribution law)

$$P(\bigcup_{i=1}^{r+1} A_i)=\sum_{i=1}^{r} P(A_i)-\sum_{1\le i<j\le r}\sum P(A_i\cap A_j)+...+(-1)^{r-1}P(A_1\cap A_2\cap...\cap A_r)+ \quad \text{(From (ii)}$$

$$P(A_{r+1})-P\{(\bigcup_{i=1}^{r} (A_i\cap A_{r+1})\}$$

$$= \sum_{i=1}^{r+1} P(A_i) - \sum_{1 \le i < j \le r} \sum P(A_i \cap A_j) + ... + (-1)^{r-1} P(A_1 \cap A_2 \cap ... \cap A_r) - \sum_{i=1}^{r} P(A_i \cap A_{r+1})$$

$$- \sum_{1 \le i < j \le r} \sum P(A_i \cap A_j \cap A_{r+1}) + ... + (-1)^{r-1} P(A_1 \cap A_2 \cap ... \cap A_r \cap A_{r+1}) \}$$

from equation (ii)

$$P(\bigcup_{i=1}^{r+1} A_i) = \sum_{i=1}^{r+1} P(A_i) - \sum_{1 \le i < j \le r} \sum P(A_i \cap A_j) + \sum_{i=1}^{r} P(A_i \cap A_{r+1}) + ... + (-1)^r P(A_1 \cap A_2 \cap ... \cap A_{r+1})$$

$$P(\bigcup_{i=1}^{r+1} A_i) = \sum_{i=1}^{r+1} P(A_i) - \sum_{1 \le i < j \le r+1} \sum P(A_i \cap A_j) + ... + (-1)^r P(A_1 \cap A_2 \cap ... \cap A_{r+1})$$

Hence by the principle of mathematical induction, it follows that (*) is true for all +ve integral values on n.

Example 12: Two dice are tossed. Find the probability of getting 'an even number on the first die or a total of 8'.

Solution: In a random toss of two dice, sample space S is given by

$S=\{1,2,3,4,5,6\} X \{,1,2,3,4,5,6\} \Rightarrow n(s)=6\times 6=36$

Let us define the events

A: Getting an even number on the first dice.

B: the sum of the points obtained on the two dice 8.

These events are represented by the following subset of S.

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

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

Also $A \cap B=\{(2,6),(6,2),(4,4)\} \Rightarrow n(A \cap B)=3$

$$\therefore \quad P(A) = \frac{n(A)}{n(S)} = \frac{18}{36} = \frac{1}{2}$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{5}{36}$$

and $P(A \cap B) = \dfrac{n(A \cap B)}{n(S)} = \dfrac{3}{36} = \dfrac{1}{12}$

Hence the required probability is given by:

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

$$= \frac{1}{2} + \frac{5}{36} - \frac{3}{36} = \frac{5}{9}.$$

Example 13: An integer is chosen at random from two hundred digits. What is the probability that the integer is divisible by 6 or 8?

Solution: The sample space of random experiment is

S=1,2,3,………199,200 $\Rightarrow$ *n(S)=200*

The event A: 'integer chosen is divisible by 6' has the sample points given by:

A= 6, 12, 18……….198 $\Rightarrow$ *n(A)=198/6=33* $\therefore P(A) = \frac{n(A)}{n(S)} = \frac{33}{200}$

Similarly the event B: 'integer chosen is divisible by 8' has the sample points given by:

B=8, 16,24,…,200 $\Rightarrow$ *n(B)=200/8=25*

$$\therefore P(B) = \frac{n(B)}{n(S)} = \frac{25}{200}$$

The LCM of 6 and 8 is 24. Hence a no. is divisible by both 6 and 8 if it is divisible by 24.

$\therefore A \cap B = 24, 48, 72, \ldots, 192 \Rightarrow n(A \cap B) = \frac{192}{24} = 8 \Rightarrow P(A \cap B) = \frac{8}{200}$

Hence the required probability is:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{33}{200} + \frac{25}{200} - \frac{8}{200} = \frac{1}{4}.$$

Example 14: The probability that a student passes a physics test is 2/3 and probability that he passes both a Physics test and an English test is 14/45. The probability that he passes at least one that is 4/5.What is the probability that he passes the English test?

Solution: Let us define the following Events:

A: the student passes a Physics test;

B: the student passes an English test;

So, $P(A)=2/3$, $P(A \cap B)=14/45$; $P(A \cup B)=4/5$ and we want, $P(B)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B), \quad \frac{4}{5} = \frac{2}{5} + P(B) - \frac{14}{45};$$

$$\therefore P(B) = \frac{4}{5} + \frac{14}{45} - \frac{2}{5} = \frac{4}{9}.$$

Example 15: An investment consultant predicts that the odds against the price of a certain stock will go up during the next week is 2:1 and the odds in favour of the price remaining the same are 1:3. What is the probability that the price of the stock will go down during the next week?

Solution: Let A denote the event that 'stock price will go up' and B be the event 'Stock price will remains same.

Then *P(A)=1/(2+1)=1/3* and *P(B)=1/(1+3)=1/4*

(∴P Stock price will either go up or remain same)

is given by : $P(A \cup B) = P(A) + P(B) = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$

Hence, the probability that stock price will go down is given by:

$P(\overline{A} \cap \overline{B}) = =1-P(A \cup B)=1-7/12=5/12.$

Example 16: An MBA applies for a job in two firms X and Y. The probability of his being selected in firm X is 0.7 and being rejected at Y is 0.5. The probability of at least one of his applications being rejected is 0.6. What is probability that he will be selected in one of the firms?

Solution: Let A and B denotes the events that the person is selected in firms X and Y respectively. Then in the usual notations, we are given:

$P(A) = 0.7 \Rightarrow P(\overline{A}) = 1 - 0.7 = 0.3$

$P(B)=0.5 \Rightarrow P(\overline{B}) = 1-0.5=0.5$

and $P(\overline{A} \cup \overline{B}) = 0.6 = P(\overline{A}) + P(\overline{B}) - P(\overline{A} \cap \overline{B})$ ….(ii)

The probability that the persons will be selected in one of the two firms X and Y is given by:

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

$=1-0.3-0.5+0.6=0.8$ (from i)

Example 17: 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 newspaper A, B and C respectively. Then we are given:

$P(E) = \frac{20}{100}, \quad P(F) = \frac{16}{100}, \quad P(G) = \frac{14}{100},$

$P(E \cap F) = \frac{8}{100}, \; P(E \cap G) = \frac{5}{100}, \quad P(F \cap G) = \frac{4}{100}$

and $P(E \cap F \cap G) = \frac{2}{100}$

The required probability that an adult reads at least one of the newspapers (By Addition Theorem) is given by:

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

$$= \frac{20}{100} + \frac{16}{100} + \frac{14}{100} - \frac{8}{100} - \frac{4}{100} - \frac{5}{100} + \frac{2}{100}$$

$$= \frac{35}{100} = 0.35$$

Hence 35% of adult population reads at least one of the newspapers.

Example 18: A card is drawn at random from a pack of 52 cards. Find the probability that the drawn card is either a club or an ace of diamond.

Solution : Let A : Event of drawing a card of cluband

B:Event of drawing an ace of diamond

The probability of drawing a card of club $P(A) = \frac{13}{52}$

The probability of drawing an ace of diamond $P(B) = \frac{1}{52}$

Since the events are mutually exclusive, the probability of the drawn card being a club or an ace of diamond is:

$$P(A \bigcup B) = P(A) + P(B) = \frac{13}{52} + \frac{1}{52} = \frac{7}{26} \ .$$

Problems:

1. From a pack of 52 playing cards, one card is drawn at random. What is the probability that it is a jack of spade or queen of heart?
2. 25 lottery tickets are marked with first 25 numerals. A ticket is drawn at random. Find the probability that it is multiple of 5 or 7.
3. Find the probability of getting either a number multiple of 3 or a prime number when a fair die is thrown.
4. There are 40 pages in a book. A page is opened at random. Find the probability that the number of this opened page is a multiple of 3 or 5.
5. A card is drawn from a pack of 52 playing cards, find the probability that the drawn card is an ace or a red colour card.
6. Two dice are thrown together. Find the probability that the sum of the numbers turned up is either 6 or 8.

7. A card is drawn from a pack of 52 playing cards. Find the probability that it is either a king or a red card.

1.6 Boole's Inequality

Statement: For n events $A_1, A_2, \ldots, A_n$, we have

$$(a)\ P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$$

$$(b)\ P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

Proof: (a)$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq 1$

$\Rightarrow P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1$ ....(i)

Hence (a) is true for $n=2$

Let us now suppose that (a) is true for $n=r$ (say) that

$$P\left(\bigcap_{i=1}^{r} A_i\right) \geq \sum_{i=1}^{r} P(A_i) - (r-1)$$

Then for $n=r+1$

$$P\left(\bigcap_{i=1}^{r+1} A_i\right) = P\left(\bigcap_{i=1}^{r} A_i \cap A_{r+1}\right) \quad .....(ii)$$

$$\geq \sum_{i=1}^{r} P(A_i) - (r-1) + P(A_{r+1}) - 1$$

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1) \qquad \text{(is true for n=r+1 also)}$$

The results now follow by the principle of mathematical induction.

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

We know, $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

$\leq P(A_1) + P(A_2)$ ...(iii)

Hence (b) is true for $n=2$.

Let us now suppose that (b) is true for n=r, so that $P(\bigcup_{i=1}^{r} A_i) \le \sum_{i=1}^{r} P(A_i)$ ....(iv)

Now, $P(\bigcup_{i=1}^{r+1} A_i) = P(\bigcup_{i=1}^{r} A_i \cup A_{r+1})$

$P(\bigcup_{i=1}^{r+1} A_i) \le P(\bigcup_{i=1}^{r} (A_i) + P(A_{r+1})$ (using (iii))

$P(\bigcup_{i=1}^{n} A_i) \le \sum_{i=1}^{n} P(A_i)$ (using (iv))(is true for n=r+1 also)

Hence by mathematical induction we conclude that (b) is true for all positive integral values of n.

1.7 Conditional Probability

In many situations we have the information about the occurrence of an event A and are required to find out the probability of the occurrence of another event B. The probability attached to such an event is called the conditional probability and is denoted by P (A|B) or in other words, probability of A given that B has occurred. For example, if we want to find the probability of an ace of spade if we know that card drawn from a pack of cards is black. Let us consider another problem relating to dairy plant. There are two lots of full cream pouches A and B, each containing some defective pouches. A coin is tossed and if it turns up with its head upside lot A is selected and if it turns with tail up, lot B is selected. In this problem, we are interested to know the probability of the event that a milk pouch selected from the lot obtained in this manner is defective.

Definition: The conditional probability of *B* given that event *A* has occurred is defined as

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

Let us consider the experiment of throwing of a die once. The sample space of thisexperiment is {1, 2, 3, 4, 5, 6}.

Let E_1: an even number shows up and E_2: multiple of 3 show up.

Then E_1: {2, 4, 6} and E_2: {3, 6}. Hence, P (E_1) = 3/6 =and P (E_2) = 2/6 =1/3.

In order to find the probability of occurrence of E_2 when it is given that E_1 has occurred,we know that in a single throw of die 2 or 4or 6 has come up. Out of these only 6 is favorable to E_2. So the probability of occurrence of E_2 when it is given that E_1 has occurred is equal to 1/3. This probability of E_2 when E_1 has occurred is written as P ($E_2|E_1$). Here we find that P ($E_2|E_1$) =P (E_2). Let us consider the event E_3: a number greater than 3 shows up then E_3:{4,5,6} and P(E_3)=3/6=1/2 Out of 2,4 and 6, two numbers namely 4 and 6 are favorable to E_3. Therefore, P ($E_3|E_1$) =2/3. The events of the type E_1 and E_2 are called independent events as the occurrence or non-occurrence of E_1 does not affect the probability of occurrence or non-occurrence of E_2. The events E_1 and E_3 are not independent.

Examples 19: From a city population, the probability of selecting (i) a male or a smoker is 7/10 (ii) a male smoker is 2/5 and (iii) a male, if a smoker is already selected is 2/3. Find the probability of selecting (a) a non-smoker (b) a male and (c) a smoker, if a male is first selected.

Solution: Define the following events:

A: a male is selected B: a smoker is selected we are given:

$$P(A \cup B) = \frac{7}{10},\ P(A \cap B) = \frac{2}{5},\ P(A \mid B) = \frac{2}{3}$$

(a) The probability of selecting a non-smoker is

$$P(\overline{B}) = 1 - P(B) = 1 - \frac{P(A \cap B)}{P(A \mid B)}\ [\because P(A|B) = P(A \cap B)/P(B)\ ]$$

$$= 1 - \frac{3}{5} = \frac{2}{5}$$

(b) The probability of selecting a male

(By Addition Theorem) is

$$P(A) = P(A \cup B) + P(A \cap B) - P(B) = \frac{7}{10} + \frac{2}{5} - \frac{3}{5} = \frac{1}{2}.$$

(c) The probability of selecting a smoker if a male is first selected is:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{2/5}{1/2} = \frac{4}{5}.$$

Example 20: Two computers A and B are to be marked. A salesman who is assigned the job of finding customers for them has 60% and 40% chances respectively of succeeding in case of computer A and B. The two computers can be sold independently. Given that he was able to sell at least one computer, what is the probability that computer A has been sold?

Solution: Let E denote the event that computer A is marketed and F denote the event that Computer B is marked. We are given:

$$P(E) = \frac{60}{100} = 0.60 \Rightarrow P(\overline{E}) = 0.40 \text{ and } P(F) = \frac{40}{100} = 0.40 \Rightarrow P(\overline{F}) = 0.60$$

Required Probability= $P[E \mid (E \cup F)] = \dfrac{P[E \cap (E \cap F)]}{P(E \cup F)}$

$$\frac{P(E)}{1 - P(\overline{E} \cap \overline{F})} = \frac{P(E)}{1 - P(\overline{E})P(\overline{F})} = \frac{0.6}{1 - 0.4 \times 0.6} = 0.79.$$

Example 21: An urn contains 4 red and 7 blue balls. Two balls are drawn one by one without replacement. Find the probability of getting 2 red balls.

Solution: Let A be the event that first ball drawn is red and B be the event that the second ball drawn is red.

$\therefore P(A) = \frac{4}{11}$ and $P(B \mid A) = \frac{3}{10}$ (it is given that one red ball has already been drawn)

$\therefore$ The required probability=

$$P(A \text{ and } B) = P(A)P(B \mid A) = \frac{4}{11} \times \frac{3}{10} = \frac{6}{55}.$$

Problems:

1. From a pack of 52 cards, two cards are drawn at random one after the other with replacement. What is the probability that both cards are kings?
2. A bag contains 4 red balls, 3 white balls and 5 black balls. Two balls are drawn one after the other with replacement. Find the probability that first is red and the second is black.

1.8 Multiplication Theorem of Probability

Theorem: For two events A and B, $P(A \cap B) = P(A).\, P(B|A) \quad P(A)>0$

$$= P(B).P(A|B) \quad , P(B)>0$$

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

Proof: Let A and B be the events associated with the sample space S of a random experiment with exhaustive number of outcomes (sample points) N, i.e., $n(S) = N$. Then by definition

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} \qquad (1)$$

For the conditional event $A|B$ (i.e., the happening of A under the condition that B has happened), the favorable outcomes (sample points) must be out of the sample points of B. In other words, for the event $A|B$, the sample space is B and hence

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

Similarly, we have

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

On multiplying and dividing equation (1) by n(A), we get

$$P(A \cap B) = \frac{n(A)}{n(S)} \times \frac{n(A \cap B}{n(A)}$$
$$= P(A).P(B \mid A)$$

Also

$$P(A \cap B) = \frac{n(B)}{n(S)} \times \frac{n(A \cap B}{n(B)}$$
$$= P(B).P(A \mid B).$$

Generalization

The multiplication theorem of probability can be extended to more than two events. Thus, for three events A_1, A_2 and A_3 we have

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

For n events $A_1, A_2,..., A_n$, we have

$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)\ P(A_2|A_1)\ P(A_3|A_1 \cap A_2)...\ P(A_n|A_1 \cap A_2 \cap ... \cap A_{n-1})$

Example 22: A bag contains 5 white and 8 red balls. Two successive drawings of 3 balls are made such that (a) the balls are replaced before the second drawing, and (b) the balls are not replaced before the second draw. Find the probability that the first drawing will give 3 white and the second 3 red balls in each case.

Solution:(a) When balls are replaced.

Total balls in the bag = 8 + 5 = 13

3 balls can be drawn out of total of 13 balls in $^{13}C_3$ ways.

3 white balls can be drawn out of 5 white balls in $^{5}C_3$ ways.

Probability of 3 white balls $= P(3W) = \frac{^5C_3}{^{13}C_3} = \frac{10}{286}$.

Since the balls are replaced after the first draw so again there are 13 balls in the bag 3 red balls can be drawn out of 8 red balls in $^{8}C_3$ ways.

Probability of 3 red balls $= P(3R) = \frac{^8C_3}{^{13}C_3} = \frac{56}{286}$

Since the events are independent, the required probability is:

$$P(3W \text{ and 3R}) = \frac{^5C_3}{^{13}C_3} \times \frac{^8C_3}{^{13}C_3} = \frac{10}{286} \times \frac{56}{286} = \frac{140}{20449}.$$

(b) When the balls are not replaced before second draw

Total balls in the bag = 8 + 5 = 13

3 balls can be drawn out of 13 balls in ${}^{13}C_3$ ways.

3 white balls can be drawn out of 5 white balls in ${}^{5}C_3$ ways.

The probability of drawing 3 white balls = $P(3W) = \frac{{}^{5}C_3}{{}^{13}C_3}$

After the first draw, balls left are 10, 3 balls can be drawn out of 10 balls in ${}^{10}C_3$ ways.

3 red balls can be drawn out of 8 balls in ${}^{8}C_3$ ways. Probability of drawing 3 red balls = $\frac{{}^{8}C_3}{{}^{10}C_3}$.

Since both the events are dependent, the required probability is:

$$P(3W \text{ and } 3\text{R}) = P(3W) \times P(3R \mid 3W) = \frac{{}^{5}C_3}{{}^{13}C_3} \times \frac{{}^{8}C_3}{{}^{10}C_3} = \frac{5}{143} \times \frac{7}{15} = \frac{7}{449}.$$

Example 23. A bag contains 5 white and 3 red balls and four balls are successively drawn and are not replaced. What is the chance that (i) white and red balls appear alternatively and (ii) red and white balls appear alternatively?

Solution (i) The probability of drawing a white ball = 5/8

The probability of drawing a red ball = 3/7

The probability of drawing a white ball = 4/6 and the probability of drawing a red ball = 2/5

Since the events are dependent, therefore the required probability is:

P(W and R and W and R)= $P(W \cap R \cap W \cap R)$

=P(W).P(R|W).P(W|WR).P(R|WRW)

$$= \frac{5}{8} \times \frac{3}{7} \times \frac{4}{6} \times \frac{2}{5} = \frac{1}{14}$$

(ii) The probability of drawing a red ball = 3/8 and the probability of drawing a white ball = 5/7

The probability of drawing a red ball = 2/6 and the probability of drawing a white ball = 4/5

Since the events are dependent, therefore the required probability is:

P(R and W and R and W)= $P(R \cap W \cap R \cap W)$

=P(R).P(W|R).P(R|RW).P(W|RWR)

$$= \frac{3}{8} \times \frac{5}{7} \times \frac{2}{6} \times \frac{4}{5} = \frac{1}{14}.$$

Example 24: (Huyghen's Problem) A and B throw alternatively witha pair of balanced dice. A wins if he throws a sum of six points before B throws a sum of seven points while B wins if he throws sum ofseven points before A throws a sum of six pints. If A begins the game, show that his probability of winning is 30/61.

Solution: Let of A_i denote the event *A's* throwing ,'6' in the ith throw, $i=1,2,3,....$ and Let of B_i denote the event *B*'s throwing ,'7' in the ith throw, $i=1,2,3,....;$ with a pair of dice. Then $\overline{A}_i$ and $\overline{B}_i$ are the complementary events.

'6' can be obtained with two dice in the following ways:

(1,5), (5,1), (2,4), (4,2), (3,3), i.e., in 5 different ways.

$$\therefore P(A_i) = \frac{5}{36} \Rightarrow P(\overline{A}) = 1 - P(A) = 1 - \frac{5}{36} = \frac{31}{36}, i = 1, 2, 3....$$

'7' can be obtained with two dice in the following ways:

(1,6), (6,1), (2,5), (5,2), (3,4), (4,3), i.e., in 6 different ways.

$$\therefore \mathrm{P}(\mathrm{A}_i) = \frac{1}{6} \Rightarrow P(\overline{A}) = 1 - P(A) = 1 - \frac{1}{6} = \frac{5}{6}, i = 1, 2, 3....$$

If *A* start the game, he will win in the following ways:

(i) A_1 happens, (ii) $\overline{A}1 \cap \overline{B}2 \cap A_3$ (iii) $\overline{A}_1 \cap \overline{B}_2 \cap \overline{A}_3 \cap \overline{B}_4 \cap A_5$ happens and so on.

Hence by addition theorem of probability of *A*'s winning, $P(A)$ is given by:

$$P(A) = P(i) + P(ii) + P(iii) + ...$$

$$= P(A_1) + P(\overline{A}1 \cap \overline{B}2 \cap A_3) + P(\overline{A}_1 \cap \overline{B}_2 \cap \overline{A}_3 \cap \overline{B}_4 \cap A_5) + ...$$

$$= P(A_1) + P(\overline{A}1)P(\overline{B}_2)P(A_3) + P(\overline{A}_1)P(\overline{B}_2)P(\overline{A}_3)P(\overline{B}_4)P(A_5) + ...$$

$$= \frac{5}{36} + \frac{31}{36} \times \frac{5}{6} \times \frac{5}{36} + \frac{31}{36} \times \frac{5}{6} \times \frac{31}{36} \times \frac{5}{6} \times \frac{5}{36} + ...$$

$$= \frac{\frac{5}{36}}{1 - \frac{31}{36} \times \frac{5}{6}} = \frac{30}{61}.$$

Problems:

1. A coin is tossed once. If it shows head, it is tossed again and if it shows tail, then a dice is tossed. Let E_1 be the event: the first throw of coin shows tail and E_2 be the event: the dice shows a number greater than 4. Find P ($E_2|E_1$).

2. A bag contains 5 white and 4 black balls. A ball is drawn from this bag and is replaced and then second draw of a ball is made. What is the probability that two balls are of different colors.

3. A can hit a target 4 times in 5 shots. B 3 times in 4 shots and C twice in 3 shots. They fire a volley. What is the probability that

   (a)Two shots hit the target

   (b) At least two shots hit the target.

4. Three small sized Herds A, B and C consist of 3 cows and 1 buffalo, 2 cows and 2 buffaloes, 1 cow and 3 buffaloes, respectively. Find the probability of selecting one cow and two buffaloes from three Herds.

1.9 Independent Event

Two or more events are said to be independent if the happening or non-happening of any one of them, does not, in any way, affect the happening of others. For Example: Throwing of two dice say die 1 and die 2. It is obvious that the occurrence of certain number of dots on the die 1 has nothing to do with a similar event for the die 2 the two are quite independent of each other.

Definition: An event A is said to be independent (or statistically independent) of another event B, if the conditional probability of A given B, i.e., $P(A|B)$ is equal to the unconditional probability of B, i.e., if $P(A|B)=P(A)$.

Multiplication Theorem of Probability for Independent Event

Theorem: If A and B are two events withpositive probabilities $P(A)\neq 0$, $P(B)\neq 0$, then A and B are independent if and only if $P(A\cap B)=P(A).P(B)$ …(*)

Proof: We have $P(A\cap B)=P(A).P(B|A)=P(B)P(A|B\ ;P(A)\neq 0,\ P(B)\neq 0$ ... (i)

If A and B are independent, i.e., A is independent of B and B is independent of A, then, we have, $P(A|B) = P(A)$and $P(B|A)=P(B)$ …(ii)

From (i) and (ii), we get $P(A\cap B)= P(A).P(B)$ are required conversely, if (*) holds, then we get

$P(A\cap B)/P(B) = P(A) \Rightarrow P(A|B)= P(A),\quad P(A\cap B)/P(A)=P(B) \Rightarrow P(B|A)=P(B)$ … (iii)

(iii) implies that A and B are independent events. Hence, for independent events A and B, the probability that both of these occur simultaneously is the product of their respective probabilities.

This rule is known as the Multiplication Rule of Probability.

Extension of Multiplication Theorem of Probability to n Events

Theorem: For n events $A_1,A_2,\ldots A_n$, we have

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

where $P(A_i\mid A_j\cap A_k\cap\ldots\cap A_l)$ represents the conditional probability of the event A_i given that the events $A_j, A_k, \ldots, A_l$ have already happened.

Proof: For two events A_1 and A_2, $P(A_1\cap A_2)=P(A_1)P(A_2|A_1)$

We have for three events A_1, A_2 and A_3

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

$=P(A_1)P((A_2\cap A_3)|A_1)\quad (\because P(A\cap B)=P(A)P(B|A))$

$=P(A_1)\ P(A_2|A_1)P(A_3|(A2\cap A_3))$

Thus we find that (A) is true for n =2 and n=3

Let us suppose that (A) is true for n=k, so that

$$P(A_1\cap A_2\cap\ldots\cap A_k)=P(A_1)P(A_2\mid A_1)P(A_3\mid A_1\cap A_2)\ldots\times P(A_k\mid A_1\cap A_2\cap\ldots\cap A_{k-1})\ \ldots(B)$$

Now $P((A_1 \cap A_2 \cap ... \cap A_k) \cap A_{k+1}) = P(A_1 \cap A_2 \cap A_3 \cap ... \cap A_k) P(A_{k+1} | A_1 \cap A_2 \cap ... \cap A_k)$

$$= P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2)... \times P(A_k | A_1 \cap A_2 \cap ... \cap A_{k-1}) \times P(A_{k+1} | A_1 \cap A_2 \cap ... \cap A_k)$$

(using B)

Thus (A) is true for $n=k+1$ also since (A) is true for $n=2$ and $n=3$, by the principle of mathematical induction. It follows that (A) is true for all positive integral value of n.

Extension of Multiplication Theorem of Probability for n Independent Events

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

$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2)P(A_3)...P(A_n)$ --- (C)

Proof: If $A_1, A_2, ..., A_n$ are independent events then

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

Hence from equation (c), we get $P(A_1 \cap A_2 \cap A_3 \cap ... \cap A_n)=P(A_1)P(A_2)....P(A_n)$

Conversely if (C) holds, then from (A) and (C), we get

$P(A_2)= P(A_2|A_1)$;

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

$\Rightarrow A_1, A_2, A_3, ..., A_n$ are independent events.

Theorem: For Any three events A,B and C

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

Proof: We have $P(A \cup B)=P(A)+P(B)-P(A \cap B)$

$$\Rightarrow P[(A \cap C) \cup (B \cap C)]=P(A \cap C)+P(B \cap C)-P(A \cap B \cap C)$$

Dividing both sides by P(C), we get

$$\frac{P[(A \cap C) \cup (B \cap C)]}{P(C)} = \frac{P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)}{P(C)}$$

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

$$\frac{P[(A \cup B) \cap C)]}{P(C)} = P(A|C) + P(B|C) - P(A \cap B|C)$$

$$\Rightarrow \frac{P[(A \cup B) \cap C)]}{P(C)} = P(A|C) + P(B|C) - P(A \cap B|C).$$

Theorem: For any three events A,B and C,

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

Proof:

$$P(A \cap \bar{B}|C) + P(A \cap B|C) = P(A \cap \bar{B} \cap C)/P(C) + P(A \cap B \cap C)/P(C)$$
$$= \frac{P(A \cap \bar{B} \cap C) + P(A \cap B \cap C)}{P(C)}$$
$$= \frac{P(A \cap C)}{P(C)} = P(A|C)$$

1.10 Pairwise Independent Events

Consider *n* events $A_1, A_2, A_3, \ldots, A_n$ defined on the same sample space so that $P(A_i)>0,\ i=1,2,\ldots n$. These events are said to be pair wise independent if every pair of two events is independent.

Definition: The events $A_1, A_2, A_3, \ldots, A_n$ are said to be pairwise independent if and only if

$$P(A_i \cap A_j) = P(A_i)P(A_j);\ i \neq j = 1,2,3,\ldots,n$$

1.11 Mutually Independent Events

The events in S (Sample Space) are said to be mutually independent if the probability of the simultaneous occurrence of (any) finite number of them is equal to the product of their respective probabilities.

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

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

$P(A_{i1} \cap A_{i2} \ldots \cap A_{ik}) = P(A_{i1})P(A_{i2})\ldots P(A_{ik});\ k = 1,2,3,\ldots,n$

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

Proof: We are required to prove:

$P[(A \cup B) \cap C] = P(A \cup B)P(C)$

$L.H.S. = P[(A \cap C) \cup (B \cap C)]$ *(By Distributive law)*

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

$= P(A)P(C) + P(B)P(C) - P(A)P(B)P(C)$ [$\because$ *A,B, and C are mutually independent*]

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

$= P(C)P(A \cup B) = R.H.S.$

Hence (A∪B) and C are independent.

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

Proof: We are given,

$$P(A\cap B)=P(A)P(B), P(B\cap C)=P(B)P(C), P(A\cap C)=P(A)P(C) \quad ....(i)$$

$P[A\cap(B\cup C)]=P(A)P(B\cup C)$

Now, $P[A\cap(B\cup C)=P[(A\cap B)\cup(A\cap C)]$

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

$$=P(A)P(B)+P(A).P(C)-P(A\cap B\cap C) \quad \text{...(ii)} \quad \text{(from (i))}$$

and $P(A)P(B\cup C)=P(A)[P(B)+P(C)-P(B\cap C)]$

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

from (ii) and (iii), on using (i), we get

$P(A\cap B\cap C)=P(A)P(B\cap C)=P(A)P(B)P(C)$ (from (i))

Hence A, B and C are mutually independent.

1.12 Law of Total Probability

Statement:– Let S be the sample space and $E_1, E_2, ..., E_n$ be n mutually exclusive and exhaustive events with $P(E_i)\neq 0; i=1,2,....,n$. Let A be any event which is a subset of $E_1\cup E_2\cup ......\cup E_n$ (i.e., at least one of the events $E_1, E_2 ......E_n$) with $P(A)>0$ then

$P(A)=P(E_1)P(A|E_1)+P(E_2)P(A|E_2)+...+P(E_n)P(A|E_n)$

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

Proof: As A is a subset of $E_1\cup E_2\cup ......\cup E_n$,

$$\therefore A=A\cap(E_1\cup E_2 ......\cup E_n) \quad (\because \text{if } A \text{ is subset of } B, \text{then } A=A\cap B)$$

$\Rightarrow A=(A\cap E_1)\cup(A\cap E_2)\cup......\cup(A\cap E_n)$

(Distributive property of set theory)

$A=(E_1\cap A)\cup(E_2\cap A)\cup.....\cup(E_n\cap A)$

$\Rightarrow P(A)=P[(A\cap E_1)\cup(A\cap E_2)\cup......\cup(A\cap E_n)]$

$P(A\cap E_1)+P(A\cap E_2)+....+P(A\cap E_n)$

[$\because E_1\cup E_2\cup......\cup E_n$ and hence $E_1\cap A$ $E_2\cap A$ ....$E_n\cap A$ are mutually exclusive].

(using multiplication theorem for dependent events)

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

Hence proved.

Example 25: There are two bags. First bag contains 5 red, 6 white balls and the second bag contains 3 red, 4 white balls. One bag is selected at random and a ball is drawn from it. What is the probability that it is (i) red (ii) white.

Solution: Let E_1 be the event first bag is selected and E_2 be the event that second bag is selected.

$\therefore P(E_1)=P(E_2)=1/2$

(i) Let A be the event of getting a red ball from the selected bag.

$$\therefore \quad P(A|E_1) = 5/11, \quad \text{and } P(A|E_2)= 3/7$$

Thus, the required probability is given by

$P(A)=P(E_1)P(A|E1)+P(E_2)P(A|E2)$

$=(1/2)\times(5/11)+1/2\times3/7 =5/22+3/14=(35+33)/154$

$=68/154=34/77.$

(ii) Let W be the event of getting a white ball from the selected bag.

$\therefore\ P(W|E_1)= 6/11, \quad and\ P(W|E_2)= 4/7$

Thus, the required probability is given by

$P(W)=P(E_1)\ P(W|E_1)+P(E_2)\ P(W|E_2)$

$=(1/2)\times(6/11)+(1/2\times(4/7)$

$=3/11+2/7=(21+22)/77$

$=43/77.$

Example 26: The probabilities of selection of 3 persons forthe post of a principal in a newly co-education in the college are 0.2, 0.3 and 0.5, respectively. Find the probability that co-education is introduced in the college.

Solution: Let E_1, E_2, E_3 be the events of selection of first, second and third person for the post of a principal respectively. Let A be the event that co-education is introduced.

$\therefore P(E_1) =4/9,\ P(E_2) =3/9,\ P(E_3)=2/9$,

$P(A|E_1)=0.2, P(A|E_2)=0.3, P(A|E_3)=0.5$

Thus the required probability is

$P(A)=P(E_1)\ P(A|E_1)+P(E_2)\ P(A|E2)+P(E_3)\ P(A|E_3)$

$=4/9\times0.2+3/9\times0.3+2/3\times0.5 =0.89+0.99+19=2.79=0.3$

1.13 Baye's Theorem

Statement: Let S be the sample space and $E_1, E_2, \ldots, E_n$ be n mutually exclusive and exhaustive events with $P(E_i) \neq 0; i=1,2\ldots.n$. Let A be any event which is a subset of $E_1 \cup E_2 \cup \ldots \ldots \cup E_n$ (i.e. at least one of the events $E_1 \cup E_2 \cup \ldots \ldots \cup E_n$) with $P(A)>0$, then

$$P(E_i \mid A) = \frac{P(E_i)P(A \mid E_i)}{P(A)}, \text{ i=1,2,,....n}$$

where $P(A)=P(E_1)\, P(A|E_1)+P(E_2)\, P(A|E_2)+\dots\dots+P(E_n)\, P(A|E_n)$

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

Since $(A \cap E_i) \subset E_i, (i=1,2,3,...,n)$ are mutually disjoint events, we have by addition theorem of probability:

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

by multiplication theorem of probability.

Also we have

$$P(A \cap E_i) = P(A)P(E_i \mid A)$$

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

Remarks:

1. The probabilities E_i are known as prior probabilities because they exist before we gain any information from experiment itself.
2. The probabilities $P(A|E_i)$, i=1,2,3,…,n are called likelihood because they indicate how likely the event A under consideration is to occur, given each and every a prior probability.
3. The probability $P(E_i|A)$ are caller posterior probabilities because they are determined after the results of the experiment are known.

Example 27: The contents of urns I, II and III are 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 from it. They happen to be white and red. What is the probability that they come from urns I, II and III?

Solution: Let E_1, E_2 and E_3 denote the events that the urn I, II and III is chosen, respectively, and let A be the event that the two balls taken from the selected urn are white & red.

then $P(E_1)=P(E_2)=P(E_3)=1/3$, $P(A|E_1)=\frac{1\times 3}{6C_2}=\frac{1}{5}$,

$P(A|E_2)=\frac{2\times 1}{14C_2}=1/3$ and $P(A/E_3)=\frac{4\times 3}{12C_2}=\frac{2}{11}$

$$\therefore P(E_2 \mid A)=\frac{P(E_2)P(A \mid E_2)}{\sum_{i=1}^{3} P(E_i)P(A \mid E_i)}=\frac{\frac{1}{3}\times\frac{1}{3}}{\frac{1}{3}(\frac{1}{5}+\frac{1}{3}+\frac{1}{11})}=\frac{55}{118}$$

Similarly

$$P(E_3/A)=\frac{\frac{1}{3}\times\frac{2}{3}}{\frac{1}{3}(\frac{1}{5}+\frac{1}{3}+\frac{2}{11})}=\frac{30}{118}$$

$$P(E_1|A)=\frac{\frac{1}{3}\times\frac{1}{5}}{\frac{1}{3}(\frac{1}{5}+\frac{1}{3}+\frac{2}{11})}=\frac{33}{118}.$$

Example 28: A speaks truth 4 out of 5 times. A die is tossed. He reports that there is a six. What is the chance that actually there were six?

Solution: Let us define the following events

E_1: *A* speaks truth; E_2:A tells a lie

A: A reports a six

Give: $P(E_1)=4/5$, $P(E_2)=1/5$, $P(A|E_1)=1/6$, $P(A|E_2)=5/6$

The required probability is

$$P(E_1 \mid A)=\frac{P(E_1)P(A \mid E_1)}{\sum_{i=1}^{2} P(E_i)P(A \mid E_i)}$$

$$=\frac{\frac{4}{5}\times\frac{1}{6}}{\frac{4}{5}\times\frac{1}{6}+\frac{1}{5}+\frac{5}{6}}=\frac{4}{9}.$$

Example 29: A letter is known to have come either from TATANAGAR or from CALCUTTA. On the envelope just two consecutive letters TA are visible. What is the probability that the letter came from CALCUTTA?

Solution – Let E_1 and E_2 denote the events that the letter came from TATANAGAR and CALCUTTA respectively. Let A denote the event that two consecutive visible letters on the envelope are TA.

We have

$P(E_1)=P(E_2)=1/2, \quad P(A|E_1)=2/8 and$

$P(A|E_2)=1/7$

Using the Baye's theorem we get

$$P(E_2 \mid A) = \frac{P(E_2)P(A \mid E_2)}{\sum_{i=1}^{2} P(E_i)P(A \mid E_i)}$$

$$= \frac{\frac{1}{2} \times \frac{1}{7}}{\frac{2}{8} \times \frac{1}{2} + \frac{1}{2} + \frac{1}{7}} = \frac{4}{11}.$$

Problems:

1. The chance that doctor *A* will diagnose a disease X correctly is 60%. The chances that a patient will die by his treatment after correct diagnosis is 40% and the chance of death by wrong diagnosis is 70%. A patient of doctor *A*, who had disease *X*, died. What is the chance that his disease was diagnosed correctly?
2. A and B are two weak students of statistics and their chance of solving a problem in statistics correctly are $\frac{1}{6}$ and $\frac{1}{8}$ respectively. Of the probability of their making a common error is find $\frac{1}{525}$ and they obtain the same answer the probability that their answer is correct.

CHAPTER-2

RANDOM VARIABLES, MATHEMATICAL EXPECTATION AND MOMENT GENERATING FUNCTION

Structure

2.1 Introduction

In the previous units, we have studied the assignment and computations of probabilities of events in detail. Also, we were interested in knowing the occurrence of outcomes. In this chapter, we will study the important concept of a random variable and its probability distribution. It has been a general notion that if an experiment or a trial is conducted under identical conditions, values so obtained would be similar. But this is not always true. The observations are recorded about a factor or character under study e.g. fat, TS, moisture content etc., in a dairy product. Also, when ‘n’ coins are tossed, one may be interested in knowing the number of heads obtained. When a pair of dice is tossed, one may seek information about the sum of points. These can take different values, and the factor or character is termed as variable. These observations vary even though the experiment has been conducted under identical conditions. Therefore, we have a set of results or outcomes (Sample points) of a random experiment. In the present unit, we will be interested in the numbers associated with such outcomes of the random experiments and this leads to study the concept of random variable.

2.1.1 Random Variable

A random variable (r.v.) is defined as a real number X connected with the outcome of a random experiment E. In other words, it is a rule that assigns a real number to each outcome (Sample points) is called a random variable. A r.v. has the following properties:

(i) Each particular value of the random variables can be assigned some probability.

(ii) Uniting all the probabilities associated with all the different values of a random variable gives the value 1.

For example, if E consists of three tosses of a coin, we may consider random variable X which denotes the number of heads (0, 1, 2 or 3)

Outcome:	HHH	HTH	THH	THH	HTT	THT	TTH	TTT
Value of X :	3	2	2	2	1	1	1	0

Thus, to every outcome (ω) there correspond a real number X((ω). Since the points of the sample space corresponds to outcomes, this means that a real number, which we denote by $X(\omega)$, is defined for each $\omega \in S$ and let us denote them by $\omega_1, \omega_2, ..., \omega_8$ i.e. $X(\omega_1) = 3$, $X(\omega_2)=2...$, $X(\omega_8) = 0$. Thus, we define a random variable as a real-valued function whose domain is the sample space associated with a random experiment and range is the real line. Generally, one-dimensional random variable is denoted by capital letters X,Y, Z, ..., etc.

Example 1: If a pair of fair dice is tossed then S= {1,2,3,4,5,6} ×{1,2,3,4,5,6} and n(S)=36

Let X be a random variable with image set

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

$$P(X=1) = P\{(1,1)\} = \frac{1}{36}$$

$$P(X=2) = P\{(2,1),(2,2),(1,2)\} = \frac{3}{36}$$

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

$$P(X=4) = P\{(4,1),(4,2),(4,3),(4,4),(3,4),(2,4),(1,4)\} = \frac{7}{36}$$

Similarly $P(X=5) = \frac{9}{36}$

and $P(X=6) = \frac{11}{36}$.

Some Results on Random Variable

Some of the fundamental results and theorems on random variables (without proof).

Consider the probability space (S, B, P), where S is the sample space, B is the σ-field of subset of S and P is the probability function on B.

1. If X_1 and X_2 are random variable and c is a constant then cX_1, X_1+X_2, X_1X_2 are also random variables.
2. If X is a random variable then

   (i) 1/X, where $\frac{1}{X}(\omega) = \infty$ if $X(\omega)=0$

   (ii) $X_+(\omega) = \max\{0, X(\omega)\}$

   (iii) $X_-(\omega) = -\min\{0, X(\omega)\}$

   (iv) $|X|$

   are random variables.
3. If X_1 and X_2 are random variables, then (i) $\max\{X_1, X_2\}$ and (ii) $\min\{X_1, X_2\}$ are random variables.
4. If X is a r.v. and f(.) is a continuous function, then f(X) is a r.v.
5. If X is a r.v. and f(.) is a increasing function , then f(X) is a r.v.

2.2 Distribution Function

Definition: Let X be a r.v., the function F defined for all real x by

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

is called the distribution function (d.f.) of the r.v. X. As X can take any real value, therefore the domain of the distribution function is the set of real number and F(x) is the probability value, therefore the range of the distribution function is [0,1].

Remark: A distribution function is also called the cumulative distribution function. Sometimes, the notation $F_X(x)$ is used to represent distribution function associated with the random variable X.

Properties of Distribution Function:

1. If F is the d.f. of the r.v. X and if a<b then

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

Proof: Using addition theorem of probability

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

$\Rightarrow P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a).$

P(a≤X≤b)=

Cor.1

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

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

Similarly,

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

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

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

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

2. If F is d.f. of one-imensional r.v. X , then

(i) $0 \leq F(x) \leq 1$

(ii) $F(x) \leq F(y)$, if $x < y$

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

3. If F is d.f. of one-dimensional r.v. X , then

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

2.3 Discrete Random Variable

If a random variable X assumes only a finite or countable set of values and for which the value which the variable takes depends on chance, it is called a discrete random variable. Here, countable number means the values which have one-to-one correspondence with the set of natural numbers.

In other words, a real-valued function defined on a discrete sample space is called a discrete random variable. In case of discrete random variable, we usually talk of values at a point. Generally, it represents counted data. For example, number of defective milk pouches in a milk plant, marks obtained in a test, number of accidents per month, number of successes in 'n' trials etc.

2.3.1 Probability Mass Function

A discrete random variable assumes each of its values with a certain probability, i.e. each possible value of the random variable has an associated probability. If X is a discrete random variable with distinct values $x_1, x_2, ..., x_n, ...$, then the function p(x) defined as:

$$p_X(x) = \begin{cases} p(X = x_i) = p_i, \text{if } x = x_i \\ 0, \quad \text{if } x \neq x_i; i = 1, 2, 3 \ldots \end{cases}$$

is called the probability mass function of r.v. X. The number $p(x_i);\ i = 1, 2, ...$ must satisfy the following conditions:

(a) $p(x_i) \geq 0 \quad \forall\, i$ i.e. all p_i'sare non-negative

(b) $\sum_{i=1}^{\infty} p(x_i) = 1, i.e.$ total probability is one

The set of values which X takes is called spectrum of the random variable. This function p_i = P (X=X_i) or p(x) is called the probability function or probability mass function (p.m.f.) of the random variable X and set of all possible ordered pairs {x, p(x)} is called the probability distribution of the random variable X.

Example 2: Toss of a coin, S = {H, T}, let X be the r.v. defined by

X(H) = 1 ; if head occurs

X(T) = 0 ; if tail occurs

If the coin is fair, the probability function is P{(H)} = $\frac{1}{2}$ and P{(T)} = $\frac{1}{2}$ and we can speak of the probability distribution of the r.v. X as

$$P(X=1) = P\{(H)\} = \frac{1}{2} \text{ and } P(X=0) = P\{(T)\} = \frac{1}{2}.$$

Example 3: A r.v. X assumes the values -2,-1, 0, 1, 2 such that $P[X=-2] = P[X=-1] = P[X=1] = P[X=2]$, and $P[X<0] = P[X=0] = P[X>0]$. Find the probability mass function.

Solution: As $P[X<0] = P[X=0] = P[X>0]$.

$\therefore P[X=-1] + P[X=-2] = P[X=0] = P[X=1] + P[X=2]$

$\Rightarrow p + p = P[X=0] = p + p$

$\Rightarrow P[X=0] = 2p$

Now, as $P[X<0] + P[X=0] + P[X>0] = 1$

$P[X=-1] + P[X=-2] + P[X=0] + P[X=1] + P[X=2] = 1$

$p + p + 2p + p + p = 1$

$6p = 1$

$p = \frac{1}{6}$

$P[X=-1] = P[X=-2] = P[X=1] = P[X=2] = \frac{1}{6}$

X	-2	-1	0	1	2
P(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

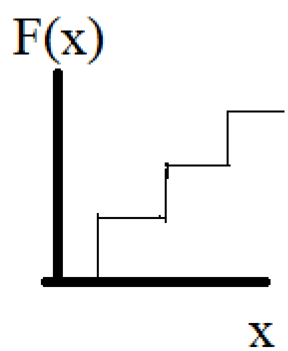
2.3.2 Discrete Distribution Function

In this case, there are a countable number of points $x_1, x_2, \ldots, x_n, \ldots$ and number $p_i \geq 0$, and $\sum_{i=1}^{\infty} p_i = 1$ such that $F(x) = \sum_{i : x_i < x} p_i$

If X takes on only a finite number of values $x_1, x_2, \ldots, x_n$, then the distribution function is given by

$$F(x) = \begin{cases} 0; & -\infty < x < x_1 \\ p(x_1); & x_1 \leq x < x_2 \\ p(x_1) + p(x_2); & x_2 \leq x < x_n \\ \ldots & \\ p(x_1) + \ldots + p(x_n); & x_n \leq x < \infty \end{cases}$$

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



Theorem: Prove that $P(x_j) = P(X=x_j) = F(x_j) - F(x_{j-1})$, where F is the d.f. of X.

Proof: Let $x_1, x_2, \ldots$ be the values of a r.v. X. Then, we have

$$F(x_j) = P(X \leq x_j) = \sum_{i=1}^{j} P(X = x_i) = \sum_{i=1}^{j} P(x_i)$$

$$F(x_{j-1}) = P(X \leq x_{j-1}) = \sum_{i=1}^{j-1} P(X = x_i) = \sum_{i=1}^{j-1} P(x_i)$$

$\Rightarrow F(x_j) - F(x_{j-1}) = p(x_j)$

Thus, given the distribution function of discrete random variable, we can compute its probability mass function.

Example 4: A r.v. X has following probability function:

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

(i) Find k

(ii) Evaluate P(X<6),P(≥6) and P(0<X<5),

(iii) $P(X\leq a) > \frac{1}{2}$, Find the minimum value of a and

(iv) Determine the distribution function of X.

Solution:

(i) Since $\sum_{x=0}^{7} p(x) = 1$, $k+2k+2k+3k+k^2+2k^2+7k^2 = 1$

$\Rightarrow 10k^2 + 9k - 1 = 0$

$\Rightarrow (10k - 1)(k + 1) = 0$

$\Rightarrow k = \frac{1}{10}$ or $k = -1$

But since p(x) cannot be negative, k=-1 is rejected. Hence $k = \frac{1}{10}$.

(ii) $P(X<6) = P(X = 0) + P(X = 1) + \ldots + P(X = 5)$

$$= \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} + \frac{1}{100} = \frac{81}{100} \Rightarrow P(X \geq 6) = 1 - P(X < 6) = 1 - \frac{81}{100} = \frac{19}{100}$$

$P(0 < X < 5) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

$$= 8k = \frac{4}{5}.$$

(iii) For $P(X \leq a) > \frac{1}{2}$. The value of a is 4

(iv) The distribution function $F_X(x)$ of X is given in the adjoining table

X	$F_X(x)$= P(X≤x)
0	0
1	k = 1/10
2	3k = 3/10
3	5k = 5/10
4	8k = 4/5
5	$8k+k^2$ = 81/100
6	$8k+3k^2$ = 83/100
7	$9k+10k^2$ = 1

Example 5: If $p(x) = \begin{cases} \frac{x}{15}, & x = 1,2,3,4,5 \\ 0, & \text{elsewhere} \end{cases}$

Find (i) P{X=1 or 2} and

(ii) $P\{\frac{1}{2} < X < \frac{5}{2} \mid X > 1\}$

Solution: (i) P{X = 1 or 2} = P(X = 1) + P(X = 2)

$$= \frac{1}{15} + \frac{2}{15} = \frac{1}{5}$$

(ii) $P\{\frac{1}{2} < X < \frac{5}{2} \mid X > 1\} =$

$$\frac{P\left\{(\frac{1}{2} < X < \frac{5}{2}) \cap (X > 1)\right\}}{P(X > 1)} = \frac{P\{(X = 1or2) \cap (X > 1)\}}{P(X > 1)}$$

$$= \frac{P(X = 2)}{1 - P(X = 1)} = \frac{\frac{2}{15}}{1 - (\frac{1}{15})} = \frac{1}{7}.$$

Problems:

1. Find the probability distribution of the number of heads when three fair coins are tossed simultaneously.
2. For the following probability distribution of a discrete r.v. X, find

X	0	1	2	3	4	5
P(x)	0	c	C	2c	3c	c

(i) the constant c

(ii) $P[X \leq 3]$ and

(iii) $P[1 < X < 4]$

3. Two dice are rolled. Let X denotes the random variable which counts the total number of points on the upturned faces. Construct a table giving the non-zero values of the probability mass function and draw the probability curve. Also, find the distribution function of X.

4. An urn contains 5 white and 4 red balls are drawn one by one with replacement. Find the probability distribution of the number of red balls.

5. Four defective milk pouches are accidentally mixed with sixteen good ones and by looking at them it is not possible to differentiate between them. Three milk pouches are drawn at random from the lot. Find the probability distribution of X, the number of defective milk pouches.

6. An experiment consists of three independent tosses of a fair coin. Let X = the number of heads,Y = the number of head runs, Z = the length of head runs, a head run being defined as consecutive occurrence of at least two heads, its length then being the number of heads occurring together in three tosses of the coin.

Find the probability function of

(i) X

(ii) Y

(iii) Z

(iv) X+Y

(v) XY

and constructed probability tables and draw their probability charts.

2.4. Continuous random variable

A random variable is said to be continuous if it can assume an infinite and uncountable set of values (integrals as well as fractional) between certain limits. A continuous random variable is in which different values cannot be put in one to one correspondence with a set of positive integers. For example, weight of calf at the age of six months might take any possible value in the interval of 160 kg to 260 kg, say 189 kg or 189.4356 kg; likewise milk yield of cows in a herd etc. In case of continuous random variable, we usually talk of values in a particular interval. Continuous random variables represent measured data.

Example: age, height, weight

2.4.1 ProbabilityDensity Function

In case of a continuous random variable X, we talk about probability in an interval (x,x+δx). If f(x) is a continuous function of x, f(x) dx gives the probability i.e. $\lim_{\delta x \to 0} \frac{P(x \le X \le x + \delta x)}{\delta x}$,

that the random variable X, takes value in a small interval of magnitude dx i.e. $\left(x - \frac{dx}{2}\right)$ and $\left(x + \frac{dx}{2}\right)$, then f(x) is called the probability density function (p.d.f.) of a random variable X. It is also known as frequency function because it also gives the proportion of units lying in theinterval $\left(x - \frac{dx}{2}\right)$ and $\left(x + \frac{dx}{2}\right)$. If x has range $[\alpha, \beta]$, $f(x) \ge 0 \ \forall \ x \in [\alpha, \beta]$ and $\int_{\alpha}^{\beta} f(x)dx = 1$

The expression f(x) dx usually written as dF(x) is known as the probability differential and the curve y= f(x) is known as the probability density curve.

Remarks: 1. The probability density function has the same property as the probability mass function. So $f(x) \ge 0$ and sum of the probabilities of all possible values that random variable can take, has to be 1.

2. An essential property of a continuous r.v. is that there is zero probability that it takes any specified numerical values, but the probability that it takes a value in specified intervals is non zero and is calculable as a definite integral of the probability density function of the r.v. and hence the probability that a continuous r.v. X will lie between two values a and b is given by

$$P(a<X<b) = \int_{a}^{b} f(x)dx$$

It is to be noted that in case of discrete r.v., the probability at a point can be zero while in case of continuous r.v., it is zero.

Properties: p.d.f. of a r.v. X, usually denoted by $f_X(x)$ or simply by f(x) has the following properties

(i) $f(x) \ge 0$

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

(iii) The probability P(E) given by P(E) = $\int_{E} f(x)$ is well defined for any event E.

2.4.2 Various Measures of Central Tendency, Dispersion, Skewness and Kurtosis for Continuous Probability Distribution

The formulae for these measures in case of discrete frequency distribution can be easily extended to the case of continuous probability distribution by simply replacing $P_i = \frac{f_i}{N}$ by f(x)dx , x_i by x and the summation over 'i' by integration over the specified range of the variable X.

Let $f_X(x)$ or $f(x)$ be the p.d.f of a r.v. X, where X is defined from a to b. Then

(i) Arithmetic Mean $= \int_a^b xf(x)$ (3)

(ii) Harmonic mean H is given by :

$$\frac{1}{H} = \int_a^b \frac{1}{x} f(x)dx \quad (4)$$

(iii) Geometric Mean G is given by:

$$= \log G = \int_a^b \log x.f(x)dx \quad (5)$$

(iv) Moment about origin, $\mu'_r = \int_a^b x^r f(x)dx$ (6)

Moment about the point X = A, $\mu'_r = \int_a^b (x-A)^r f(x)dx$ (6a)

Moment about Mean, $\mu_r = \int_a^b (x-\text{mean})^r f(x)dx$ (6b)

In particular from (3) and (6)

$$\mu'_1 = \text{Mean} = \int_a^b xf(x)dx$$

$$\mu'_2 = \int_a^b x^2 f(x)dx$$

Hence $\mu_2 = \mu'_2 - \mu'^2_2 = \int_a^b x^2 f(x)dx - \left(\int_a^b xf(x)dx\right)^2$ (6c)

From (6) on putting r = 3 and 4 respectively, we get the values of μ'_3 and μ'_4. Then third and fourth order moment is given by:

$$\left.\begin{aligned} \Rightarrow \mu_3 &= \mu'_3 - 3\mu'_2\,\mu'_1 + 2\mu'^3_3 \\ \Rightarrow \mu_4 &= \mu'_4 - 4\mu'_3\,\mu'_1 + 6\mu'^2_1\,\mu'_2 - 3\mu'^4_1 \end{aligned}\right\} \quad (6d)$$

(v) Median:

If M is the median, then $\int_a^M f(x)dx = \int_M^b f(x)dx = \frac{1}{2}$

Thus solving for M, we get the value of median.

(vi) Mean Deviation:

Mean deviation about the mean μ'_1 is given by

$$\text{M.D.} = \int_a^b |x - \text{mean}| f(x)dx$$

In general, mean deviation about an average 'A' is given b:

$$\text{M.D. about 'A'} = \int_a^b |x - A| f(x)dx$$

(vii) Quartiles and Deciles: Q_1 and Q_3 are given by the equations

$$\int_a^{Q1} f(x)dx = \frac{1}{4} \text{ and}$$

$$\int_a^{Q3} f(x)dx = \frac{3}{4}$$

$D_i = i^{th}$ decile is given by:

$$\int_a^{D_i} f(x)dx = \frac{i}{10}; i = 1, 2, ..., 9$$

(viii) Mode: Mode is the value of x for which f(x) is maximum. Mode is thus the solution of $f'(x) = 0$, $f''(x) < 0$ provided it lies in [a, b].

Example 6: The diameter of an electric cable, say X is assumed to be a continuous random variable with p.d.f.:

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

(i) Check that f(x) is p.d.f. and

(ii) Determine a number b s.t. $P(X < b) = P(X > b)$

Solution: (i) Obviously, for $0 \le x \le 1$, $f(x) \ge 0$

Now

$$\int_0^1 f(x)dx = 6\int_0^1 x(1-x)dx = 6\int_0^1 (x-x^2)dx = 1$$

Hence f(x) is the p.d.f. of r.v. X.

(ii) P(X<b) = P(X>b) $\Rightarrow \int_0^b f(x)dx = \int_b^1 f(x)dx$

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

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

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

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

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

$\therefore$ 2b-1 = 0 $\Rightarrow b = 1/2$

or $(2b^2-2b-1) = 0 \Rightarrow b = \dfrac{2 \pm \sqrt{4+8}}{4} = \dfrac{1 \pm \sqrt{3}}{2}$

Hence b = 1/2 is the only real value lying between 0 and 1.

Example 7: Calculate the standard deviation and mean deviation from mean if the frequency function f(x) has the form:

$$f(x) = \begin{cases} \dfrac{3+2x}{18}, & \text{for } 2 \le x \le 4 \\ 0, & \text{otherwise} \end{cases}$$

Solution: We have,

Mean =

$$\mu'_1 = \int_{-\infty}^{\infty} xf(x)dx = \int_2^4 x\left(\frac{3+2x}{18}\right)dx = \frac{83}{27}$$

$$\Rightarrow \mu'_2 = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_2^4 x^2\left(\frac{3+2x}{18}\right)dx = \frac{88}{9}$$

$\therefore$ Variance = $\mu'_2 - \mu'^2_1 = \left\{\frac{88}{9} - \left(\frac{83}{27}\right)^2\right\} = \frac{239}{729}$

$$\Rightarrow \text{S.D.} = \sqrt{\frac{239}{729}} = 0.57$$

Mean Deviation =

$$\int_{-\infty}^{\infty} |x - \bar{x}| f(x)dx = \int_2^4 \left|x - \frac{83}{27}\right| \left(\frac{3+2x}{18}\right)dx$$

$$= \int_2^{83/27} \left|\frac{83}{27} - x\right| \left(\frac{3+2x}{18}\right)dx + \int_{83/27}^{4} \left|x - \frac{83}{27}\right| \left(\frac{3+2x}{18}\right)dx = 0.49$$

(on simplification)

Hence mean deviation and standard deviation are 0.49 and 0.57 respectively.

Problems:

1. A continuous r.v. X has a p.d.ff(x) = $3x^2, 0<x\leq 1$, find a and b such that

(i) $P(X\leq a) = P(X>a)$

(ii) $P(X>b) = 0.05$

2. Suppose that the life in hours of a certain part of radio tube is a continuous r.v. X with p.d.f. given by:

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

(i) What is the probability that none of three such tubes in a given radio set will have to be replaced during the first 150 hours of operations?

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

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

3. The kms X in thousands of kms which car owners get with a certain kind of tyre is a random variable having probability density function:

$$\mathbf{f}(x) = \begin{cases} \frac{1}{20} e^{-x/20}, & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Find the probability that one of these tyres will last

(i) atmost 10,000 kms

(ii) anywhere from 16,000 to 24,000 kms

(iii) atleast 30,000 kms.

2.4.3 Continuous Distribution Function

Definition: If X is a continuous random variable with the p.d.f. f(x), then the function

$$F_X(x) = P[X \le x] = \int_{-\infty}^{x} f(t)dt, \quad -\infty < x < \infty$$

is called the continuous distribution function (d.f.) or sometimes the cumulative distribution function (c.d.f.) of the random variable X.

As we know

$$f(x) = \lim_{\delta x \to 0} \frac{P[x \le X \le x + \delta x]}{\delta x}$$

$$= \lim_{\delta x \to 0} \frac{P[X \le x + \delta x] - P[X \le x]}{\delta x}$$

$$= \lim_{\delta x \to 0} \frac{F(x + \delta x) - F(x)}{\delta x}$$

$$\Rightarrow f(x) = F'(x)$$

$$dF = f(x)dx$$

Here, dF is known as the probability differential.

So, $F(x) = \int_{-\infty}^{x} f(x)dx.$

Properties:

(i) $0 \le F(x) \le 1, -\infty < x < \infty$

(ii) F(x) is non-decreasing function of x

(iii) $F(-\infty) = \lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{-\infty} f(x)dx = 0$

$$F(\infty) = \lim_{x \to +\infty} F(x) = \lim_{x \to \infty} \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1$$

$$\Rightarrow 0 \le F(x) \le 1$$

(iv) F(x) is a continuous function of x from the right.

(v) The points of discontinuity of F(x) are at the most countable.

(vi) $P(a \le X \le b) = \int_a^b f(x)dx = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = P(X \le b) - P(X \le a) = F(b) - F(a)$

(vii) $$F'(x) = \frac{dF}{dx} = f(x) = dF(x) = f(x)dx$$

dF(x) is known as probability differential of X.

Example 8: Verify that the following is a distribution function:

$$F(x) = \begin{cases} 0, & x < -a \\ \frac{1}{2}(\frac{x}{a}+1), & -a \le x \le a \\ 1, & x > a \end{cases}$$

Solution: Obviously the properties (i), (ii), (iii) and (iv) are satisfied. Also we observe that F(x) is continuous at x=a and x= -a as well.

Now

$$\frac{d}{dx}F(x) = \begin{cases} \frac{1}{2a}, & -a \le x \le a \\ 0, & \text{otherwise} \end{cases}$$

$= f(x)$

In order that F(x) is a distribution function, f(x) must be a p.d.f. Thus we have to show that

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

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-a}^{a} f(x)dx = \frac{1}{2a}\int_{-a}^{a} 1.dx = 1$$

HenceF(x) is a distribution function.

Example 9: The diameter, says X of an electric cable, is assumed to be a continuous r.v. with p.d.f.:

$f(x) = 6x(1-x) \quad ; \ 0 \le x \le 1$

(i) Obtained an expression for the c.d.f. of X.

(ii) Compute $P\left(X \le \frac{1}{2} \middle| \frac{1}{3} \le X \le \frac{2}{3}\right)$.

Solution:

(i) $$F(x) = \begin{cases} 0, & \text{if } x \le 0 \\ \int_0^x 6t(1-t)dt = (3x^2 - 2x^3), & 0 \le x \le 1 \\ 1, & \text{if } x > 1 \end{cases}$$

(ii) $$P(X \le \frac{1}{2} | \frac{1}{3} \le X \le \frac{2}{3}) = \frac{P(\frac{1}{3} \le X \le \frac{1}{2})}{P(\frac{1}{3} \le X \le \frac{2}{3})} = \frac{\int_{1/3}^{1/2} 6x(1-x)dx}{\int_{1/3}^{2/3} 6x(1-x)dx} = \frac{11/54}{13/27} = \frac{11}{26}.$$

Problems:

1. Let X be a continuous r.v. with p.d.f. given by:

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

(i) Determine the constant k

(ii) Determine F(x), the c.d.f. and

(iii) If X_1, X_2 and X_3 are three independent observations from X, what is the probability that exactly one of these three numbers is larger than 1.5?

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

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

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

3. Let X be a continuous r.v. with p.d.f. given by:

$$f(x) = \begin{cases} \frac{x}{2}; & 0 \le x < 1 \\ \frac{1}{2}; & 1 \le x < 2 \\ \frac{1}{2}(3-x); & 2 \le x < 3 \\ 0; & \textit{elsewhere} \end{cases}$$

Determine F(x).

2.5 Two-Dimensional Random Variable

Let S be a sample space associated with a random experiment E. Let X and Y be two random variables defined on S, then the pair (X,Y) is called a two-dimensional random variable. The value of (X,Y) at a point is given by the ordered pair of real numbers (X(s), Y(s)) = (x, y) where X(s) = x, Y(s) = y.

The probability of the event $\{X \le a, Y \le b\}$ will be denoted by $P(X \le a, Y \le b)$

Let $A = \{a < X \le b\}$

$B = \{c < Y \le d\}$ be two event, then the event

$\{a < X \le b, c < Y \le d\} = \{a < X \le b\} \cap \{c < Y \le d\}$

$= A \cap B$

$\therefore \quad P\{a < X \le b, c < Y \le d\} = P(A \cap B)$

Remarks:

1. A two-dimensional r.v. is called discrete if it takes at most a countable number of points in R^2
2. A two-dimensional r.v. is called continuous if it takes infinite or uncountable number of points in R^2.

2.5.1 Joint Probability Mass Function of Two-Dimensional Random Variable

Let X and Y be r.v. on a sample space S with respective image sets

$X(S) = (x_1, x_2, x_3, ..., x_n)$ and $Y(S) = (y_1, y_2, y_3, ..., y_m)$

We make the product set

$X(S) \times Y(S) = \{X_1, X_2, X_3, ..., X_n\} \times \{Y_1, Y_2, Y_3, ..., Y_m\}$ into a probability space.

The function p on $X(S) \times Y(S)$ defined as

$$p_{ij} = P\left(X = x_i \cap Y = y_j\right) = p(x_i, y_j)$$

is called the joint probability function of X and Y and is usually represented in the form of the following table:

X \ Y	y_1 y_2 ………………… y_j …………………… y_m	Total
x_1 x_2 .. x_i . . . x_n	p_{11} p_{12} ……………………… p_{1j} …………… p_{1m} p_{21} p_{22} ………… p_{2j} …………… p_{2m} . . p_{i1} p_{i2} …………… p_{ij} ……………… p_{im} . . p_{n1} p_{n2} ………… p_{nj} ……………… p_{nm}	$p_{1\cdot}$ $p_{2\cdot}$ . . $p_{i\cdot}$ . . $p_{n\cdot}$
Total	$p_{\cdot 1}$ $p_{\cdot 2}$ ………………… $p_{\cdot j}$ ……………… $p_{\cdot m}$	1

Definition: If (X, Y) is a two-dimensional discrete r.v., then the joint discrete function of (X, Y) also called the joint probability mass function of (X, Y) denoted by p_{XY} is defined as:

$p_{XY}(x_i, y_j) = P(X = x_i, Y = y_j)$; for a value (x_i, y_j) of (X,Y) and

$p_{XY}(x_i, y_j) = 0$; otherwise

and satisfying the following properties:

1. $p(x_i, y_j) \geq 0$

2. It may be noted that $\sum \sum p_{XY}(x_i, y_j) = 1$

(a) Marginal Probability Function

Let (X,Y) be a discrete two-dimensional r.v. Then the probability distribution of X is determined as follows:

$$p_X(x_i) = P(X = x_i)$$
$$= P(X = x_i \cap Y = y_1) + P(X = x_i \cap Y = y_2) + \ldots + P(X = x_i \cap Y = y_m)$$
$$= p_{i1} + p_{i2} + p_{i3} + \ldots + p_{im} = \sum_{j=1}^{m} p_{ij} = \sum_{j=1}^{m} p_{XY}(x_i, y_j) = p_{i.}$$

and is known as marginal probability mass function or discrete marginal density function of X.

Also

$$\sum_{i=1}^{n} p_{i.} = p_{1.} + p_{2.} + p_{3.} + \cdots + p_{m.} = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij}(x_i, y_j) = 1$$

Similarly,

$$p_Y(y_j) = P(Y = y_j) = \sum_{i=1}^{n} p_{ij} = \sum_{i=1}^{n} p(x_i, y_j) = p_{.j}$$

which is the marginal probability mass function of Y.

(b) Conditional Probability Function

Definition: Let (X,Y) be a discrete two-dimensional r.v. Then, the conditional probability density function or the conditional probability mass function of X, given Y = y denoted by $p_{X|Y}(x|y)$ is defined as:

$$p_{X|Y}(x|y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$, provided $P(Y=y) \neq 0$

A necessary and sufficient condition for the discrete r.v. X and Y to be independent is:

$P(X= x_i,Y= y_j) = P(X= x_i)P(Y= y_j)$ for all values (x_i, y_j) of (X,Y).

Example 10: The following table represents the joint probability distribution of the discrete r.v. (X,Y).

X \ Y	1	2
1	0.1	0.2
2	0.1	0.3
3	0.2	0.1

Find:

(i) The marginal distribution.

(ii) The conditional distribution of X given Y=1

(iii) $P[(X+Y)<4]$

Solution: (i) To find the marginal distribution, we have to find the marginal totals,

X \ Y	1	2	p(x) (Totals)
1	0.1	0.2	0.3
2	0.1	0.3	0.4
3	0.2	0.1	0.3
P(y)	0.4	0.6	1

Thus the marginal probability distribution of X is

X	1	2	3
p(x)	0.3	0.4	0.3

and the marginal probability distribution of Y is

Y	1	2
p(x)	0.4	0.6

(i) As $P[X=1|Y=1]=\dfrac{P[X=1,Y=1]}{P[Y=1]}=\dfrac{0.1}{0.4}=\dfrac{1}{4}$

$$P[X=2|Y=1]=\frac{P[X=2,Y=1]}{P[Y=1]}=\frac{0.1}{0.4}=\frac{1}{4}$$

$$P[X=3|Y=1]=\frac{P[X=3,Y=1]}{P[Y=1]}=\frac{0.2}{0.4}=\frac{1}{2}$$

The conditional distribution of X given Y=1 is

X	1	2	3
P[X=x|Y=1]	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

(ii) As the values of (X,Y) which satisfy X+Y<4 are (1,1),(1,2) and (2,1) only.

$$\therefore P[(X+Y)<4]=P[X=1,Y=1]+P[X=1,Y=2]+P[X=2,Y=1]$$
$$=0.1+0.2+0.1=0.4$$

Problem:

1. Two discrete random variables X and Y have $P[X=0,Y=0]=\frac{2}{9}$, $P[X=0,Y=1]=\frac{1}{9}$, $P[X=1,Y=0]=\frac{1}{9}$, and $P[X=1,Y=1]=\frac{5}{9}$. Examine whether X and Y are independent?

2.5.2 Joint Density Function

From the joint distribution function $F_{XY}(x,y)$ of two-dimensional continuous r.v., we get the joint probability density function by differentiation as follow:

$$f_{XY}(x,y)=\frac{\partial^2 F(x,y)}{\partial x \partial y}=\lim_{\substack{\delta x\to 0\\ \delta y\to 0}}\frac{P(x\le X\le x+\delta x, y\le Y\le y+\delta y)}{\delta x\delta y}$$

We will discuss the two-dimensional distribution function later.

(a) Marginal Density Function

The marginal density function of X and Y can be obtained in the following manner also:

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

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

(b) Conditional Probability Density Function

The conditional probability density function of Y given X, for two r.v. X and Y which are jointly continuously distributed is defined as follows , for two real numbers x and y:

$$f_{Y|X}(y\,|\,x)=\frac{\partial F_{Y|X}(y\,|\,x)}{\partial y}$$

Remark: If we know the joint p.d.f. of two-dimensional random variable, we can find individual density function but converse is not true.

2.5.3 Two-Dimensional Distribution Function

The distribution function of the two-dimensional r.v. (X,Y) is real valued function F defined for all x and y by the relation:

$F_{XY}(x, y) = P(X \le x, Y \le y)$

Properties of Joint Distribution Function

(i) $F(-\infty, y) = 0 = F(x, -\infty), F(\infty, \infty) = 1$

(ii) If the density function $f(x, y)$ is continuous, $\frac{\partial^2 F}{\partial x \partial y} = f(x, y)$.

where $f(x,y)$ is the joint density function which we will discuss later.

(a) Marginal Distribution Function

Marginal distribution function of X and Y w.r.t the joint distribution function $F_{XY}(x, y)$

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

Similarly

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

where $F_X(x)$ is termed as marginal distribution function of X and $F_Y(y)$ is called marginal distribution function of Y w.r.t. the joint marginal function $F_{XY}(x,y)$.

In the case of jointly discrete random variable the marginal distribution function are given as

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

$$F_Y(y) = \sum_x P(X = x, Y \le y)$$

Similarly in the case of jointly continuous r.v., the marginal distribution functions are given as

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

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

(b) Conditional Distribution Function

The joint distribution function $F_{XY}(x,y)$ for any real number x and y is given by:

$F_{XY}(x,y) = P(X \le x, Y \le y)$

Using conditional probabilities we may now write

$$F_{XY}(x,y) = \int_{-\infty}^{x} P(A|X=x)\,dF_X(x)$$

where A be the event ($Y \le y$) s.t. the event A is said to occur when Y assumes values upto and inclusive of y.

The conditional distribution function $F_{X|Y}(x \mid y) = P(Y \le y \mid X = x)$

Also

$$F_{XY}(x,y) = \int_{-\infty}^{x} F_{Y|X}(y \mid x)\,dF_X(x)$$

Similarly

$$F_{XY}(x,y) = \int_{-\infty}^{y} F_{X|Y}(x \mid y)\,dF_Y(y)$$

2.5.4 Independent Random Variables

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

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

2.5.5 Generalization to n-Dimensional Random Variable

(A) Joint and Marginal Probability Mass Function

(I) The joint p.m.f. of $(X_1, X_2, \ldots, X_n)$ is defined as

$$p_{X_1, X_2 \ldots X_n}(x_1, x_2, \ldots x_n) = P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n]$$

$$= P[\bigcap_{i=1}^{n} (X_i = x_i)]$$

where

(i) $p(x_1, x_2, x_3, \ldots, x_n) \ge 0 \quad \forall \quad (x_1, x_2, \ldots x_n) \in R^n$ and

(ii) $\sum_{(x_1, x_2, x_3, \ldots, x_n)} p(x_1, x_2, x_3, \ldots, x_n) = 1$

(II) The marginal p.m.f. of any r.v. say X_i is

$$p_{X_i}(x_i) = \sum_{\substack{x_1, x_2, x_3, \ldots, x_n \\ \text{except } x_i}} p\left(x_1, x_2, x_3, \ldots, x_n\right)$$

(B) Joint and Marginal Probability Density Function

(I) The joint p.d.f. of (X_1, X_2, X_3,…, X_n) is given by

$$f_{X_1, X_2, X_3, \ldots, X_n}\left(x_1, x_2, x_3, \ldots, x_n\right) = \lim_{\substack{dx_1 \to 0 \\ dx_2 \to 0 \\ \vdots \\ dx_n \to 0}} P\left[\bigcap_{i=1}^{n}\left(x_i < X_i < x_i + dx_i\right)\right]$$

where

(i) $f(x_1, x_2, x_3, \ldots, x_n) \geq 0 \quad \forall \quad (x_1, x_2, \ldots x_n) \in R^n$

(ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots x_n)\, dx_1 dx_2 \ldots dx_n = 1$

(II) The marginal p.d.f. of any variable says X_i (integral over all variables except x_i)

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$$

Example 11: A random observation on a bivariate population (X,Y) can yield one of the following pairs of values with probabilities noted against them:

For each observation pair	Probability
(1,1);(2,1);(3,3);(4,3)	1/20
(3,1);(4,1);(1,2);(2,2);(3,2);(4,2);(1,3);(2,3)	1/10

Find the probability that Y=2 given that X=4. Also find the probability that Y=2. Examine if the two events X=4 and Y=2 are independent.

Solution: $P(Y=2) = P\{(1,2) \cup (2,2) \cup (3,2) \cup (4,2)\} = 4/10 = 2/5$

$$P(X=4) = P\{(4,1) \cup (4,2) \cup (4,3)\} = \frac{1}{10} + \frac{1}{10} + \frac{1}{20} = \frac{1}{4}$$

P(X=4,Y=2) = P{(4,2)} = 1/10

$$P(Y=2\,|\,X=4)=\frac{P(X=4\cap Y=2)}{P(X=4)}=\frac{1/10}{1/4}=\frac{2}{5}$$

Now P(X=4).P(Y=2) = $\frac{1}{4}.\frac{2}{5}=\frac{1}{10}=P(X=4\cap Y=2)$

Hence the events X=4 and Y=2 are independent.

Example 12: For the adjoining bivariate probability distribution of X and Y.

X \ Y	1	2	3	4	5	6
0	0	0	1/32	2/32	2/32	3/32
1	1/16	1/16	1/8	1/8	1/8	1/8
2	1/32	1/32	1/64	1/64	0	2/64

(i)P(X≤1,Y=2)

(ii)P(X≤1)

(iii) P(Y≤3)

(iv) P(X<3,Y≤4)

Solution: The marginal distributions are given below:

X \ Y	1	2	3	4	5	6	$P_X(x)$
0	0	0	1/32	2/32	2/32	3/32	8/32
1	1/16	1/16	1/8	1/8	1/8	1/8	10/16
2	1/32	1/32	1/64	1/64	0	2/64	8/64
$P_Y(y)$	3/32	3/32	11/64	13/64	6/32	16/64	$\sum p(x)=1$ $\sum p(y)=1$

(i) P(X≤1,Y=2) = P(X=0,Y=2)+ P(X=1,Y=2)

$$=0+\frac{1}{16}=\frac{1}{16}.$$

(ii) $P(X \le 1) = P(X=0) + P(X=1)$

$$= \frac{8}{32} + \frac{10}{16} = \frac{7}{8}$$

(iii) $P(Y \le 3) = P(Y=1) + P(Y=2) + P(Y=3)$

$$= \frac{3}{32} + \frac{3}{32} + \frac{11}{64} = \frac{23}{64}$$

(iv) $P(X<3, Y \le 4) = P(X=0, Y \le 4) + P(X=1, Y \le 4) + P(X=2, Y \le 4)$

$$= \left(\frac{1}{32} + \frac{2}{32}\right) + \left(\frac{1}{16} + \frac{1}{16} + \frac{1}{8} + \frac{1}{8}\right) + \left(\frac{1}{32} + \frac{1}{32} + \frac{1}{64} + \frac{1}{64}\right) = \frac{9}{64}.$$

Example 13: Suppose that two-dimensional continuous r.v. (X,Y) has joint p.d.f given by :

$$f(x,y) = \begin{cases} 6x^2 y, 0 < x < 1, 0 < y < 1 \\ 0 \quad , \quad \text{elsewhere} \end{cases}$$

(i)Verify that $\int_0^1 \int_0^1 f(x,y)dxdy = 1$

(ii) Find $P\left(0 < X < \frac{3}{4}, \frac{1}{3} < Y < 2\right), P(X+Y<1), P(X>Y)$ and $P(X<1 | Y<2)$

Solution: (i) $\int_0^1 \int_0^1 f(x,y)dxdy = 6\int_0^1 \int_0^1 x^2 ydxdy = 6\int_0^1 x^2 \left|\frac{y^2}{2}\right|_0^1 dx = 3\int_0^1 x^2 dx = \left|x^3\right|_0^1 = 1$

(ii) $P\left(0 < X < \frac{3}{4}, \frac{1}{3} < Y < 2\right) =$

$$6\int_0^{3/4} \int_{1/3}^{1} x^2 ydxdy + \int_0^{3/4} \int_1^2 0dxdy = 6\int_0^{3/4} x^2 \left|\frac{y^2}{2}\right|_{1/3}^1 dx = \frac{8}{9}\int_0^{3/4} 3x^2 dx = \frac{8}{9}\left|x^3\right|_0^{3/4} = \frac{3}{8}$$

$$P(X+Y<1) = 6\int_0^1 \int_0^{1-x} x^2 ydxdy = 6\int_0^1 x^2 \left.\frac{y^2}{2}\right|_0^{1-x} dx$$

$$= \int_0^1 3x^2(1-x)^2 dx = \frac{1}{10}$$

P(X>Y) =

$$6\int_0^1\int_0^x x^2 y\,dx\,dy = \int_0^1 x^2 \left| y^2 \right|_0^x dx = \int_0^1 3x^4 dx = \frac{3}{5}$$

$$P\left(X < 1 \middle| Y < 2\right) = \frac{P\left(X < 1 \cap Y < 2\right)}{P(Y < 2)}$$

where $P\left(X < 1 \cap Y < 2\right) = 6\int_0^1\int_0^1 x^2 y\,dx\,dy + \int_0^1\int_1^2 0\,dx\,dy = 1$

and

$$P\left(Y < 2\right) = \int_0^1\int_0^2 f(x, y)\,dx\,dy$$

$$= 6\int_0^1\int_0^1 x^2 y\,dx\,dy + \int_0^1\int_1^2 0.dx\,dy = 1$$

$$\therefore P\left(X < 1 \middle| Y < 2\right) = \frac{P\left(X < 1 \cap Y < 2\right)}{P(Y < 2)} = 1$$

Example 14: Let joint distribution of X and Y is given by:

$$f(x, y) = 4xye^{-\left(x^2+y^2\right)} \quad ; x\geq 0, y\geq 0$$

Test whether X and Y are independent. For the above joint distribution, find the conditional density of X given Y=y.

Solution: Given joint p.d.f. of X and Y is:

$$f(x, y) = 4xye^{-\left(x^2+y^2\right)} \quad ; x\geq 0, y\geq 0$$

Marginal density of X is given by:

$$f_X(x) = \int_0^\infty f_{XY}(x,y)dy = \int_0^\infty 4xye^{-\left(x^2+y^2\right)}dy$$

$$= 4xe^{-x^2}\int_0^\infty ye^{-y^2}dy$$

$$= 4xe^{-x^2}\int_0^\infty ye^{-t}\frac{dt}{2}$$

$$= 2xe^{-x^2}\left|-e^{-t}\right|_0^\infty$$

$$\therefore f_X(x) = 2xe^{-x^2};x \geq 0$$

Similarly, the marginal p.d.f. of sY is given by

$$f_Y(y) = \int_0^\infty f_{XY}(x,y)dx$$

$$= 2ye^{-y^2};y \geq 0$$

Since $f_{XY}(x,y) = f_X(x).f_Y(y)$, X and Y are independently distributed .

The conditional distribution of X for given Y is given by:

$$f_{XY}\left(X = x \middle| Y = y\right) = \frac{f(x,y)}{f_Y(y)} = 2xe^{-x^2};x \geq 0$$

Example 15: If the joint distribution of X and Y is given by:

$$F_{XY}(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y)} & ; x > 0, y > 0 \\ 0 & ;\ \text{elsewhere} \end{cases}$$

(a) Find the marginal densities of X and Y

(b) Are X and Y independent?

(c) Find $P\left(X \leq 1 \cap Y \leq 1\right)$ and $P\left(X + Y \leq 1\right)$

Solution:

(a) The joint p.d.f. of the r.v.'s (X,Y) is given by

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y} = \frac{\partial}{\partial x}\left[e^{-y} - e^{-(x+y)}\right] \qquad (1)$$

$$= \begin{cases} e^{-(x+y)}; x \geq 0, y \geq 0 \\ 0 \quad ; \text{ otherwise} \end{cases}$$

We have $f_{XY}(x,y) = e^{-x} \cdot e^{-y} = f_X(x) f_Y(y)$ (2)

where

$$\left. \begin{aligned} f_X(x) &= e^{-x}, x \geq 0 \\ f_Y(y) &= e^{-y}, y \geq 0 \end{aligned} \right\} \quad (3)$$

(b) From (2) X and Y are independent and from (3) we get marginal p.d.f's of X and Y.

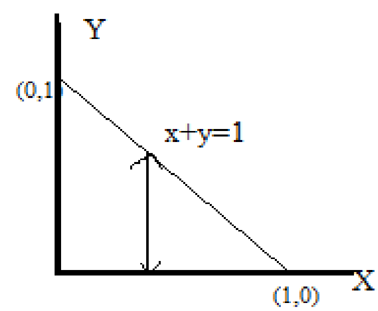
(c) $P(X \leq 1 \cap Y \leq 1) = \int_0^1 \int_0^1 f(x,y) dx dy$

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

$$P(X + Y \leq 1) = \int_{x+y \leq 1} \int f(x,y) dx dy$$

$$= \int_0^1 \left\{ \int_0^{1-x} f(x,y) dy \right\} dx = \int_0^1 \left\{ \int_0^{1-x} e^{-(x+y)} dy \right\} dx$$

$$= \int_0^1 e^{-x} \left\{ 1 - e^{-(1-x)} dy \right\} dx = 1 - 2e^{-1}$$



Problems:

1. The joint probability distribution of random variables X and Y is given by:

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

Find

(i) Marginal distribution of X and Y and

(ii) The conditional probability distribution of X given Y=1

2. The joint p.d.f. of two r.v.'s X and Y is given by:

$$f(x,y) = \frac{9(1+x+y)}{2(1+x)^4(1+y)^4} \quad ; \quad 0 \leq x < \infty, 0 \leq y < \infty$$

(i) Find Marginal distribution of X and Y and

(ii) The conditional probability distribution of Y for X=x.

3. Given

$$f(x,y) = e^{-(x+y)}; \quad x \in [0,\infty) \text{ and } y \in [0,\infty)$$

Are X and Y independent?

Find

(i) P(X>1)

(ii) P(X<Y|X<2Y)

(iii) P(1<X+Y<2)

4. A two-dimensional random variable *(X,Y)* have the joint density

$$f(x,y) = \begin{cases} 8xy, & 0 < x < y < 1 \\ 0, & otherwise \end{cases}$$

(i) Find $P\left[X < \frac{1}{2} \cap Y < \frac{1}{4}\right]$

(ii) Find the marginal and conditional distribution

(iii) Are X and Y independent?

2.6 Mathematical Expectation

A very important concept in probability and statistics is that of the mathematical expectation, expected value, or briefly the expectation, of a random variable. Let X denotes a discrete random variable which assumes values $x_1, x_2, x_3, ..., x_n$ with corresponding probabilities $p_1, p_2, p_3, ..., p_n$ where $p_1 + p_2 + p_3 + ... + p_n = 1$, the mathematical expectation ofX or simply the expectation of X, denoted by E(X) is defined as:

$$E[X] = \mu = \sum_{i=1}^{n} x_1 p_1 + x_2 p_2 + x_3 p_3 + ... + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

i.e. it is the sum of the product of different possible values of x and the corresponding probabilities.

The mathematical expression for computing continuous r.v. X with probability density function (p.d.f.) f(x) is, however as follow

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

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

Hence, mathematical expectation of a random variable is equal to its arithmetic mean. The mean, or expectation of X gives a single value that acts as a representative or average of the values of X,and for this reason, it is often called a measure of central tendency. It is also denoted by μ .

Example 16: Find the expectation of the number on an unbiased die when thrown.

Solution: Let X be the random variable representing the number on a die when thrown.

Therefore X can take the values 1,2,3,4,5,6, with

$$P[X=1]=P[X=2]=P[X=3]=P[X=4]=P[X=5]=P[X=6]=\frac{1}{6}$$

Thus, the probability distribution of X is given by

X:	1	2	3	4	5	6
p(x):	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Hence, the expectation of number on the die is

$$E(X)=\sum_{i=1}^{6}x_i p_i=1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=\frac{21}{6}=\frac{7}{2}.$$

Problems:

1. Two cards are drawn successively with replacement from a well-shuffled pack of 52 cards. Find the expected value for the number of aces.
2. The distribution of a continuous random variable X is defined by

   $$f(x)=\begin{cases} x^3; & 0<x<1 \\ (2-x)^3; & 1<x<2 \\ 0; & elsewhere \end{cases}$$

   Obtain the expected value of X.
3. A fair coin is tossed until a tail appears. What is the expectation of the number of tosses?

Expected Value of Function of a Random Variable

Consider ar.v. X with p.d.f. (p.m.f.) f(x) and distribution function F(x). If g(.) is a function of r.v. X, then we define

$$E[g(X)]=\int_{-\infty}^{\infty} g(x)f(x)dx \quad \text{(for continuous r.v.)} \qquad (1)$$

$$E[g(X)]=\sum_{x} g(x)f(x) \quad \text{(for discrete r.v.)}$$

Note: For higher dimensions, the above results are $E(X)=\int_{-\infty}^{\infty} xf_X(x)dx$ and $E(Y)=\int_{-\infty}^{\infty} yf_Y(y)dy$

where f_X and f_Y are marginal density function. If X and Y have a joint p.d.f. f(x, y) and Z = h(x, y) is a r.v. for some function h and if E(Z) exists then

$$E(Z)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)f(x,y)dxdy \qquad (3)$$

$$E(Z)=\sum_x\sum_y h(x,y)f(x,y) \qquad (3a)$$

Particular Case:

1. If we take $g(X)=X^r$, r being a positive integer in eq(1)

   $$E\left(X^r\right)=\int_{-\infty}^{\infty} x^r f(x)dx=\mu'_r \quad \text{(about origin)}$$

   where μ'_r, the r[th] moment of the probability distribution

   μ'_1 (about origin) = E(X) = mean or $\bar{x}$ (4)

   μ'_2 (about origin) = E(X^2) and

   $$\mu_2=\mu'_2-\left(\mu'_1\right)^2=E(X^2)-\{E(X)\}^2 \quad (4a)$$

2. If g(X) = $[X-E(X)]^r$ = $(X-\overline{X})^r$ then from (1)

   $$E\left(X-E(X)\right)^r=\int_{-\infty}^{\infty}\left(x-E(X)\right)^r f(x)dx=\int_{-\infty}^{\infty}\left(x-\overline{x}\right)^r f(x)dx$$

   which is μ_r, the r[th] moment about mean put r=2

   $$\mu_2=E\left(X-E(X)\right)^2=\int_{-\infty}^{\infty}\left(x-\overline{x}\right)^r f(x)dx$$

3. Taking g(x) = constant = c in (1)

   E(c) = $E(c)=\int_{-\infty}^{\infty} cf(x)dx=c\int_{-\infty}^{\infty} f(x)dx=c$

   E(c) = c

Properties of Expectation

Property 1: Addition Theorem of Expectation

If X and Y are random variables, then E(X+Y) = E(X) + E(Y), provided all the expectations exist.

Proof: By definition

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

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) \, dy$$

where $f_X(x)$ and $f_Y(y)$ are marginal p.d.f.

$$E(X+Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f_{XY}(x,y)\,dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,y)\,dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{XY}(x,y)\,dxdy$$

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

$$= \int_{-\infty}^{\infty} x f_X(x)\,dx + \int_{-\infty}^{\infty} y f_y(y)\,dy$$

= E(X) + E(Y)

Hence proved.

Generalization: If $X_1, X_2, X_3, \ldots, X_n$ are random variables then

$$E(X_1+X_2+X_3+\ldots+X_n) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n)$$

i.e. $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$, if all the expectations exist.

Property 2: Multiplication Theorem of Expectations

If X and Y are independent random variables then

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

Proof: $E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x,y)\,dxdy$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x) f_Y(y)\,dxdy$$

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

= E(X)E(Y), provided X and Y are independent.

Generalization: If $X_1, X_2, X_3,\ldots,X_n$ are independent random variables then

$E(X_1X_2X_3\ldots X_n) = E(X_1)E(X_2)E(X_3)\ldots E(X_n)$ i.e.

$E\left(\prod_{i=1}^{n} Xi\right) = \prod_{i=1}^{n} E(Xi)$, provided all the expectations exist.

Property 3: If X is a r.v. and 'a' is constant, then

(i) E[a h(X)] = a E[h(X)]

(ii) E[h(X) + a] = E[h(X)] + a

where h(X) , a function of X is a r.v. and all the expectations exist.

Property 4: If X is a r.v., a and b are constants, then

E(aX + b) = aE(X) + b, provided all the expectations exist.

Property 5: Expectation of a Linear Combination of Random Variable

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

$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$, provided all the expectations exist.

Property 6: If $X \geq 0$ then $E(X) \geq 0$

Property 7: If X and Y are two random variables s.t. $Y \leq X$ then $E(Y) \leq E(X)$, provided all the expectations exist.

Property 8: If μ'_r exist, then μ'_s exist for all $1 \leq s \leq r$

Mathematically, If E (X^r) exists, then E (X^s) exist for all $1 \leq s \leq r$ i.e.$E(X^r)<\infty \Rightarrow E(X^s)<\infty$ for all $1\leq s\leq r$.

Example 17: Given the following probability distribution

X	-2	-1	0	1	2
p(x)	0.25	0.30	0	0.30	0.25

Find (i) E(X) (ii) E(2X+3) (iii) $E(X^2)$

Solution:

(i) $E(X) = \sum_{i=1}^{5} x_i p_i = x_1 p_1 + x_2 p_2 + ... + x_5 p_5$

$= (-2)(0.15) + (-1)(0.30) + (0)(0) + (1)(0.30) + (2)(0.25)$

$= -0.3-0.3+0.3+0.5 = 0.2$

(ii) $E(2X+3) = 2E(X)+3 = 2(0.2)+3 = 0.4+3 = 3.4$

(iii) $E(X^2) = \sum_{i=1}^{5} x_i^2 p_i = x_1^2 p_1 + x_2^2 p_2 + ... + x_5^2 p_5$

$= (-2)^2(0.15) + (-1)^2(0.30) + (0)^2(0) + (1)^2(0.30) + (2)^2(0.25)$

$= 0.6 + 0.3 + 0 + 0.3 + 1 = 2.2.$

Problems:

1. If X is random variable with mean μ and standard deviation σ then what is the expectation of $Z = \frac{X - \mu}{\sigma}$?
2. Two cards are drawn one by one with replacement from 8 cards numbered from 1 to 8. Find the expectation of the product of the numbers on the drawn cards.

2.7 Variance and Covariance

The variance of a r.v. is defined by

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

The variance is a nonnegative number. The positive square root of the variance is called the standard deviation and is given by

$$\sigma_X = \sqrt{Var(X)} = \sqrt{E[(X-\mu)^2]}$$

If X is a discrete random variable taking the values $x_1, x_2, ..., x_n$ and having probability function p(x), then variance is given by

$$\text{Var(X)} = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i)$$

If X is a continuous random variable having density function f(x), then the variance is given by

$$\text{Var(X)} = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

Remark: The variance (or the standard deviation) is a measure of the dispersion, or scatter, of the values of the random variable about the mean. If the values tend to be concentrated near the mean, the variance is small; while if the values tend to be distributed far from the mean, the variance is large.

Theorem: If X is a r.v., then $V(aX + b) = a^2V(X)$, where a and b are constants

Proof: Let Y = aX + b, then E(Y) = aE(X) + b

$\Rightarrow Y - E(Y) = a[X - E(X)]$

Squaring and taking expectation of both sides we get

$E[Y - E(Y)]^2 = a^2E[X - E(X)]^2$

$\Rightarrow V(Y) = a^2V(X)$ or $V(aX + b) = a^2V(X)$, where V(X) is written for variance of X.

Cor.

(i) If b = 0, then $V(aX) = a^2V(X)$

$\Rightarrow$Variance is not independent of change of scale.

(ii) If a = 0, then V(b) = 0

$\Rightarrow$Variance of a constant is zero.

(iii) If a = 1, then V(X + b) = V(X)

$\Rightarrow$Variance is independent of change of origin.

Example 18: If X and Y are independent random variable with variance 2 and 3 respectively. Find the variance of 3X+4Y.

Solution: V (3X+4Y)=9V(X)+16V(Y)

=9*2+16*3=66

Problem:

Suppose that X is a random variable for which E(X) = 10 and V(X) = 25. Find the positive values of a and b such that Y=aX-b has expectation 0 and variance 1.

Covariance:

If X and Y are two random variables, then covariance between them is defined as

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

which can be simplified as

$$\begin{aligned} &= E\left[XY - XE(Y) - YE(X) + E(X)E(Y)\right] \\ &= E(XY) - E(Y)E(X) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

Cor.

If X and Y are independent then E(XY) = E(X)E(Y) and hence Cov(X, Y) = 0

- Cov (aX, bY) = ab Cov(X, Y)
- Cov (X + a, Y + b) = Cov(X, Y)
- Cov (aX + b, cY + d) = ac Cov(X, Y)
- Cov (X + Y, Z) = Cov(X, Z) + Cov(Y, Z)

Variance of a Linear Combination of Random Variable

LetX_1, X_2, X_3,…,X_n be n random variables then

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

Proof: Let $U = a_1X_1 + a_2X_2 + ... + a_nX_n$ so that

$$E(U) = a_1E(X_1) + a_2E(X_2) + ... + a_nE(X_n)$$

Squaring and taking expectation on both sides, we get

$$U - E(U) = a_1\{X_1 - E(X_1)\} + a_2\{X_2 - E(X_2)\} + ... + a_n\{X_n - E(X_n)\}$$

$$E[U - E(U)]^2 = a_1^2E[X_1 - E(X_1)]^2 + a_2^2E[X_2 - E(X_2)]^2 + ... + a_n^2E[X_n - E(X_n)]^2$$

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

$$V(U) = a_1^2V(X_1) + a_2^2V(X_2) + ... + a_n^2V(X_n) + 2\sum_{i=1}^{n}\sum_{\substack{j=1\\ i<j}}^{n} a_i a_j Cov(X_i, X_j)$$

Hence proved.

Example 19: Two unbiased dice are thrown. Find the expected values of the sum of the numbers on them.

Solution: The probability function of X (the sum of numbers obtained on two dice) is

Value of X: x	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$$E(X) = \sum_i p_i x_i$$

$$= 2\times\frac{1}{36} + 3\times\frac{2}{36} + 4\times\frac{3}{36} + ... + 11\times\frac{2}{36} + 12\times\frac{1}{36}$$

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

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

Example 20: A gamester has a disc with a freely revolving needle. The disc is divided into 20 equal sectors by thin lines and the sectors are marked 0, 1, 2,…,19. The gamester treats 5 or any multiple of 5 as lucky numbers and zero as a special lucky number. He allows a player to whirl the needle on a charge of 10 paise. When the needle stops at the lucky number the gamester pays to the player 5 times of the sum charged. Is the game fair? What is the expectation of the player?

Solution:

Event	Favorable	p(x)	Player's Gain (x)
Lucky number	5,10,15	$\frac{3}{20}$	20-10 = 10p
Special lucky number	0	$\frac{1}{20}$	50-10 = 40p
Other number	1,2,3,4,6,7,8,9,11,12,13,14, 16,17,18,19	$\frac{16}{20}$	-10p

$$\therefore E(X) = \frac{3}{10}\times 10 + \frac{1}{10}\times 40 - \frac{16}{10}\times 10 = -\frac{9}{2} \neq 0$$

i.e. the game is not fair.

Example 21: A coin is tossed until a head appears. What is the expectation of the number of tosses required?

Solution: Let X denote the number of tosses required to get the first head. Then X can materialize in the following ways:

Event	X	Probability, p(x)
H	1	$\frac{1}{2}$
TH	2	$\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$
TTH	3	$\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{8}$
. . .	. . .	. . .

$$\therefore E(X)=\sum_{x=1}^{\infty} xp(x)=1\times\frac{1}{2}+2\times\frac{1}{4}+3\times\frac{1}{8}+4\times\frac{1}{16}+... \quad (1)$$

This is an arithmetic-geometric series with ratio of GP being r = $\frac{1}{2}$

Now

$$S=1.\frac{1}{2}+2.\frac{1}{4}+3.\frac{1}{8}+4.\frac{1}{16}+...$$
$$\frac{S}{2}=\frac{1}{4}+2.\frac{1}{8}+3.\frac{1}{16}+...$$
$$\therefore\left(1-\frac{1}{2}\right)S=\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+...$$
$$\Rightarrow\frac{1}{2}S=\frac{\frac{1}{2}}{1-\frac{1}{2}}=1$$
$$\Rightarrow S=2$$

Thus, from (1) E(X) = 2.

Example 22: Two random variables X and Y have the following joint probability density function:

$$f(x,y)=\begin{cases}2-x-y; 0\le x\le 1, 0\le y\le 1\\ 0; otherwise\end{cases}$$

Find

(i) Marginal probability density functions of X and Y

(ii) Conditional density functions

(iii) Var(X) and Var(Y)

(iv) Covariance between X and Y.

Solution:

(i) $$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 (2-x-y)dy = \frac{3}{2} - x$$

$$\therefore f_X(x) = \begin{cases} \frac{3}{2} - x; & 0<x<1 \\ 0; & otherwise \end{cases}$$

Similarly

$$f_Y(x) = \begin{cases} \frac{3}{2} - y; & 0<y<1 \\ 0; & otherwise \end{cases}$$

(i) $$f_{X|Y}(x,y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{(2-x-y)}{(\frac{3}{2}-y)}, 0<(x,y)<1$$

$$f_{Y|X}(x,y) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{(2-x-y)}{(\frac{3}{2}-y)}, 0<(x,y)<1$$

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

$$E(Y) = \int_0^1 y f_Y(y)dy = \int_0^1 y\left(\frac{3}{2}-y\right)dy = \frac{5}{12}$$

$$E(X^2) = \int_0^1 x^2\left(\frac{3}{2}-x\right)dx = \left|\frac{3}{6}x^3 - \frac{x^4}{4}\right|_0^1 = \frac{1}{4}$$

$$V(X) = E(X^2) - \left\{E(X)\right\}^2 = \frac{1}{4} - \frac{25}{144} = \frac{11}{144}$$

Similarly $V(Y) = \frac{11}{144}$.

(iii) $E(XY) = \int_0^1 \int_0^1 xy(2-x-y)dxdy$

$$= \int_0^1 \left| 2\frac{x^2y}{2} - \frac{x^3y}{3} - \frac{x^2y^2}{2} \right|_{x=0}^{x=1} dy$$

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

$\therefore Cov(X,Y) = E(XY) - E(X)E(Y)$

$= \frac{1}{6} - \frac{5}{12} \cdot \frac{5}{12} = -\frac{1}{144}$.

Problems:

1. An urn contains 7 white and 3 red balls. Two balls are drawn together, at random from this urn. Compute the probability that neither of them is white. Find also the probability of getting one white and one red ball. Hence compute the expected number of white balls drawn.
2. A box contains 2^n tickets among which ${}^nC_i$ tickets bear the number i; i = 0, 1, 2, 3, …, n. A group of m tickets is drawn. What is the expectation of the sum of their number?
3. Let

   $$f(x,y) = \begin{cases} 8xy, & 0 < x < y < 1 \\ 0, & elsewhere \end{cases}$$

   Find

   $(a) E(Y|X = x)$

   $(b) E(XY|X = x)$

   $(c) Var(Y|X = x)$

Mathematical Expectation of Two-Dimensional Random Variable

The mathematical expectation of a function g(x,y) of two dimensional r.v. (X, Y) with p.d.f. f(x, y) is given by:

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dxdy$$

(If X and Y are continuous variables)

$$E[g(X,Y)] = \sum_i \sum_j g(x_i y_j) P(X = x_i \cap Y = y_j)$$

(If X and Y discrete variables) provided the expectation exists.

Conditional Expectation

Discrete Case:

$$E\left\{g(X,Y)\middle|Y=y_j\right\}=\sum_{i=1}^{\infty}g(x_i,y_j)P(X=x_i\middle|Y=y_j)$$

$$=\sum_{i=1}^{\infty}\frac{g(x_i,y_j)P(X=x_i\cap Y=y_j)}{P(Y=y_j)}$$

Continuous Case:

$$E\left\{g(X,Y)\middle|Y=y\right\}=\int_{-\infty}^{\infty}g(x,y)f_{X|Y}(x|y)dx$$

$$=\int_{-\infty}^{\infty}\frac{g(x,y)f(x,y)dx}{f_Y(y)}$$

2.8 Moment Generating Function and Their Properties

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

$$M_X(t)=E\left(e^{tX}\right)=\begin{cases}\int e^{tx}f(x)\,dx, & for\ continuous\ distribution\\ \sum e^{tx}f(x), & for\ discrete\ distribution\end{cases} \qquad (1)$$

Let us assume that r.h.s. of eq(1) is absolutely convergent for some positive number h such that –h<t<h. Thus

$$M_X(t)=E\left(e^{tX}\right)$$

$$=E(1+tX+\frac{t^2X^2}{2!}+\ldots+\frac{t^rX^r}{r!}+\ldots)$$

$$=1+tE(X)+\frac{t^2E(X^2)}{2!}+\ldots+\frac{t^rE(X^r)}{r!}+\ldots)$$

$$=1+t\mu_1'+\frac{t^2}{2!}\mu_2'+\ldots+\frac{t^r}{r!}\mu_r'+\ldots \qquad 1(a)$$

$$=\sum_{r=0}^{\infty}\frac{t^r}{r!}\mu_r' \qquad (1b)$$

where

$$\mu_r' = E(X^r) = \begin{cases} \int_{-\infty}^{\infty} x^r f(x)dx, & \textit{for continuous distribution} \\ \sum_x x^r f(x), & \textit{for discrete distribution} \end{cases}$$

is the r^{th} moment of X about origin. Since the coefficients in this expansion enable us to find the moments, the reason for the name moment generating function is apparent

Differentiating eq(1a) w.r.t. 't' r times and then putting t=0, we get

$$\left|\frac{d^r}{dt^r}\{M_X(t)\}\right|_{t=0} = \left|\frac{\mu_r'}{r!}.r! + \mu_{r+1}'.t + \mu_{r+2}'.\frac{t^2}{2!} + \ldots\right|_{t=0}$$

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

In general, the moment generating function X about the point X = a is defined as:

$$M_X(t)(\text{about } X = a) = E\left(e^{t(X-a)}\right)$$

$$= E\left\{1 + t(X-a) + \frac{t^2(X-a)^2}{2!} + \ldots + \frac{t^r(X-a)^r}{r!} + \ldots\right\}$$

$$= 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \ldots + \frac{t^r}{r!}\mu_r' + \ldots$$

where μ_r' = {E(X-a)^r} is the r^{th} moment about the point X=a.

Some Limitations of Moment Generating Function

1. A random variable X may have no moments although its m.g.f. exists.
2. A r.v. X can have m.g.f. and some moments, yet the m.g.f. does not generate the moments.
3. A r.v. X can have all or some moments, but m.g.f. does not exists except perhaps at one point.

Properties of Moment Generating Function

Property1: $M_{cX}(t) = M_X(ct)$, c being a constant

Proof: By definition

L.H.S. $= M_{cX}(t) = E(e^{tcX})$

R.H.S. $= M_X(ct) = E(e^{ctX})$ = L.H.S.

Property2: The moment generating function of the sum of a number of independent r.v.'s is equal to the product of their respective moment generating functions.

Symbolically, if X_1, X_2,…,X_n are independent random variables, then moment generating function of their sum $X_1+X_2+\ldots+X_n$ is given by: $M_{X_1+X_2+\ldots+X_n}(t)=M_{X_1}(t)M_{X_2}(t)\ldots M_{X_n}(t)$

Proof: By definition

$$M_{X_1+X_2+\ldots+X_n}(t)=E[e^{t(X_1+X_2+\ldots+X_n)}]$$

$$M_{X_1+X_2+\ldots+X_n}(t)=E[e^{tX_1}e^{tX_2}\ldots e^{tX_n}]$$

$$M_{X_1+X_2+\ldots+X_n}(t)=E[e^{tX_1}]E[e^{tX_2}]\ldots E[e^{tX_n}]$$

$$M_{X_1+X_2+\ldots+X_n}(t)=M_{X_1}(t)M_{X_2}(t)\ldots M_{X_n}(t)$$

Hence proved.

Property 3: Effect of change of origin and scale of m.g.f.

Let us transform X to the new variable U by changing both the origin and scale in X as follow:

$U=\frac{X-a}{h}$, where a and h are constants.

Moment generating function of U (about origin) is given by

$$M_U(t)=E(e^{tU})=E\left[\exp\left\{\frac{t(X-a)}{h}\right\}\right]$$

$$=E\left[e^{\frac{tX}{h}}e^{-\frac{at}{h}}\right]=e^{-\frac{at}{h}}E(e^{\frac{tX}{h}})=e^{-\frac{at}{h}}M_X\left(\frac{t}{h}\right)$$

where $M_X(t)$ is the m.g.f. of X about origin.

Remark: In particular, if we take

$a=E(X)=\mu$ and $h=\sigma=\sigma_X$ then

$U=\frac{X-E(X)}{\sigma_X}=\frac{X-\mu}{\sigma}=Z$ is known as a standard variate.

Thus the m.g.f. of a standard variate Z is given by

$$M_Z(t) = e^{-\frac{\mu t}{\sigma}} M_X(\frac{t}{\sigma})$$

Uniqueness Theorem of Moment Generating Function

The moment generating function of a distribution if it exists, uniquely determines the distribution. This implies that corresponding to a given probability distribution, there is only one m.g.f. (provides it exists) and corresponding to a given m.g.f., there is only one probability distribution.

Hence $M_X(t) = M_Y(t)$ implies X and Y are identically distribution.

Example 23: If the moments of variate X are defined by

$E(X^r) = 0.6$; r = 1,2,3,…

Show that P(X = 0) = 0.4

P(X = 1) = 0.6, P(X ≥2)= 0

Solution: The m.g.f. of variate X is:

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!}\mu'_r = 1 + \sum_{r=1}^{\infty} \frac{t^r}{r!}(0.6)$$

$$= 0.4 + 0.6 \sum_{r=0}^{\infty} \frac{t^r}{r!} = 0.4 + 0.6e^t \qquad (1)$$

But $M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} P(X = x)$

$$= P(X = 0) + e^t . P(X = 1) + \sum_{x=2}^{\infty} e^{tx} P(X = x) \qquad (2)$$

From (1) and (2), we get

P(X = 0) = 0.4, P(X = 1) = 0.6; P(X ≥2) = 0

Example 24: Find the moment generating function of the r.v. whose moments are:

$$\mu'_r = (r+1)!2^r$$

Solution: The m.g.f. is given by:

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!}\mu'_r = \sum_{r=0}^{\infty} \frac{t^r}{r!}(r+1)!2^r = \sum_{r=0}^{\infty} (r+1)(2t)^r$$

$$M_X(t) = 1 + 2(2t) + 3(2t)^2 + 4(2t)^3 + ...$$

$$M_X(t) = (1-2t)^{-2}$$

CHAPTER-3

DISCRETE AND CONTINUOUS DISTRIBUTIONS

Structure

3.1 Discrete Distributions

The probability distribution gives us a law according to which different values of random variable are distributed with specified probability with some definite law. Probability distributions are generally divided into two classes. A discrete probability distribution (applicable to the scenarios where the set of possible outcomes is discrete, such as a coin toss or a roll of dice) can be encoded by a discrete list of the probabilities of the outcomes, known as a probability mass function. On the other hand, a continuous probability distribution (applicable to the scenarios where the set of possible outcomes can take on values in a continuous range (e.g. real numbers), such as the temperature on a given day) is typically described by probability density functions (with the probability of any individual outcome actually being 0).

In this section we discuss some special discrete probability distributions such as Bernoulli, Binomial, Poisson, Geometric etc., of a random variable that is successfully applied in a wide variety of decision situations.

3.1.1 Discrete Uniform Distribution

Definition: A r.v. X is said to have a discrete uniform distribution over the range [1, n] if its p.m.f. is expressed as follows:

$$P(X = x) = \begin{cases} \frac{1}{n}; & \text{for} \quad x = 1, 2, ..., n \\ 0; & \text{otherwise} \end{cases} \qquad \dots (1)$$

Here n is known as the parameter of the distribution and lies in the set of all positive integers. Equation (1) is also called a discrete rectangular distribution.

Moments:

$$\mu_1' = E(X) = \frac{1}{n}\sum_{i=1}^{n} i = \frac{n+1}{2} = \text{Mean}$$

$$\mu_2' = E(X^2) = \frac{1}{n}\sum_{i=1}^{n} i^2 = \frac{(n+1)(2n+1)}{6}$$

$$\mu_3' = E(X^3) = \frac{1}{n}\sum_{i=1}^{n} i^3 = \frac{n(n+1)^2}{4}$$

Variance = $V(X) = E(X^2) - (E(X))^2 = \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}$

The moment generating function (m.g.f.) of X is:

$$M_X(t) = E(e^{tX}) = \frac{1}{n}\sum_{x=1}^{n} e^{tx} = \frac{e^t(1-e^{nt})}{n(1-e^t)} \qquad \text{(Using sum of G.P. formula)}$$

3.1.2 Bernoulli Distribution

The Bernoulli distribution, named after the Swiss Mathematician Jacques Bernoulli (1654–1705), describes a probabilistic experiment where a trial has two possible outcomes, a success or a failure with reference of presence or absence of a particular attribute and characteristic. The probability of presence of attributes is called success which is denoted by p while the absence is known as failure which is denoted by q. For example, getting head in tossing of coin may be treated as success while tail is failure.

Definition: A r.v. X is said to have a Bernoulli distribution with parameter p if its p.m.f. is given by:

$$P(x,p) = \begin{cases} q; & \text{if } x=0 \quad \text{(failure)} \\ p; & \text{if } x=1 \quad \text{(success)} \end{cases}$$

and zero elsewhere. The parameter p satisfies $0 \le p \le 1$ and (1-p) is denoted as q.

Remark: The Bernoulli distribution is useful whenever a random experiment has only two possible outcomes, which may be labelled as success and failure.

Moments of Bernoulli Distribution

The r^{th} moment about origin is:

$$\mu_r' = E(X^r) = 0^r.q + 1^r p = p; \; r = 1, 2, ...$$

$$\Rightarrow \mu_1' = E(X) = p$$

and $\mu_2' = E(X^2) = p$ so that

$$\mu_2 = \mathrm{Var}(X) = \mu_2' - \mu_1'^2 = p - p^2 = p(1-p) = pq$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$$
$$= p(1-p)(1-2p)$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2^1(\mu_1')^2 - 3(\mu_1')^4$$
$$= p(1-p)(1-3p+3p^2)$$

The m.g.f. of Bernoulli variate is

$$M_X(t) = e^{0t}P(X=0) + e^{1t}P(X=1) = q + pe^t$$

Example 1: Let X be random variable having Bernoulli distribution with parameter p=0.3. Find the mean and variance.

Solution: Here p=0.3 and q=0.7

Mean=0.3

Variance=pq= (0.3)(0.7)=0.21

3.1.3 Binomial Distribution

Suppose, we have an experiment such as tossing a coin or throwing a die repeatedly or choosing a marble from an urn repeatedly. Each toss or selection is called a *trial.* In any single trial there will be a probability associated with a particular event such as head on the coin, 4 on the die, or selection of a red marble. In some cases this probability will not change from one trial to the next (as in tossing a coin or die). Such trials are then said to be independentand are often called Bernoulli trials. If Bernoulli trial is repeated a finite number of times then their probability distribution is called Binomial distribution. Binomial distribution was discovered by J. Bernoulli (1654-1705) and was first published eight years after his death, i.e. in 1713 and is also known as "Bernoulli distribution for n trials". Binomial distribution is applicable for a random experiment comprising a finite number of Bernoulli trial with constant probability of success for each trial.

Definition: The Binomial model has three defining properties:

- Bernoulli trials are conducted n times,
- The trials are independent,
- The probability of success p does not change between trials.

A r.v. X is said to follow Binomial distribution if it assumes only non-negative values and its probability mass function is given by:

$$P(X=x) = p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & ; x = 0,1,2,\ldots,n; q = 1-p \\ 0; \text{otherwise} \end{cases}$$

Here two independent constants ‘n’ and ‘p’ are known as the parameters of the distribution. ‘n’ is also known as the degree of the binomial distribution.‘p’ is the probability of success in each Bernoulli trial and q=1-p is the probability of failure in each trial.

Any random variable which follows Binomial distribution is known as binomial variate.

Notation: X ~ B(n, p) to denote the random variable X follows Binomial distribution with parameters n and p. The probability p(x) is also sometimes denoted as B(x, n, p).

Remark:

1. **Physical condition for Binomial distribution**
   - (i) Each trail results in two exhaustive and mutually disjoint outcomes, termed as success and failure.
   - (ii) The number of trials ‘n’ is finite.
   - (iii) The probability of success ‘p’ is constant for each trial.
   - (iv) The trials are independent of each other
2. **Additive property of Binomial distribution:**

   If X and Y are two binomial distributed independent random variables with parameters (n_1, p) and (n_2, p) respectively, then their sum also follows a binomial distribution with parameters n_1+n_2 and p. But if the probability of success is not same for the two random variables then this property does not hold.

Example 2: In a Binomial distribution consisting of 5 independent trials, probability of 1 and 2 successes are 0.4096 and 0.2048 respectively. Find the parameter ‘p’ of the distribution.

Solution: Let X ~ B(n, p).

By definition of Binomial distribution, we have

n = 5, p(1) = 0.4096, p(2) = 0.2048

According to Binomial probability law:

$$P(X = x) = p(x) = \binom{5}{x} p^x (1-p)^{5-x}, x = 0,1,2,...5$$

Now

$$p(1) = \binom{5}{1} p(1-p)^4 = 0.4096 \qquad (1)$$

$$p(2) = \binom{5}{2} p^2(1-p)^3 = 0.2048 \qquad (2)$$

Dividing (1) by (2) we get

$$\frac{\binom{5}{1}p(1-p)^4}{\binom{5}{2}p^2(1-p)^3} = \frac{0.4096}{0.2048} \Rightarrow \frac{5(1-p)}{10p} = 2$$

$$\Rightarrow p = \frac{1}{5} = 0.2$$

Example 3: Ten coins are thrown simultaneously. Find the probability of getting atleast seven heads.

Solution: Letp = probability of getting a head = $\frac{1}{2}$

q = probability of not getting a head = $\frac{1}{2}$

The probability of getting x heads in a random throw of 10 coins is:

$$p(x) = \binom{10}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{10-x} = \binom{10}{x}\left(\frac{1}{2}\right)^{10}; x = 0,1,2,...,10$$

∴ Probability of getting at least seven heads is given by

$$P(X \geq 7) = p(7) + p(8) + p(9) + p(10)$$

$$\left(\frac{1}{2}\right)^{10}\left\{\binom{10}{7}+\binom{10}{8}+\binom{10}{9}+\binom{10}{10}\right\} = \frac{120+45+10+1}{1024} = \frac{176}{1024}$$

Example 4: An irregular six-faced die is thrown and the expectation that in 10 throws it will give five even numbers is twice the expectation that it will give four even numbers. How many items in 10,000 sets of 10 throws each, would you expect it to give no even number?

Solution: let p be the probability of getting an even number in a thrown of a die. Then the probability of getting x even numbers in ten throws of a die is given by:

$$P(X = x) = \binom{10}{x}p^x q^{10-x}; x = 0,1,2,...,10$$

We are given that:

$$P(X = 5) = 2P(X = 4)$$

$$\Rightarrow \binom{10}{5}p^5q^5 = 2\binom{10}{4}p^4q^6$$

$$\Rightarrow \frac{10!p}{5!5!} = \frac{2(10!)q}{4!6!} \Rightarrow \frac{p}{5} = \frac{2q}{6} \Rightarrow 3p = 5(1-p)$$

$$\Rightarrow p = \frac{5}{8} \text{or } \ q = \frac{3}{8}$$

Thus $P(X = x) = \binom{10}{x}\left(\frac{5}{8}\right)^{x}\left(\frac{3}{8}\right)^{10-x}$

Hence the required number of times that in 10,000 sets of 10 throws each, we get no even number

$= 10,000 \times P(X = 0)$

$= 10,000 \times \left(\frac{3}{8}\right)^{10} = .5499.$

Problems:

1. A and B play a game in which their chances of winning are in the ratio 3:2. Find A's chance of winning atleast three games out of the five games played.
2. With the usual notations, find p for a Binomial variate X, if n = 6 and 9P(X=4) = P(X=2).
3. Let X and Y be two independent random variables such that $X \sim B(4, 0.7)$ and $Y \sim B(3, 0.8)$ Find $P[X + Y \leq 1]$.

Moments of Binomial Distribution

$$\mu'_1 = E(X) = \sum_{x=0}^{n} x\binom{n}{x}p^x q^{n-x} = np\sum_{x=0}^{n}\binom{n-1}{x-1}p^{x-1}q^{n-x} = np(q+p)^{n-1} = np$$

$$\left\{\because \binom{n}{x} = \frac{n}{x}\binom{n-1}{x-1} = \frac{n}{x}.\frac{n-1}{x-1}\binom{n-2}{x-2} = \frac{n}{x}.\frac{n-1}{x-1}.\frac{n-2}{x-2}\binom{n-3}{x-3} \text{ and so on.}\right\}$$

Thus the mean of the Binomial distribution is np.

$$\mu'_2 = E(X^2) = \sum_{x=0}^{n} x^2\binom{n}{x}p^x q^{n-x}$$

$$= \sum_{x=0}^{n}\left\{x(x-1)+x\right\}\frac{n(n-1)}{x(x-1)}.\binom{n-2}{x-2}p^x q^{n-x}$$

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

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

$$= n(n-1)p^2 + np$$

$$\mu'_3 = E(X^3) = \sum_{x=0}^{n} x^3\binom{n}{x}p^x q^{n-x}$$

$$= \sum_{x=0}^{n} \left\{ x(x-1)(x-2) + 3x(x-1) + x \right\} \binom{n}{x} p^x q^{n-x}$$

$$= n(n-1)(n-2)p^3 \sum_{x=3}^{n} \binom{n-3}{x-3} p^{x-3} q^{n-x} + 3n(n-1)p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} q^{n-x} + np$$

$$= n(n-1)(n-2)p^3 (q+p)^{n-3} + 3n(n-1)p^2 (q+p)^{n-2} + np$$

$$= n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np$$

Similarly

$$\mu'_4 = E(X^4) = \sum_{x=0}^{n} x^4 \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} \left(x(x-1)(x-2)(x-3) + 6x(x-1)(x-2) + 7x(x-1) + x \right) \binom{n}{x} p^x q^{n-x}$$

$$= n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np$$

Central Moments of Binomial Distribution

$$\mu_2 = \mu'_2 - \mu'^2_1 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p) = npq$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1$$

$$= \left\{ n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np \right\} - 3np\left\{ n(n-1)p^2 + np \right\} + 2(np)^3$$

$$= np(-3np^2 + 3np + 2p^2 - 3p + 1 - 3npq)$$

$$= np(2p^2 - 3p + 1) = np(2p^2 - 2p + q) = npq(1-2p)$$

$$= npq(q + p - 2p) = npq(q-p)$$

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1 = npq[1 + 3(n-2)pq] \text{(on simplification)}$$

Hence

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{n^2p^2q^2(q-p)^2}{n^3p^3q^3} = \frac{(q-p)^2}{npq} = \frac{(1-2p)^2}{npq}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{npq(1+3(n-2)pq)}{n^2p^2q^2} = \frac{(1+3(n-2)pq)}{npq} = 3 + \frac{1-6pq}{npq}$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{q-p}{\sqrt{npq}} = \frac{1-2p}{\sqrt{npq}}$$

$$\gamma_2 = \beta_2 - 3 = \frac{1-6pq}{npq}$$

Remarks:

(i) Variance = $npq<np$ = mean

Hence, for Binomial distribution variance is less than mean.

(ii) Standard deviation is $\sqrt{npq}$

Example 5: The mean and variance of Binomial distribution are 4 and 4/3 respectively. Find $P(X\geq 1)$

Solution: Let $X \sim B(n, p)$.

Then we are given:

Mean = $np = 4$ (1)

$Var(X) = npq = 4/3$ (2)

Dividing, we get $\frac{np}{npq} = \frac{4}{4/3} \Rightarrow q = \frac{1}{3} \Rightarrow p = \frac{2}{3}$

Substituting in (1), we obtained

$$n = \frac{4}{p} = \frac{4\times 3}{2} = 6$$

$$\therefore P(X \geq 1) = 1 - P(X=0) = 1 - q^n = 1 - \left(\frac{1}{3}\right)^6 = 1 - \frac{1}{729} = 0.99863$$

Problems:

1. For a Binomial distribution with $p = ¼$ and $n = 10$. Find mean and variance.
2. If $X \sim B(n, p)$. find p if $n = 6$ and $9P[X=4] = P[X=2]$

Recurrence Relation for the Moment of Binomial Distribution

By definition, we have

$$\mu_r = E\left[X - E(X)\right]^r = \sum_{x=0}^{n}(x-np)\binom{n}{x}p^x q^{n-x}$$

Differentiating w.r.t. to p, we get

$$\frac{d\mu_r}{dp}=\sum_{x=0}^{n}\binom{n}{x}\left[-nr(x-np)^{r-1}p^xq^{n-x}+(x-np)^r\left(xp^{x-1}q^{n-x}-(n-x)p^xq^{n-x-1}\right)\right]$$

$$=-nr\sum_{x=0}^{n}\binom{n}{x}(x-np)^{r-1}p^xq^{n-x}+\sum_{x=0}^{n}\binom{n}{x}(x-np)^r p^xq^{n-x}\left(\frac{x}{p}-\frac{n-x}{q}\right)$$

$$=-nr\sum_{x=0}^{n}\binom{n}{x}(x-np)^{r-1}p(x)+\sum_{x=0}^{n}\binom{n}{x}(x-np)^r p(x)\frac{(x-np)}{pq}$$

$$=-nr\sum_{x=0}^{n}\binom{n}{x}(x-np)^{r-1}p(x)+\frac{1}{pq}\sum_{x=0}^{n}\binom{n}{x}(x-np)^{r+1}p(x)$$

$$=-nr\mu_{r-1}+\frac{1}{pq}\mu_{r+1}$$

$$\Rightarrow \mu_{r+1}=pq\left(nr\mu_{r-1}+\frac{d\mu_r}{dp}\right)$$

For r=1, 2

$$\mu_2=pq\left(\mu_0+\frac{d\mu_1}{dp}\right)=npq$$

$$\mu_3=npq(q-p)$$

Moment Generating Function of Binomial Distribution

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

$$M_X(t)=E\left(e^{tX}\right)=\sum_{x=0}^{n}e^{tx}\binom{n}{x}p^xq^{n-x}=\sum_{x=0}^{n}\binom{n}{x}(pe^t)^xq^{n-x}=\left(q+pe^t\right)^n$$

m.g.f. about Mean of Binomial Distribution

$$E\left[e^{t(X-np)}\right]=e^{-tnp}.\left(E^{tX}\right)=e^{-tnp}.M_X(t)=e^{-tnp}\left(q+pe^t\right)^n=\left(qe^{-pt}+pe^{tq}\right)^n$$

$$=\left[q\left\{1-pt+\frac{p^2t^2}{2!}-\frac{p^3t^3}{3!}+...\right\}+p\left\{1+tq+\frac{q^2t^2}{2!}+\frac{q^3t^3}{3!}+...\right\}\right]^n$$

$$=\left[\left(q+p\right)+\frac{t^2}{2!}pq(q+p)+\frac{t^3}{3!}pq(q^2-p^2)+\frac{t^4}{4!}pq(q^3+p^3)+...\right]^n$$

$$=\left[1+\left\{\frac{t^2}{2!}pq+\frac{t^3}{3!}pq(q-p)+\frac{t^4}{4!}pq(1-3pq)+...\right\}\right]^n$$

$$= \left[\begin{array}{l} 1 + \binom{n}{1} \left\{ \frac{t^2}{2!} pq + \frac{t^3}{3!} pq(q-p) + \frac{t^4}{4!} pq(1-3pq) + \ldots \right\} \\ + \binom{n}{2} \left\{ \frac{t^2}{2!} pq + \frac{t^3}{3!} pq(q-p) + \frac{t^4}{4!} pq(1-3pq) + \ldots \right\}^2 + \ldots \end{array} \right]$$

$$\mu_2 = \text{coefficinet of } \frac{t^2}{2!} = npq, \quad \mu_3 = \text{coefficinet of } \frac{t^3}{3!} = npq(q-p)$$

Additive property of Binomial Distribution

Let $X \sim B(n_1, p_1)$ and $Y \sim B(n_2, p_2)$ be the independent reabndom variables.Then

$$M_X(t) = \left(q_1 + p_1 e^t\right)^{n_1}, \quad M_X(t) = \left(q_2 + p_2 e^t\right)^{n_2} \ldots(1)$$

Then we have,

$$M_{X+Y}(t) = \left(q_1 + p_1 e^t\right)^{n_1} . \left(q_2 + p_2 e^t\right)^{n_2} \qquad \ldots(2)$$

The equation (2) cannot be expressed in the for $\left(q + pe^t\right)^n$, from uniqueness theorem of m.g.f.'s it follows that X+Y is not a binomial distribution. Hence, in general the sum of two independent random variates is not a binomial variate. In the above case, if we take $p_1 = p_2 = p$, then we can say that X+Y is a binomial variate.

Fitting of Binomial Distribution

To fit a Binomial distribution, we need the observed data which is obtained from repeated trials of given experiment. Process of finding the probabilities corresponding to each value of the Binomial variable becomes easy if we use the recurrence relation for the probabilities of Binomial distribution.

Recurrence Relation

We have studied that Binomial probability function is

$$p(x) = P[X = x] = {}^nC_x p^x q^{n-x} \qquad (1)$$

If we replace x by x+1, we set

$$p(x+1) = P[X = x+1] = {}^nC_{x+1} p^{x+1} q^{n-(x+1)} \qquad (2)$$

Dividing (2) by (1), we have

$$\frac{p(x+1)}{p(x)} = \frac{{}^nC_{x+1} p^{x+1} q^{n-x-1}}{{}^nC_x p^x q^{n-x}} = \frac{n!x!(n-x)!}{n!(x+1)x!(n-x-1)!} \times \frac{p}{q}$$

$$= \frac{n!(n-x)(n-x-1)!}{n!(x+1)!(n-x-1)!} \times \frac{p}{q} = \frac{n-x}{x+1} \times \frac{p}{q}$$

$$\Rightarrow p(x+1) = \frac{n-x}{x+1} \times \frac{p}{q} p(x) \qquad (3)$$

Putting x = 0, 1, 2, 3,… in this equation we get p(1) in terms of p(0), p(2) in terms of p(1), … and so on.

Process of Fitting Binomial Distribution

We first find the mean from the given frequency distribution and equate it to np. From this, we can find the value of p, after that we obtain $p(0) = q^n$, where q = 1-p

Then the recurrence relation i.e. p(x+1) = $\left(\frac{n-x}{x+1}\right)p(x)$ is applied to find the value of p(1), p(2),…. In this way, the binomial distribution is fitted to the given data. Thus fitting of a binomial distribution involves comparing the observed frequencies with the expected frequencies to see how best the observed results fit with the theoretical (expected) results.

Example 6: Four coins were tossed and number of heads noted. The experiment is repeated 200 times. The number of tosses showing 0,1 2, 3 and 4 heads were found. Fit a binomial distribution as under. Fit a binomial distribution to these observed results assuming that the nature of the coins is not known.

Number of heads	0	1	2	3	4
Number of tosses	15	35	90	40	20

Solution: Here n = 4, N = 200

First, we obtained the mean of the given frequency distribution as follows:

Number of heads (X)	Number of tosses (f)	f_X
0	15	0
1	35	35
2	90	180
3	40	120
4	20	80
Total	200	415

$$\therefore \text{Mean} = \frac{\sum f(x)}{\sum f} = \frac{415}{200} = 2.075$$

As mean for Binomial distribution is np.

$\therefore np = 2.075$

$$\Rightarrow p = \frac{2.075}{4} = 0.5188$$

$$\Rightarrow P(X = 0) = q^4 = (0.4812)^4 = 0.0536$$

Now, using the recurrence relation

$$p(x+1) = \frac{n-x}{x+1} \cdot \frac{p}{q} p(x); x = 0,1,2,3,4$$

Numberof heads X	$\left(\frac{n-x}{x+1}\right)\frac{p}{q} = \frac{4-x}{x+1}\left(\frac{0.5188}{0.4812}\right)$ $= \frac{4-x}{x+1}(1.07814)$	p(x)	Expected theoretical frequency of f(x)
0	4.31256	P(0)=0.0536	10.72≈11
1	1.61721	P(1)=0.23115	46.23≈46
2	0.71876	P(2)=0.37382	74.76≈75
3	0.26954	P(3)=0.26869	53.73≈ 54
4	0	P(4)=0.0724	14.48 ≈14

Problem: Seven coins are tossed and number of heads noted. The experiment is repeated 128 times and the following distribution is obtained:

Number of heads	0	1	2	3	4	5	6	7
Frequencies	7	6	19	35	30	23	7	1

Fit a binomial distribution to these observed results assuming that the nature of the coins is not known.

3.1.4 Poisson Distribution

Sometimes, we are interested in checking whether event is occurring at random points in time or space. Poisson distribution was discovered by Simeon Denis Poisson (1781-1840). Poisson distribution is used to calculate the probability of having a specified number of occurrences of an event over a fixed interval of time or space. It provides a good model when the data count from an experiment is small i.e. the number of observations is rare during a given time period.

Examples:

Some examples of random variables which usually follow the Poisson distribution are:

1. The number of misprints on a page of a book.
2. The number of people in a community living to 100 years of age
3. The number of wrong telephone numbers dialled in a day.
4. The number of customers entering a shop on a given day.
5. The number of α-particles discharged in a fixed period of time from some radioactive source.

Poisson distribution is a limiting case of the Binomial distribution under the following conditions

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

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

(iii) $np = \lambda$ is finite.

Thus, $p = \frac{\lambda}{n}, q = \left(1 - \frac{\lambda}{n}\right)$, where λ is a positive real number.

The probability of x successes in a series of n independent trials is

$$B(x;n,p) = \binom{n}{x} p^x q^{n-x}; x = 0,1,2,...,n$$

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

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

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

$$\therefore \lim_{n\to\infty} B(x;n,p) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0,1,2...$$

(without using sterling approximation.)

Note: Poisson distribution can also be derived using sterling approximation.

Definition: A r.v. X is said to follow a Poisson distribution if it assumes only non-negative values and its probability mass function is given by:

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

Here λ is known as the parameter of the distribution.

Remarks:

(i) If X follows Poisson distribution with parameter λ then we shall use the notation [$X \sim P(\lambda)$]

(ii) If X and Y are two independent Poisson variate with parameters λ_1 and λ_2 respectively, then X+Y is also a Poisson variate with parameter $\lambda_1+\lambda_2$. This is known as additive property of Poisson distribution.

Moments of Poisson Distribution

r^{th} order moment about origin of Poisson variates is

$$\mu_r = E(X^r) = \sum_{x=0}^{\infty} x^r p(x) = \sum_{x=0}^{\infty} x^r \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$\mu_1 = E(X) = \sum_{x=0}^{\infty} xp(x) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= \lambda e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + ...\right) = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Hence Mean = λ

$$\mu'_2 = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^{x}}{x!} = \sum_{x=0}^{\infty} (x(x-1)+x)\frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= e^{-\lambda}\sum_{x=0}^{\infty} x(x-1)\frac{\lambda^x}{x!} + e^{-\lambda}\sum_{x=0}^{\infty} x\frac{\lambda^x}{x!}$$

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

$$\Rightarrow \mu'_2 = \lambda^2 + \lambda$$

$$\therefore V(X) = \mu_2 = \mu'_2 - \mu'^2_1 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

$$\mu'_3 = E(X^3) = \sum_{x=0}^{\infty} x(x-1)(x-2)e^{-\lambda}\frac{\lambda^x}{x!}$$

$$= e^{-\lambda}\sum_{x=0}^{\infty} x(x-1)(x-2)\frac{\lambda^x}{x!} + 3e^{-\lambda}\sum_{x=0}^{\infty} x(x-1)\frac{\lambda^x}{x!} + e^{-\lambda}\sum_{x=0}^{\infty} x\frac{\lambda^x}{x!}$$

$$= \lambda^3 e^{-\lambda}\left[\sum_{x=3}^{\infty}\frac{\lambda^{x-3}}{(x-3)!}\right] + 3\lambda^2 e^{-\lambda}\left[\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!}\right] + \lambda = \lambda^3 e^{-\lambda}e^{\lambda} + 3\lambda^2 e^{-\lambda}e^{\lambda} + \lambda$$

$$= \lambda^3 + 3\lambda^2 + \lambda$$

$$\mu'_4 = E(X^4) = \sum_{x=0}^{\infty} x^4 p(x,\lambda)$$

$$= \sum_{x=0}^{\infty}\{x(x-1)(x-2)(x-3) + 6x(x-1)(x-2) + 7x(x-1) + x\}e^{-\lambda}\frac{\lambda^x}{x!}$$

$$= e^{-\lambda}\lambda^4\sum_{x=4}^{\infty}\frac{\lambda^{x-4}}{(x-4)!} + 6e^{-\lambda}\lambda^3\sum_{x=3}^{\infty}\frac{\lambda^{x-3}}{(x-3)!} + 7e^{-\lambda}\lambda^2\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!} + \lambda$$

$$= \lambda^4(e^{-\lambda}e^{\lambda}) + 6\lambda^3 e^{-\lambda}e^{\lambda} + 7\lambda^3 e^{-\lambda}e^{\lambda} + \lambda = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$$

Third and fourth order central moment is

$$\mu_2 = \mu'_2 - \mu'^2_1 = (\lambda^2 + \lambda) - \lambda^2 = \lambda$$

$$\mu_3 = \lambda$$

$$\mu_4 = 3\lambda^2 + \lambda$$

Co-efficient of skewness and kurtosis are given by

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\lambda^2}{\lambda^3} = \frac{1}{\lambda} \text{ and}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{\lambda}$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{1}{\sqrt{\lambda}} \quad \text{and}$$

$$\gamma_2 = \beta_2 - 3 = \frac{1}{\lambda}$$

Hence the Poison distribution is always a skewed distribution. Preceding to the limit as $\lambda \to \infty, \beta_1 = 0, \beta_2 = 3$.

Note: In Poisson distribution mean =variance.

Recurrence Relation for Moments of the Poisson Distribution

We know

$$\mu_r = E\left[X - E(X)\right]^r = \sum_{x=0}^{\infty}(x-\lambda)^r\, p(x,\lambda) = \sum_{x=0}^{\infty}(x-\lambda)^r \frac{e^{-\lambda}\lambda^x}{x!}$$

Differentiating w.r.to λ , we get

$$\frac{d\mu_r}{d\lambda} = \sum_{x=0}^{\infty} r(x-\lambda)^{r-1}\frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty}\frac{(x-\lambda)^r}{x!}\left\{x\lambda^{x-1}e^{-\lambda} - \lambda^x e^{-\lambda}\right\}$$

$$= -r\sum_{x=0}^{\infty}(x-\lambda)^{r-1}\frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty}\frac{(x-\lambda)^r}{x!}\left\{\lambda^{x-1}e^{-\lambda}(x-\lambda)\right\}$$

$$= -r\sum_{x=0}^{\infty}(x-\lambda)^{r-1}\frac{e^{-\lambda}\lambda^x}{x!} + \frac{1}{\lambda}\sum_{x=0}^{\infty}(x-\lambda)^{r+1}\frac{\lambda^x e^{-\lambda}}{x!}$$

$$= -r\mu_{r-1} + \frac{1}{\lambda}\mu_{r+1}$$

$$\Rightarrow \mu_{r+1} = r\lambda\mu_{r-1} + \lambda\frac{d\mu_r}{d\lambda}$$

Putting r=1,2, and 3 sucessively, we get

$$\mu_2 = \lambda\mu_0 + \lambda\frac{d\mu_1}{d\lambda} = \lambda,\ \ \mu_3 = 2\lambda\mu_1 + \lambda\frac{d\mu_2}{d\lambda} = \lambda,\ \mu_4 = 3\lambda\mu_2 + \lambda\frac{d\mu_3}{d\lambda} = 3\lambda^2 + \lambda$$

m.g.f. of Poisson Distribution

$$M_X(t) = E\left(e^{tX}\right) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda}\left\{1 + \lambda e^t + \frac{\left(\lambda e^t\right)^2}{2!} + ...\right\} = e^{\lambda\left(e^t - 1\right)}$$

Additive or Reproductive Property of Poisson Variates

The sum of independent Poisson variates is also a Poisson variate. If $X_i(i=1,2,3...,n)$ are independent Poisson variates with parameters $\lambda_i; i=1,2,...,n$ respectively, then $\sum_{i=1}^{n} X_i$ is also a Poisson variate with parameter $\sum_{i=1}^{n}\lambda_i$.

Proof:

$$M_{X_i}(t) = e^{\lambda_i(e^t -)}; \; i = 1, 2, 3, \ldots, n$$

$$M_{X_1 + X_2 + \ldots + X_n}(t) = M_{X_1}(t).M_{X_2}(t)\ldots M_{X_n}(t)$$

$$= e^{\lambda_1(e^t - 1)} e^{\lambda_2(e^t - 1)} \ldots e^{\lambda_n(e^t - 1)} = e^{(\lambda_1 + \lambda_2 \ldots + \lambda_n)(e^t - 1)}$$

which is the m.g.f. of a Poisson variatewith parameter $\lambda_1 + \lambda_2 \ldots + \lambda_n$.

Hence by Uniqueness theorem of m.g.f.'s $\sum_{i=1}^{n} X_i$ is also a Poisson Variate with parameter $\sum_{i=1}^{n} \lambda_i$

Example 7: If the probability that an individual suffers a bad reaction from an injection of a given serum is 0.001, determine the probability that out of 1500 individuals

(i) Exactly 3

(ii) More than 2,

individuals suffer from bad reaction

Solution: Let X be the Poisson variate, "number of individuals suffering from bad reaction". Then

n = 1500, p = 0.001

$\therefore \lambda = np = (1500)(0.001) = 1.5$

By Poisson distribution,

$$P[X = x] = \frac{e^{-\lambda} \lambda^x}{x!}; \; x = 0, 1, 2 \ldots$$

$$= \frac{e^{-1.5} (1.5)^x}{x!}; \; x = 0, 1, 2 \ldots$$

Thus

(i) $$P[X = 3] = \frac{e^{-1.5} (1.5)^3}{3!} = 0.1255$$

$$[e^{-0.5} = 0.6065, e^{-1} = 0.3679 \Rightarrow e^{-1.5} = (0.3679)(0.6065) = 0.2231]$$

(ii) $$P[X > 2] = 1 - P[X \le 2] = 1 - [P(X = 2) + P(X = 1) + P(X = 0)]$$

$$= 1 - \left[\frac{e^{-1.5}(1.5)^2}{2!} + \frac{e^{-1.5}(1.5)^1}{1!} + \frac{e^{-1.5}(1.5)^0}{0!} \right]$$

$$= 1 - e^{-1.5} \left[\frac{2.25}{2} + 1.5 + 1 \right] = 1 - 3.625e^{-1.5} = 0.1913$$

Example 8: If the mean of a Poisson distribution is 1.44, find the values of variance and the central moments of order 3 and 4.

Solution: Mean = 1.44 = λ

Variance = λ = 1.44

$$\mu_3 = \lambda = 1.44$$

$$\mu_4 = 3\lambda^2 + \lambda = 3(1.44)^2 + 1.44 = 7.66$$

Example 9: If a Poisson variate X is s.t. P(X=1) = 2P(X=2). Find the mean and variance of the distribution.

Solution: Let λ be the mean of the distribution, hence by Poisson distribution.

$$P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!}; x = 0,1,2..$$

Now

$$P[X = 1] = 2P[X = 2]$$

$$\Rightarrow \frac{e^{-\lambda}\lambda}{1!} = 2\frac{e^{-\lambda}\lambda^2}{2!}$$

$$\Rightarrow \lambda = \lambda^2$$

$$\Rightarrow \lambda^2 - \lambda = 0$$

$$\Rightarrow \lambda = 0,1$$

But λ = 0 is rejected.

{$\because$ if λ = 0 then either n = 0 or p = 0 which implies that Poisson distribution does not exist in this case}

$\therefore \lambda$ = 1. Hence Mean = λ = 1 and variance = 1.

Problems:

1. A car hire firm has two cars, which it hires out day by day. The number of demands for a car on each day is distributed as a Poisson distribution with mean 1.5. Calculate the proportion of days on which

(i) Neither car is used and

(ii) The proportion of days on which some demand is refused.

2. In a Poisson frequency distribution, frequency corresponding to 3 successes is 2/3 times frequency corresponding to 4 successes. Find the mean and standard deviation of the distribution.

Note: Poisson frequency distribution

$$f(x) = NP[X=x] = N\frac{e^{-\lambda}\lambda^{x}}{x}; x = 0,1,2,... \text{\{If an experiment is repeated N times\}}$$

Fitting of Poisson Distribution

To fit a Poisson distribution to the observed data, we find the theoretical frequencies corresponding to each value of the Poisson variate. Process of finding the probabilities corresponding to each value of the Poisson variate becomes easy if we use recurrence relation for the probability of Poisson distribution.

Recurrence Formula

For a Poisson distribution with parameter λ, we have

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad (1)$$

Changing x to x+1, we have

$$p(x+1) = \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!} \qquad (2)$$

Dividing (2) by (1), we have

$$\frac{p(x+1)}{p(x)} = \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!}.\frac{x!}{e^{-\lambda}\lambda^{x}} = \frac{\lambda}{x+1}$$

$$\Rightarrow p(x+1) = \frac{\lambda}{x+1}p(x)$$

This is the recurrence relation for probability of Poisson distribution. After obtaining the value of p(0) using Poisson probability function i.e.

$p(0) = \frac{e^{-\lambda}\lambda^{0}}{(0)!} = e^{-\lambda}$, we obtain p(1), p(2) and so on….

Example 10: After correcting 50 pages of the proof of a book, the proof reader finds that there are on the average, 2 error per 5 pages. How many pages would one expect to find with 0,1,2,3 and 4 errors, in 1000 pages of the first print of the book?

Solution: Let the random variable X represents the number of error per page. Then the mean number of error per page is given by: $\lambda = \frac{2}{5} = 0.4$. The probability of x error per page is given by:

$$P(X = x) = \frac{e^{-\lambda}\lambda^{-x}}{x!} = \frac{e^{-0.4}(0.4)^{-x}}{x!} \quad ; x = 0,1,2,...$$

Expected number of pages with x errors per page in a book of N=1000 pages are:

$$f(x) = N \times P(X = x) = 1000 \times \frac{e^{-0.4}(0.4)^{-x}}{x!} \quad ; x = 0,1,2,...$$

No. of error per page (x)	Probability p(x)	Expected number of pages f(x)=Np(x, λ)
0	$p(0) = e^{-0.4} = 0.6703$	670.3≃670
1	$p(1) = \frac{e^{-0.4}}{0+1} p(0) = 0.26812$	268 .12≃ 268
2	$p(1) = \frac{e^{-0.4}}{1+1} p(1) = 0.053624$	53.62454≃54
3	$p(1) = \frac{e^{-0.4}}{2+1} p(2) = 0.0071298$	7.1298≃7
4	$p(1) = \frac{e^{-0.4}}{3+1} p(3) = 0.00071298$	0.71298≃1

Problems:

1. Fit a Poisson distribution to the following data which gives the number of datasets in a sample of clover seeds:

Number of doddens(x) :	0	1	2	3	4	5	6	7	8
Observed Frequency(f) :	56	156	132	92	37	22	4	0	1

2. If X has a Poisson distribution such that $P(X = 1) = P(X = 2)$, evaluate P(X=4).

3. The following data give frequency of aircraft accidents experienced by 2,546 pilots during a four-year period:

No. of accidents(x) :	0	1	2	3	4	5
Frequencies(f) :	2036	422	71	13	3	1

Calculate mean number of accidents and fit a Poisson distribution.

4. In a certain factory turning out fountain pens, there is a small chance, $\frac{1}{500}$, for any pen to be defective. The pens are supplied in packets of 10. Calculate the approximate number of packets containing (i) one defective (ii) two defective pens in a consignment of 20000 packets.

3.1.5 Geometric Distribution

This distribution is used when we have to wait first success after failure so many times. A r.v. is said to have geometric distribution if it assumes only non-negative values and its p.m.f. is given by:

$$P(X=x)=\begin{cases} q^x p; x=0,1,2,\ldots,; 0<p\le 1; q=1-p \\ 0; \qquad otherwise \end{cases}$$

The random variable X represents the number of Bernoulli trials up to which the first success occurs. This is the number of trials having failure before first success. Here p is the probability of success in a single trial.

Clearly assignment of probabilities is permissible, since

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

Example 11: An unbiased die is cast until 6 appears. What is the probability that it must be cast more than five times?

Solution: Let p be the probability of a success i.e. getting 6 in a throw of the die

$$\therefore\ p=\frac{1}{6} \quad \text{and } q=1-\frac{1}{6}=\frac{5}{6}$$

Let X be the number of failure preceding the first success.

$$P[X=x]=q^x p;\ \text{for } x=0,1,2,\ldots$$

P[The number of failure preceding the first success is at least 5]= $P[X\ge 5]$

$$=P[X=5]+P[X=6]+P[X=7]+\ldots$$

$$=\left(\frac{5}{6}\right)^5\frac{1}{6}+\left(\frac{5}{6}\right)^6\left(\frac{1}{6}\right)+\left(\frac{5}{6}\right)^7\frac{1}{6}+\ldots$$

$$=\left(\frac{5}{6}\right)^5\left(\frac{1}{6}\right)\left(1+\frac{5}{6}+\left(\frac{5}{6}\right)^2+\left(\frac{5}{6}\right)^3+\ldots\right)=\left(\frac{5}{6}\right)^5$$

Moments of Geometric Distribution

Mean= $\mu_1' = \sum_{x=0}^{\infty} xP(X=x) = \sum_{x=1}^{\infty} xq^x p = pq\sum_{x=1}^{\infty} xq^{x-1} = pq(1-q)^{-2} = \frac{q}{p}$

Variance= $E(X^2) - [E(X)]^2 = E[X(X-1)] + E(X) - [E(X)]^2$

$$= \quad E[X(X-1)] = \sum_{x=1}^{\infty} x(x-1)q^x p = \sum_{x=2}^{\infty} x(x-1)q^x p = 2pq^2 \sum_{x=1}^{\infty} \frac{x(x-1)}{2} q^{x-2} = 2pq^2(1-q)^{-3} = \frac{2q^2}{p^2}$$

$$\therefore \; V(x) = \mu_2 = \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2} = \frac{q}{p^2}$$

Note: The variance of geometric distribution is always greater than mean.

Moment Generating Function

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} q^x p = p\sum_{x=0}^{\infty} (e^t q)^x$$

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

Remark: In Geometric distribution, Variance= $\frac{q}{p^2} = \frac{\text{Mean}}{p}$ >Mean

Hence for Geometric distribution variance is greater than mean.

Problems:

1. Probability of hitting a target in any attempt is 0.6, what is the probability that it would be hit on fifth attempt?
2. Determine the geometric distribution for which the mean is 3 and variance is 4.

Lack of Memory

This is the only discrete distribution which has the property of 'memoryless' or' forgetfulness'. For example, in a random experiment satisfying geometric distribution the wait upto 3 trials for first success does not affect the probability that one has to wait for a further 5 trials if it is given that the first three trials are failure.

Suppose an event occurs at one of the trail 1, 2,3,4,…and the occurrence time X has a geometric distribution with probability p. Let X is the number of trials preceding to the successful attempt. Thus

$$P[X \geq j] = P[X=j] + P[X=j+1] + P[X=j+2] + ...$$

$$= q^j p + q^{j+1} p + q^{j+2} p + ...$$

$$= q^j p[1 + q + q^2 + ...]$$

$$= q^j p\left[\frac{1}{1-q}\right]$$

$$= q^j$$

Now

$$P[X \geq j+k \mid X \geq j] = \frac{P[X \geq j+k \mid X \geq j]}{P[X \geq j]}$$
$$= \frac{P[(X \geq j+k) \cap (X \geq j)]}{P[X \geq j]}$$
$$= \frac{P[X \geq j+k]}{P[X \geq j]} = \frac{q^{j+k}}{q^j}$$
$$= q^k = P[X \geq k]$$

So, $P[X \geq j+k \mid X \geq j] = P[X \geq k]$

The above result reveals that the conditional probability of at least first j+k trials are unsuccessful before the first success given that at least first j trial were unsuccessful, is the same as the probability that the first k trials were unsuccessful. So, the probability to get first success remains same if we start counting of k unsuccessful trials from anywhere provided all the trials preceding to it are unsuccessful, i.e. future does not depend on the past, it depends only on the present. This distribution is called memoryless or forgetfulness because it forgets the preceding trials.

3.2 Continuous Distributions

In this section, we will discuss some univariate continuous distributions such as uniform distribution, normal distribution, and exponential distribution in details.

3.2.1 Uniform Distribution

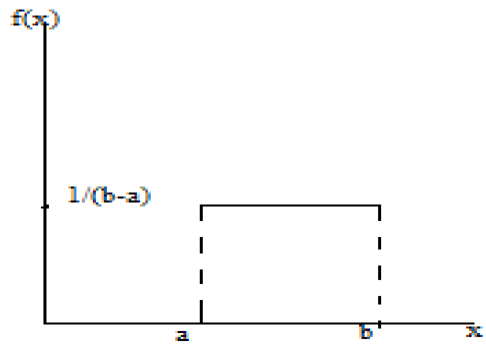
A random variable X is said to be uniformly distributed in $a \leq x \leq b$ if its density function is

$$f(x;a,b) = \begin{cases} \frac{1}{b-a}; & a \leq x \leq b \\ 0; & otherwise \end{cases}$$

Here a$(-\infty < a < \infty)$ and b$(-\infty < b < \infty)$ are the parameters of the distribution. A random variable which follows uniform distribution is written as $X \sim U[a,b]$

Remarks:

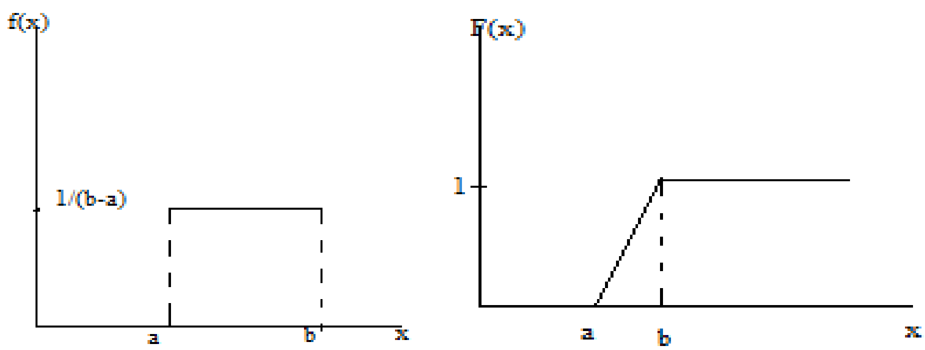
1. This distribution is called uniform distribution because the probability has a uniform value over the interval [a,b].
2. This distribution is also called rectangular distribution because if we draw the graph *y=f(x)* over x-axis and between the ordinates *x=a* and *x=b*, it describe a rectangle as shown in below figure.



3. A cumulative distribution function F(x) is given by:

$$F(x) = \begin{cases} 0; x \le a \\ \dfrac{x-a}{b-a}; a < x < b \\ 1; x \ge b \end{cases}$$

4. The graphs of uniform p.d.f. f(x) and the corresponding cumulative distribution function F(x) are given below



5. For a rectangular or uniform variate X in (-a, a), the p.d.f. is given by:

$$f(x) = \begin{cases} \dfrac{1}{2a}; -a < x < a \\ 0; otherwise \end{cases}$$

Moments of Rectangular Distribution

Let X ~ U[a, b]

$$\mu'_r = \int_a^b x^r f(x)dx = \frac{1}{b-a}\int_a^b x^r dx = \frac{1}{b-a}\left(\frac{b^{r+1} - a^{r+1}}{r+1}\right)$$

In particular for r=1

$$\text{mean} = \mu'_1 = \frac{1}{b-a}\left(\frac{b^2-a^2}{2}\right) = \frac{b+a}{2}$$

$$\mu'_2 = \frac{1}{b-a}\left(\frac{b^3-a^3}{3}\right) = \frac{b^2+ab+a^2}{3}$$

$$\text{Variance} = \mu'_2 - \mu'^2_1 = \frac{(b^2+ab+a^2)}{3} - \left\{\frac{(b+a)}{2}\right\}^2 = \frac{(b-a)^2}{12}$$

Moments Generating Function of Rectangular Distribution

$$M_X(t) = \int_a^b e^{tx} f(x)dx = \int_a^b \frac{e^{tx}}{b-a}dx = \frac{e^{bt}-e^{at}}{t(b-a)}, t \neq 0$$

Mean Deviation about Mean (η)

η of Rectangular distribution is given by:

$$\eta = E\left|X - Mean\right| = \int_a^b \left|x - Mean\right| f(x)dx = \frac{1}{b-a}\int_a^b \left|x - \frac{a+b}{2}\right| dx$$

$$\Rightarrow \eta = \frac{1}{b-a}\int_{\frac{-(b-a)}{2}}^{\frac{b-a}{2}} \left|t\right| dt$$

where $t = x - \dfrac{a+b}{2}$

$$\Rightarrow \eta = \frac{1}{b-a}.2\int_0^{\frac{b-a}{2}} tdt = \frac{b-a}{4}$$

Example 12: If X is uniformly distributed with mean 1 and Variance 4/3, find P(X<0).

Solution: Let X ~ U[a, b], so that

$$p(x) = \frac{1}{b-a}, a < x < b$$

We are given:

Mean=

$$\frac{1}{2}(b+a) = 1 \Rightarrow b+a = 2$$

and Variance $= \frac{1}{12}(b-a)^2 = \frac{4}{3} \Rightarrow b-a = \pm 4$

solving, we get a = -1 and b = 3; (a<b)

$$\therefore\ p(x) = \frac{1}{4}; -1 < x < 3$$

$$P(X < 0) = \int_{-1}^{0} p(x)dx = \frac{1}{4}\left|x\right|_{-1}^{0} = \frac{1}{4}$$

Example 13: 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 atleast twenty minutes?

Solution: Let the r.v. X denotes the waiting time for the next train. Under the assumption that a man arrives at the random, X is distributed uniformly on (0, 30) with p.d.f.

$$f(x) = \begin{cases} \frac{1}{30}; 0 < x < 30 \\ 0; otherwise \end{cases}$$

The probability that he has to wait atleast 20 minutes is given by:

$$P(X \geq 20) = \int_{20}^{30} f(x)dx = \frac{1}{30}\int_{20}^{30} 1.dx = \frac{1}{30}(30-20) = \frac{1}{3}$$

Problems:

1. If X is uniformly distributed with mean 2 and variance 12. Find P[X<3].
2. A random variable X has a uniform distribution over (-2,2). Find k for which $P[X > k] = \frac{1}{2}$
3. Metro trains are scheduled every 5 minutes at a certain station. A person comes to the station at a random time. Let the random variable X count the number of minutes he/she has to wait for the next train. Assume X has a uniform distribution over the interval (0,5). Find the probability that he/she has to wait at least 3 minutes for the train.

3.2.2 Exponential Distribution

The exponential distribution occurs in many different connections such as the radioactive or particle decays or the time between events in a Poisson process where events happen at a constant rate. So exponential distribution serves as a good model whenever there is a waiting time involved for a specific event to occur e.g. waiting time for a failure to occur in a machine.

Definition: A r.v. X is said to have an exponential distribution with parameter θ>0, if its p.d.f. is given by:

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

The cumulative distribution function F(x) is given by:

$$F(x) = \int_0^x f(u)du = \theta \int_0^x \exp(-\theta u)du$$

$$F(x) = \begin{cases} 1-\exp(-\theta x); x \geq 0 \\ 0; otherwise \end{cases}$$

Moment Generating Function of Exponential Distribution

$$M_X(t) = E(e^{tx}) = \theta \int_0^\infty e^{tx} e^{-\theta x} dx = \theta \int_0^\infty \exp\{-(\theta - t)x\}dx$$

$$= \frac{\theta}{\theta - t} = \left(1 - \frac{t}{\theta}\right)^{-1} = \sum_{r=0}^{\infty} \left(\frac{t}{\theta}\right)^r ; \theta > t$$

$\therefore \mu_r' = E(X^r) =$ coeff. of $\frac{t^r}{r!}$ in $M_X(t) = \frac{r!}{\theta^r}$, r=1,2,3,…

Mean = $\mu'_1 = \frac{1}{\theta}, \mu'_2 = \frac{2}{\theta^2}$ and variance = $\mu_2 = \mu'_2 - \mu'^2_2 = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2}$

Hence if X ~ exp(θ), then Mean = $\frac{1}{\theta}$ and variance = $\frac{1}{\theta^2}$

Remark: Variance $= \frac{1}{\theta^2} = \frac{1}{\theta} . \frac{1}{\theta} = \frac{Mean}{\theta}$

∴ Variance > Mean, if 0< θ<1

Variance = Mean, if θ = 1

Variance < Mean, if θ>1

Theorem: If $X_1, X_2, \ldots, X_n$ are independent r.v.'s, X_i having an exponential distribution with parameter θ_i; $i = 1, 2, 3, \ldots, n$, then $Z = min(X_1, X_2, \ldots, X_n)$ has exponential distribution with parameter $\sum_{i=1}^{n} \theta_i$.

Proof: Cumulative distribution of Z is

$$G_Z(Z) = P(Z \le z) = 1 - P(Z > z)$$

$$= 1 - P[\min(X_1, X_2, \ldots, X_n) > z]$$

$$= 1 - P(X_i > z; i = 1, 2, ...n)$$

$$= 1 - \prod_{i=1}^{n} P(X_i > z) \qquad (\because X_1, X_2, \ldots, X_n \text{ are independent})$$

$$G_Z(z) = 1 - \prod_{i=1}^{n} [1 - P(X_i \le z)] = 1 - \prod_{i=1}^{n} [1 - F_{X_i}(z)]$$

(where F is the distribution function of X_i)

$$G_Z(z) = 1 - \prod_{i=1}^{n} [1 - (1 - e^{-\theta_i z})] = \begin{cases} 1 - \exp\{(-\sum_{i=1}^{n} \theta_i) z\}; z > 0 \\ 0; \qquad otherwise \end{cases}$$

$$\therefore g_Z(z) = \frac{d}{dz} G(z) = \begin{cases} \left(\sum_{i=1}^{n} \theta_i\right) \exp\left\{\left(-\sum_{i=1}^{n} \theta_i\right) z\right\}; z > 0 \\ 0 \qquad ; otherwise \end{cases}$$

$\Rightarrow Z = \min(X_1, X_2, ..., X_n)$ is an exponential variate with parameter $\sum_{i=1}^{n} \theta_i$

Memoryless Property of Exponential Distribution

Exponential distribution is the only continuous distribution which has the memoryless property given by:

$$P[X \le x + a \mid X \ge a] = P[X \le x] \text{ for all } x, \text{ where } Y = X - a$$

i.e. the conditional probability of waiting upto the time 'x+a' given that it exceeds 'a' is same as the probability of waiting upto the time 'x'.

Proof: The p.d.f of the exponential distribution with parameter θ is

$$f(x) = \theta \exp(-\theta x); \theta > 0, 0 < x < \infty$$

We have

$$P(Y \le x \cap X \ge a) = P(X - a \le x \cap X \ge a) \text{ where } Y = X\text{-}a$$

$$= P(X \le a + x \cap X \ge a) = P(a \le X \le a + x)$$

$$= \theta \int_a^{a+x} e^{-\theta x} dx = e^{-a\theta}(1 - e^{-\theta x})$$

and

$$P(X \geq a) = \theta \int_a^{\infty} e^{-\theta x} dx = e^{-a\theta}$$

$$\therefore P(Y \leq x \mid X \geq a) = \frac{P(Y \leq x \cap X \geq a)}{P(X \geq a)} = 1 - e^{-\theta x} \quad (1)$$

Also

$$P(X \leq x) = \theta \int_0^{x} e^{-\theta x} dx = 1 - e^{-ax} \quad (2)$$

From (1) and (2), we get

$$P(Y \leq x \mid X \geq a) = P(X \leq x)$$

i.e. $P(X \leq x + a \mid x \geq a) = P(X \leq x)$

i.e. exponential distribution lacks memory.

Example 14: Telephone calls arrive at a switchboard following an exponential distribution with parameter $\lambda = 12$ per hour. If we are at the switchboard, what is the probability that the waiting time for a call is

(i) atleast 15 minutes

(ii) not more than 10 minutes.

Solution: Let X be the waiting time(in hours) for a call

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

$$\Rightarrow F(x) = P[X \leq x] = 1 - e^{-\lambda x} = 1 - e^{-12x} \quad (1)$$

[where λ = 12]

Now

(i) P[waiting time is atleast 15 minutes]

$$= \text{P[waiting time is atleast } \frac{1}{4} \text{ hours]}$$

$$= P\left[X \geq \frac{1}{4}\right] = 1 - P\left[X < \frac{1}{4}\right]$$

$$= 1 - \left[1 - e^{-12 \times \frac{1}{4}}\right]$$

$$= e^{-3} = 0.0498$$

(ii) P[waiting time more than 10 minutes]

= P[waiting time not more than $\frac{1}{6}$ hours]

$$= P\left[X \leq \frac{1}{6}\right] = 1 - e^{-12 \times \frac{1}{6}}$$

$$= 1 - e^{-2} = 1 - (0.1353) = 0.8647$$

Problems:

1. Show that for the exponential distribution $f(x) = Ae^{-x},\ 0 \leq x < \infty$, mean and variance are equal.
2. Find the value of k>0 for which the function given by:

   $f(x) = 2e^{-kx},\ x \geq 0$

   follows an exponential distribution.

3.2.3 Normal Distribution

The concept of normal distribution was initially discovered by English mathematician Abraham De Moivre (1667-1754) in 1733. The Normal Distribution is the most important and most widely used continuous probability distribution. It is the cornerstone of the application of statistical inference in analysis of data because the distributions of several important sample statistics tend towards a Normal distribution as the sample size increases. Empirical studies have indicated that the Normal distribution provides an adequate approximation to the distributions of many physical variables. Specific examples include meteorological data, such as temperature and rainfall, measurements on living organisms, scores on aptitude tests, physical measurements of manufactured parts, weights of contents of food packages, volumes of liquids in bottles/cans, instrumentation errors and other deviations from established norms, and so on.

The normal distribution has a unique position in probability theory because it can be used as an approximation to most of the other distributions.

Examples of Normal Distribution

1. The age at first calving of cows belonging to the same breed and living under similar environmental conditions tend to normal frequency distribution.
2. The milk yield of cows in a large herd tends to follow a normal frequency distribution.
3. The chemical constituents of milk like fat, SNF, protein etc., for large samples follow normal distribution.

Definition: A r.v. X is said to have a normal distribution with parameters μ (called mean) and σ^2 (called variance) if its p.d.f. is given by the probability law:

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

which may also be written as

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad ; -\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma > 0$$

where μ and σ are the parameters.

Remarks:

1. It is customary to write X is normally distributed as $N(\mu, \sigma^2)$ and is expressed by $X \sim N(\mu, \sigma^2)$
2. If $X \sim N(\mu, \sigma^2)$, then $Z=\frac{X-\mu}{\sigma}$ is a standard normal variate with E(Z)=0 and Var(Z) = 1 and we write $Z \sim N(0, 1)$
3. The p.d.f. of standard normal variate Z is given by

$$\varphi(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} ; -\infty < z < \infty$$

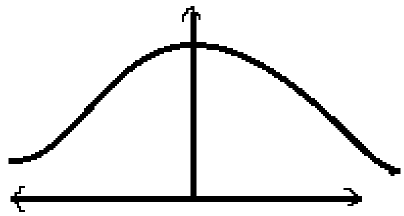
and the corresponding distribution function, denoted by

$$\phi(z)=P(Z\le z)=\int_{-\infty}^{z}\varphi(u)du=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-u^2/2}du$$

Results:

1. $\phi(-Z)=1-\phi(Z), Z>0$
2. $P(a\le X\le b)=\phi(\frac{b-\mu}{\sigma})-\phi(\frac{a-\mu}{\sigma})$ where $X \sim N(\mu, \sigma^2)$
3. The graph of f(x) is famous "bell-shaped" curve.

The top of the bell is directly above the mean μ. For larger values of σ, the curve tends to flattened for small values of σ, and it has a sharp peak.



The normal distribution has various properties and a large number of applications. The theory of estimation of population parameters and testing of hypothesis on the basis of sample statistics have also been developed using the concept of normal distribution as most of the sampling distribution tends to normally for large samples.

Normal Distribution as a Limiting Case of Binomial Distribution

Normal distribution as a limiting case of binomial distribution under the following conditions:

1. n, the number of trials, is indefinitely large i.e., $n \to \infty$
2. neither p nor q is too close to zero.

Under these conditions, the Binomial distribution can be closely associated by a normal distribution with a standardized variable given by

$$Z = \frac{X - np}{\sqrt{npq}}, X = 0, 1, 2, ..., n \qquad (1)$$

The p.m.f. of the Binomial distribution with parameters n and p is given by:

$$p(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}; x = 0, 1, 2, ..., n \qquad (2)$$

When X = 0, $Z = \frac{-np}{\sqrt{npq}} = -\sqrt{\frac{np}{q}}$ and

When X = n, $Z = \frac{n - np}{\sqrt{npq}} = \sqrt{\frac{nq}{p}}$

Thus in the limit as $n \to \infty$, Z takes the value from $-\infty$ to ∞.

Using Sterling's approximation to r! for larger r.

$$\lim_{r \to \infty} r! \simeq \sqrt{2\pi} e^{-r} . r^{(r+\frac{1}{2})}$$

We have

$$\lim p(x) = \lim\left[\frac{\sqrt{2\pi}e^{-n}n^{n+\frac{1}{2}}p^x q^{n-x}}{\sqrt{2\pi}e^{-x}x^{x+\frac{1}{2}}\sqrt{2\pi}e^{-(n-x)}(n-x)^{n-x+\frac{1}{2}}}\right]$$

$$= \lim\left[\frac{(np)^{x+\frac{1}{2}}(nq)^{n-x+\frac{1}{2}}}{\sqrt{2\pi}\sqrt{npq}\,x^{x+\frac{1}{2}}(n-x)^{n-x+\frac{1}{2}}}\right]$$

$$= \lim\left[\frac{1}{\sqrt{2\pi}\sqrt{npq}}\frac{(np)^{x+\frac{1}{2}}}{x^{x+\frac{1}{2}}}\frac{(nq)^{n-x+\frac{1}{2}}}{(n-x)^{n-x+\frac{1}{2}}}\right] = \lim\left[\frac{1}{\sqrt{2\pi}\sqrt{npq}}\left(\frac{np}{x}\right)^{x+\frac{1}{2}}\left(\frac{nq}{n-x}\right)^{n-x+\frac{1}{2}}\right] \qquad (3)$$

From (1), we get

$$X = np + Z\sqrt{npq} \Rightarrow \frac{X}{np} = 1 + Z\sqrt{\frac{q}{np}} \qquad (4)$$

Further

$$n - X = n - np - Z\sqrt{npq} = nq - Z\sqrt{npq} \Rightarrow \frac{n-X}{nq} = 1 - Z\sqrt{\frac{p}{nq}} \qquad (5)$$

Also $dZ = \dfrac{1}{\sqrt{npq}}dx$

Hence, the probability differential from equation (3) is

$$dG(z) = g(z)dz = \lim_{n\to\infty}\left(\frac{1}{\sqrt{2\pi}} \times \frac{1}{N}\right)dz$$

where $N = \left(\dfrac{x}{np}\right)^{x+\frac{1}{2}}\left(\dfrac{n-x}{nq}\right)^{n-x+\frac{1}{2}}$

$$\Rightarrow \log N = \left(x + \frac{1}{2}\right)\log\left(\frac{x}{np}\right) + \left(n - x - \frac{1}{2}\right)\log\left(\frac{n-x}{nq}\right) \quad \left(\text{Using } Z = \frac{X-np}{\sqrt{npq}}\right)$$

$$= \left(np + z\sqrt{npq} + \frac{1}{2}\right)\log\left\{1 + z\sqrt{\frac{q}{np}}\right\} + \left(nq - z\sqrt{npq} + \frac{1}{2}\right)\log\left\{1 - z\sqrt{\frac{p}{nq}}\right\} \text{ (using (4) and (5))}$$

$$=\left(np+z\sqrt{npq}+\frac{1}{2}\right)\left\{z\sqrt{\frac{q}{np}}-\frac{1}{2}z^2\left(\frac{q}{np}\right)+\frac{1}{3}z^3\left(\frac{q}{np}\right)^{3/2}-\ldots\right\}+\left(nq-z\sqrt{npq}+\frac{1}{2}\right)\left\{-z\sqrt{\frac{p}{nq}}-\frac{1}{2}z^2\left(\frac{p}{nq}\right)-\frac{1}{3}z^3\left(\frac{p}{nq}\right)^{3/2}-\ldots\right\}$$

$$=\left[\begin{array}{l}\left\{z\sqrt{npq}-\frac{1}{2}qZ^2+\frac{1}{3}z^3\left(\frac{q^{3/2}}{\sqrt{np}}\right)+z^2q-\frac{1}{2}z^3\frac{q^{3/2}}{\sqrt{np}}+\frac{1}{2}z\sqrt{\frac{q}{np}}-\frac{1}{4}z^2\frac{q}{np}+\ldots\right\}\\+\left\{-z\sqrt{npq}-\frac{1}{2}pz^2-\frac{1}{3}z^3\left(\frac{p^{3/2}}{\sqrt{nq}}\right)+z^2p-\frac{1}{2}z^3\frac{p^{3/2}}{\sqrt{nq}}-\frac{1}{2}z\sqrt{\frac{p}{nq}}-\frac{1}{4}z^2\frac{p}{nq}+\ldots\right\}\end{array}\right]$$

$$=\left[-\frac{1}{2}z^2(p+q)+z^2(p+q)+\frac{z}{2\sqrt{n}}\left(\frac{q}{p}+\frac{p}{q}\right)+O(n^{-1/2})\right]$$ Substituting in (6)

$$=\frac{z^2}{2}+O(n^{-1/2})\rightarrow\frac{z^2}{2}\Rightarrow\lim_{n\to\infty}N=e^{z^2/2}$$

we get

$$dG(z)=g(z)dz=\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz;-\infty<z<\infty \qquad (6a)$$

Hence probability function of Z is

$$g(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2};-\infty<z<\infty \qquad (6b)$$

From (6b) using standard variate $Z=\frac{X-\mu}{\sigma}$

We have $f_X(x)=\begin{cases}\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}, & -\infty<x<\infty\\ 0, & otherwise\end{cases}$

Note: Normal distribution can also be obtained as a limiting case of Poisson distribution with the parameter $\lambda\rightarrow\infty$.

Example 15: If X ~ N(40, 25), then write down the p.d.f. of X.

Solution: X ~ N(40, 25) in usual notations,

$\mu=40,\sigma^2=25\Rightarrow\sigma=\pm\sqrt{25}=5\ \{\because\ \sigma>0\}$

Now, the p.d.f. of random variable X is given by

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}=\frac{1}{5\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-40}{5}\right)^2},-\infty<x<\infty$$

Problems:

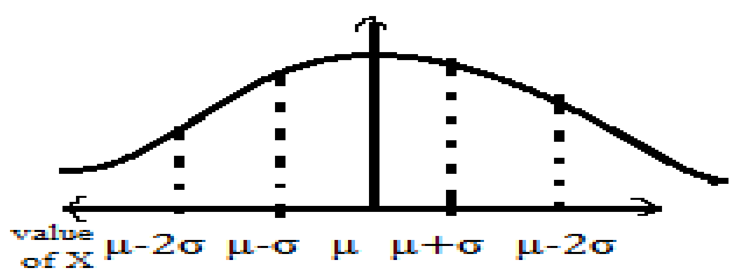
1. Write down the pdf of r.v. X ~ $N(\frac{1}{2}, \frac{4}{9})$.

2. Find the mean and variance if the p.d.f. is $f(x) = \frac{1}{2\sqrt{2\pi}} e^{\frac{-x^2}{8}}, \quad -\infty < x < \infty$

Chief Characteristics of Normal Distribution:

The normal probability distribution with mean μ and variance σ^2 has the following properties:

1. The curve of the normal distribution is bell shaped.
2. The curve of the distribution is completely symmetrical about X = μ.
3. For normal distribution, mean = median = mode.
4. f(x) being the probability can never be negative and hence no portion of the curve lies below x-axis.
5. Though x-axis becomes closer and closer to the normal curve as the magnitude of the value of x goes towards ∞ or -∞, yet it never touches it.
6. Normal curve has only one mode.



7. Central moments of normal distribution are

   $\mu_1 = 0, \mu_2 = \sigma^2, \mu_3 = 0, \mu_4 = 3\sigma^4$

   and

   $\beta_1 = 0, \beta_2 = 3$

   $\Rightarrow \mu_{2r+1} = 0, \mu_{2r} = 1.3.5...(2r-1)\sigma^{2r}$

   (r = 0, 1, 2, …)

   (all odd order central moments are 0 for normal distribution)

8. For normal curve Q_3-Median = Median-Q_1

9. Quartile deviation (Q.D.) $= \frac{Q_3 - Q_1}{2} \approx \frac{2}{3}$ of the standard deviation.

10. Mean deviation ≈ 4/5 of the standard deviation.

11. Q.D:M.D:S.D = $\frac{2}{3}\sigma : \frac{4}{5}\sigma : \sigma = 10:12:15$

12. The point of inflexion of the curve are

$X = \mu \pm \sigma$

13. If $X_1, X_2, X_3, \ldots X_n$ are independent normal variables then the linear combination of $a_1X_1 + a_2X_2 + a_3X_3 ... + a_nX_n$ is also normal variable with

mean $a_1\mu_1 + a_2\mu_2 + a_3\mu_3 + ... + a_n\mu_n$ and variance $a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + a_3^2\sigma_3^2 + \ldots a_n^2\sigma_n^2$

14. Particularly sum or difference of two independent normal variate is also normal variate.

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

$$X - Y \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

15. Area property:

$$P[\mu - \sigma < X < \mu + \sigma] = \int_{\mu-\sigma}^{\mu+\sigma} f(x)dx = 0.6827$$

$$P[\mu - 2\sigma < X < \mu + 2\sigma] = \int_{\mu-2\sigma}^{\mu+2\sigma} f(x)dx = 0.9544$$

$$P[\mu - 3\sigma < X < \mu + 3\sigma] = \int_{\mu-3\sigma}^{\mu+3\sigma} f(x)dx = 0.9973$$

Mode and Median of Normal Distribution

Mode: Mode is the value of x for which f(x) is maximum, i.e. mode is the solution of

$$f'(x) = 0 , f''(x) < 0$$

Now let $X \sim N(\mu, \sigma^2)$, the p.d.f. of X is

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

Taking log on both side

$$\log f(x) = \log\frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2}(x-\mu)^2 \log e$$

$$= \log \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2}(x-\mu)^2$$

Differentiating w.r.t. x

$$\frac{1}{f(x)} f^{'}(x) = 0 - \frac{1}{2\sigma^2} 2(x-\mu) = -\frac{x-\mu}{\sigma^2}$$

$$\Rightarrow f^{'}(x) = -\frac{(x-\mu)}{\sigma^2} f(x) \qquad (2)$$

For maximum or minimum

$$f^{'}(x) = 0 \Rightarrow -\left(\frac{x-\mu}{\sigma^2}\right) f(x) = 0$$

$$\Rightarrow x - \mu = 0 \Rightarrow x = \mu$$

as f(x) is never zero.

Now differentiate (2) w.r.t. 'x', we have

$$f^{''}(x)\Big|_{x=\mu} = f^{'}(x)\frac{(x-\mu)}{\sigma}\Big|_{x=\mu} - \frac{1}{\sigma^2} f(x)\Big|_{x=\mu}$$

$$f^{''}(x)\Big|_{x=\mu} = \frac{-f(\mu)}{\sigma^2} < 0$$

$\therefore x = \mu$ is Mode.

Median: We know that median M divides the distribution into two equal parts.

$$\int_{-\infty}^{M} f(x)dx = \int_{M}^{\infty} f(x)dx = \frac{1}{2} \Rightarrow \int_{-\infty}^{M} f(x)dx = \frac{1}{2}$$

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

In the first integral, let us put

$$\frac{x-\mu}{\sigma} = z$$

$$\therefore dx = \sigma dz$$

Also when $x = \mu \Rightarrow z = 0$ and when $x \to -\infty \Rightarrow z \to -\infty$

We have

$$\Rightarrow \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz + \int_{\mu}^{M} f(x)dx = \frac{1}{2}$$

$$\Rightarrow \frac{1}{2} + \int_{\mu}^{M} f(x)dx = \frac{1}{2}$$

$$\Rightarrow \int_{\mu}^{M} f(x)dx = 0$$

$$\Rightarrow M = \mu$$

as f(x)≠0

Hence Mean = Median = Mode = μ

Mean Deviation about Mean

Mean deviation about mean for normal distribution is

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

Put $\frac{x-\mu}{\sigma} = z \Rightarrow dz = \frac{dx}{\sigma}$

M.D. about mean $== \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z| e^{-\frac{1}{2}z^2} dz$

M.D. about mean $= \frac{2\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} |z| e^{-\frac{1}{2}z^2} dz$

{∵ If f(x) is even function $\int_{-a}^{a} f(x)dx = 2\int_{0}^{a} f(x)dx$ }

Put $\frac{z^2}{2} = t \Rightarrow z^2 = 2t \Rightarrow 2zdz = 2dt \Rightarrow zdz = dt$

M.D. about Mean $= \frac{2\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-t} dt = \sqrt{\frac{2}{\pi}}\sigma[-0+1] = \sqrt{\frac{2}{\pi}}\sigma$

M.D. about Mean = $\sqrt{\frac{2}{\pi}}\sigma \approx \frac{4}{5}\sigma$ (approx.)

m.g.f. of Normal Distribution

The m.g.f. (about origin) is given by:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{t(\mu + \sigma z)\} \exp(-z^2/2)dz \qquad \left(z = \frac{x-\mu}{\sigma}\right)$$

$$= e^{\mu t} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(z^2 - 2t\sigma z)\right\}dz$$

$$= e^{\mu t} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\{(z-\sigma t)^2 - \sigma^2 t^2\}\right]dz$$

$$= e^{\mu t + t^2\sigma^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(z-\sigma t)^2\right]dz$$

$$= e^{\mu t + t^2\sigma^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-u^2/2)du$$

Hence $M_X(t) = e^{\mu t + t^2\sigma^2/2}$ (1)

Note:

m.g.f. of Standard Normal Variate: If $X \sim N(\mu, \sigma^2)$, then standard normal variate is given by

$Z = (X - \mu)/\sigma$

$M_Z(t) = e^{-\mu t/\sigma} M_X(t/\sigma) = \exp(-\mu t/\sigma)\exp\{(\mu t/\sigma) + (t^2/\sigma^2)(\sigma^2/2)\} = \exp(t^2/2)$ (1a)

Cumulant Generating Function (C.G.F) of Normal Distribution

The c.g.f of normal distribution is given by:

$$K_X(t) = \log_e M_X(t) = \log_e(e^{\mu t + t^2\sigma^2/2}) = \mu t + \frac{t^2\sigma^2}{2}$$

$\therefore$Mean = κ_1 = coefficient of t in $K_X(t)$ = μ

Variance = κ_2 = coefficient of $\frac{t^2}{2!}$ in $K_X(t)$ = σ^2

and κ_r = coefficient of $\frac{t^r}{r!}$ in $K_X(t)$ = 0; r=3,4…

Thus $\mu_3 = \kappa_3 = 0$ and $\mu_4 = \kappa_4 + 3\kappa_2^2 = 3\sigma^4$

Hence $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$ and $\beta_2 = \frac{\mu_4}{\mu_2^2} = 3$ (2)

Moments of Normal Distribution: Odd order moments about mean are given by:

$$\mu_{2n+1} = \int_{-\infty}^{\infty} (x-\mu)^{2n+1} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^{2n+1} \exp\left\{-(x-\mu)^2 / 2\sigma^2\right\} dx$$

$$\mu_{2n+1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1} \exp\left\{-(z)^2 / 2\right\} dz \qquad \left(z = \frac{x-\mu}{\sigma}\right)$$

$$\mu_{2n+1} = \frac{\sigma^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (z)^{2n+1} \exp\left\{-(z)^2 / 2\right\} dz = 0 \qquad (3)$$

Since the integrand $z^{2n+1}e^{-z^2/2}$ is an odd function of z.

Even order moments about mean are given by:

$$\mu_{2n} = \int_{-\infty}^{\infty} (x-\mu)^{2n} f(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n} \exp\left\{-(z)^2 / 2\right\} dx$$

$$= \frac{\sigma^{2n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (z)^{2n} \exp\left\{-(z)^2 / 2\right\} dz = \frac{\sigma^{2n}}{\sqrt{2\pi}} 2\int_{0}^{\infty} (z)^{2n} \exp\left\{-(z)^2 / 2\right\} dz$$

$$= \frac{\sigma^{2n} 2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-t} (2t)^n \frac{dt}{\sqrt{2t}} = 0, \quad (t = \frac{z^2}{2})$$

$$= \frac{\sigma^{2n} 2^n}{\sqrt{\pi}} \int_{0}^{\infty} e^{-t} t^{(n+1/2)-1} dt = 0$$

$$\Rightarrow \mu_{2n} = \frac{\sigma^{2n} 2^n}{\sqrt{\pi}} \Gamma\left(n + \frac{1}{2}\right)$$

Changing n to (n-1), we get

$$\mu_{2n-2} = \frac{2^{n-1}.\sigma^{2n-2}}{\sqrt{\pi}} \Gamma\left(n - \frac{1}{2}\right)$$

$$\therefore \frac{\mu_{2n}}{\mu_{2n-2}} = 2\sigma^2 . \frac{\Gamma\left(n + \frac{1}{2}\right)}{\Gamma\left(n - \frac{1}{2}\right)} = 2\sigma^2 \left(n - \frac{1}{2}\right)$$

$$\Rightarrow \mu_{2n} = \sigma^2 (2n-1) \mu_{2n-2} \qquad (4) \quad \left[\because \Gamma(r) = (r-1)\Gamma(r-1)\right]$$

Which gives the recurrence relationfor the moments of normal distribution. From(4), we have

$$\mu_{2n} = [(2n-1)\sigma^2][(2n-3)\sigma^2]\mu_{2n-4}$$
$$= [(2n-1)\sigma^2][(2n-3)\sigma^2][(2n-5)\sigma^2]\mu_{2n-6}$$
$$= [(2n-1)\sigma^2][(2n-3)\sigma^2][(2n-5)\sigma^2]...(3\sigma^2)(1\sigma^2)\mu_0$$
$$= 1.3.5...(2n-1)\sigma^{2n} \quad (5)$$

From (3) and (5), we conclude that for the normal distribution all odd order moments about mean vanish and even order moments about mean are given by (5).

A Linear Combination of Independent Normal Variates is also a Normal Variate:

Let $X_i, (i = 1, 2, 3, ..., n)$ be n independent normal variates with mean μ_i and variance σ_i^2 respectively. Then

$$M_{X_i}(t) = \exp\left\{\mu_i t + (t^2\sigma_i^2/2)\right\} \quad (7)$$

The m.g.f. of their linear combination $\sum_{i=1}^{n} a_i X_i$, where $a_i, a_2, ..., a_n$ are constants is given by

$$M_{\sum a_i X_i}(t) = \prod_{i=1}^{n} M_{a_i X_i}(t) \qquad (\because X_i\text{'s are independent})$$

$$= M_{X_1}(a_1 t).M_{X_2}(a_2 t)...M_{X_n}(a_n t) \quad [\because M_{cX}(t) = M_X(ct)] \quad (8)$$

From (7), we have $M_{X_i}(a_i t) = e^{\mu_i a_i t + t^2 a_i^2 \sigma_i^2/2}$

$$\therefore M_{\sum_i a_i X_i}(t) = e^{\mu_1 a_1 t + t^2 a_1^2 \sigma_1^2/2} \times e^{\mu_2 a_2 t + t^2 a_2^2 \sigma_2^2/2} \times ... \times e^{\mu_n a_n t + t^2 a_n^2 \sigma_n^2/2} \quad \text{(From 8)}$$

$$= \exp\left[(\sum_{i=1}^{n} a_i \mu_i)t + t^2(\sum_{i=1}^{n} a_i^2 \sigma_i^2)/2\right]$$

which is the m.g.f. of a normal variate with mean $(\sum_{i=1}^{n} a_i \mu_i)$ and variance $(\sum_{i=1}^{n} a_i^2 \sigma_i^2)$.

Hence by uniqueness theorem of m.g.f.,

$$(\sum_{i=1}^{n} a_i X_i) \sim N\left[\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right] \quad (8a)$$

Remarks:

1. If we take $a_1 = a_2 = 1, a_3 = a_4 = ... = 0,$ then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

If we take $a_1 = 1, a_2 = -1, a_3 = a_4 = ... = 0,$ then $X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

Thus we see that the sum as well as the difference of two independent normal variates is also a normal variable. This result provides a sharp contrast to the Poisson distribution, in which case though the sum of two independent Poisson variates is a Poisson variate, the difference is not a Poisson variate.

2. If we take $a_1 = a_2 = a_n = ... = 1,$ then $\sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$ (8b)

i.e., the sum of independent normal variates is also a normal variate, which establishes the *additive property* of the normal distribution.

3. If $X_i; i = 1, 2, ..., n$ are identically and independently distributed as $N(\mu, \sigma^2)$ and if we take $a_1 = a_2 = ... = a_n = 1/n$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\frac{1}{n}\sum_{i=1}^{n} \mu, \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2,\right) \Rightarrow \overline{X} \sim N(\mu, \sigma^2/n), \quad \text{where} \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

This leads to the following important conclusion:

If $X_i; i = 1, 2, ..., n$ are identically and independently distributed normal variates with mean μ and variance σ^2, then their mean $\overline{X} \sim N(\mu, \sigma^2/n)$.

Area Property (Normal probability integral)

If $X \sim N(\mu, \sigma^2)$, then the probabilitythat random value of X will lie between $X = \mu$ and $X = x_1$ is given by:

$$P(\mu < X < x_1) = \int_{\mu}^{x_1} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{x_1} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Put $\frac{X - \mu}{\sigma} = Z \Rightarrow X - \mu = \sigma Z$

When $X = \mu$, $Z = 0$ and when $X = x_1$, $Z = \frac{x_1 - \mu}{\sigma} = Z_1$

$$\therefore P(\mu < X < x_1) = P(0 < Z < z_1) = \frac{1}{\sqrt{2\pi}} \int_{0}^{z_1} e^{-\frac{Z^2}{2}} dZ = \int_{0}^{z_1} \varphi(Z)dZ$$

where $\varphi(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}}$ is the probability function of standard normal variate. The define integral $\int_0^{z_1} \varphi(Z)dZ$ is known as standard normal probability integral and given 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 gap of 0.01 in a table gives at the end of the chapter.

In particular, the probability that a random value of X lies in the interval (μ-σ, μ+σ) is given by:

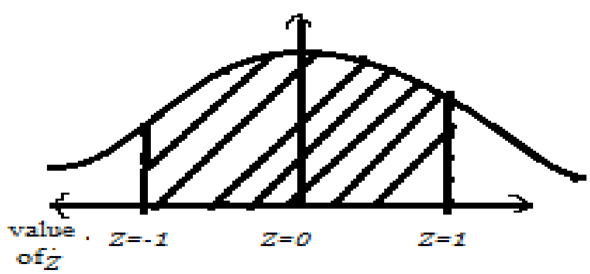
$$P(\mu - \sigma < X < \mu + \sigma) = \int_{\mu+\sigma}^{\mu-\sigma} f(x)dx$$

$$\Rightarrow P(-1 < Z < 1) = \int_{-1}^{1} \varphi(Z)dZ \quad \left[\begin{array}{l} \because Z = \frac{X-\mu}{\sigma}, \text{when X} = \mu\text{-}\sigma, Z = \frac{\mu-\sigma-\mu}{\sigma} = -1 \\ \text{when X} = \mu+\sigma,\ Z = \frac{\mu+\sigma-\mu}{\sigma} = 1 \end{array}\right]$$

$$= 2\int_0^1 \varphi(Z)dZ \quad [\text{By symmetry}]$$

$$= 2 \times 0.3413 = 0.6826$$

Similarly



$$\Rightarrow P(\mu - 2\sigma < X < \mu + \sigma) = \int_{\mu-\sigma}^{\mu+\sigma} f(x)dx = 2\int_0^2 \varphi(Z)dZ = 2 \times 0.4772 = 0.9544$$

And $P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < Z < 3) = \int_{-3}^{3} \varphi(Z)dZ = 2\int_0^3 \varphi(Z)dZ$

$$2\times0.49865 = 0.9973$$

Thus the probability that a normal variate X lies outside the range μ±3σ is given by:

$$P(|X - \mu| > 3\sigma) = P(|Z| > 3) = 1 - P(-3 \le Z \le 3) = 0.0027$$

Thus, all probability of normal variate we should expect to lie within the range $\mu\pm3\sigma$, though theoretically, it may range from $-\infty$ to ∞. From the discussion, it is clear that while solving numerical problems, we need to transform the given normal variate into a standard normal variate. The reason behind this is that area under different normal curves being infinitely many cannot be made available whereas the standard normal curve is one and hence table for each area under this curve can be made available.

Example 16: For a certain normal distribution, the first moment about 10 is 40 and the fourth moment about 50 is 48. What is the arithmetic mean and standard deviation of the distribution?

Solution: We know that if μ_1' is the first moment about the point X=A, the arithmetic mean is given by: Mean = A+ μ_1'

We are given: μ_1' (about the point X = 10) = 40

$\Rightarrow$ Mean = 10+40 = 50

Also μ_4' (about the point X = 50) = 48, i.e. $\mu_4 = 48$

But for a normal distribution with standard deviation σ

$$\mu_4 = 3\sigma^4 \Rightarrow 3\sigma^4 = 48 \Rightarrow \sigma = 2$$

Example 17: X is a normal variable with mean 30 and S.D. 5. Find the probabilities that

(i) $26 \le X \le 40$

(ii) $X \ge 45$

(iii) $|X - 30| > 5$

Solution: Here $\mu = 30$ and $\sigma = 5$

(i) When X = 26, $Z = \dfrac{X-\mu}{\sigma} = \dfrac{26-30}{5} = -0.8$

and when X = 40, $Z = \dfrac{40-30}{5} = 2$

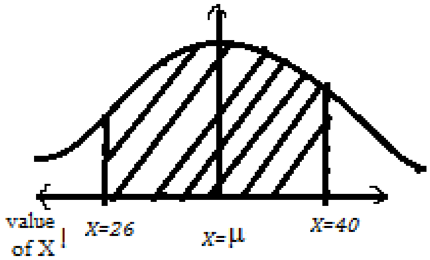
$$\therefore P(26 \le X \le 40) = P(-0.8 \le Z \le 2) = P(-0.8 \le Z \le 0) + P(0 \le Z \le 2)$$

$$= P(0 \le Z \le 0.8) + (0 \le Z \le 2)$$

$$= 0.2881 + 0.4772 = 0.7653$$

(ii) When x = 45,

$$Z = \frac{45-30}{5} = 3$$



$$\therefore P(X \geq 45) = P(Z \geq 3)$$

$$= 0.5 - P(0 \leq Z \leq 3) = 0.5 - 0.49865 = 0.00135$$

(iii)

$$P(|X - 30| \leq 5) = P(25 \leq X \leq 35) = P(-1 \leq Z \leq 1)$$

$$= 2P(0 \leq Z \leq 1) = 2 \times 0.3413 = 0.6826$$

$$\therefore P(|X - 30| > 5) = 1 - P(|X - 30|) \leq 5 = 1 - 0.6826 = 0.3174$$

Example 18: Two independent random variate X and Y are both normally distributed with means 1 and 2 and standard derivations 3 and 4 respectively. If Z = X-Y, write the probability density function of Z. Also state the median, S.D. and mean of the distribution of Z. Find probability (Z+1≤0).

Solution: Since X ~ N(1, 9) and Y ~ N(2, 16) are independent,

Z = X - Y ~ N(1 - 2, 9 + 16)

i.e. Z = X - Y ~ N(-1, 25)

Hence p.d.f. of z is given by

$$p(z) = \frac{1}{5\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{z+1}{5}\right)^2\right], -\infty < z < \infty$$

For the distribution of Z

Median = mean = -1 and S.D. = √25 = 5

and P(Z+1<0) = P(Z≤-1) = P(U≤0) = 0.5 $\left[U = \frac{Z+1}{5} \sim N(0,1)\right]$

Example 19: If the r.v. X is normally distributed with mean 80 and standard deviation 5, then find

(i) P[X>95] (ii)P[X<72](iii) P[60.5<X<90]

Solution: Here $X \sim N(80, 25)$

Mean= $\mu = 80$ and variance $= \sigma^2 = 25$

If Z is the standard normal variate then $Z = \dfrac{X-\mu}{\sigma} = \dfrac{X-80}{5}$

(i) For X=95,

$$Z = \frac{95-80}{5} = 3$$

$$\therefore P[X > 95] = P[Z > 3]$$

$$= 0.5 - P[0 < Z < 3]$$

$$=0.5-0.4987=0.0013$$

(ii) For X=72,

$$Z = \frac{72-80}{5} = -1.6$$

$$\therefore P[X < 72] = P[Z < -1.6]$$

$$= P[Z > 1.6]$$

$$=0.5-P[0 < Z < 1.6]=0.0548$$

(iii) X=60.5, $Z = \dfrac{60.5-80}{5} = -3.9$

X=90, $Z = \dfrac{90-80}{5} = 2$

$$\therefore P[60.5 < X < 90] = P[-3.9 < Z < 2]$$

$$= P[-3.9 < Z < 0] + P[0 < Z < 2]$$

$$=P[0 < Z < 3.9] + P[0 < Z < 2]$$

$$=0.5000+0.4772$$

$$=0.9772$$

Example 20: If X, and Y are independent normal variates with means 6, 7 and variances 9, 16 respectively, determine λ such that $P(2X + Y \le \lambda) = P(4X - 3Y \ge 4\lambda)$.

Solution: Since X and Y are independent normal variates, then

$$U = 2X + Y \sim N(2\times 6 + 7, 4\times 9 + 16), \quad \text{i.e.,} \quad U \sim N(19, 52)$$

$$V = 4X - 3Y \sim N(4\times 6 - 3\times 7, 16\times 9 + 9\times 16), \quad \text{i.e.,} \quad V \sim N(3, 288)$$

and $P(2X+Y \le \lambda) = P(U \le \lambda) = P(Z \le \frac{\lambda - 19}{\sqrt{52}})$, where $Z \sim N(0,1)$

and $P(4X-3Y \ge 4\lambda) = P(V \ge 4\lambda) = P(Z \ge \frac{4\lambda - 3}{12\sqrt{2}})$, where $Z \sim N(0,1)$

Now $P(2X+Y \le \lambda) = P(4X-3Y \ge 4\lambda)$ (given)

$$\Rightarrow P(Z \le \frac{\lambda-19}{\sqrt{52}}) = P(Z \ge \frac{4\lambda-3}{12\sqrt{2}}) \Rightarrow \frac{\lambda-19}{\sqrt{52}} = -\left(\frac{4\lambda-3}{12\sqrt{2}}\right)$$

[Since, $P(Z \le a) = P(Z \ge b) \Rightarrow a = -b$, because normal probability curve is symmetric about Z=0]

$$\Rightarrow \frac{\lambda-19}{\sqrt{13}} = \frac{3-4\lambda}{6\sqrt{2}} \Rightarrow (6\sqrt{2}+4\sqrt{13})\lambda = 114\sqrt{2}+3\sqrt{13} \Rightarrow \lambda = \frac{114\sqrt{2}+3\sqrt{13}}{6\sqrt{2}+4\sqrt{13}}$$

Example 21: In a university the mean weight of 1000 male students is 60kg and standard deviation is 16kg.

(a) Find the number of male students having their weights less than 55kg

(b) What is the lowest weight of the 100 heaviest male students?

(Assuming that the weights are normally distributed)

Solution: Let X be a normal variate, "The weight of the male students of the university". Here, we are given that mean = 60kg and standard deviation = 16kg, therefore

$X \sim N(60, 256)$

We know that if $X \sim N(\mu, \sigma^2)$, then the standard normal variate is given by $Z = \frac{X-\mu}{\sigma}$

Hence, for the given information, $Z = \frac{X-60}{16}$

(a)For $X = 55, Z = \frac{55-60}{16} = -0.3125 \simeq -0.31$

Therefore,

$$P[X < 55] = P[Z < -0.31] = P[Z > 0.31]$$
$$= 0.5 - P[0 < Z < 0.31] \quad [\text{area on both sides of } Z = 0 \text{ is } 0.5]$$
$$= 0.5 - 0.1217 = 0.3783 \quad [\text{using table area under normal curve}]$$

∴ Number of male students having weight less than 55 is

$= 1000 \times 3783 = 378$

(b) Let x_1 be the lowest weight amongst 100 heaviest students.

Now for

$$X = x_1, Z = \frac{x_1 - 60}{16} = z_1 \text{(say)}$$

$$P[X \geq x_1] = \frac{100}{1000} = 0.1 \quad \text{(see fig5.2)}$$

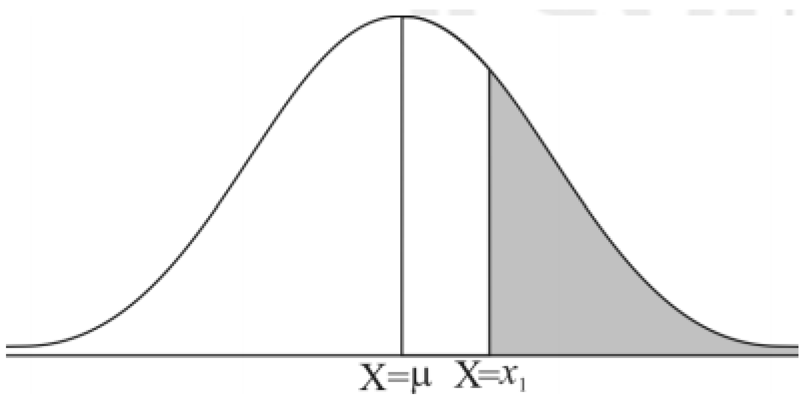
$$\Rightarrow P[Z \geq z_1] = 0.1$$

$$\Rightarrow P[0 \leq Z \leq z_1] = 0.5 - 0.1 = 0.4$$

$$\Rightarrow z_1 = 1.28$$

$$\Rightarrow x_1 = 60 + 16 \times 1.28 = 60 + 20.48 = 80.48$$

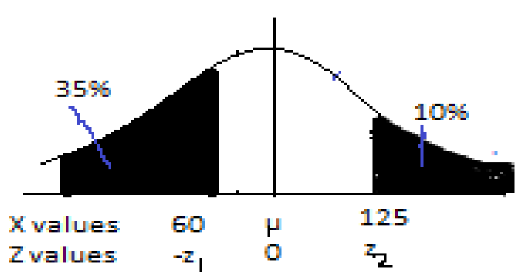
Therefore, the lowest weight of 100 heaviest male students is 80.48kg.



(*area representing the 100 heaviest male students*)

Example 22: In a normal distribution 10% of the items are over 125 and 35% are under 60. Find the mean and standard deviation of the distribution.

Solution: Let $X \sim N(\mu, \sigma^2)$, where μ and σ^2 are unknown and are to be obtained.



(*Area representing the items under 60 and over 125*)

Here we are given

$P[X > 125] = 0.1$ and $P[X < 60] = 0.35$ [see fig]

We know that if $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma}$

For X = 60, $Z = \dfrac{60-\mu}{\sigma} = -z_1$ (say) (1)

[-ve sign is taken because $P[Z < 0] = P[Z > 0] = 0.5$]

For X = 125, $Z = \dfrac{125-\mu}{\sigma} = z_2$ (say) (2)

Now

$$P[X < 60] = P[Z < -z_1] = 0.35$$
$$\Rightarrow P[Z > z_1] = 0.35 \quad \text{[by symmetry of normal curve]}$$
$$\Rightarrow 0.5 - P[0 < Z < z_1] = 0.35$$
$$\Rightarrow P[0 < Z < z_1] = 0.15$$
$$\Rightarrow z_1 = 0.39$$

[from the table area under normal curve]

And

$$P[X > 125] = p[Z > z_2] = 0.10$$
$$\Rightarrow 0.5 - P[0 < Z < z_2] = 0.10$$
$$\Rightarrow P[0 < Z < z_2] = 0.40$$
$$\Rightarrow z_2 = 1.28 \quad \text{(from the table)}$$

Putting the values of z_1 and z_2 in eq. (1) and (2), we get

$$\frac{60-\mu}{\sigma} = -0.39 \quad (3)$$

$$\frac{125-\mu}{\sigma} = 1.28 \quad (4)$$

(4)-(3) gives

$$\frac{125-\mu-60+\mu}{\sigma} = 1.28 + 0.39$$

$$\frac{65}{\sigma} = 1.67 \Rightarrow \sigma = \frac{65}{1.67} = 38.92$$

From eq. (4), $\mu = 125 - 1.28\sigma \Rightarrow \mu = 125 - 1.28 \times 38.92 = 75.18$

Hence μ= mean =75.18; σ = S.D. = 38.92

Problems:

1. X is normally distributed and the mean of X is 12 and S.D. is 4.

(a) Find out the probability of the following:

(i) $X \geq 20$

(ii) $X \leq 20$

(iii) $0 \leq X \leq 12$

(b) Find x' when $P(X > x') = 0.24$

2. If $X \sim N(30,16)$, then find α in each case

$(i) P[X > \alpha] = 0.2492$

$(ii) P[X < \alpha] = 0.0496$

3. Average lactation yield for 1000 cows maintained at a farm is 1700 kg and their standard deviation is 85 kg. A cow is considered as high yielder if it has a lactation yield greater than 1900 kg and poor yielder if it has lactation yield less than 1600 kg. Find the number of high yielding and poor yielding cows.

4. If 100 true coins are thrown, how would you obtain an approximation for the probability of getting (i) 55 heads, (ii) 55 or more heads, using tables of Area of normal probability function?

5. In a particular branch of a bank, it is noted that the duration/waiting time of the customers for being served by the teller is normally distributed with mean 5.5 minutes and standard deviation 0.6 minutes. Find the probability that a customer has to wait

a) between 4.2 and 4.5 minutes, (b) for less than 5.2 minutes, and (c) more than 6.8 minutes

Importance of Normal Distribution

Normal distribution plays a very important role in statistics because

(i) Most of the discrete probability distributions occurring in practice e.g., Binomial and Poisson can be approximated to normal distribution as n number of trials tends to increase.

(ii) Even if a variable is not normally distributed, it can be sometimes brought to normal by a simple mathematical transformation, if the distribution of X is skewed, the distribution of $\sqrt{x}$ or log x might come out to be normal.

(iii) If $X \sim N(\mu,\sigma^2)$ then $P[\mu - 3\sigma < x < \mu + 3\sigma] = 0.9973 \Rightarrow [|Z|>3] = 1 - 0.9973 = 0.0027$. Thus the probability of standard normal variate going outside the limits 3 is practically zero. This property of normal distribution forms the basis of entire large sample theory.

(iv) Many of the sampling distribution e.g., students t, Snedecors F, Chi square distributions etc tend to normality for large samples. Further, the proof of all the tests of significance in the sample is based upon the fundamental assumptions that the populations from which the samples have been drawn are normal.

(v) The whole theory of exact sample (small sample) tests viz. t , χ^2, F etc, is based on the fundamental assumption that the parent population from which the samples have been drawn follows normal distribution.

(vi) Normal distribution finds large applications in statistical quality control in industry for setting up of control limits.

(vii) Theory of normal curves can be applied to the graduation of the curve which is not normal.

Fitting of Normal Distribution

In order to fit normal distribution to the given data we first calculate the mean μ and standard deviation σ from the given data. Then the normal curve fitted to the given data is given by:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-(x-\mu)^2/2\sigma^2\right\},-\infty<x<\infty$$

To calculate the expected normal frequencies we first find the standard normal variates corresponding to the lower limits of each of the class intervals, i.e., we compute $z_i=(x_i'-\mu)/\sigma$, where x_i' is the lower limits of the *i*th class intervals. Then the areas under the normal curve to the left of the ordinates at $z=z_i$, say $\Phi(z_i)=P(Z\le z_i)$ are computed from the tables. Finally, the areas for the successive class intervals are obtained by subtraction, viz, $\Phi(z_{i+1})-\Phi(z_i),(i=1,2,...)$ and on multiplying these areas by N , we get the expected normal frequencies.

Example 23: Obtained the equation of the normal curve that may be fitted to the following data:

Class	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100
frequency	3	21	150	335	326	135	26	4

Also obtained the expected normal frequencies.

Solution: For the given data, $N=1000,\mu=79.945$ and $\sigma=5.545$

Hence the equation of the normal curve fitted to the given data is:

$$f(x)=\frac{1000}{\sqrt{2\pi}\times5.545}\exp\left\{-\frac{1}{2}\left(\frac{(x-79.945)}{5.545}\right)^2\right\}$$

class	Lower class Boundary (X')	$z = \frac{X' - \mu}{\sigma}$	$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du$	$\Delta\Phi(Z) = \Phi_{Z+1} - \Phi_Z$	Expected frequency $= N\Delta\Phi(z)$
Below 60	-∞	-∞	0	0.000112	0.12≡0
60-65	60	-3.663	0.000112	0.002914	2.914≡3
65-70	65	-2.745	0.003026	0.031044	31.044≡31
70-75	70	-1.826	0.034070	0.147870	147.870≡148
75-80	75	-0.908	0.181940	0.322050	322.050≡322
80-85	80	0.010	0.503990	0.919300	319.300≡319
85-90	85	0.928	0.823290	0.144072	144.072≡144
90-95	90	1.487	0.997154	0.029792	29.792≡30
95-100	95	2.675	0.997154	0.002733	2.733≡3
100 and over	100	3.683	0.999887		
Total					1,000

Problems:

1. For a certain normal distribution, the first moment about 10 is 40 and the fourth moment about 50 is 48. What is the arithmetic mean and standard deviation of the distribution?
2. X is normally distributed and the mean of X is 12 and S.D is 4.

   (a) Find out the probability 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_0' and x_1', when $P(x_0' < X < x_1') = 0.50$ and $P(X > x_1') = 0.25$

Value of $e^{-\lambda}$ **(for computing Poisson Probabilities)**

($0<\lambda<I$)

λ	0	1	2	3	4	5	6	7	8	9
0.0	1.0000	0.9900	0.9802	0.9704	0.9608	0.9512	0.9418	0.9324	0.9231	0.9139
0.1	0.9048	0.8958	0.8860	0.8781	0.8694	0.8607	0.8521	0.8437	0.8353	0.8270
0.2	0.7187	0.8106	0.8025	0.7945	0.7866	0.7788	0.7711	0.7634	0.7558	0.7483
0.3	0.7408	0.7334	0.7261	0.7189	0.7118	0.7047	0.6970	0.6907	0.6839	0.6771
0.4	0.6703	0.6636	0.6570	0.6505	0.6440	0.6376	0.6313	0.6250	0.6188	0.6125
0.5	0.6065	0.6005	0.5945	0.5886	0.5827	0.5770	0.5712	0.5655	0.5599	0.5543
0.6	0.5448	0.5434	0.5379	0.5326	0.5278	0.5220	0.5160	0.5113	0.5066	0.5016
0.7	0.4966	0.4916	0.4868	0.4810	0.4771	0.4724	0.4670	0.4630	0.4584	0.4538
0.8	0.4493	0.4449	0.4404	0.4360	0.4317	0.4274	0.4232	0.4190	0.4148	0.4107
0.9	0.4066	0.4026	0.3985	0.3946	0.3906	0.3867	0.3791	0.3791	0.3753	0.3716
(λ=1,2,3,… , 10)										
λ	1	2	3	4	5	6	7	8	9	10
$e^{-\lambda}$	0.3679	0.1353	0.0498	0.0183	0.0070	0.0028	0.0009	0.0004	0.0001	0.00004

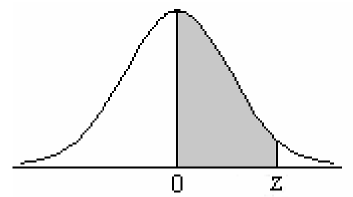
Note: To obtain values of $e^{-\lambda}$ for other values of λ, use the laws of exponents i.e.

$e^{-(a+b)} = e^{-a}. e^{-b}$

e.g. $e^{-2.25}= e^{-2}. e^{-0.25}$=(0.1353).(0.7788)=0.1054

TABLE1 Normal Curve Areas:

The entries in the body of the table correspond to the area shaded under the normal curve.



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359	.0000
0.1	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753	0.0398
0.2	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141	0.0793
0.3	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517	0.1179
0.4	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879	0.1554
0.5	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224	0.1915
0.6	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549	0.2257
0.7	0.2611	0.2642	0.2673	0.2703	0.2734	0.2764	0.2794	0.2823	0.2852	0.2580
0.8	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133	0.2881
0.9	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389	0.3159
1.0	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621	0.3413
1.1	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830	0.3643
1.2	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015	0.3849
1.3	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177	0.4032
1.4	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319	0.4192
1.5	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441	0.4332
1.6	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545	0.4452
1.7	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633	0.4554
1.8	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706	0.4641
1.9	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767	0.4713
2.0	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817	0.4772

2.1	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857	0.4821
2.2	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890	0.4861
2.3	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916	0.4893
2.4	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936	0.4918
2.5	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952	0.4938
2.6	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964	0.4953
2.7	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974	0.4965
2.8	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981	0.4974
2.9	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986	0.4981
3.0	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990	0.4987
3.1	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993	0.4990
3.2	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995	0.4993
3.3	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997	0.4995
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998	0.4997
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4998
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

CHAPTER-4

SAMPLING THEORY

Structure

4.1 Introduction

The science of statistics may broadly be studied under the two heads descriptive and inductive. So far we have confined ourselves to descriptive statistics which help us in describing the characteristics of numerical data. In other parts i.e., inductive statistics are also known as statistical inference which is termed as logic of drawing valid statistical conclusions about the population in any statistical investigation based on examining a part of the population known as Sample. It is scientifically drawn from the population. In all the spheres of life (such as economic, social, scientific, industry etc.) the need for statistical investigation and data analysis is increasing day by day. There are two methods of collection of statistical data i.e. census and sample method**.** Under census method, information related to the entire field of investigation or units of the population is collected; whereas under sample method, rather than collecting information about all the units of the population, information relating to only selected units is collected.

4.1.1 Basic Definitions

Population: In any statistical investigation, our interest lies in the assessment of the general magnitude and the study of variation with respect to one or more characteristics relating to individuals belonging to a group. The group of individuals under study is called population or universe. In general, we can say that a universe or population means the entire field under investigation about which knowledge is sought. It is a collection of objects, animate or inanimate or anything conceivable pertaining to certain characteristics that we want to study/test. The collection of objects could be cities, students, factories, etc. A population can be of two kinds (i) Finite and (ii) Infinite. In a finite population, number of items is definite such as number of students or teachers in a college, daily milk yield of 500 milch animals in a livestock farm. On the other hand, an infinite population has infinite number of items e.g. the population of pressures at various points in the atmosphere, the population of real numbers between 0 and 1, the population of all integers, number of water drops in an ocean, number of leaves on a tree or number of hairs on the head etc.

Sample: In the real world, its tough task to get complete information about the population. Hence, we draw a sample out of that population and derive the same statistical measures mentioned above. Therefore, a finite subset of the population, selected from it by using scientific procedure with the objective of investigating its properties is called a sample. In other words, selected or sorted units from the population are known as a sample. Thus, sample means some units selected out of a population that represents it. For example, if an investigator selects 100 animals from 2000 animals in a herd then these 100 animals will be termed as a sample and number of the individuals in the sample is called a sample size.

Sampling: The process of selecting a sample is called sampling. It is a tool that enables us to conclude the characteristics of the population after studying only those items which are included in the sample. Sampling is quite often used in our day-to-day practical life. For example: in a shop, we assess the quality of sugar, wheat or any other commodity by taking a handful of it from the bag and then decide to purchase it or not. A housewife normally tests whether food is properly cooked and contains the proper quantity of salt. The main objective of sampling is

- To obtain the maximum information about the characteristics of the population with the available sources e.g. time, money, manpower etc.
- To obtain best estimates of the population parameter

Types of Sampling

There are various methods of sampling that may be used singly or along with others. The choice of appropriate sampling design is of paramount importance in the execution of a sample survey and is generally made keeping in view the objectives and scope of the inquiry and the nature of the population to be sampled. The sampling techniques may be broadly classified as follows:

1. Purposive or Subjective or Judgment Sampling
2. Probability Sampling
3. Simple random sampling
4. Stratified Random Sampling
5. Cluster Sampling
6. Multi-stage sampling

1. Purposive or Subjective or Judgment Sampling

In this method of sampling, the choice of the sample items depends exclusively on the judgment of the investigator. The desired number of sample units are selected deliberately or purposively depending upon the object of the inquiry so that only the important items representing the true characteristics of the population are included in the sample. Purposive sampling is one in which the sample units are selected with a definite purpose in view. This type of sampling suffers from the drawback of favouritism and nepotism depending upon the beliefs and prejudices of the investigator and thus does not give a true representation to the population.

2. Probability Sampling

Probability sampling provides a scientific technique of drawing samples from the population according to some laws of chance in which each unit in the universe has some definite pre-assigned probability of being selected in the sample. The selection of the sample based on the theory of probability is also known as the random selection and the probability sampling is also called Random Sampling. Different types of sampling are:

(i) Each sample unit has an equal chance of being selected

(ii) Sampling units have varying probability of being selected

(iii) Probability of selection of a unit is proportional to the sample size.

3. Simple Random Sampling

Simple random sampling (S.R.S.) is the technique in which the sample is drawn in such a way that every unit in the population has an equal and independent chance of being included in the sample. Suppose we take a sample of size n from a finite population of size N. Then there are ${}^{N}C_{n}$ possible samples. A S.R.S. is a technique of selecting the sample in which each of ${}^{N}C_{n}$ samples has an equal chance or probability p $= 1/{}^{N}C_{n}$ of being selected.

4. Stratified Random Sampling

Stratified random sampling is one where the population is divided into mutually exhaustive strata or sub-groups and then a simple random sample is selected within each strata or sub-group e.g. cows in a big herd can be divided into different strata based on breed, age groups, body weight groups, lactation length, lactation order, daily/lactation milk yield groups, etc. The criterion used for the stratification of the universe into various strata is known as the stratifying factor. In general, geographical, sociological or economic characteristics form the basis of stratification of the given population. Some of the commonly used stratifying factors are age, sex, income, occupation, educational level, geographical area, and economic status, etc.

5. Cluster Sampling

When the population size is very large, the previously mentioned sampling methods lead to several difficulties. The sampling frame is not available and it is too expensive and time consuming to prepare it. The other difficulties are firstly the high cost and administrative difficulty of surveying widely scattered sampling units and secondly the elementary units may not be easily identifiable and locatable. In such cases, cluster sampling is useful. In this case, the total population is divided, depending upon on problem under study, into some recognizable sub-divisions which are termed as clusters. A specified number of clusters are selected at random, and the observation is made on all the units in the sampled clusters. We then observe, measure and interview for every unit in the selected clusters. The clusters are called the primary units. Cluster sampling is also known as area sampling.For example, the cluster may be consisting of all households in a village and hence there are as many clusters as the number of villages in a district. It may be noted that the cluster is a heterogeneous sub-population whereas the stratum is a homogeneous sub-population.

6. Multistage Sampling

Instead of enumerating all the sampling units in the selected clusters, one can obtain better and more efficient estimators by resorting to sub-sampling within the clusters. The technique is called two-stage sampling, clusters being termed as primary units and the units within the clusters as secondary units. The above technique may be generalized to what is called multistage sampling. As the name suggests, multistage sampling refers to a sampling technique that is carried out in various stages. Here the population is regarded as a secondary stage unit in which we are interested. For example, if we are interested in obtaining consisting of severalof primary units each of which is further composed of severalof a sample of, say, n households from a particular state the first stage units may be districts and second stage units may be villages in the districts and third stage units will be households in the villages. Each stage thus results in a reduction of the sample size. Multistage sampling is more flexible as compared to other methods of sampling.

4.2 Parameter and Statistic

The statistical constants of the population like mean (μ), variance (σ^2), skewness (β_1), kurtosis (β_2), correlation coefficient (ρ), etc., are known as parameters. In other words, the terms which are used for studying the measures of the population are known as parameters while others that are used for sample are known as statistics. Therefore, a statistic is a summary description of a characteristic or measure of the sample. The sample statistic is used as an estimate of the population parameter. Also, in order to avoid verbal confusion with the statistical constants of the population, viz., mean (μ), variance (σ^2), etc., which are usually referred to as parameters, statistical measures computed from the sample observations alone, e.g., mean ($\bar{x}$), variance(s^2), etc., have been termed by Professor R.A. Fisher as statistics.

Let us consider a finite population of N units and let $Y_1, Y_2, Y_3, ..., Y_N$ be the observations on the N units in the population.

$$\text{Mean }(\mu) = \frac{1}{N}\left(Y_1+Y_2+...+Y_N\right) = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

$$\sigma^2 = \frac{1}{N}\left[\left(Y_1-\mu\right)^2+\left(Y_2-\mu\right)^2+...+\left(Y_N-\mu\right)^2\right]$$
$$= \frac{1}{N}\sum_{i=1}^{N}\left(Y_i-\mu\right)^2$$

Suppose, we draw a sample of size n from this population. Let $X_1, X_2, X_3, ---, X_n$ be the observations on the sample units. Then we can compute sample mean $\left(\overline{X}\right)$ and sample variance (s^2) as given below:

$$\text{Mean }(\overline{X}) = \frac{1}{n}\left(X_1+X_2+...+X_n\right) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$s^2 = \frac{1}{n}\left[\left(X_1-\overline{X}\right)^2+\left(X_2-\overline{X}\right)^2+...+\left(X_n-\overline{X}\right)^2\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$

In practice, parameter values are not known and the estimates based on the sample values are generally used. Thus statistics which may be regarded as an estimate of the parameter, obtained from the sample, is a function of the sample values only. It may be pointed out that a statistic, as it is based on sample values and as there are multiple choices of the samples that can be drawn from a population, varies from sample to sample.

Remark: Now onwards, μ and σ^2 will refer to the population mean and variance, respectively while the sample mean and variance will be denoted by $\overline{X}$ and s^2respectively.

4.3 Sampling Distribution of a Statistic

A Sampling Distribution is a probability distribution of a statistic obtained through a large number of samples drawn from a specific population. For example: Suppose a simple random sample of five hospitals is to be drawn from a population of 20 hospitals. The possibilities could be, (20, 19, 18, 17, 16) or (1, 2,4,7,8) or any of the 15,504 (using $20C_5$ combinations) different samples of size 5 can be drawn. If we draw a sample of size n from a given finite population of size N, then the total number of possible samples is:

$$^{N}C_{n} = \frac{N!}{n!(N-n)!} = \text{k(say)}$$

For each of these samples we can compute some statistic $t = t(X_1, X_2, ..., X_n)$ e.g. mean $\overline{X}$, the variance s^2, etc., as given below.The set of values of the statistic so obtained one for each sample constitutes what is called the sampling distribution of the statistic. For example, the values $t_1, t_2, ..., t_k$ determine the sampling distribution of the statistic t.In other words, the statistic t can be regarded as a random variable which can take values $t_1, t_2, ..., t_k$ and we can compute various statistical constants like mean, variance, skewness, and kurtosis etc.,

Table

Sample number	**Statistic**		
	t	$\overline{X}$	s^2
1	t_1	$\overline{X}_1$	s_1^2
2	t_2	$\overline{X}_2$	s_2^2
3	t_3	$\overline{X}_3$	s_3^2
.	.	.	.
.	.	.	.
.	.	.	.
			s_k^2
k	t_k	$\overline{X}_k$	

for its distribution e.g. the mean and variance of the sampling distribution of the statistic are given by

$$\bar{t} = \frac{1}{k}\sum_{i=1}^{k} t_i ; \overline{\overline{X}} = \frac{1}{k}\sum_{i=1}^{k} \overline{X}_i$$

$$V(t) = \frac{1}{k}\sum_{i=1}^{k} (t_i - \bar{t})^2 ; V(\overline{X}) = \frac{1}{k}\sum_{i=1}^{k} \left(\overline{X}_i - \overline{\overline{X}}\right)^2$$

Remark: In general, the mean of the sampling distribution will be approximately equivalent to the population mean i.e. $E(\bar{x}) = \mu$

4.4 Standard Error

The standard deviation of the sampling distribution of a statistic is known as its standard error. It is very similar to the standard deviation. Both are measures of spread. The higher the number, the more spread out your data is. To put it simply, the two terms are essentially equal, but there is one important difference. While the standard error uses statistic (sample data), standard deviation use parameters (population data). The standard error tells you how far your sample statistic (like the sample mean) deviates from the actual population parameter. The larger your sample size, the smaller the SE. In other words, if we increase the sample size, our sample mean will become closer to the population mean. The standard error of a statistic t is given by:

$$S.E.(t)=\sqrt{V(t)}=\sqrt{\frac{1}{k}\sum_{i=1}^{k}\left(t_i-\bar{t}\right)^2}$$

The standard errors of some of the well known statistics, for large samples, are given below where n is the sample size, σ^2 the population variance, P the population proportion and Q= 1-P, n_1 and n_2 represent sizes of two independent random samples.

Table

Sr. No.	Statistic	Standard Error
1.	Sample mean $\overline{X}$	$\frac{\sigma}{\sqrt{n}}$
2.	Sample proportion p	$\sqrt{\frac{PQ}{n}}$
3.	Sample standard deviation	$\sqrt{\frac{\sigma^2}{2n}}$
4.	Sample variance (s^2)	$\sigma^2\sqrt{\frac{2}{n}}$
5.	Sample correlation coefficient(r)	$\frac{\left(1-p^2\right)}{\sqrt{n}}$
6.	Difference between two sample means $\left(\overline{X}_1-\overline{X}_2\right)$	$\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
7.	Difference between two sample Standard deviation (s_1-s_2)	$\sqrt{\frac{\sigma_1^2}{2n_1}+\frac{\sigma_2^2}{2n_2}}$
8.	Difference between two sample proportion (p_1-p_2)	$\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}$

4.4.1 Utility of Standard Error

1. S.E. plays a very important role in large sample theory and forms the basis of the testing of hypothesis. Thus, if the discrepancy between the observed and expected (hypothetical) value of a statistic is greater than or equal to Z_α times S.E., the hypothesis is rejected at α level of significance otherwise the deviation is not regarded as significant and is considered as due to fluctuations of sampling or chance causes.

2. The magnitude of S.E. gives an index of the precision of the estimate of parameter. The reciprocal of the S.E. is taken as the measure of reliability or precision of the sample e.g. S.E. of sample mean and sample proportion are $\frac{\sigma}{\sqrt{n}}$ and $\sqrt{\frac{PQ}{n}}$ respectively, which vary inversely as the square root of the sample size. Thus in order to double the precision which amounts to reducing the S.E. to one half, the sample size has to be increased four times.

3. S.E. enables us to determine the probable limits/confidence limits within which the population parameter may be expected to lie.

4.5 Statistical Hypothesis

A statistical hypothesis is an assumption or statement about a population parameter. This assumption may or may not be true, which we want to test based on evidence from a random sample. Hypothesis testing refers to the formal procedures used by statisticians to accept or reject statistical hypotheses. In other words, it is a tentative conclusion logically drawn concerning any parameter of the population. For example, the average fat percentage of milk of Red Sindhi Cow is 5%, the average quantity of milk filled in the pouches by an automatic machine is 500ml.

Before starting any investigation**,** we need to set some assumptions and presumptions which subsequently in the study must be proved or disproved. The hypothesis is based on observation which is discussed below:

4.5.1 Null Hypothesis

The technique of randomization used for the selection of sample units makes the test of significance valid for us. For applying the test of significance, we first set up a hypothesis-a definite statement about the population parameter. Such a hypothesis, which is usually a hypothesis of no difference, is called null hypothesis. According to Prof. R. A. Fisher, a hypothesis which is tested for possible rejection under the assumption that it is true is usually called Null Hypothesis and is denoted by H_0. The common way of stating a hypothesis is that there is no difference between the two values, namely the population mean and the sample mean. The term no difference means that the difference, if any, is merely due to sampling fluctuations. Thus, if the statistical test shows that the difference is significant, the hypothesis is rejected. To test whether there is any difference between the two populations we shall assume that there is no difference. Similarly, to test whether there is a relationship between two variates, we assume there is no relationship. So a hypothesis is an assumption concerning the parameter of the population. The reason is that a hypothesis can be rejected but cannot be proved. Rejection of no difference will mean a difference, while the rejection of no relationship will imply a relationship.

Example:

1. If we want to test that the average milk production of Karan Swiss cows in lactation is 3000 litres thenthe null hypothesis may be expressed symbolically as H_o: $\mu = 3000$ litres.

2. In case of a single statistic, H_0 will be that the sample statistic does not differ significantly from the hypothetical parameter value and in the case of two statistics, H_0 will be that the sample statistics do not differ significantly.

3. Let us consider the 'light bulbs' problem. Let us suppose that the bulbs manufactured under some standard manufacturing process have an average life of μ hours and it is proposed to test a new procedure for manufacturing light bulbs, those manufactured by standard process and those manufactured by the new process.

In this problem, the following three hypotheses may be set up:

(1) New process is better than the standard process.

(2) New process is inferior to the standard process.

(3) There is no difference between the two processes.

The first two statements appear to be biased since they reflect a preferential attitude to one or the other of the two processes. Hence the best course is to adopt the hypothesis of no difference, as stated in (3). This suggests that the statistician should take up the neutral or null attitude regarding the outcome of the test.

4.5.2 Alternative Hypothesis

Any hypothesis which is complementary to the null hypothesis is called an Alternative hypothesis. In other words, we can say that the sample result is different, i.e. greater or lower than the hypothetical value of population. It is usually denoted by H_1. For example: If we want to test the null hypothesis that the population has a specified mean μ_0, (say), $H_0 = \mu_0$ then the alternative hypothesis could be:

(1) $H_1: \mu \neq \mu_0$ (i.e., $\mu > \mu_0$ or $\mu < \mu_0$)

(2) $H_1: \mu > \mu_0$, (3) $H_1: \mu < \mu_0$

The alternative hypothesis in (1) is known as a two-tailed alternative and the alternatives in (2) and (3) are known as right tailed and left tailed alternatives respectively.

Remark: The setting of alternative hypothesis is very important since it enables us to decide whether we have to use a single-tailed test or two-tailed test. For example: In the example of light bulbs, alternative hypothesis could be,

(1) $H_1: \mu_1 > \mu_0$ or $\mu_1 < \mu_0$ or $\mu_1 \neq \mu_0$

Example: We want to test if college students take less than five years to graduate from college, on the average.

$H_0 : \mu \geq 5, H_1 : \mu < 5$

4.5.3 Simple Hypothesis

If the statistical hypothesis completely specifies the population or distribution, it is called a simple hypothesis. In this hypothesis, all parameters associated with the distribution are stated or a specified with a particular value. For instance, if the height of the students in a school is distributed normally with $\sigma^2 = 6$ and the hypothesis that the mean stands equivalent to 70 implying $H_0 : \mu = 70$. This stands to be the simple hypothesis as variance and mean both completely specify the normal distribution. In general, a simple hypothesis reflects that $\theta = \theta_0$ where θ_0 is the specified value of θ (θ may represent $\mu, \sigma, \mu_1 - \mu_2$).

4.5.4 Composite Hypothesis

A hypothesis which is not simple, i.e., in which not all of the parameters are specified is called a composite hypothesis. For instance, if we hypothesize that $H_0 : \mu > 62$ and $\sigma^2 = 4$ or $H_0 : \mu = 62$ and $\sigma^2 < 4$, the hypothesis becomes a composite hypothesis because we cannot know the exact distribution of the population in either case. Obviously, the parameters $\mu > 62$ and $\sigma^2 < 4$ have more than one value and no specified values are being assigned. The general form of a composite hypothesis is $\theta \le \theta_0$ or $\theta \ge \theta_0$; that is, the parameter θ does not exceed or does not fall short of a specified value θ_0. The concept of simple and composite hypotheses applies to both the hypothesis.

4.6 Two Types of Errors

The main objective in sampling theory is to draw valid inferences about the population parameters on the basis of the sample results. The decision to accept or reject the null hypothesis H_0 is made on the basis of the information supplied by the observed sample observations. When we perform a hypothesis test, there are four outcomes depending on the actual truth (or falseness) of the null hypothesis H_0 and the decision to reject or not. These four possible outcomes are

1. The decision is to not reject H_0 when, in fact, H_0 is true (correct decision).
2. The decision is to reject H_0 when, in fact, H_0 is true (incorrect decision known as a Type I error).
3. The decision is to not reject H_0 when, in fact, H_0 is false (incorrect decision known as a Type II error).
4. The decision is to reject H_0 when, in fact, H_0 is false (correct decision whose probability is called the Power of the Test).

Each error occurs with some probability. As such we are liable to commit the following two types of errors.

Type I Error: Reject H_0 when it is true.

Type II Error: Accept H_0 when it is wrong, i.e., accept H_0 when H_1 is true. If we write

P(Type I Error) = Probability of rejecting H_0 when H_0 is true $= \alpha$ and

P(Type II Error)= Probability of accepting H_0 when H_0 is false $= \beta$

Then α and β are called the sizes of type I error and type II error, respectively. In practice, type I error amounts to rejecting a lot when it is good and type II error may be regarded as accepting the lot when it is bad. Thus P{Reject a lot when it is good} $= \alpha$ and P{Accept a lot when it is bad}$= \beta$, where α and β are referred to as producer's risk and consumer's risk respectively.

Example: Suppose the null hypothesis, H_0, is: The victim of an automobile accident is alive when he arrives at the emergency room of a hospital.

Type I error: The emergency crew concludes that the victim is dead when, in fact, the victim is alive.

Type II error: The emergency crew concludes that the victim is alive when, in fact, the victim is dead.

α = probability that the emergency crew thinks the victim is dead when, in fact, he is really alive

= P(Type I error)

β = probability that the emergency crew thinks the victim is alive when, in fact, he is dead

= P(Type II error)

The error with the greater consequence is the Type I error. (If the emergency crew thinks the victim is dead, they will not treat him).

4.7 Critical Region and Level of Significance

A region corresponding to a statistic t in the sample space S which amount to rejection of H_0 is termed as critical region of rejection. If ω is the critical region and if $t= t(x_1, x_2, ,x_3,... x_n)$ sample of size n then

$$P\left(t \in \omega \mid H_0\right) = \alpha, \quad P\left(t \in \bar{\omega} \mid H_1\right) = \beta$$

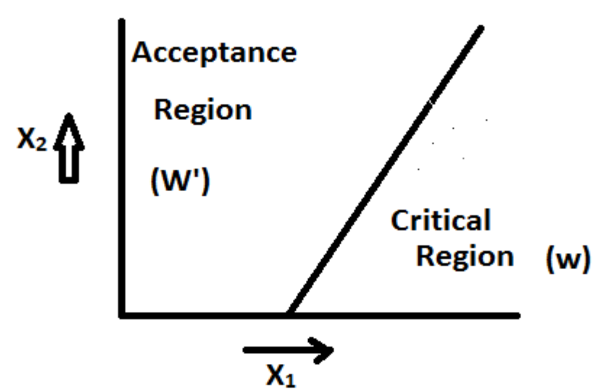
where $\bar{\omega}$, the complementary set of ω, is called the accepting region.

We have $\omega \cup \bar{\omega} = S$ and $\omega \cap \bar{\omega} = \phi$

The probability α that a random value of the statistic t belong to the critical region is known as the level of significance. In other words, level of significance is the size of the type I error (or the maximum producer's risk). The level of significance usually employed in testing are 5% and 1%. The level of significance is always fixed in advanced before collecting the sample information.

Level of significance: α, the probability of type I error, is known as the level of significance of the test. It is also called the size of the critical region.

Critical region: Let $x_1, x_2, \ldots, x_n$ be the sample observations denoted by A. All the value of A will be aggregate of a sample and they constitute a space called the sample space, which is denoted by S. Since the sample values $x_1, x_2, \ldots x_n$ can be taken as a point in n-dimensional space, we specify some region of the n-dimensional space and see whether this point lies within this region or outside this region. We divide the whole sample space S into two disjoint parts W and S-W or Ẇ or W'. The null hypothesis H_0 is rejected if the observed sample point falls in W and if falls in W' we reject H_1 and accept H_0. The region of rejection of H_0 when H_0 is true is that region of the outcome set where H_0 is rejected if the sample point falls in the region W and is called critical region. Evidently, the size of the critical region is α, the probability of committing type I error (discuss below).



Suppose, if the test is based on a sample of size 2 then the outcome set or sample space is the first quadrant in two-dimensional space and a test criterion will enable us to separate our outcome set into complementary subsets,W and W'.If the sample point falls in the subset W, H_0 is rejected, otherwise H_0 is accepted. This is shown in the diagram.

4.8 One-Tailed and Two-Tailed Tests

In any hypothetical test, the critical region is represented by a portion of the area under the probability curve of the sampling distribution of the test statistic. An alternatives hypothesis may be one-sided or two-sided (right-tailed or left-tailed). A one-sided hypothesis claims that a parameter is either larger or smaller than the value given by the null hypothesis. A two-sided hypothesis claims that a parameter is simply not equal to the value given by the null hypothesis - the direction does not matter.

For example: A test for testing the mean of a population H_0: $\mu = \mu_0$ against the alternative hypothesis: H_1: $\mu > \mu_0$ (right-tailed) or H_1: $\mu < \mu_0$ (left-tailed), is a single-tailed test.

In the right-tailed test (H_1: $\mu > \mu_0$), the critical region lies entirely in the right tail of the sampling distribution of $\bar{x}$, while for the left-tailed test (H_1: $\mu < \mu_0$), the critical region is entirely in the left tail of the distribution. A test of statistical hypothesis where the alternative hypothesis is two-tailed such as: H_0: $\mu=\mu_0$, against the alternative hypothesis H_1: $\mu \neq \mu_0$ ($\mu > \mu_0$ or $\mu < \mu_0$), is known as two-tailed test and in such a case the critical region is given by the portion of the area lying in both tails of the probability curve of the test statistic. In particular problem, whether one-tail or two-tailed test is to be applied depends entirely on the nature of the alternative hypothesis. If the alternative hypothesis is two-tailed, we apply two-tailed and if alternative hypothesis is one-tailed, we apply one-tailed test.

4.9 Critical Values or Significant Values

The value of test statistic which separates the critical (or rejection) region and the acceptance region is called the critical value or significant value. It depends upon:

(1) The level of significance used, and

(2) The alternative hypothesis, whether it is two tailed or single tailed.

As has been pointed out earlier, for large samples, the standardized variable corresponding to the statistic t, viz.,

$$Z = \frac{t - E(t)}{S.E(t)} \sim N(0,1), \quad .... (*)$$

Asymptotically as n $\to \infty$. The value of Z given by (*) under the null hypothesis is known as test statistic. The critical value of the test statistic at level of significance α for a two- tailed test is given by Z_α is determined by the equation:

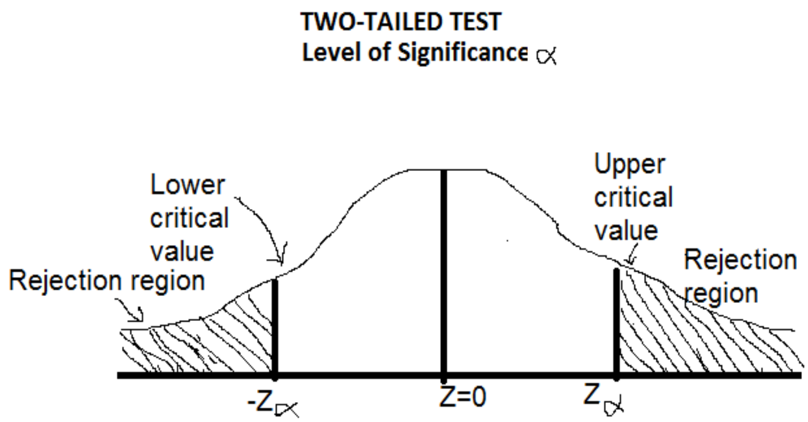
$$P(|Z| > z_\alpha) = \alpha \qquad .....(1)$$

i.e., z_α is the value so that the total area of critical region of both tails is α . Since normal probability curve is a symmetrical curve, from (1), we get

$P(Z > z_\alpha) + P(Z<-z_\alpha)=\alpha \Rightarrow P(Z > z_\alpha) + P(Z > z_\alpha) = \alpha$ [By Symmetry]

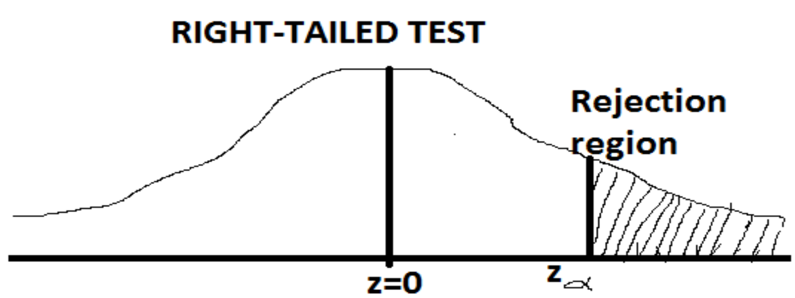
$\Rightarrow \quad 2\,P(Z > z_\alpha)= \alpha, \Rightarrow P(Z > z_\alpha)= \alpha/2$

In other words, the area of each tail is α /**2.** Thus z_α is the value such that area to the right of z_α is $\alpha/2$ and to the left of $(-z_\alpha)$ is $\alpha/2$, as shown in the diagram:



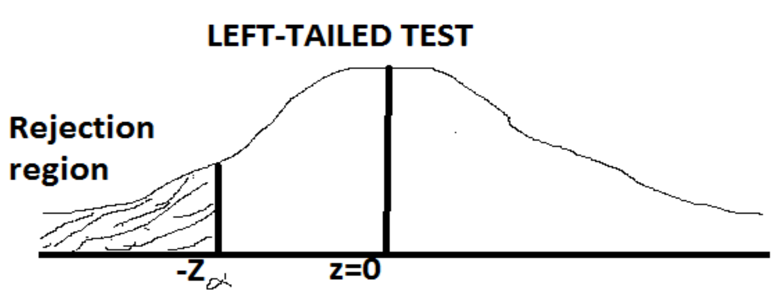
In case of single-tail alternative, the critical value Z_α is determined so that area to the right of it (for right-tailed test) is α and for left-tailed test the total area to the left of $(-Z_\alpha)$ is α (see the Diagrams), i.e.,

For Right-tailed test: $P(Z > z_\alpha) = \alpha$ ....(2)



s

For Left-tailed test: $P(Z <-z_\alpha) = \alpha$ .....(3)



Thus the significance or critical value of Z for a single-tailed test (left or right) at level of significance 'α' is same as the critical value of Z at for a two-tailed test at level of significance '2α'. So, the critical values of Z at commonly used levels of significance for both two-tailed and single-tailed tests are given below. These values have been obtained from equation (1), (2) and (3) on using the Normal Probability Table is**:**

Normal Probability Table

Critical value (Z_α)	Level of significance(α) 1%	5%	10%
Two-tailed test	$\|Z_\alpha\|$=2.58	$\|Z_\alpha\|$=1.96	$\|Z_\alpha\|$=1.645
Right-tailed test	Z_α =2.33	Z_α =1.645	Z_α=1.28
Left-tailed test	Z_α =-2.33	Z_α =-1.645	Z_α=-1.28

Note: - It is to be noted if n is very small i.e. n is less than 30, the above sampling distribution of the test statistic will not be normal. In this case, we have to use other test e.g. t-test.

4.10 Tests of Significance

In the previous section, we encountered a problem to decide whether our sample observations have come from a postulated population or not. Based on sample observations, a test is performed to decide whether the postulated hypothesis is accepted or rejected and this involves a certain amount of risk. The amount of risk is termed as a level of significance. When the hypothesis is rejected, we consider it as a significant result and when a reverse situation is encountered, we consider it as a non-significant result. We have seen that for large values of n, the number of trials, almost all the distributions e.g., Binomial, Poisson etc. are very closely approximated by Normal distribution and in this case we apply Normal Deviate test (Z-test). In cases where the population variance(s) is/are known, we use Z-test. The distribution of Z is always normal with mean zero and variance one. In this subsection, we shall be studying the problem relating to test of significance for large samples only. In statistics, a sample is said to be large if its size exceeds 30.

To perform testing of any statistical hypothesis such as test of significance of mean, difference of mean, single proportion and large sample, we need to set some pre-assumptions. Thus, we follow a procedure which is given below:

Steps of Test of Significance

1. Null hypothesis: Set up the null hypothesis H_0
2. Alternative Hypothesis: Set up the alternative hypothesis H_1 which enables us to decide whether we have to use single-tailed or two-tailed test.
3. Level of Significance: Choose the appropriate level of significance.

4. Test Statistic: Compute the test statistic

$$Z=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}, \text{ under } H_0$$

5. Conclusion: We compare the computed value of Z with the significant value z_α at level of significance, $'\alpha'$.

If $|z|<z_\alpha$, i.e. if value of Z is less than the significant value z_α, we say it is not significant. If $|z|>z_\alpha$, i.e., if the computed value of test statistic is greater than the critical value, then we say result of sample data is significant and the null hypothesis is rejected at the level of significance α.

4.10.1 Test of Significance for Large Samples

As we know for large samples, most of the distribution approximated by normal distribution and also sample variance approaches population variance and is deemed to be almost equal to population variance. Thus in this case, we apply the normal test which is based upon the following fundamental property of normal probability curve.

If $X \sim N(\mu, \sigma^2)$, then $Z=\frac{X-\mu}{\sigma}=\frac{X-E(X)}{\sqrt{V(X)}} \sim N(0,1)$

It should be noted that for all probability, we should expect that standard normal variate lie between -3 to 3. Also the significant values of Z at 5% and 1% levels of significance for two-tailed test are 1.96 and 2.58, respectively.

From normal probability tables, we have

$P[-3\le Z\le 3]=P[|Z| \le 3] = 0.9973 \Rightarrow P[|Z| > 3] = 1-P[|Z| \le 3] = 0.0027$. Thus, the value of Z=3 is regarded as critical or significant value at all levels of significance. Thus if $|Z| > 3$, H_0 is always rejected. If $|Z| \le 3$, we test its significance at certain level of significance usually at 5% and sometimes at 1% level of significance. Also $P[|Z| >1.96]=0.05$ and $P[|Z| >2.58]=0.01$.Thus, significant values of Z at 5% and 1% level of significance are 1.96 and 2.58 respectively. If $|Z| >1.96$, H_0 is rejected at 5% level of significance if $|Z| <1.96$, H_0 may be retained at 5% level of significance. Similarly $|Z|>2.58$, H_0 is rejected at 1% level of significance and if $|Z| <2.58$, H_0 is retained at 1% level of significance. In the following sections we shall discuss the large sample (normal) tests for attributes and variables.

4.10.2 Test of Significance for Single Proportion

If the observations on various items or objects are categorized into two classes c_1 and c_2 (binomial population), viz. defective or not defective item, we often want to test the hypothesis, whether the proportion of items in a particular class, viz., defective items is P_0 or not. For example, the management of a dairy plant is interested in knowing that whether the population of leaked pouches filled by automatic milk filling machine is one percent. Thus for binomial population, the hypothesis wewant to test is whether the sample proportion is representative of the Population proportion $P = P_0$ against H_1: $P\neq P_0$ or H_1: $P>P_0$ or H_1: $P<P_0$ can be tested by Z-test where P is the actual proportion of items in the

population belonging to class c_1. Proportions are mostly based on large samples and hence Z-test is applied.

If X is the number of successes in n independent trials with constant probability P of success for each trial then E (X) = P and V (X) = nPQ, where Q = 1−P. It is known that for large n, the Binomial distribution tends to Normal distribution. Hence, for large n, X~N (nP, nPQ). Therefore, Z statistic for single proportion is given by

$$Z = \frac{X - E(X)}{SE(X)} = \frac{X - E(X)}{\sqrt{V(X)}}$$

$$Z = \frac{(X - nP)}{\sqrt{nPQ}} \sim N(0,1)$$

and we can apply a normal test.

If in a sample of size n, X be the number of persons possessing the given attributes then observed proportion of successes $\frac{X}{n} = p$

$$E(p) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}nP = P$$

$$V(p) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \frac{1}{n^2}nPQ = \frac{PQ}{n}$$

$$S.E.(p) = \sqrt{\frac{PQ}{n}}$$

Since X and consequently X/n is asymptotically normal for large n, the normal test for the proportion of success becomes.

$$Z = \frac{p - E(p)}{SE(p)} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1)$$

Since the probable limits for a normal variate X are $E(X) \pm 3\sqrt{V(X)}$, the probable limits for the proportion of successes are:

$E(p) \pm 3S.E.(p) \quad i.e., \; p \pm 2\sqrt{PQ/n}$

If P is not known then taking p as an estimate of P, the probable limits for the proportion in the population are: $p \pm 3\sqrt{pq/n}$. However, the limits for P at level of significance α are given by:

$p \pm z_\alpha \sqrt{pq/n}$, where z_α is the significant value of Z at level of significance α .

Example 1: A random sample of 500 pineapples was taken from a large consignment and 65 were found to be bad. Show that the S.E. of the proportion of bad ones in a sample of this size is 0.015 and deduce that the percentage of bad pineapples in the consignment almost certainly lies between 8.5 and 17.5.

Solution: Here we are given: n=500

X =number of bad pineapples in the sample =65

p = proportion of bad pineapples in the sample =65/500 = 0.13 $\Rightarrow$q = 1-p = 0.87

since P, the proportion of bad pineapples in the consignment is not known, we may take: $\hat{P} = p = 0.13, \hat{Q} = q = 0.87$

S.E. of proportion = $\sqrt{\hat{P}\hat{Q}/n} = \sqrt{0.13 \times 0.87/500} = 0.015$

Thus, the limits for the proportion of bad pineapples in the consignment are: $\hat{P} \pm 3\sqrt{\hat{P}\hat{Q}/n} = 0.130 \pm 3 \times 0.015 = 0.130 \pm 0.045 = (0.085, 0.175)$

Hence the percentage of bad pineapples in the consignment lies almost certainly between 8.5 and 17.5.

Example 2: In a large consignment of baby food packets, a random sample of 100 packets revealed that 5 packets were leaking. Test whether the sample comes from the population (large consignment) containing 3 percent leaked packets.

Solution: In this example n=100, X=5, P=0.03, $p = X/n = 5/100 = 0.05$.

H_0: P = 0.03 .i.e., the proportion of the leaked pouches in the population is 3 per cent

H_1: P $\neq$ 0.03.

Here, we shall use standard normal deviate (Z) test for single proportion as under

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.05-0.03}{\sqrt{\frac{(0.03)(0.97)}{100}}} = \frac{0.02}{0.01706} = 1.17$$

Since calculated value of Z statistic is less than 1.96 therefore H_0 is not rejected at 5% level of significance which implies that the sample is representative of the population (large consignment) of packets containing 3% leaked packets.

Example 3: A random sample of 500 apples was taken from a large consignment and 60 were found to be bad. Obtained the 98% confidence limits for the percentage of bad apples in the consignment.

Solution: We have

p = proportion of bad apples in the sample =60/500 =0.12

Since significant value of Z at 98% confidence coefficient (level of significance 2%) is 2.33, (from normal tables), 98% confidence limits for population proportion are:

$p \pm 2.33\sqrt{pq/n} = 0.12 \pm 2.33\sqrt{0.12 \times 0.88/500} = 0.12 \pm 2.33 \times \sqrt{0.0002112} = 0.12 \pm 2.33 \times 0.01453$
=(0.08615,0.15385)

Hence 98% confidence limits for percentage of bad apples in the consignment are (8.61,15.38).

Example 4: A normal population has a mean of 0.1 and standard deviation of 2.1. Find the probability that mean of a sample of size 900 will be negative.

Solution: Here we are given that $X \sim N(\mu,\sigma^2)$, where $\mu = 0.1$ and $\sigma = 2.1$ and n =900. Since $X \sim N(\mu,\sigma^2)$, the sample mean $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$. The standard normal variate corresponding to $\bar{x}$ is given by:

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x}-0.1}{2.1/30} = \frac{\bar{x}-0.1}{0.07} \Rightarrow \bar{x} = 0.1 + 0.07Z, \text{ where } Z \sim N(0,1)$$

The required probability p, that the sample mean is negative is given by:

$$p = P(\bar{x}<0) = P(0.1+0.07Z<0) = P(Z<\frac{-0.10}{0.07})$$

$$= P(Z<-1.43) = P(Z\geq 1.43) = 0.5 - P(0<Z<1.43) = 0.5 - 0.4236 = 0.0764.\text{(From normal probability table)}$$

Example 5: The guaranteed average life of a certain type of electric light bulbs 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 to the guaranteed average more than 2.5%. What must be the minimum size of the sample?

Solution: Here μ=1000 hrs, σ = 125 hrs.

Since we do want the sample mean to be less than the guaranteed average mean by more than 2.5%, we should have ($\bar{x}$>1000-2.5% of 1000 $\Rightarrow \bar{x}$>1000-25=975

Let n be the given sample size.Then

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \text{ since sample is large .}$$

We want $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} > \frac{975-1000}{\sigma/\sqrt{n}} = -\frac{\sqrt{n}}{5}$

According to the given condition:

$$P(Z>-\sqrt{n}/5) = 0.90 \Rightarrow\Rightarrow P(0<Z<\sqrt{n}/5) = 0.40$$

$\sqrt{n}/5$ =1.28 (From normal probability table)

$\therefore$ n =25 × $(1.28)^2$=41 (approx)

4.10.3 Test of Significance for Difference of Proportions

Suppose we want to compare two distinct populations with respect to the prevalence of a certain of attribute, say A, among their members. Let X_1, X_2 be the member of persons possessing the given attribute A in random samples of sizes n_1 and n_2 from the two populations, respectively. Then sample proportions are given by: $p_1=X_1/n_1$ and $p_2= X_2/n_2$. If P_1 and P_2 population proportions, then$E(p_1)=P_1$, $E(p_2)=P_2$ and $V(p_1)=(P_1Q_1)/n_1$and $V(p_2)=(P_2Q_2)/n_2$

Since for large samples, p_1 and p_2 are independently and asymptotically normally distributed, (p_1-p_2) is also normally distributed. Then the standard variable corresponding to the difference (p_1-p_2) is given by:

$$Z = \frac{(p_{1-} - p_2) - E(p_{1-} - p_2)}{\sqrt{V(p_{1-} - p_2)}} \sim N(0,1)$$

Under the null hypothesis, $H_0:P_1=P_2$, i.e., there is no significant difference between the sample proportions, we have

$E(p_1-p_2) = E(p_1)- E(p_2)= p_1-p_2=0$

Also $V(p_1-p_2)=V(p_1)+V(p_2)$, (Under H_0)

The covariance term $Cov(p_1,p_2)$ vanishes, since sample proportions are independent.

$$\Rightarrow V(p_1 - p_2) = \frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2} = PQ(\frac{1}{n_1} + \frac{1}{n_2}),$$ [$\because$ under $H_0:P_1=P_2=P$ (say), and $Q_1=Q_2=Q$]

Hence, under $H_0:P_1=P_2$, the test statistic for the difference of proportions becomes:

$$Z = \frac{p_1 - p_2}{\sqrt{PQ(\frac{1}{n_1} + \frac{1}{n_2})}} \sim N(0,1)$$

In general, we do not have any information as to the proportion of A's in the populations from which the samples have been taken. Under $H_0:P_1=P_2=P$(say), an unbiased estimate of the population P, based on both the samples is given by:

$$\hat{P} = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} = \frac{X_1 + X_2}{n_1 + n_2}$$

The estimate is unbiased, since

$$E(\hat{P}) = \frac{1}{n_1 + n_2}E[n_1p_1 + n_2p_2] = \frac{1}{n_1 + n_2}[n_1E(p_1) + n_2E(p_2))$$

$= \frac{1}{n_1+n_2}(n_1P_1+n_2P_2) = P$

So this is the required test statistic.

Example 6: Random samples of 400 men and 600 women were asked whether they would like to have a flyover near 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 against that they are not, at 5% level.

Solution: Null Hypothesis H_0:P_1=P_2=P(say), i.e., there is no significant difference between the opinions of men and women as far as proposal of flyover is concerned.

Alternative hypothesis, H_1:$P_1 \neq P_2$ (two-tailed)

We are given: n_1=400, X_1=Number of men favouring the proposal =200

n_2=600, X_2=Number of women favouring the proposal = 325

$\therefore p_1$=proportion of men favouring the proposal in the sample =X_1 /n_1 =200/400 = 0.5

p_2 = proportion of women favouring the proposal in the sample = X_2/n_2 = 325/600 = 0.541

Test Statistic: Since samples are large, the test statistic under the null hypothesis, H_0 is:

$$Z=\frac{p_1-p_2}{\sqrt{\hat{P}\hat{Q}(\frac{1}{n_1}+\frac{1}{n_2})}} \sim N(0,1)\text{, where}$$

$$\hat{P}=\frac{n_1p_1+n_2p_2}{n_1+n_2}=\frac{X_1+X_2}{n_1+n_2}=\frac{200+325}{400+600}=0.525 \qquad \Rightarrow \hat{Q}=1-\hat{P} = 1-0.525 = 0.475$$

$$\therefore \quad Z=\frac{0.500-0.541}{\sqrt{0.525\times0.475\times(\frac{1}{400}+\frac{1}{600})}}=\frac{-0.041}{\sqrt{0.001039}}=\frac{-0.041}{0.0323}=-1.269$$

Conclusion: Since $|Z|$=1.269 which is less than 1.96, it is not significant at 5% level of significance. Hence H_0 may be accepted at 5% level of significance and we may conclude that men and women do not differ significantly as regards proposal of flyover is concerned.

Example 7: Before an increase in excise duty on tea, 800 persons out of a sample of 1000 persons were found to be tea drinkers. After an increase in duty, 800 people were tea drinkers in a sample of 1200 people. Using standard error of proportion, state whether there is a significant decrease in the consumption of tea after the increase in excise duty?

Solution: In usual notations, we have n_1 = 1000; n_2 = 1200.

p_1 = sample proportion of tea drinkers before increase in excise duty=800/1000 = 0.80

p_2 = sample proportion of tea drinkers after increase in excise duty = 800/1200 = 0.67

Null hypothesis, H_0:P_1=P_2, i.e., there is no significant difference in the consumption of tea before and after the increase in excise duty.

Alternative hypothesis, H_1:P_1>P_2 (right-tailed alternative).

Test statistic. Under the null hypothesis, the test statistic is:

$$Z=\frac{p_1-p_2}{\sqrt{\hat{P}\hat{Q}(\frac{1}{n_1}+\frac{1}{n_2})}} \sim N(0,1)\,, \quad \text{(since samples are large)}$$

Where $\hat{P}=\frac{n_1p_1+n_2p_2}{n_1+n_2}=\frac{800+800}{1000+1200}=\frac{16}{22}$ and $\hat{Q}=1-\hat{P}$ =6/22

$$\therefore Z=\frac{0.80-0.67}{\sqrt{\left(\frac{16}{22}\times\frac{6}{22}\left(\frac{1}{1000}+\frac{1}{1200}\right)\right)}}=\frac{0.13}{0.019}$$

$$\therefore Z=\frac{0.80-0.67}{\left(\sqrt{\frac{16}{22}\times\frac{6}{22}\left(\frac{1}{1000}+\frac{1}{1200}\right)}\right)}=\frac{0.13}{0.019} = 6.842$$

Conclusion: Since Z is much greater than 1.645 as well as 2.33 (since test is one tailed), it is highly significant at both 5% and 1% levels of significance. Hence, we reject the null hypothesis H_0 and conclude that there is a significant decrease in the consumption of tea after increase in the excise duty.

4.10.4 Test for Significance of Single Mean

We have seen that if X_i (i=1, 2,..., n) is a random sample of size n from a normal population with mean μ and variance σ^2, then the sample mean $\overline{X}$ is distributed normally with mean μ and variance σ^2/n i.e., $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.Thus for large samples normal variate corresponding to $\overline{X}$ is

$$Z=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

In test of significance for a single mean we deal the following situations

1. To test if the mean of the population has a specified value (μ_0) and null hypothesis in this case will be H_0: $\mu = \mu_0$ i.e., the population has a specified mean value.
2. To test whether the sample mean differs significantly from the hypothetical value of population mean with null hypothesis as there is no difference between sample mean $\overline{X}$ and population mean (μ).

3. To test if the given random sample has been drawn from a population with specified mean μ_0 and variance σ^2 with null hypothesis the sample has been drawn from a normal population with specified mean μ_0 and variance σ^2.

In all the above three situations, the test statistic is given by

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

If $|Z| < 1.96$, H_0 is not rejected at 5% level of significance which implies that there is no significant difference between sample mean and population mean and whatever difference is there, it exists due to fluctuation of sampling.

$|Z| > 1.96$, H_0 is rejected at 5% level of significance which implies that there is a significant difference between sample mean and population mean. The above situations are illustrated by following examples:

Example 8: A random sample of 100 students gave a mean weight of 64 kg with a standard deviation of 16 kg. Test the hypothesis that the mean weight in the population is 60 kg.

Solution: In this example, n=100, μ =60 kg., $\overline{X}$ =64 kg., σ =16

H_0: μ =60 kg. , i.e. the mean weight in the population is 60 kg.

We shall use standard normal deviate (z) test for single mean as under:

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{64 - 60}{16 / \sqrt{100}} = 2.5$$

Since calculated value of Z statistic is more than 1.96, it is significant at 5% level of significance. Therefore, H_0 is rejected at all levels of significance which implies that mean weight of population is not 60 kg.

Example 9: A sample of 50 cows in a herd has average lactation yield 1290 litres. Test whether the sample has been drawn from the population having herd average lactation yield of 1350 litres with a standard deviation of 65 litres.

Solution: In this example, n=50, μ =1350 litres, $\overline{X}$ =1290, σ =65

H_0: μ =1350 litres i.e., the mean lactation milk yield of the cows in the population is 1350

H_1: $\mu \neq 1350$ litres

We shall use standard normal deviate (Z) test for single mean as under:

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{1290 - 1350}{65 / \sqrt{50}} = -6.53 \Rightarrow |Z| = 6.53$$

Since calculated value of Z statistic is more than 3, it is significant at all levels of significance. Therefore, H_0 is rejected at all levels of significance which implies that the sample has not been drawn from the population having mean lactation milk yield as 1350 litres or there is a significant difference between sample mean and population mean.

4.10.5 Test of significance for Difference of Means

Let X_1 be the mean of a sample of size n_1 drawn from a population with mean μ_1 and variance σ_1^2 and let X_2 be the mean of an independent sample of size n_2 drawn from another population with mean μ_2 and variance σ_2^2. Since sample sizes are large.

$$\overline{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \text{ and } \overline{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

Also $\left(\overline{X}_1 - \overline{X}_2\right)$, being the difference in means of two independent normal variates is also a normal variate. The standard normal variate corresponding to $\overline{X}_1 - \overline{X}_2$ is given by

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - E\left(\overline{X}_1 - \overline{X}_2\right)}{\sqrt{V\left(\overline{X}_1 - \overline{X}_2\right)}} \sim N(0,1)$$

Under the null hypothesis $H_0: \mu_1 = \mu_2$ i.e., the two population means are equal, we get

$$E\left(\overline{X}_1 - \overline{X}_2\right) = E\left(\overline{X}_1\right) - E\left(\overline{X}_2\right) = \mu_1 - \mu_2 = 0$$

$$V\left(\overline{X}_1 - \overline{X}_2\right) = V\left(\overline{X}_1\right) - V\left(\overline{X}_2\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

The covariance terms vanish, since the sample means $\overline{X}_1$ and $\overline{X}_2$ are independent.

Thus under $H_0: \mu_1 = \mu_2$, the Z statistic is given by

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Here σ_1^2 and σ_2^2 are assumed to be known. If they are unknown then their estimates provided by corresponding sample variances s_1^2 and s_2^2 respectively are used, i.e., $\hat{\sigma}_1^2 = s_1^2$ and $\hat{\sigma}_2^2 = s_2^2$, thus, in this case the test statistic becomes

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$$

Remarks: If we want to test whether the two independent samples have come from the same population i.e., if $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (with common S.D. σ), then under $H_0: \mu_1 = \mu_2$

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

If the common variance σ^2 is not known, then we use its estimate based on both the samples which is given by

$$\hat{\sigma}^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

Example 10: In a certain factory there are two independent processes manufacturing the same item. The average weight in a sample of 100 items produced from one process is found to be 50g with a standard deviation of 5g while the corresponding figures in a sample of 75 items from the other process are 52g and 6g respectively. Is the difference between two means significant?

Solution: In this example, $n_1 = 100,\ \overline{X}_1 = 50g,\ s_1 = 5g$ and $n_2 = 75,\ \overline{X}_2 = 52g,\ s_2 = 6g$.

Let μ_1 andμ_2 be the population mean of the weight of items manufactured by two independent processes.

H_0: $\mu_1 = \mu_2$, i.e., mean weights of the items manufactured by two independent processes inthe population is same.

H_0: $\mu_1 \neq \mu_2$

Here, we shall use standard normal deviate test (Z-test) for calculating difference between two means as under

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{50 - 52}{\sqrt{\frac{25}{100} + \frac{36}{75}}} = \frac{-2}{0.8544} = -2.34$$

$$\Rightarrow |Z| = 2.34$$

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{50 - 52}{\sqrt{\frac{25}{100} + \frac{36}{75}}} = -\frac{2}{0.8544} = -2.34 \Rightarrow |z| = 2.34$$

Since calculated value of Z statistic is more than 1.96, therefore, H_0 is rejected at 5% level of significance which implies that there is a significant difference between mean weights of the items obtained from two manufacturing processes.

Example 11: The means of two single large samples of 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of standard deviation 2.5 inches? (test at 5% level of significance).

Solution: In usual notations, we are given: n_1=1000, n_2=2000, $\overline{x_1}$=67.5 inches, $\overline{x_2}$=68.0 inches.

Null hypothesis, H_0: $\mu_1=\mu_2$ and σ =2.5 inches, i.e., the samples have been drawn from the same population of standard deviation 2.5 inches.

Alternative hypothesis H_1: $\mu_1 \neq \mu_2$(two-tailed)

Test statistic: Under H_0, the test statistic is:

$$Z=\frac{(\overline{x_1}-\overline{x_2})}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}} \sim N(0,1) \qquad \text{(since samples are large)}$$

$$\text{Now, } Z=\frac{67.5-68}{2.5\times\sqrt{\frac{1}{1000}+\frac{1}{2000}}}=-5.1$$

Conclusion: Since $|Z|>3$, the value is highly significant and we reject the null hypothesis and conclude that samples are certainly not form the same population with standard deviation 2.5.

Example 12: The average hourly wage of a sample of 150 workers in a plant 'A' was Rs. 2.56 with standard deviation of Rs. 1.08. The average hourly wage of a sample of 200 workers in plant 'B' was Rs. 2.87 with a standard deviation of Rs. 1.28. Can an applicant safely assume that the hourly wages paid by plant 'B' are higher than those paid by plant 'A'?

Solution: Let $X_1$and X_2 denote the hourly wages (in Rs.) of workers in plant 'A' and plant 'B' respectively. Then in usual notations we are given:

n_1=150, n_2=2000, $\overline{x_1}$=2.56, $\overline{x_2}$=2.87

$s_1=1.08=\widehat{\sigma_1}$, $s_2=1.28=\widehat{\sigma_2}$ (since sample are large)

Null hypothesis, H_0: $\mu_1=\mu_2$,i.e., there is no significance difference between mean level of wages of workers in plant A and plant B

Alternative hypothesis, H_1: $\mu_2>\mu_1$ (left –tailed test)

Test statistic: under H_0, the test statistic is :

$$Z=\frac{(\overline{x_1}-\overline{x_2})}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \sim N(0,1)$$

$$Z=\frac{2.56-2.87}{\sqrt{\left(\frac{(1.08)^2}{150}+\frac{(1.28)^2}{200}\right)}}=\frac{-0.31}{\sqrt{0.016}}=\frac{-0.31}{0.126}=-2.46$$

Critical region: For a one tailed test, the critical value of Z at 5% level of significance is 1.645. The critical region for left-tailed test thus consist of all values of Z≤-1.645.

Conclusion: Since calculated value of Z (-2.46) is less then critical value (-1.645), it is significant at 5% level of significant. Hence the null hypothesis rejected at 5% level of significant and we conclude that the average hourly wages paid by plant B are certainly higher than those paid by plant A.

Problems:

(Problems based on all test of significance)

1. A sample of 900 members has a mean 3.4cms. and s.d 261cms. Is the sample from a large population of mean 3.25cms. and s.d. 2.61cms.?If the population is normal and its mean is unknown, find the 95% and 98% fiducial limits of true mean.
2. A survey is proposed to be conducted to know the annual earnings of the old statistics graduates of Delhi University. How large should the sample be taken in order to estimate the mean monthly earnings within plus and minus Rs. 10000 at 95% confidence level? The standard deviation of the annual earning of the entire population is known to be Rs. 30000.
3. In a survey of buying habits, 400 women shoppers are chosen at random in super market A located in certain section of the city. Their average weekly food expenditure is Rs.250 with a standard deviation of Rs.40. For 400 women shoppers chosen at random in super market B in other section of the city, the average weekly food expenditure is Rs. 220 with a standard deviation of Rs. 55test at 1% level of significance whether the average weekly food expenditure of the two populations of shoppers are equal.
4. In a sample of 1000 people in Maharashtra, 540 are rice eaters and the rest wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance?
5. A die is thrown 900 times and a throw 3 or 4 is observed 3240 times. Show that the die cannot be regarded as an unbiased one and find the limits between which the probability of a thrown of 3 or 4 lies.
6. In a year there are 956 births in a town A, of which 52.5% were males while if town A and B combined, this proportion in a total of 1406 births was 0.496. Is their any significant difference in the proportion of males birth in the two towns?
7. In two large populations, there are 30 and 25 percent respectively of blue-eyed people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations?