UNIT - I

Random Variables

Introduction

Consider an experiment of throwing a coin twice. The outcomes {HH, HT, TH, TT} consider the sample space. Each of these outcome can be associated with a number by specifying a rule of association with a number by specifying a rule of association (eg. The number of heads). Such a rule of association is called a random variable. We denote a random variable by the capital letter (X, Y, etc) and any particular value of the random variable by x and y.

Thus a random variable X can be considered as a function that maps all elements in the sample space S into points on the real line. The notation X(S)=x means that x is the value associated with the outcomes S by the Random variable X.

PROBABILITY

Probability (or) Chance: Probably, Chances, Likely, Possible - The terms convey the same meaning.

Example:

1. Probably your method is correct
2. The chances of getting ranks Ram and Babu are equal.
3. It is likely that Ram may not come for taking his classes today.
4. It is possible to reach the college by 8.30am.

Ordinary Language: The word probability means uncertainty about happening.

Mathematics or Statistics: A numerical measure of uncertainty is practiced by the important branch of statistics is called the Theory of Probability.

Day to Day Life:

- **Certainty** - Every day the sun rises in the east
- **Impossibility** - It is possible to live without water
- **Uncertainty** - Probably Raman gets that job. In the theory of probability, we represent certainty by 1, impossibility by 0 and uncertainty by a positive fraction which lies between 0 and 1.

Applications: There is no area in social, physical (or) natural sciences where the probability theory is not used.

- It is the base of the fundamental laws of statistics.
- It gives solutions to betting of games.

• It is extensively used in business situations characterized by uncertainty.

• It is essential tool in statistical inference and forms the basis of the Decision Theory.

Random Experiment (or) Trial and Event (or) Cases: Consider an experiment of throwing a coin. When tossing a coin, we may get a head or tail. Here tossing of a coin is a trial and getting a head or tail is an event.

Throwing of a die is a trial and getting 1 or 2 or 3 or 4 or 5 or 6 is an event.

Favorable Events: The number of outcomes favorable to an event in an experiment is the number of outcomes which entail the happening of the event

Example: In tossing 2 coins the cases favorable to the event of getting a head are HT, TH, and HH.

Exhaustive Events: The total number of possible outcomes in any trial is known as exhaustive events.

Example: In tossing a coin the possible outcomes are getting a head or tail. Hence we have 2 exhaustive events in throwing a coin.

Mutually Exclusive Event: Two events are said to be mutually exclusive when the occurrence of one affects the occurrence of the other. In other words, if A & B are mutually exclusive events and if A happens then B will not happen and vice versa.

Example: In tossing a coin the events head or tail are mutually exclusive, since both tail & head cannot appear in the same time

Equally Likely Events: Two events are said to be equally likely if one of them cannot be expected in preference to the other.

Example: In tossing a coin, head or tail are equally likely event

Independent Event : Two events are said to be independent when the actual happening of one does not influence in any way the happening of the other.

Example : In tossing a coin, the event of getting a head in the 1st toss is independent of getting a head in the 2nd toss, 3rd toss, etc.

Mathematical Definition of Probability: If P is the notation for probability of happening of the event, then

$$P(A) = \frac{number\ of\ favourable\ case\ to\ A}{Total\ number\ of\ out\ comes} = \frac{m}{n}$$

Statistical Definition of Probability: If n trials, an event E happens m times, then $P(E) = \lim_{n\to\infty} \frac{m}{n}$

Axiomatic Definition of Probability:

1. For any event A, $P(A) \geq 0$.
2. $P(S) = 1$
3. If $A_1, A_2, A_3, \dots, A_n$ are finite number of disjoint events of S, then $P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \cdots = \sum P(A_i)$

ADDITION LAW OF PROBABILITY

Case (i): When events are mutually exclusive

If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$.

Case (ii): When events are not mutually exclusive

If A and B are any two events, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

THEOREM : ADDITION LAW OF PROBABILITY

If A and B are any two events and are not disjoint, then $\mathbf{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$

THEOREM : MULTIPLICATION LAW OF PROBABILITY

For two events A and B, $\mathbf{P(A \cap B) = P(A)\,P(B/A) = P(B)\,P(A/B)}$, $\mathbf{P(A) > 0, P(B) > 0}$ where $\mathbf{P(B/A)}$ represents the conditional probability of occurrence of B when the event A has already happened and $\mathbf{P(A/B)}$ is the conditional probability of happening A, given that B has already happened.

RANDOM VARIABLE

In this sample space each of these outcomes can be associated with a number by specifying a rule of association. Such a rule of association is called a random variables.

Eg : Number of heads

We denote random variable by the letter (X, Y, etc) and any particular value of the random variable by x or y.

S = {HH, HT, TH, TT} X(S) = {2, 1, 1, 0}

Thus a random X can be the considered as a fun. That maps all elements in the sample space S into points on the real line. The notation X(S) = x means that x is the value associated with outcome s by the R.V.X.

Example

In the experiment of throwing a coin twice the sample space S is S = {HH,HT,TH,TT}. Let X be a random variable chosen such that X(S) = x (the number of heads).

Note:

Any random variable whose only possible values are 0 and 1 is called a Bernoulli random variable.

Types of Random variables : (i) Discrete random variable (ii) Continuous random variable

DISCRETE RANDOM VARIABLE

Definition : A discrete random variable is a R.V.X whose possible values consitute finite set of values or countably infinite set of values.

Examples

All the R.V.'s from Example : 1 are discrete R.V's

Remark

The meaning of $P(X \leq a)$ is simply the probability of the set of outcomes 'S' in the sample space for which $X(s) \leq a$. Or $P(X \leq a) = P\{S : X(S) \leq a\}$

In the above example: 1 we should write

$P(X \leq 1) = P(HH, HT, TH) = 3/4$

Here $P(X \leq 1) = 3/4$ means the probability of the R.V.X (the number of heads) is less than or equal to 1 is ¾ .

Distribution function of the random variable X or cumulative distribution of the random variable X

Def :

The distribution function of a random variable X defined in $(-\infty, \infty)$ is given by

$F(x) = P(X \leq x) = P\{s : X(s) \leq x\}$

Note:

Let the random variable X takes values x1, x2, ….., xn with probabilities P1, P2, ….., Pn and let x1< x2< ….. <xn

Then we have

$F(x) = P(X < x1) = 0, -\infty < x < x,$

$F(x) = P(X < x1) = 0, P(X < x1) + P(X = x1) = 0 + p1 = p1$

$F(x) = P(X < x2) = 0, P(X < x1) + P(X = x1) + P(X = x2) = p1 + p2$

$F(x) = P(X < xn) = P(X < x1) + P(X = x1) + \ldots + P(X = xn)$

$= p1 + p2 + \ldots + pn = 1$

PROPERTIES OF DISTRIBUTION FUNCTIONS

Property : 1 $P(a < X \leq b) = F(b) - F(a)$, where $F(x) = P(X \leq x)$

Property : 2 $P(a \leq X \leq b) = P(X = a) + F(b) - F(a)$

Property : 3 $P(a < X < b) = P(a < X \leq b) - P(X = b)$

$= F(b) - F(a) - P(X = b)$ by prob (1)

PROBABILITY MASS FUNCTION (OR) PROBABILITY FUNCTION

Let X be a one dimenstional discrete R.V. which takes the values x1, x2, …… To each possible outcome 'x_i' we can associate a number p_i. i.e.,

$P(X (x_i) = pi$ called the probability of x_i. The number $p_i = P(x_i)$ satisfies the following conditions.

(i) $p(x_i) \geq 0, \forall_i$ (ii) $\sum_{i=1}^{\infty} p(x_i) = 1$

The function p(x) satisfying the above two conditions is called the probability mass function (or) probability distribution of the R.V.X. The probability distribution {xi, pi} can be displayed in the form of table as shown below.

$X = x_i$	x_1	x_2	…….	x_i
$P(X = x_i) = p_i$	p_1	p_2	…….	p_i

Notation

Let 'S' be a sample space. The set of all outcomes 'S' in S such that X(S) = x is denoted by writing X = x.

$P(X = x) = P\{S : X(s) = x\}$

|||ly $P(x \leq a) = P\{S : X() \in (-\infty, a)\}$

and $P(a < x \leq b) = P\{s : X(s) \in (a, b)\}$

$P(X = a \text{ or } X = b) = P\{(X = a) \cup (X = b)\}$

$P(X = a \text{ and } X = b) = P\{(X = a) \cap (X = b)\}$ and so on.

Theorem :1 If X1 and X2 are random variable and K is a constant then KX_1, $X_1 + X_2$, X_1X_2, $K_1X_1 + K_2X_2$, X_1-X_2 are also random variables.

Theorem :2

If 'X' is a random variable and f(•) is a continuous function, then f(X) is a random variable.

Note

If F(x) is the distribution function of one dimensional random variable then

I. $0 \leq F(x) \leq 1$
II. If $x < y$, then $F(x) \leq F(y)$
III. $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$
IV. $F(\infty) = \lim_{x \to \infty} F(x) = 1$
V. If 'X' is a discrete R.V. taking values x_1, x_2, x_3
Where $x_1 < x_2 < x_{i-1}\ x_i$ …….. then
$P(X = x_i) = F(x_i) - F(x_{i-1})$

Example:

A random variable X has the following probability function

Values of X	0	1	2	3	4	5	6	7	8
Probability P(X)	a	3a	5a	7a	9a	11a	13a	15a	17a

(i) Determine the value of 'a'

(ii) Find P(X<3), P(X≥3), P(0<X<5)

(iii) Find the distribution function of X.

Solution

Table 1

Values of X	0	1	2	3	4	5	6	7	8
p(x)	a	3a	5a	7a	9a	11a	13a	15a	17a

(i) We know that if p(x) is the probability of mass function then

$$\sum_{i=0}^{8} p(x_i) = 1$$

p(0) + p(1) + p(2) + p(3) + p(4) + p(5) + p(6) + p(7) + p(8) = 1

a + 3a + 5a + 7a + 9a + 11a + 13a + 15a + 17a = 1

81 a = 1

a = 1/81

put a = 1/81 in table 1, e get table 2

Table 2

X = x	0	1	2	3	4	5	6	7	8
P(x)	1/81	3/81	5/81	7/81	9/81	11/81	13/81	15/81	17/81

(ii) $P(X < 3) = p(0) + p(1) + p(2)$

$= 1/81 + 3/81 + 5/81 = 9/81$

(ii) $P(X \geq 3) = 1 - p(X < 3)$

$= 1 - 9/81 = 72/81$

(iii) $P(0 < x < 5) = p(1) + p(2) + p(3) + p(4)$ here 0 & 5 are not include

$= 3/81 + 5/81 + 7/81 + 9/81$

$$= \frac{3 + 5 + 7 + 8 + 9}{81} = \frac{24}{81}$$

(iv) To find the distribution function of X using table 2, we get

X = x	$F(X) = P(x \le x)$
0	$F(0) = p(0) = 1/81$
1	$F(1) = P(X \le 1) = p(0) + p(1)$ $= 1/81 + 3/81 = 4/81$
2	$F(2) = P(X \le 2) = p(0) + p(1) + p(2)$ $= 4/81 + 5/81 = 9/81$
3	$F(3) = P(X \le 3) = p(0) + p(1) + p(2) + p(3)$ $= 9/81 + 7/81 = 16/81$
4	$F(4) = P(X \le 4) = p(0) + p(1) + \dots + p(4)$ $= 16/81 + 9/81 = 25/81$
5	$F(5) = P(X \le 5) = p(0) + p(1) + \dots + p(4) + p(5)$ $= 2/81 + 11/81 = 36/81$
6	$F(6) = P(X \le 6) = p(0) + p(1) + \dots + p(6)$ $= 36/81 + 13/81 = 49/81$
7	$F(7) = P(X \le 7) = p(0) + p(1) + \dots + p(6) + p(7)$ $= 49/81 + 15/81 = 64/81$
8	$F(8) = P(X \le 8) = p(0) + p(1) + \dots + p(6) + p(7) + p(8)$ $= 64/81 + 17/81 = 81/81 = 1$

CONTINUOUS RANDOM VARIABLE

Def : A R.V.'X' which takes all possible values in a given internal is called a continuous random variable.

Example : Age, height, weight are continuous R.V.'s.

PROBABILITY DENSITY FUNCTION

Consider a continuous R.V. 'X' specified on a certain interval (a, b) (which can also be a infinite interval $(-\infty, \infty)$).

If there is a function y = f(x) such that

$$\lim_{\Delta x \to 0} \frac{P(x < X < x + \Delta x)}{\Delta x} = f(x)$$

Then this function f(x) is termed as the probability density function (or) simply density function of the R.V. 'X'.

It is also called the frequency function, distribution density or the probability density function.

The curve y = f(x) is called the probability curve of the distribution curve.

Remark

If f(x) is p.d.f of the R.V.X then the probability that a value of the R.V. X will fall in some interval (a, b) is equal to the definite integral of the function f(x) a to b.

$$P(a < x < b) = \int_a^b f(x)dx \quad \text{(or)}$$

$$P(a \le X \le b) = \int_a^b f(x)dx$$

PROPERTIES OF P.D.F

The p.d.f f(x) of a R.V.X has the following properties

1. In the case of discrete R.V. the probability at a point say at x = c is not zero. But in the case of a continuous R.V.X the probability at a point is always zero.

$$P(X = c) = \int_{-\infty}^{\infty} f(x)dx = [x]_c^C = C - C = 0$$

2. If x is a continuous R.V. then we have p(a ≤ X ≤ b) = p(a ≤ X < b)= p(a < X V b)

IMPORTANT DEFINITIONS INTERMS OF P.D.F

If f(x) is the p.d.f of a random variable 'X' which is defined in the interval (a, b) then

i	Arithmetic mean	$\int_a^b x\, f(x)dx$
ii	Harmonic mean	$\int_a^b \frac{1}{x} f(x)dx$
iii	Geometric mean 'G' log G	$\int_a^b \log x\, f(x)dx$
iv	Moments about origin	$\int_a^b x^r\, f(x)dx$
v	Moments about any point A	$\int_a^b (x-A)^r\, f(x)dx$
vi	Moment about mean μ_r	$\int_a^b (x-\text{mean})^r\, f(x)dx$
vii	Variance μ_2	$\int_a^b (x-\text{mean})^2\, f(x)dx$
viii	Mean deviation about the mean is M.D.	$\int_a^b \lvert x-\text{mean}\rvert\, f(x)dx$

Mathematical Expectations

Def : Let 'X' be a continuous random variable with probability density function f(x). Then the mathematical expectation of 'X' is denoted by E(X) and is given by

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

It is denoted by

$$\mu_r' = \int_{-\infty}^{\infty} x^r f(x) dx$$

Thus

$$\mu_1' = E(X) \qquad (\mu_1' \text{ about origin})$$

$$\mu_2' = E(X^2) \qquad (\mu_2' \text{ about origin})$$

$$\therefore \text{Mean} = \overline{X} = \mu_1' = E(X)$$

And

$$\text{Variance} = \mu_2' - \mu_2'^2$$

$$\text{Variance} = E(X^2) - [E(X)]^2 \qquad \text{(a)}$$

* r^{th} moment (abut mean)

Now

$$E\{X - E(X)\}^r = \int_{-\infty}^{\infty} \{x - E(X)\}^r f(x) dx$$

$$= \int_{-\infty}^{\infty} \{x - \overline{X}\}^r f(x) dx$$

Thus

$$\mu_r = \int_{-\infty}^{\infty} \{x - \overline{X}\}^r f(x) dx \qquad \text{(b)}$$

$$\text{Where} \quad \mu_r = E[X - E(X)^r]$$

This gives the r^{th} moment about mean and it is denoted by μ_r

Put r = 1 in (B) we get

$$\mu_r = \int_{-\infty}^{\infty} \{x - \overline{X}\} f(x)\,dx$$

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

$$= \overline{X} - \overline{X} \int_{-\infty}^{\infty} f(x)\,dx \qquad \left[\because \int_{-\infty}^{\infty} f(x)\,dx = 1\right]$$

$$= \overline{X} - \overline{X}$$

$$\mu_1 = 0$$

Put r = 2 in (B), we get

$$\mu_2 = \int_{-\infty}^{\infty} (x - \overline{X})^2 f(x)\,dx$$

$$\text{Variance} = \mu_2 = E[X - E(X)]^2$$

Which gives the variance interms of expectations.

Note

Let g(x) = K (Constant), then

$$E\left[g(X)\right] = E(K) = \int_{-\infty}^{\infty} K\, f(x)\,dx$$

$$= K \int_{-\infty}^{\infty} f(x)\,dx \qquad \left[\because \int_{-\infty}^{\infty} f(x)\,dx = 1\right]$$

$$= K \,.\, 1 = K$$

Thus $E(K) = K \Rightarrow E[\text{a constant}] = \text{constant}$.

EXPECTATIONS (Discrete R.V.'s)

Let 'X' be a discrete random variable with P.M.F p(x)

Then

$$E(X) = \sum_x x\, p(x)$$

For discrete random variables 'X'

$$E(X^r) = \sum_x x^r\, p(x) \qquad \text{(by def)}$$

If we denote

$$E(X^r) = \mu_r'$$

Then

$$\mu_r' = E[X^r] = \sum_x x^r p(x)$$

Put r = 1, we get

$$\text{Mean } \mu_r' = \sum x\, p(x)$$

Put r = 2, we get

$$\mu_2' = E[X^2] = \sum_x x^2 p(x)$$

$$\therefore \mu_2 = \mu_2' - \mu_1'^2 = E(X^2) - \{E(X)\}^2$$

The r^{th} moment about mean

$$\mu_r' = E[\{X - E(X)\}^r]$$

$$= \sum_x (x - \overline{X})^r p(x), \qquad E(X) = \overline{X}$$

Put r = 2, we get

$$\text{Variance} = \mu_2 = \sum_x ((x - \overline{X})^2 p(x)$$

ADDITION THEOREM (EXPECTATION)

Theorem 1

If X and Y are two continuous random variable with pdf fx(x) and fy(y) then

E(X+Y) = E(X) + E(Y)

MULTIPLICATION THEOREM OF EXPECTATION

Theorem 2

If X and Y are independent random variables,

Then E(XY) = E(X) . E(Y)

Note :

If X1, X2, ……, Xn are 'n' independent random variables, then

E[X1, X2, ……, Xn] = E(X1), E(X2), ……, E(Xn)

Theorem 3

If ‘X’ is a random variable with pdf f(x) and ‘a’ is a constant, then

(i) E[a G(x)] = a E[G(x)]

(ii) E[G(x)+a] = E[G(x)+a]

Where G(X) is a function of ‘X’ which is also a random variable.

Theorem 4

If ‘X’ is a random variable with p.d.f. f(x) and ‘a’ and ‘b’ are constants, then E[ax + b] = a E(X) + b

Cor 1:

If we take a = 1 and b = –E(X) = – X , then we get

E(X- X) = E(X) – E(X) = 0

Note

$$E\left(\frac{1}{X}\right) \neq \frac{1}{E(X)}$$

$$E[\log (x)] \neq \log E(X)$$

$$E(X^2) \neq [E(X)]^2$$

EXPECTATION OF A LINEAR COMBINATION OF RANDOM VARIABLES

Let X1, X2, ……, Xn be any ‘n’ random variable and if a1, a2 , ……, an are constants, then
E[a1X1 + a2X2 + ……+ anXn] = a1E(X1) + a2E(X2)+ ……+ anE(Xn)

Result

If X is a random variable, then

Var (aX + b) = a^2Var(X) ‘a’ and ‘b’ are constants.

Covariance :

If X and Y are random variables, then covariance between them is defined as Cov(X, Y) = E{[X - E(X)] [Y - E(Y)]}

Cov(X, Y) = E(XY) – E(X) . E(Y) (A)

If X and Y are independent, then

E(XY) = E(X) E(Y)

Sub (B) in (A), we get Cov (X, Y) = 0

∴ If X and Y are independent, then

Cov (X, Y) = 0

Note

(i) Cov(aX, bY) = ab Cov(X, Y)

(ii) Cov(X+a, Y+b) = Cov(X, Y)

(iii) Cov(aX+b, cY+d) = ac Cov(X, Y)

(iv) Var (X1 + X2) = Var(X1) + Var(X2) + 2 Cov(X1, X2)

If X1, X2 are independent

Var (X1+ X2) = Var(X1) + Var(X2)

EXPECTATION TABLE

Discrete R.V's	**Continuous R.V's**
1. $E(X) = \sum x\, p(x)$	1. $E(X) = \int_{-\infty}^{\infty} x f(x) dx$
2. $E(X^r) = \mu_r' = \sum_x x^r p(x)$	2. $E(X^r) = \mu_r' = \int_{-\infty}^{\infty} x^r f(x) dx$
3. Mean $= \mu_r' = \sum x\, p(x)$	3. Mean $= \mu_r' = \int_{-\infty}^{\infty} x\, f(x) dx$
4. $\mu_2' = \sum x^2\, p(x)$	4. $\mu_2' = \int_{-\infty}^{\infty} x^2\, f(x) dx$
5. Variance $= \mu_2' - \mu_1'^2 = E(X^2) - \{E(X)\}^2$	5. Variance $= \mu_2' - \mu_1'^2 = E(X^2) - \{E(X)\}^2$

SOLVED PROBLEMS ON DISCRETE R.V'S

Example :1

When die is thrown, 'X' denotes the number that turns up. Find E(X), $E(X^2)$ and Var (X).

Solution

Let 'X' be the R.V. denoting the number that turns up in a die. 'X' takes values 1, 2, 3, 4, 5, 6 and with probability 1/6 for each

X = x	1	2	3	4	5	6
p(x)	1/6	1/6	1/6	1/6	1/6	1/6
	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$	$p(x_6)$

Now

$$
\begin{aligned}
E(X) &= \sum_{i=1}^{6} x_i\, p(x_i) \\
&= x_1 p(x_1) + x_2 p(x_2) + x_3 p(x_3) + x_4 p(x_4) + x_5 p(x_5) + x_6 p(x_6) \\
&= 1 \text{ x } (1/6) + 1 \text{ x } (1/6) + 3 \text{ x } (1/6) + 4 \text{ x } (1/6) + 5 \text{ x } (1/6) + 6 \text{ x } (1/6) \\
&= 21/6 \quad = \quad 7/2 \qquad (1)
\end{aligned}
$$

$$
\begin{aligned}
E(X) &= \sum_{i=1}^{6} x_i\, p(x_p) \\
&= x_1^2 p(x_1) + x_2^2 p(x_2) + x_3^2 p(x_3) + x_4^2 p(x_4) + x_5^2 p(x_5) + x_6 p(x_6) \\
&= 1(1/6) + 4(1/6) + 9(1/6) + 16(1/6) + 25(1/6) + 36(1/6) \\
&= \frac{1+4+9+16+25+36}{6} \quad = \quad \frac{91}{6} \qquad (2)
\end{aligned}
$$

Variance (X) $= \text{Var }(X) = E(X^2) - [E(X)]^2$

$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} \quad = \frac{35}{12}$$

Example :2

Find the value of (i) C (ii) mean of the following distribution given

$$f(x) = \begin{cases} C(x - x^2), & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Solution

$$\text{Given } f(x) = \begin{cases} C(x - x^2), & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

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

$$\int_0^1 C(x - x^2)\,dx = 1 \qquad \text{[using (1)] } [\therefore 0<x<1]$$

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

$$C\left[\frac{1}{2} - \frac{1}{3}\right] = 1$$

$$C\left[\frac{3-2}{6}\right] = 1$$

$$\frac{C}{6} = 1 \qquad C = 6 \qquad (2)$$

Sub (2) in (1), $f(x) = 6(x - x^2)$, $0 < x < 1$ (3)

$$\text{Mean} = E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$= \int_{0}^{1} x\, 6(x - x^2)\, dx \qquad \text{[from (3)]} \qquad [\therefore 0 < x < 1]$$

$$= \int_{0}^{1} (6x^2 - x^3)\, dx$$

$$= \left[\frac{6x^3}{3} - \frac{6x^4}{4}\right]_0^1$$

$\therefore$ Mean = ½

Mean	C
½	6

CONTINUOUS DISTRIBUTION FUNCTION

Def :

If f(x) is a p.d.f. of a continuous random variable 'X', then the function

$$F_X(x) = F(x) = P(X \le x) = \int_{-\infty}^{\infty} f(x)\, dx, \ -\infty < x < \infty$$

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

* PROPERTIES OF CDF OF A R.V. 'X'

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

(ii) $\underset{x \to -\infty}{\text{Lt}} F(x) = 0, \qquad \underset{x \to -\infty}{\text{Lt}} F(x) = 1$

(iii) $P(a \le X \le b) = \int_a^b f(x)\, dx = F(b) - F(a)$

(iv) $F'(x) = \dfrac{dF(x)}{dx} = f(x) \ge 0$

(v) $P(X = x_i) = F(x_i) - F(x_i - 1)$

Example :1.4.1

Given the p.d.f. of a continuous random variable 'X' follows

$$f(x)=\begin{cases} 6x(1-x), & 0<x<1 \\ 0 & \text{otherwise} \end{cases}, \text{ find c.d.f. for 'X'}$$

Solution

$$\text{Given } f(x)=\begin{cases} 6x(1-x), & 0<x<1 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{The c.d.f is } F(x)=\int_{-\infty}^{x} f(x)\,dx,\ -\infty<x<\infty$$

(i) When $x < 0$, then

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

$$= \int_{-\infty}^{x} 0\,dx \quad = 0$$

(ii) When $0< x < 1$, then

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

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

$$= 0 + \int_{0}^{x} 6x(1-x)\,dx \quad = 6\int_{0}^{x} x(1-x)\,dx = 6\left[\frac{x^2}{2}-\frac{x^3}{3}\right]_0^x$$

$$= 3x^2 - 2x^3$$

(iii) When $x > 1$, then

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

$$= \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} 6x(1-x)\,dx + \int_{0}^{x} 0\,dx$$

$$= 6\int_{0}^{1} (x-x^2)\,dx \quad = 1$$

Using (1), (2) & (3) we get

$$F(x)=\begin{cases} 0, & x<0 \\ 3x^2-2x^3, & 0<x<1 \\ 1, & x>1 \end{cases}$$

Example:1.4.2

(i) If $f(x)=\begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ defined as follows a density function ?

(ii) If so determine the probability that the variate having this density will fall in the interval (1, 2).

Solution

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

(a) In $(0, \infty)$, e^{-x} is +ve

$\therefore f(x) \geq 0$ in $(0, \infty)$

(b) $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} f(x)\,dx$

$= \int_{-\infty}^{0} 0\,dx + \int_{0}^{\infty} e^{-x}\,dx$

$= \left[-e^{-x}\right]_0^{\infty} \qquad = -e^{-\infty} + 1$

$= 1$

Hence f(x) is a p.d.f

(ii) We know that

$P(a \leq X \leq b) \qquad = \int_a^b f(x)\,dx$

$P(1 \leq X \leq 2) \qquad = \int_1^2 f(x)\,dx \qquad = \int_1^2 e^{-x}\,dx = [-e^{-x}]_{+1}^{2}$

$= \int_1^2 e^{-x}\,dx = [-e^{-x}]_{+1}^{2}$

$= -e^{-2} + e^{-1} \qquad = -0.135 + 0.368 \qquad = 0.233$

Example:1.4..3

A probability curve y = f(x) has a range from 0 to ∞. If $f(x) = e^{-x}$, find the mean and variance and the third moment about mean.

Solution

$$\text{Mean} = \int_0^\infty x\, f(x)\, dx$$

$$= \int_0^\infty x e^{-x}\, dx \qquad = \left[x[-e^{-x}]-[e^{-x}]\right]_0^\infty$$

Mean = 1

$$\text{Variance } \mu_2 = \int_0^\infty (x-\text{Mean})^2 f(x)dx$$

$$= \int_0^\infty (x-1)^2 e^{-x}\, dx$$

$$\mu_2 = 1$$

Third moment about mean

$$\mu_3 = \int_a^b (x-\text{Mean})^3 f(x)dx$$

Here a = 0, b = ∞

$$\mu_3 = \int_a^b (x-1)^3 e^{-x}\, dx$$

$$= \left\{(x-1)^3(-e^{-x}) - 3(x-1)^2(e^{-x}) + 6(x-1)(-e^{-x}) - 6(e^{-x})\right\}_0^\infty$$

$$= -1 + 3 - 6 + 6 = 2$$

$$\mu_3 = 2$$

MOMENT GENERATING FUNCTION

Def : The moment generating function (MGF) of a random variable 'X' (about origin) whose probability function f(x) is given by

$$M_X(t) = E[e^{tX}]$$

$$= \begin{cases} \displaystyle\int_{x=-\infty}^{\infty} e^{tx} f(x)dx, \text{ for a continuous probably function} \\ \displaystyle\sum_{x=-\infty}^{\infty} e^{tx} p(x), \text{ for a discrete probably function} \end{cases}$$

Where t is real parameter and the integration or summation being extended to the entire range of x.

Example :1.5.1

Prove that the r^{th} moment of the R.V. 'X' about origin is $M_X(t) = \int_{r=0}^{\infty} \frac{t^r}{r!} \mu_r'$

Proof

$$\text{WKT } M_X(t) = E(e^{tX})$$

$$= E\left[1 + \frac{tX}{1!} + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + ... + + \frac{(tX)^r}{r!} + ...\right]$$

$$= E[1] + t\,E(X) + \frac{t^2}{2!}E(X^2) + ... + \frac{t^r}{r!}E(X^r) + ...$$

$$M_X(t) = 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + ... + \frac{t^r}{r!}\mu_r' + ...$$

[using $\mu_r' = E(X^r)$]

Thus r^{th} moment = coefficient of $\frac{t^r}{r!}$

Note

1. The above results gives MGF interms of moments.
2. Since $M_X(t)$ generates moments, it is known as moment generating function.

Example:1.5.2

Find μ_1' and μ_2' from $M_X(t)$

Proof

$$\text{WKT } M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!}\mu_r'$$

$$M_X(t) = \mu_0' + \frac{t}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + ... + \frac{t^r}{r!}\mu_r' \qquad \text{(A)}$$

Differenting (A) W.R.T 't', we get

$$M_X'(t) = \mu_1' + \frac{2t}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + ... \qquad \text{(B)}$$

Put $t = 0$ in (B), we get

$$M_X'(0) = \mu_1' = \text{Mean}$$

$$\text{Mean} = M_1'(0) \quad \text{(or)} \quad \left[\frac{d}{dt}(M_X(t))\right]_{t=0}$$

$$M_X''(t) = \mu_2' + t\,\mu_3' + ...$$

Put $t = 0$ in (B)

$$M_X''(0) = \mu_2' \quad \text{(or)} \quad \left[\frac{d^2}{dt^2}(M_X(t))\right]_{t=0}$$

In general $\mu_r' = \left[\frac{d^r}{dt^r}(M_X(t))\right]_{t=0}$

Example :1.5.3

Obtain the MGF of X about the point X = a.

Proof

The moment generating function of X about the point X = a is $M_X(t) \quad = E[e^{t(X-a)}]$

$$= E\left[1 + t(X-a) + \frac{t^2}{2!}(X-a)^2 + ... + \frac{t^r}{r!}(X-a)^r + ...\right]$$

$$\left[\begin{array}{l}\text{Formula} \\ e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + ...\end{array}\right]$$

$$= E(1) + E[t(X-a)] + E[\frac{t^2}{2!}(X-a)^2] + ... + E[\frac{t^r}{r!}(X-a)^r] + ...$$

$$= 1 + t\,E(X-a) + \frac{t^2}{2!}E(X-a)^2 + ... + \frac{t^r}{r!}E(X-a)^r + ...$$

$$= 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + ... + \frac{t^r}{r!}\mu_r' + ... \qquad \text{Where } \mu_r' = E[(X-a)^r]$$

$$[M_X(t)]_{x=a} \quad = 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + ... + \frac{t^r}{r!}\mu_r' + ...$$

Result:

$M_{CX}(t) = E[e^{tcx}]$ (1)

$M_X(t) = E[e^{ctx}]$ (2)

From (1) & (2) we get

$$M_{CX}(t) = M_X(ct)$$

Example :1.5.4

If X_1, X_2, ….., X_n are independent variables, then prove that

$$M_{X_1+X_2+....+X_n}(t) \quad = E[e^{t(X_1+X_2+....+X_n)}]$$

$$= E[e^{tX_1}.e^{tX_2}.....e^{tX_n}]$$

$$= E(e^{tX_1}).E(e^{tX_2}).....E(e^{tX_n})$$

[∴ $X_1, X_2, \ldots, X_n$ are independent]

$$= M_{X_1}(t).M_{X_2}(t)..........M_{X_n}(t)$$

Example:1.5.5

Prove that if $\cup = \frac{X-a}{h}$, then $M_{\cup}(t) = e^{\frac{-at}{h}}.M_X^{\left(\frac{t}{h}\right)}$, where a, h are constants.

Proof

By definition

$$M_{\cup}(t) = E\left[e^{tu}\right] \qquad \because \left[M_X(t) = E[e^{tx}]\right]$$

$$= E\left[e^{t\left(\frac{X-a}{h}\right)}\right]$$

$$= E\left[e^{\frac{tX}{n}-\frac{ta}{n}}\right]$$

$$= E[e^{\frac{tX}{h}}]\ \ E[e^{\frac{-ta}{h}}]$$

$$= e^{\frac{-ta}{h}}\ E[e^{\frac{tX}{h}}] \qquad \text{[by def]}$$

$$= e^{\frac{-ta}{h}}\ .\ M_X^{\left(\frac{t}{h}\right)}$$

$\therefore M_{\cup}(t) = e^{\frac{-at}{h}}.M_X\left(\frac{t}{h}\right)$, where $\cup = \frac{X-a}{h}$ and $M_X(t)$ is the MGF about origin.

Example:1.5.6

Find the MGF for the distribution where

$$f(x) = \begin{cases} \frac{2}{3} & \text{at } x = 1 \\ \frac{1}{3} & \text{at } x = 2 \\ 0 & \text{otherwise} \end{cases}$$

Solution

Given $\quad f(1) = \frac{2}{3}$

$f(2) = \frac{1}{3}$

$f(3) = f(4) = \ldots\ldots = 0$

MGF of a R.V. 'X' is given by

$$M_X(t) \qquad = E[e^{tx}]$$

$$= \sum_{x=0}^{\infty} e^{tx} f(x)$$

$= e^0 f(0) + e^t f(1) + e^{2t} f(2) + \ldots\ldots$

$= 0 + e^t f(2/3) + e^{2t} f(1/3) + 0$

$= 2/3e^t + 1/3e^{2t}$

$\therefore$ MGF is $M_X(t) = \frac{e^t}{3}[2 + e^t]$

UNIT-2 THEORETICAL DISTRIBUTION

There are two theoretical distribution (i) Discrete distribution (ii) Continuous Distribution

The important discrete distribution of a random variable 'X' are

1. Binomial Distribution
2. Poisson Distribution
3. Geometric Distribution

BINOMIAL DISTRIBUTION

Def : A random variable X is said to follow binomial distribution if its probability law is given by P(x) = p(X = x successes) = nCx px q^{n-x} Where x = 0, 1, 2, ……., n, p+q = 1

Note

Assumptions in Binomial distribution

i) There are only two possible outcomes for each trail (success or failure).

ii) The probability of a success is the same for each trail.

iii) There are 'n' trails, where 'n' is a constant.

iv) The 'n' trails are independent.

Example :1.6.1

Find the Moment Generating Function (MGF) of a binomial distribution about origin.

WKT $\quad M_X(t) = \sum_{x=0}^{n} e^{tx}\, p(x)$

Let 'X' be a random variable which follows binomial distribution then MGF about origin is given by

$$E[e^{tX}] = M_X(t) = \sum_{x=0}^{n} e^{tx}\, p(x)$$

$$= \sum_{x=0}^{n} e^{tx}\, nC_x p^x q^{n-x} \qquad \left[\because p(x) = nC_x p^x q^{n-x}\right]$$

$$= \sum_{x=0}^{n} (e^{tx})\, p^x nC_x q^{n-x}$$

$$= \sum_{x=0}^{n} (pe^t)^x\, nC_x q^{n-x}$$

$$\therefore M_X(t) = (q + pe^t)^n$$

Example:1.6.2

Find the mean and variance of binomial distribution.

Solution

$$M_X(t) = (q+pe^t)^n$$

$$\therefore M_X'(t) = n(q+pe^t)^{n-1}.pe^t$$

Put $t = 0$, we get

$$M_X'(0) = n(q+p)^{n-1}.p$$

$$\text{Mean} = E(X) = np \qquad [\because (q+p)=1] \qquad [\text{Mean } M_X'(0)]$$

$$M_X''(t) = np\left[(q+pe^t)^{n-1}.e^t + e^t(n-1)(q+pe^t)^{n-2}.pe^t\right]$$

Put $t = 0$, we get

$$M_X''(t) = np\left[(q+p)^{n-1} + (n-1)(q+p)^{n-2}.p\right]$$

$$= np\left[1+(n-1)p\right]$$

$$= np + n^2p^2 - np^2$$

$$= n^2p^2 + np(1-p)$$

$$M_X''(0) = n^2p^2 + npq \qquad [\because 1-p=q]$$

$$M_X''(0) = E(X^2) = n^2p^2 + npq$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = n^2/p^2 + npq - n^2/p^2 = npq$$

$$\text{Var }(X) = npq$$

$$\text{S.D} = \sqrt{npq}$$

Example :1.6.3

Find the Moment Generating Function (MGF) of a binomial distribution about mean (np).

Solution

Wkt the MGF of a random variable X about any point 'a' is

$M_x(t)$ (about X = a) $= E[e^{t(X-a)}]$

Here 'a' is mean of the binomial distribution

$M_X(t)$ (about X = np) $= E[e^{t(X-np)}]$

$= E[e^{tX} . e^{-tnp)}]$

$= e^{-tnp} . [-[e^{-tX)}]]$

$= e^{-tnp} . (q+pe^t)\text{-}^n$

$= (e^{-tp})^n . (q + pe^t)^n$

$\therefore$ MGF about mean $= (e^{-tp})^n . (q + pe^t)^n$

Example :1.6.4

Additive property of binomial distribution.

Solution

The sum of two binomial variants is not a binomial variate.

Let X and Y be two independent binomial variates with parameter (n_1, p_1) and (n_2, p_2) respectively.

Then

$$M_X(t) = \left(q_1 + p_1 e^t\right)^{n_1}, \qquad M_Y(t) = \left(q_2 + p_2 e^t\right)^{n_2}$$

$$\therefore\ M_{X+Y}(t) = M_X(t).M_Y(t) \qquad [\because \text{X \& Y are independent R.V.'s}]$$

$$= \left(q_1 + p_1 e^t\right)^{n_1}.\left(q_2 + p_2 e^t\right)^{n_2}$$

RHS cannot be expressed in the form $\left(q + pe^t\right)^n$. Hence by uniqueness theorem of MGF X+Y is not a binomial variate. Hence in general, the sum of two binomial variates is not a binomial variate.

Example :1.6.5

If $M_X(t) = \left(q+pe^t\right)^{n_1}$, $M_Y(t) = \left(q+pe^t\right)^{n_2}$, then

$$M_{X+Y}(t) = \left(q+pe^t\right)^{n_1+n_2}$$

Problems on Binomial Distribution

1. Check whether the following data follow a binomial distribution or not. Mean = 3; variance = 4.

Solution

Given Mean $np = 3$ (1)

Variance $npr = 4$ (2)

$$\frac{(2)}{(1)} \Rightarrow \frac{np}{npq} = \frac{3}{4}$$

$$\Rightarrow q = \frac{4}{3} = 1\frac{1}{3} \text{ which is} > 1.$$

Since $q > 1$ which is not possible $(0 < q < 1)$. The given data not follow binomial distribution.

Example :1.6.5

The mean and SD of a binomial distribution are 5 and 2, determine the distribution.

Solution

Given Mean = $np = 5$ (1)

$SD = \sqrt{npq} = 2$ (2)

$$\frac{(2)}{(1)} \Rightarrow \frac{np}{npq} = \frac{4}{5} \Rightarrow q = \frac{4}{5}$$

$$\therefore p = 1 - \frac{4}{5} = \frac{1}{5} \qquad \Rightarrow \qquad p = \frac{1}{5}$$

Sub (3) in (1) we get

$n \times 1/5 = 5$

$n = 25$

$\therefore$ The binomial distribution is

$P(X = x) = p(x) \quad = nC_x p^x q^{n-x}$

$= 25C_x(1/5)^x(4/5)^{n-x}, \qquad x = 0, 1, 2, \ldots.., 25$

4.The probability of a man hitting a target is 1/3. How many times must he fire so that the probability of hitting at least once is more than 90%?

Solution :

Let X be the number of times he hits the target.P=1/3 => q=2/3. Then X follows a Binomial distribution, with P(X=x) = $nC_x p^x q^{n-x} = nC_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{n-x}$

given $P(X \geq 1)$ >0.9 To find n

- ⇨ 1-P(X<1) > 0.9
- ⇨ - P(X<1) > - 0.1 => P(X<1) < 0.1 => P(X=0) < 0.1
- ⇨ $nC_0 p^0 q^{n-0} = nC_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^{n-0}$ < 0.1 => $\left(\frac{2}{3}\right)^n$
- ⇨ n log(2/3) < log(0.1)
- ⇨ n > 5.65
- ⇨ $n \geq 6$
- ⇨ Therefore he must fire at least 6 times.

5.Out of 800 families with 4 children each, how many families would be expected to have (i) 2 boys and 2 girls (ii) at least 1 boy (iii) at most 2 girls and (iv) children of both genders. Assume equal probabilities for boys and girls.

Solution: Considering each child is a trial, n = 4.

Assuming that birth of a boy is a success, Then X ~ B (n, p)

P[X=x] = $nC_x p^x q^{n-x}$, x = 0, 1, 2, …..n.

By data, p = ½ and q= ½.

(i) P[2 boys and 2 girls] = P[X = 2] = $4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{3}{8}$

∴ No. of families having 2 boys and 2 girls = 800x$\frac{3}{8}$ = 300

(ii) P[at least 1 boy] = P[X≥1] = 1- P[X<1] = 1- P[X=0]

= 1- $4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4$ = 1- $\frac{1}{16} = \frac{15}{16}$

∴ No. of families having at least 1 boy = 800x$\frac{15}{16}$ = 750.

(iii) P[at most 2 girls] = P[exactly 0 girl, 1 girl (or) 2 girls]

= P[X=4, X=3, X=2]

$= 1- \{P[X=0] + P[X=1]\}$

$=1- \{4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4 + 4C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^3\}$

$= 1-[\left(\frac{1}{2}\right)^4 + 4 \left(\frac{1}{2}\right)^4] = 1-\frac{5}{16} = \frac{11}{16}$

$\therefore$ No. of families having at most 2 girls = $800 \text{x} \frac{11}{16} = 550$

(iv) P[children of both genders] = 1 – P[children of the same gender]

= 1 – {P (all are boys) + P (all are girls)}

$= 1 - \{P(X=4) + P(X=0)\}$

$= 1 - \{4C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^0 + 4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4\}$

$= 1 - 2\left(\frac{1}{2}\right)^4 = 1 - \frac{1}{8} = \frac{7}{8}$

$\therefore$ No. of families having children of both gender = $800 \text{x} \frac{7}{8} = 700$

Passion Distribution

Def :

A random variable X is said to follow if its probability law is given by

$$P(X = x) \qquad = p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots, \infty$$

Poisson distribution is a limiting case of binomial distribution under the following conditions or assumptions.

1. The number of trails ‘n’ should e infinitely large i.e. $n \to \infty$.
2. The probability of successes ‘p’ for each trail is infinitely small.
3. $np = \lambda$, should be finite where λ is a constant.

*** To find MGF**

$$
\begin{aligned}
M_X(t) &= E(e^{tx}) \\
&= \sum_{x=0}^{\infty} e^{tx} p(x) \\
&= \sum_{x=0}^{\infty} e^{tx} \left(\frac{\lambda^x e^{\lambda}}{x!} \right) \\
&= \sum_{x=0}^{\infty} \frac{e^{-\lambda} (\lambda e^t)^x}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\
&= e^{-\lambda} \left[1 + \lambda e^t + \frac{(\lambda e^t)^2}{2!} + \ldots\ldots \right] \\
&= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}
\end{aligned}
$$

Hence $\quad M_X(t) = e^{\lambda(e^t-1)}$

* To find Mean and Variance

WKT $\quad M_X(t) = e^{\lambda(e^t-1)}$

$\therefore \quad M_X'(t) = e^{\lambda(e^t-1)}.e^t$

$M_X'(0) = e^{-\lambda}.\lambda$

$$\mu_1' = E(X) = \sum_{x=0}^{\infty} x.p(x)$$

$$= \sum_{x=0}^{\infty} x.\frac{e^{-\lambda}\lambda^{x}}{x!} \quad = \sum_{x=0}^{\infty} \frac{x.\, e^{-\lambda}\lambda\lambda^{x-1}}{x!}$$

$$= 0 + e^{-\lambda}.\lambda \sum_{x=1}^{\infty} \frac{x.\lambda^{x-1}}{x!}$$

$$= \lambda e^{-\lambda}.\sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda e^{-\lambda}\left[1+\lambda+\frac{\lambda^{2}}{2!}+.....\right]$$

$$= \lambda e^{-\lambda}.e^{\lambda}$$

Mean $= \lambda$

$$\mu_2' \quad = E[X^2] \quad = \sum_{x=0}^{\infty} x^2.p(x) \quad = \sum_{x=0}^{\infty} x^2.\frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= \sum_{x=0}^{\infty} \{x(x-1)+x\}.\frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= \sum_{x=0}^{\infty} \frac{x(x-1)e^{-\lambda}\lambda^{x}}{x!} + \sum_{x=0}^{\infty} \frac{x.e^{-\lambda}\lambda^{x}}{x!}$$

$$= e^{-\lambda}\lambda^{2} \sum_{x=0}^{\infty} \frac{\lambda^{x-2}}{(x-2)(x-3)....1} + \lambda$$

$$= e^{-\lambda}\lambda^{2} \sum_{x=0}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda$$

$$= e^{-\lambda}\lambda^{2}\left[1+\frac{\lambda}{1!}+\frac{\lambda^{2}}{2!}+....\right]+\lambda$$

$$= \lambda^{2} + \lambda$$

Variance $\mu_2 = E(X^2) - [E(X)]^2 \quad = \lambda^2 + \lambda - \lambda^2 = \lambda$

Variance $= \lambda$

Hence Mean = Variance $=\lambda$

Note : * sum of independent Poisson Vairates is also Poisson variate.

PROBLEMS ON POISSON DISTRIBUTION

Example:1.7.1

If x is a Poisson variate such that P(X=1) = $\frac{3}{10}$ and P(X=2) = $\frac{1}{5}$, find the P(X=0) and P(X=3).

Solution

$$P(X = x) \quad = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$\therefore P(X=1) \quad = e^{-\lambda}\lambda \quad = \frac{3}{10} \qquad \text{(Given)}$$

$$= \lambda e^{-\lambda} \quad = \frac{3}{10} \qquad (1)$$

$$P(X=2) \quad = \frac{e^{-\lambda}\lambda^2}{2!} = \frac{1}{5} \qquad \text{(Given)}$$

$$\frac{e^{-\lambda}\lambda^2}{2!} = \frac{1}{5} \qquad (2)$$

$$(1) \Rightarrow e^{-\lambda}\lambda = \frac{3}{10} \qquad (3)$$

$$(2) \Rightarrow e^{-\lambda}\lambda^2 = \frac{2}{5} \qquad (4)$$

$$\frac{(3)}{(4)} \Rightarrow \frac{1}{\lambda} = \frac{3}{4}$$

$$\lambda = \frac{4}{3}$$

$$\therefore \quad P(X=0) \quad = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-4/3}$$

$$P(X=3) \quad = \frac{e^{-\lambda}\lambda^3}{3!} = \frac{e^{-4/3}(4/3)^3}{3!}$$

Example :1.7.2

If X is a Poisson variable

P(X = 2) = 9 P(X = 4) + 90 P(X=6)

Find (i) Mean if X (ii) Variance of X

Solution

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}, \; x = 0,1,2,.....$$

Given P(X = 2) = 9 P(X = 4) + 90 P(X=6)

$$\frac{e^{-\lambda}\lambda^2}{2!} = 9\frac{e^{-\lambda}\lambda^4}{4!} + 90\frac{e^{-\lambda}\lambda^6}{6!}$$

$$\frac{1}{2} = \frac{9\lambda^2}{4!} + \frac{90\lambda^4}{6!}$$

$$\frac{1}{2} = \frac{3\lambda^2}{8} + \frac{\lambda^4}{8}$$

$$1 = \frac{3\lambda^2}{4} + \frac{\lambda^4}{4}$$

$$\lambda^4 + 3\lambda^2 - 4 = 0$$

$\lambda^2 = 1$ or $\lambda^2 = -4$

$\lambda = \pm 1$ or $\lambda = \pm 2i$

$\therefore$ Mean = λ = 1, Variance = λ = 1

$\therefore$ Standard Deviation = 1

GEOMETRIC DISTRIBUTION

Def: A discrete random variable 'X' is said to follow geometric distribution, if it assumes only non-negative values and its probability mass function is given by

$P(X=x) = p(x) = q^{x-1}$; $x = 1, 2, \ldots\ldots, 0 < p < 1,$ Where $q = 1-p$

Example:1.8.1

To find MGF

$$M_X(t) = E[e^{tx}]$$

$$= \sum e^{tx} p(x)$$

$$= \sum_{x=1}^{\infty} e^{tx} q^{x-1} p$$

$$= \sum_{x=1}^{\infty} e^{tx} q^{x} q^{-1} p$$

$$= \sum_{x=1}^{\infty} e^{tx} q^{x} p / q$$

$$= p / q \sum_{x=1}^{\infty} e^{tx} q^{x}$$

$$= p / q \sum_{x=1}^{\infty} (e^{t} q)^{x}$$

$$= p / q \left[(e^{t}q)^{1} + (e^{t}q)^{2} + (e^{t}q)^{3} + \ldots \right]$$

Let $x = e^{t}q$ $= p / q \left[x + x^{2} + x^{3} + \ldots \right]$

$$= \frac{p}{q} x \left[1 + x + x^{2} + \ldots \right] \quad = \frac{p}{q} (1-x)^{-1}$$

$$= \frac{p}{q} qe^{t} \left[1 - qe^{t} \right] \quad = pe^{t} \, [1 - qe^{t}]^{-1}$$

$$\therefore M_X(t) = \frac{pe^{t}}{1 - qe^{t}}$$

*** To find the Mean & Variance**

$$M_X'(t) = \frac{(1-qe^t)pe^t - pe^t(-qe^t)}{(1-qe^t)^2} = \frac{pe^t}{(1-qe^t)^2}$$

$\therefore$ E(X) = $M_X'(0)$ = 1/p

$\therefore$ Mean = 1/p

Variance $\quad \mu_X''(t) = \frac{d}{dt}\left[\frac{pe^t}{(1-qe^t)^2}\right]$

$$= \frac{(1-qe^t)^2 pe^t - pe^t 2(1-qe^t)(-qe^t)}{(1-qe^t)^4}$$

$$= \frac{(1-qe^t)^2 pe^t + 2pe^t qe^t(1-qe^t)}{(1-qe^t)^4}$$

$$M_X''(0) = \frac{1+q}{p^2}$$

$$\text{Var (X)} = E(X^2) - [E(X)]^2 = \frac{(1+q)}{p^2} - \frac{1}{p^2} \Rightarrow \frac{q}{p^2}$$

$$\text{Var (X)} = \frac{q}{p^2}$$

Note:

Another form of geometric distribution

$P[X=x] = q^x p$; x = 0, 1, 2, ….

$$M_X(t) = \frac{p}{(1-qe^t)}$$

Mean = q/p, Variance = q/p^2

Example:1.8.2

If the MGF of X is $(5\text{-}4et)^{-1}$, find the distribution of X and P(X=5)

Solution

Let the geometric distribution be

$P(X = x) = q^x p, \qquad x = 0, 1, 2, \ldots..$

The MGF of geometric distribution is given by

$$\frac{p}{1-qe^t} \qquad (1)$$

Here $M_X(t) = (5 - 4e^t)^{-1} \Rightarrow 5^{-1}\left[1-\frac{4}{5}e^t\right]^1 \qquad (2)$

Company (1) & (2) we get $\quad q=\frac{4}{5};\ p=\frac{1}{5}$

$\therefore P(X = x) \quad = pq^x, \qquad x = 0, 1, 2, 3, \ldots\ldots.$

$$= \left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^x$$

$$P(X = 5) \quad = \left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^5 = \frac{4^5}{5^6}$$

CONTINUOUS DISTRIBUTIONS

If 'X' is a continuous random variable then we have the following distribution

1. Uniform (Rectangular Distribution)
2. Exponential Distribution
3. Normal Distribution

9.1 Uniform Distribution (Rectangular Distribution)

Def : A random variable X is set to follow uniform distribution if its

Def : A random variable X is set to follow uniform distribution if its

$$f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & \text{otherwise}\end{cases}$$

*** To find MGF**

$$\begin{aligned}M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f(x)dx\\ &= \int_a^b e^{tx}\frac{1}{b-a}dx\\ &= \frac{1}{b-a}\left[\frac{e^{tx}}{t}\right]_b^a\\ &= \frac{1}{(b-a)t}\left[e^{bx}-e^{at}\right]\end{aligned}$$

$\therefore$ The MGF of uniform distribution is

$$M_X(t) = \frac{e^{bt}-e^{at}}{(b-a)t}$$

* To find Mean and Variance

$$E(X) \quad = \int_{-\infty}^{\infty} x\, f(x)dx$$

$$= \int_a^b b_x \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x\,dx = \frac{\left(\frac{x^2}{2}\right)_a^b}{b-a}$$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2} = \frac{a+b}{2}$$

$$\text{Mean} \quad \mu_1' = \frac{a+b}{2}$$

Putting $r = 2$ in (A), we get

$$\mu_2' = \int_a^b x^2 f(x)dx = \int_a^b \frac{x^2}{b-a} dx$$

$$= \frac{a^2+ab+b^2}{3}$$

$$\therefore \text{ Variance} = \mu_2' - \mu_1'^2$$

$$= \frac{b^2+ab+b^2}{3} - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)^2}{12}$$

$$\text{Variance} = \frac{(b-a)^2}{12}$$

PROBLEMS ON UNIFORM DISTRIBUTION

Example

If X is uniformly distributed over $(-\alpha,\alpha)$, $\alpha< 0$, find α so that

(i) $P(X>1) = 1/3$

(ii) $P(|X| < 1) = P(|X| > 1)$

Solution

If X is uniformly distributed in $(-\alpha, \alpha)$, then its p.d.f. is

$$f(x) = \begin{cases} \frac{1}{2\alpha} & -\alpha < x < \alpha \\ 0 & \text{otherwise} \end{cases}$$

(i) $P(X>1) = 1/3$

$$\int_1^\alpha f(x)dx = 1/3$$

$$\int_1^\alpha \frac{1}{2\alpha} dx = 1/3$$

$$\frac{1}{2\alpha}(x)_1^\alpha = 1/3 \qquad \Rightarrow \frac{1}{2\alpha}(\alpha-1) = 1/3$$

$$\alpha = 3$$

(ii) $P(|X| < 1) = P(|X| > 1) = 1 - P(|X| < 1)$

$$P(|X| < 1) + P(|X| < 1) = 1$$

$$2\,P(|X| < 1) = 1$$

$$2\,P(-1 < X < 1) = 1$$

$$2\int_{-1}^{1} f(x)dx = 1$$

$$2\int_{-1}^{1} \frac{1}{2\alpha} dx = 1$$

$$\Rightarrow \alpha = 2$$

Note:

1. The distribution function F(x) is given by

$$F(x) = \begin{cases} 0 & -\alpha < x < \alpha \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & b < x < \infty \end{cases}$$

2. The p.d.f. of a uniform variate 'X' in (-a, a) is given by

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

THE EXPONENTIAL DISTRIBUTION

Def :A continuous random variable 'X' is said to follow an exponential distribution with parameter λ>0 if its probability density function is given by

$$F(x) = \begin{cases} \lambda e^{-\lambda x} & x > a \\ 0 & \text{otherwise} \end{cases}$$

To find MGF

Solution

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

$$= \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx \qquad = \lambda \int_0^{\infty} e^{-(\lambda - t)x} dx$$

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

$$= \frac{\lambda}{-(\lambda - t)} \left[e^{-\infty} - e^{-0} \right] \qquad = \frac{\lambda}{\lambda - t}$$

$$\therefore \text{ MGF of } x = \frac{\lambda}{\lambda - t}, \lambda > t$$

*** To find Mean and Variance**

We know that MGF is

$$M_X(t) = \frac{\lambda}{\lambda - t} = \frac{1}{1 - \frac{t}{\lambda}} = \left(1 - \frac{t}{\lambda}\right)^{-1}$$

$$= 1 + \frac{t}{\lambda} + \frac{t^2}{\lambda^2} + \ldots.. + \frac{t^r}{\lambda^r}$$

$$= 1 + \frac{t}{\lambda} + \frac{t^2}{2!}\left(\frac{2!}{\lambda^2}\right) + \ldots.. + \frac{t^r}{r!}\left(\frac{t!}{\lambda^r}\right)$$

$$M_X(t) = \sum_{r=0}^{\infty} \left(\frac{t}{\lambda}\right)^r$$

$$\therefore \text{ Mean} \qquad \mu_1' = \text{Coefficient of } \frac{t^1}{1!} = \frac{1}{\lambda}$$

$$\mu_2' = \text{Coefficient of } \frac{t^2}{2!} = \frac{2}{\lambda^2}$$

$$\text{Variance} = \mu_2 = \mu_2' - \mu_1'^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

$$\text{Variance} = \frac{1}{\lambda^2} \qquad \text{Mean} = \frac{1}{\lambda}$$

Example: 1.10.1

Let 'X' be a random variable with p.d.f

$$F(x) = \begin{cases} \frac{1}{3}e^{\frac{-x}{3}} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Find 1) $P(X > 3)$ 2) MGF of 'X'

Solution

WKT the exponential distribution is

$F(x) = \lambda e^{-\lambda x}, \quad x > 0$

Here $\lambda = \frac{1}{3}$

$P(x>3) = \int_3^\infty f(x)dx \qquad = \int_3^\infty \frac{1}{3}e^{-\frac{x}{3}}dx$

$P(X>3) = e^{-1}$

MGF is $M_X(t) \qquad = \frac{\lambda}{\lambda - t}$

$$= \frac{\frac{1}{3}}{\frac{1}{3} - t} \qquad = \frac{\frac{1}{3}}{\frac{1-3t}{3}} \qquad = \frac{1}{1-3t}$$

$M_X(t) = \frac{1}{1-3t}$

Note

If X is exponentially distributed, then

$P(X > s+t / x > s) \qquad = P(X > t)$, for any $s, t > 0$.

TUTORIAL QUESTIONS

1.It is known that the probability of an item produced by a certain machine will be defective is 0.05. If the produced items are sent to the market in packets of 20, fine the no. of packets containing at least, exactly and atmost 2 defective items in a consignment of 1000 packets using (i) Binomial distribution (ii) Poisson approximation to binomial distribution.

2. The daily consumption of milk in excess of 20,000 gallons is approximately exponentially distributed with . 3000 = θ The city has a daily stock of 35,000 gallons. What is the probability that of two days selected at random, the stock is insufficient for both days.

3. binomial variable X satisfies the relation 9P(X=4)=P(X=2) when n=6. Find the parameter p of the Binomial distribution.

4. Find the M.G.F for Poisson Distribution.

5. If X and Y are independent Poisson variates such that P(X=1)=P(X=2) and P(Y=2)=P(Y=3). Find V(X-2Y).

6. In a component manufacturing industry, there is a small probability of 1/500 for any component to be defective. The components are supplied in packets of 10. Use Poisson distribution to calculate the approximate number of packets containing (1). No defective. (2). Two defective components in a consignment of 10,000 packets.

Example :3

The mean and SD of a binomial distribution are 5 and 2, determine the distribution.

Solution

Given Mean = np = 5 (1)

$SD = \sqrt{npq} = 2$ (2)

$$\frac{(2)}{(1)} \Rightarrow \frac{np}{npq} = \frac{4}{5} \Rightarrow q = \frac{4}{5}$$

$$\therefore p = 1 - \frac{4}{5} = \frac{1}{5} \qquad \Rightarrow \qquad p = \frac{1}{5}$$

Sub (3) in (1) we get

n x 1/5 = 5

n = 25

$\therefore$ The binomial distribution is

$P(X = x) = p(x) = nC_x p^x q^{n-x}$

$= 25C_x(1/5)^x(4/5)^{n-x}$, $x = 0, 1, 2, \ldots, 25$

Example :4

If X is a Poisson variable

P(X = 2) = 9 P(X = 4) + 90 P(X=6)

Find (i) Mean if X (ii) Variance of X

Solution

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0,1,2,\ldots..$$

Given P(X = 2) = 9 P(X = 4) + 90 P(X=6)

$$\frac{e^{-\lambda}\lambda^2}{2!} = 9\frac{e^{-\lambda}\lambda^4}{4!} + 90\frac{e^{-\lambda}\lambda^6}{6!}$$

$$\frac{1}{2} = \frac{9\lambda^2}{4!} + \frac{90\lambda^4}{6!}$$

$$\frac{1}{2} = \frac{3\lambda^2}{8} + \frac{\lambda^4}{8}$$

$$1 = \frac{3\lambda^2}{4} + \frac{\lambda^4}{4}$$

$$\lambda^4 + 3\lambda^2 - 4 = 0$$

$\lambda^2 = 1$ or $\lambda^2 = -4$

$\lambda = \pm 1$ or $\lambda = \pm 2i$

$\therefore$ Mean = λ = 1, Variance = λ = 1

$\therefore$ Standard Deviation = 1

UNIT- TESTING HYPOTHESIS

1. **Define Sample**
   A finite subset of statistical individuals in a population is called sample.
2. **Define Sample Size**:
   The number of individuals in a sample is called the sample size.
3. **Define Parameter**
   Any statistical constant which is computed by considering each and every observation of the population is called as a parameter.
4. **Define Statistic**
   Any statistical constant which is computed by considering a part of the information from the population is called as a statistic.
5. **What is sampling distribution of the statistic?**
   The Probability distribution of a statistic is called a sampling distribution.
6. **Define Standard Error**
   The standard deviation of sampling distribution of a statistic is known as its standard error and it is denoted by (S.E)
7. **Explain Test of Significance**
   A very important aspect of the sampling theory is the study of tests of significance which enable us to decide on the basis of the sample results, if
   (i) The deviation between the observed sample statistic and the hypothetical parameter value is significant.
   (ii) The deviation between two sample statistics is significant.
8. **Define Null Hypothesis**
   A null hypothesis is denoted as H_0 and it is assumed to be true by carrying out the test procedure. The null hypothesis always assumes that there is no significant difference between the things that are compared.
9. **Define Alternative Hypothesis**
   It is a contradiction to the null hypothesis. It is represented as H_1. It may be an inequality (or) greater than (or) less than type. The alternate hypothesis decides the nature of test under consideration.

10. Define Type I error and Type II error.

Type I error: Rejecting H_0, when H_0 is true

Type II error: Accepting H_0, when H_0 is false.

11. What is Producer's risk and Consumer's risk?

The sizes of Type I error and Type II errors are called Producer's risk and Consumer's risk respectively.

12. Explain Level of Significance

Level of significance is the probability of type I error. It is denoted by α. It is the size of the rejection region. The significance of α is that it gives the percentage of cases in which the conclusion derived on the basis of sample will go wrong.

13. Define Critical Region

In tests of significance, the total area of the probability curve is divided into two regions (i) Acceptance and (ii) Rejection. The rejection region is also called as the critical region and its size is equal to α. Consequently the size of the acceptance region will be 1-α.

14. Define Critical Value

The value of the test statistic which divides the rejection region and the acceptance region is called the critical value. The Critical values is usually extracted from the statistical tables.

15. Define Student's t-Distribution

A random variable T is said to follow student's t-Distribution or simply t-distribution, if its pdf is given by

$f(t) = \frac{1}{\sqrt{v}\beta\left(\frac{v}{2},\frac{1}{2}\right)}\left(1+\frac{t^2}{v}\right)^{\frac{-(v+1)}{2}} \; -\infty < t < \infty$, where v denotes the number of degrees of freedom of the t-distribution.

16. Uses of t-distribution

The t-distribution is used to test the significance of the difference between

(i) The mean of a small sample and the mean of the population

(ii) The means of two small samples and

(iii) The coefficient of correlation in the small sample and that in the population, assumed zero.

17. State the properties of t – distribution

(i) The probability curve of the t-distribution is similar to the standard normal curve, and is symmetric about t=0, bell-shaped and asymptotic to the t-axis

(ii) For sufficiently large value of v, the t-distribution tends to the standard normal distribution.

(iii) The mean of the t-distribution is zero.

(iv) The variance of the t-distribution is $\frac{v}{v-2}$, if v>2 and is greater than 1, but it tends to 1 as $v \to \infty$

1. **An automatic machine fills tea in sealed tins with mean weight of tea 1 kg and standard deviation of 1gm. A random sample of 50 tins was examined, and it was found that their mean weight was 999.50g. State whether the machine is working properly or not.**

Solution: $\bar{x}$= sample mean = 999.50 μ = population mean = 1000 σ = population standard deviation = 1 n= 50 large sample

H_0= the machine work properly i.e $\bar{x} = \mu$

H_1= the machine not work properly i.e $\bar{x} \neq \mu$

LOS = 5 %

Test statistics: $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{999.50-1000}{\frac{1}{\sqrt{50}}}$ = =-3.53535 The critical value of z =1.96

Conclusion: reject the null hypothesis and accept the alternative hypothesis ie the machine not work properly as Z cal is greater than Z tab

2. **The weights of fish in a certain pond that is regularly stocked are considered to be normally distributed with a mean of 3.1 kg and a standard deviation of 1.1kg. A random sample of size 30 is selected from the pond and the sample mean is found to be 2.4kg. Is these sufficient evidence to indicate that the mean weight of the fish differs from 3.1kg? Use a 10% significance level.**

Solution: $\bar{x}$= sample mean = 2.4 μ = population mean = 3.1 σ = population standard deviation = 1.1 n= 30 large sample

H_0= The mean weight of the fish does not differ significantly i.e $\bar{x} = \mu$

H_1= The mean weight of the fish differ significantly i.e $\bar{x} \neq \mu$

LOS = 10 %

Test statistics: $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{2.4-3.1}{\frac{1.1}{\sqrt{30}}}$ = =-3.48

The critical value of z =1.645

Conclusion: reject the null hypothesis and accept the alternative hypothesis ie The mean weight of the fish differ significantly as Z cal is greater than Z tab

3. **A Stenographer claims that she can type at a rate of 120 words per minute. Can use reject her claim on the basis of 100 trials in which she demonstrates a mean of 116 words with a standard deviation of 15 words? Use 5% level of significance.**

Solution: $\bar{x}$= sample mean = 116 μ = population mean = 120 σ = population standard deviation = not given

n= 100 large sample s= Sample standard deviation = 15

H_0= The mean typing speed =120

5. **The average annual pay in 1989 was Rs 21,128 in the state of Tamil Nadu and Rs. 25,233 in the state of Maharastra. There is a difference of Rs 4,105. Suppose that a statistician believes that the difference is much less for employees in the manufacturing industry and takes an independent random sample of employees in the manufacturing industry in each state. The results are as follows:**

At the 0.05 significance level, do the data support the statistician's belief that for employees in the manufacturing industry, the mean annual salary in Tamil Nadu differs from the mean annual salary in Maharastra by less than Rs 4105?

State	$\bar{x}$	s	n
TamilNadu	21,900	3,700	150
Maharashtra	24,800	3,100	190

Solution

n1= 150 $\bar{x}1 = 21900$ s1 =3700

n2= 190 $\bar{x}2 = 24800$ s2 =3100

H_0= there is no significant difference between the average of two samples $\mu_1 - \mu_2 = 4105$

H_1= there is a significant difference between the average of two samples $\mu_1 - \mu_2 \neq 4105$

LOS = 5 %

Test statistics: $z = \frac{(x_1 - x_2) - \mu 1 - \mu 2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(21900 - 24800) - 4105}{\sqrt{\frac{3700^2}{150} + \frac{3100^2}{190}}} = -18.5995$

The critical value of z =1.96

Conclusion: reject the null hypothesis and accept the alternative hypothesis that there is a significant difference between the averages of two samples

6. **A college conducts both day and night classes intended to be identical. A sample of 100 day students' field estimation results as below:** $\overline{x_1} = 72.4 \ \ and \ \ \sigma_1 = 14.8$ **A sample of 200 night students' field examination results as below:** $\overline{x_2} = 73.9 \ \ and \ \ \sigma_2 = 17.9$ **Are the two means statistically equal at 10% level of significance?**

Solution n_1=100 $\overline{x_1} = 72.4 \ \ and \ \ \sigma_1 = 14.8$

n_2=200 $\overline{x_2} = 73.9 \ \ and \ \ \sigma_2 = 17.9$

H_0= there is no significant difference between day and night classes $\mu_1=\mu_2$

H_1= there is a significant difference between day and night classes $\mu_1 \neq \mu_2$

LOS = 5 %

Test statistics: $z= \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \dfrac{120-124}{\sqrt{\dfrac{12^2}{250} + \dfrac{14^2}{124}}} =$ -0.77025

The critical value of z =1.96

Conclusion: Z cal is less than Z critical hence accept the null hypothesis that there is no significant difference between day and night classes

An automatic machine fills tea in sealed tins with mean weight of tea 1 kg and standard deviation of 1gm. A random sample of 50 tins was examined, and it was found that their mean weight was 999.50g. State whether the machine is working properly or not.

Solution:

$\bar{x}$= sample mean = 999.50; μ = population mean = 1000;

σ= populationstandard deviation = 1; n= 50 large sample

i. H_0= The machine work properly i.e$\bar{x}$ = μ
ii. H_1= The machine not work properly i.e$\bar{x} \neq$μ
iii. LOS = 5 %
iv. Test statistics:Z = $\dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{999.50 - 1000}{\frac{1}{\sqrt{50}}}$ = =-3.53535

The critical value of z =1.96

v. Conclusion: reject the null hypothesis and accept the alternative hypothesis

i.e the machine not work properly as Z cal is greater than Z tab.

4 .Test significance of the difference between the means of the samples, drawn from two normal populations with the same SD usingthe following data :

	Size	Mean	SD
Sample-1	***100***	***61***	***4***
Sample-2	***200***	***63***	***6***

Solution:

Given

	Size	Mean	SD
Sample-1	100	61	4
Sample-2	200	63	6

We want test that the samples are drawn from a population with same SD.

i. $H_0: \mu_1 = \mu_2$

ii. $H_1: \mu_1 \neq \mu_2$

iii. LOS = 5 %

iv. Test statistics :Z = $\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{61-63}{\sqrt{\frac{16}{100} + \frac{36}{200}}} = -3.43 \Rightarrow |Z| = 3.43$

The table value of Z at 5% level is 1.96

Calculated value >The table value.

v. Conclusion: reject the null hypothesis H_0.

Two random samples of sizes 400 and 500 have mean 10.9 and 11.5 respectively. Can the samples be regarded as drawn from the same population with variance 25?

Solution: Given

	Size	Mean	Variance
Sample-1	400	10.9	25
Sample-2	500	11.5	25

We want test that the samples are drawn from a population with variance 25

i. $H_0: \mu_1 = \mu_2$

ii. $H_1: \mu_1 \neq \mu_2$

iii. LOS = 5 %

iv. Test statistics :Z = $\frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{10.9-11.5}{5\sqrt{\frac{1}{400} + \frac{1}{500}}} = -\frac{12}{\sqrt{45}} = -1.78$

$$|Z| = 1.78$$

The table value of Z at 5% level is 1.96

Calculated value < The table value.

v. Conclusion: Accept the null hypothesis H_0.

A random sample of 10 boys had the following IQs: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100.Does data support the assumption of a population mean IQ of 100?

Solution: Given n=10 and $\mu = 100$

i. H$_0$:$\mu = 100$

ii. H$_1$:$\mu = 100$

iii. LOS = 5 %

x	70	120	110	101	88	83	95	98	107	100	$\sum x$=972
x^2	70^2	120^2	110^2	101^2	88^2	83^2	95^2	98^2	107^2	100^2	$\sum x^2$ =96312

$\bar{x} = \frac{\sum x}{n} = \frac{972}{10} = 97.2$; $s^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = 183.96 \Rightarrow s = 13.5$

iv. Test statistics : $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}} = \frac{97.2-100}{\frac{13.5}{\sqrt{9}}} = -\frac{2.8}{4.5} = -0.62$

The table value of t at 5% level is 2.262

Calculated value<The table value.

v. Conclusion: Accept the null hypothesis H$_0$.

16.A random sample of 10 boys had the following IQs: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100.Does the data support the assumption of a population mean IQ of 100?

Solution: Given n=10 and $\mu = 100$

i. H$_0$:$\mu = 100$

ii. H$_1$:$\mu = 100$

iii. LOS = 5 %

x	70	120	110	101	88	83	95	98	107	100	$\sum x$=972
x^2	70^2	120^2	110^2	101^2	88^2	83^2	95^2	98^2	107^2	100^2	$\sum x^2$ =96312

$\bar{x} = \frac{\sum x}{n} = \frac{972}{10} = 97.2$; $s^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = 183.96 \Rightarrow s = 13.5$

iv. Test statistics : $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}} = \frac{97.2-100}{\frac{13.5}{\sqrt{9}}} = -\frac{2.8}{4.5} = -0.62$

The table value of t at 5% level is 2.262

Calculated value<The table value.

v. Conclusion: Accept the null hypothesis H$_0$.

17. In a sample of 8 observations, the sum of the squared deviations of items from the mean was 94.5. In another sample of 10 observations, the value was found to be 101.7. Test whether the difference in the variances is significant at 5% level.

Solution: Given

$n_1 = 8 \text{ and } n_2 = 10$; $\sum (x-\bar{x})^2 = 94.5$ and $\sum (y-\bar{y})^2 = 101.7$

The sample variances are $s_1^2 = \frac{\sum (x-\bar{x})^2}{n_1} = \frac{94.5}{8}$ and $s_2^2 = \frac{\sum (y-\bar{y})^2}{n_2} = \frac{101.7}{10}$

An automatic machine fills tea in sealed tins with mean weight of tea 1 kg and standard deviation of 1gm. A random sample of 50 tins was examined, and it was found that their mean weight was 999.50g. State whether the machine is working properly or not.

Solution:

$\bar{x}$= sample mean = 999.50; μ = population mean = 1000;

σ= populationstandard deviation = 1; n= 50 large sample

i. H_0= The machine work properly i.e$\bar{x}$ = μ
ii. H_1= The machine not work properly i.e$\bar{x} \neq$μ
iii. LOS = 5 %
iv. Test statistics:$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{999.50-1000}{\frac{1}{\sqrt{50}}}$ = =-3.53535

The critical value of z =1.96

v. Conclusion: reject the null hypothesis and accept the alternative hypothesis

i.e the machine not work properly as Z cal is greater than Z tab.

18. The following random sample are measurement of heat-producing capability(in millions of calories per ton) of specimens of coal from two mines:

Mine I : 8260 8130 8350 8070 8340
Mine II: 7950 7890 7900 8140 7920 7840

Use the a=0.01 level of significance to test whether the difference between the means of these two sample is significant?

Solution:

i. $H_0: \mu_1 = \mu_2$
ii. $H_1: \mu_1 \neq \mu_2$
iii. LOS = 1 % d.f=n_1+n_2-2=9

$$\sum x_1 = 41150 \ and \sum x_2 = 47650$$

$$\sum x_1^2 = 338727500 \ and \sum x_2^2 = 378316200$$

$$\bar{x}_1 = 8230 \ and \ \bar{x}_2 = 7940$$

$$s_1^2 = \frac{\sum x_1^2}{n_1} - \left(\frac{\sum x_1}{n_1}\right)^2 = 12600 \ and \ s_2^2 = \frac{\sum x_2^2}{n_2} - \left(\frac{\sum x_2}{n_2}\right)^2 = 9100$$

$$S^2 = \frac{n_1 s_1^2 + +n_2 s_2^2}{n_1 + n_2 - 2} = 13066.67$$

iv. Test statistics :$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{8230-7940}{114.31\sqrt{\frac{1}{5}+\frac{1}{6}}} = 4.19$

Calculated value < The table value.

v. Conclusion: Accept the null hypothesis H_0.

UNIT-IV

- **F –Test**

  The F-distribution is formed by the ratio of two independent chi-square variables divided by their respective degrees of freedom.

- Properties of F – Distribution

  1. The F – Distribution is positively skewed and its skewness decreases with increase in ν_1 & ν_2
  2. The value of F is always be positive or 0
  3. The mean of F – Distribution is $\frac{\nu_2}{\nu_2 - 2}$ $\nu_2 > 2$
  4. The shape of F – Distribution is depends upon the number of degree of freedom.

- **Testing of Hypothesis for Equality of Two Variances**

  Null hypothesis H_0: $\sigma_1^2 = \sigma_2^2$

  To test the equality of variances in two independently selected random samples drawn from two normal populations with H_0: $\sigma_1^2 = \sigma_2^2$

$$\boxed{F = \frac{s_1^2}{s_2^2}}$$ when $S_1 > S_2$

$$S_1 = \sqrt{\frac{1}{n_1 - 1}\sum (x_i - \bar{x}_1)^2}$$

$$S_2 = \sqrt{\frac{1}{n_2 - 1}\sum (x_j - \bar{x}_2)^2}$$

$\nu_1 = n_1 - 1$ & $\nu_2 = n_2 - 1$ $S_1 > S_2$

Note To test whether two independent samples have been drawn from same normal population

(i) Equality of population means using t –test

(ii) Equality of population variances using F-test

First apply F test then t-test

If sample variances are given as s_1 and s_2 the S_1 and S_2 are calculated as

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} \text{ and } S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$$

- Chi-square distribution

  Definition: A distribution obtained from the multiplying the ratio of sample variance to population variance by the degrees of freedom when random samples are selected from a normally distributed population

- Contingency Table

  Data arranged in table form for the chi-square independence test

- Expected Frequency

  The frequencies obtained by calculation.

- Goodness-of-fit Test

  A test to see if a sample comes from a population with the given distribution.

- Independence Test

  A test to see if the row and column variables are independent.

- Observed Frequency

  The frequencies obtained by observation. These are the sample frequencies.

Goodness-of-Fit Chi-Square Test

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

df = # categories – 1

Chi-Square Test for Independence

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad \text{df} = (\text{rows} - 1)(\text{columns} - 1) \qquad \text{Expected value} = \frac{\text{row sum} \times \text{column sum}}{\text{grand total}}$$

8. Define Snedecor's F-Distribution

A random variable F is said to follow snedecor's F-distribution or simply F-distribution if its pdf is given by

$$f(F) = \frac{(v_1|v_2)^{v_1/2} F^{v_1/2 - 1}}{\beta\left(\frac{v_1}{2}, \frac{v_2}{2}\right)\left(1 + \frac{v_1 F}{v_2}\right)^{(v_1+v_2)/2}}, \ F>0.$$

9. Write down the value of χ^2 for 2 X 2 contingency table with cell frequencies a, b, c and d.

The 2 X 2 contingency table for χ^2 test is $\chi^2 = \frac{(a+b+c+d)(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$

A lawn-equipment shop is considering adding a brand of lawn movers to its merchandise. The manager of the shop believes that the highest quality lawn movers are Trooper, lawn eater and Nipper, and he needs to decide whether it makes a difference which of these three shop adds to its existing merchandise. Twenty owners of each of these three types of lawn movers are randomly sampled and asked how satisfied they are with their lawn movers

Lawn mover	**Very satisfied**	**Satisfied**	**Not satisfied**	**Total**
Trooper	**11**	**6**	**3**	**20**
Lawn eater	**13**	**4**	**3**	**20**
Nipper	**13**	**6**	**1**	**20**

Are the owners of the lawn movers homogeneous in their response of the survey? Use a 5% significance level.

Solution :

H_0: Owners of lawn movers homogenous in this response to the server.

H_1: Owners of the lawn movers are not homogenous in the response to the server.

We combine satisfied and not satisfied category as not satisfied category has value less than S.

Observed	Expected	$(\text{Obse-Exp})^2/\text{Exp}$
11	12	0.0833
9	8	0.1250
13	12	0.0833
7	8	0.1250
13	12	0.0833
7	8	0.1250
Total		0.0625

Eater from X^2 table for

$df = (r - 1)(C - 1) = (3 - 1)(2 - 1) = 2$

$\alpha = 0.05$

$x_u^2 = 5.991$

$(x^2 = 0.625) < (x_u^2 = 5.991)$

H_0 is accepted. Hence the owners of the Lawn Movers are homogeneous in their response to the survey.

In an industry, 200 workers, employed for a specific job, were classified according their performance and training received /not received to test independence of a specific training and performance. The dates is Performance

	Good	Not Good	Total
Trained	**100**	**50**	**150**
Untrained	**20**	**30**	**50**
Total	**120**	**80**	**200**

5% level of significance

Solution : H_0: Worker are homogenous

H_1: Worker are not homogenous in.

We combine satisfied and not satisfied category as not satisfied category has value less than S.

Observed	Expected	$(\text{Obse-Exp})^2/\text{Exp}$
100	90	1.111111111
50	60	1.666666667
20	30	3.333333333
30	20	5
Total		11.11111111

Eater from X^2 table for

$df = (r - 1)(C - 1) = (2 - 1)(2 - 1) = 1 \quad \alpha = 0.05$

$x_u^2 = (x^2 = 11.11111111) > (x_u^2 =)$

H_0 is Rejected. Hence the the workers are not homogeneous.

To see whether silicon chip sales are independent of where the U.S economy is in the business cycle, data have been collected on the weekly sales of Zippy Chippy, a silicon Valley firm , and on whether the U.S economy was rising to a cycle peak, falling to a cycle through, of at a cycle through the results are Economically weakly chip sales:

	High	Medium	Low	Total
At Peak	**8**	**3**	**7**	**18**
Rising	**4**	**8**	**5**	**17**
Falling	**8**	**4**	**3**	**15**
Total	**20**	**15**	**15**	**50**

O_i	E_i	$(O_i - E_i)^2 / E_i$
8	7.2	0.0889
4	6.8	1.153
8	6.0	0.6667
3	5.4	1.0667
8	5.1	0.056
4	4.5	0.474
7	5.4	0.0020
5	.1	0.5
3	4.5	5.657

d.f = (r - 1) (c - 1) = 2 x 2 = 4

Table Value = 7.779 for = 0.10

). Two scientists, Dr.X and Dr. Y, find a previously unknown type of fish in a remote river. They both trap some Fish, from different location some distance apart. The weights of their fish are as follows(in kg):

Dr.X : 0.15 0.18 0.25 0.36 0.42 0.44

Dr.Y: 0.25 0.26 0.26 0.30 0.32 0.33 0.37 0.37

Do the figure support, at the5% significance level, the theory that the fish came from populations with the same variation?

ution:

vi. $H_0: \sigma_1^2 = \sigma_2^2$

vii. $H_1: \sigma_1^2 \neq \sigma_2^2$

viii. LOS = 5 % ; d.f is $F_{(5,7)}$

ix. Calculate the mean and variance

Dr. X		Dr. Y	
x	x*x	y	y*y
0.15	0.023	0.25	0.063
0.18	0.032	0.25	0.063
0.25	0.063	0.26	0.068
0.36	0.130	0.30	0.090
0.42	0.176	0.32	0.102
0.44	0.194	0.33	0.109
		0.37	0.137
		0.37	0.137
1.8	0.617	1.71	0.4939

$n_1 = 6 \; and \; n_2 = 8 \quad ; \quad \bar{x} = 0.3 \; and \; \bar{y} = 0.308$

$$s_1^2 = \frac{\Sigma x^2}{n_1} - \left(\frac{\Sigma x}{n_1}\right)^2 = 0.193 \quad and \quad s_2^2 = \frac{\Sigma y^2}{n_2} - \left(\frac{\Sigma y}{n_2}\right)^2 = 0.0024$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = 0.23 \quad and \quad S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = 0.003$$

$$\text{here } S_1^2 > S_2^2 \qquad \therefore F = \frac{S_1^2}{S_2^2} = 76.67$$

The table value of $F_{(5,7)} = 4.87$ at 5% LOS

Hence calculated value > The table value

x. Conclusion: Reject the null hypothesis H_0.

. The theory predicts the proportion of the beans in the four groups A,B,Cand D should be 9:3:3:1. In an experiment among 1600 beans, the numbers in the four groups were 882, 313, 287, and 118. Does the experimental result support the theory?

Solution:

i. H_0= The proportion of beans in the four groups should be 9:3:3:1

ii. H_1= The proportion of beans in the four groups should not be 9:3:3:1

iii. LOS = 5 %

iv. Test statistics: Given that the observed frequencies are respectively 882, 313, 287, and 118. The total observed frequencies=882+313+287+118=1600.

Under $H_0$the expected frequencies are 900, 300, 300 and 100

The Test statistics$\chi^2 = \sum\frac{(O-E)^2}{E}$

O_i	E_i	$(O_i - E_i)^2/E_i$
882	900	0.360
313	300	0.563
287	300	0.563
118	100	3.240

$$\chi^2 = 4.727$$

Degrees of freedom ν=n-1=3

From the χ^2 table at 5% level for 3 d.f =7.81

Calculated value < The table value.

iv. Conclusion: Accept the null hypothesis H_0

QUEUEING THEORY

The input (or Arrival Pattern)

(a) Basic Queueing Process:

Since the customers arrive in a random fashion. Therefore their arrival pattern can be described in terms of Prob. We assume that they arrive according to a Poisson Process i.e., the no f units arriving until any specific time has a Poisson distribution. This is the case where arrivals to the queueing systems occur at random, but at a certain average rate.

(b) Queue (or) Waiting Line

(c) Queue Discipline

It refers to the manner in which the members in a queue are choosen for service

Example:

i. First Come First Served (FIFS) (or) First In First Out (FIFO)
ii. Last Come, First Served (LCFS)
iii. Service In Random Order (SIRO)
iv. General Service Discipline (GD)

1.1 TRANSIENT STATE

A Queueing system is said to be in transient state when its operating characteristics are dependent on time. A queueing system is in transient system when the Prob. distribution of arrivals waiting time & servicing time of the customers are dependent.

1.2 STEADY STATE:

If the operating characteristics become independent of time, the queueing system is said to be in a steady state. Thus a queueing system acquires steady state, when the Prob. distribution of arrivals are independent of time. This state occurs in the long run of the system.

2 TYPES OF QUEUEING MODELS

There are several types of queueing models. Some of them are

1. Single Queue - Single Server Point -infinite capacity
2. Single Queue - Single Server Point -finite capacity

Note

P = Traffic intensity or utilization factor which represents the proportion of time the servers are busy = λ/4.

MODEL I (M/M/1): $(\infty/\boldsymbol{FIFO})$

$[\boldsymbol{SINGLE\ SERVER\ WITH\ INFINITE\ CAPACITY\ QUEUE}]$

(i) Probability that there are n customers in the system = $p_n = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right)$

(ii) Probability that the system is busy = $\frac{\lambda}{\mu}$

(iii) Probability that the system is empty $= p_0 = 1 - \frac{\lambda}{\mu}$

(iv) The probability that the waiting time of a customers in the system is exceeds‘t’.

$$p[w > t] = e^{-(\mu-\lambda)t}$$

(v) The probability that the waiting time of a customers in the system is exceeds‘t’ $= \frac{\lambda}{\mu} e^{-(\mu-\lambda)t}$

(vi) The expected number of customers in the system $= L_s = \frac{\boldsymbol{\lambda}}{\boldsymbol{\mu - \lambda}}$

(vii) The expected number of customers in the queue $= L_q = \frac{\boldsymbol{\lambda^2}}{\boldsymbol{\mu(\mu - \lambda)}}$

(viii) The average waiting time of a customer in the $system = W_s = \frac{\mathbf{1}}{\boldsymbol{\mu - \lambda}}$

(ix) The average waiting time of a customer in the queue $= w_q = \frac{\boldsymbol{\lambda}}{\boldsymbol{\mu(\mu - \lambda)}}$

(x) probability that the no. of customers in the system exceeds $'\mathrm{k}' = P[N > k] = \left(\frac{\lambda}{\mu}\right)^{k+1}$

1. Write down the little’s formula.

(i). $L_s = \lambda W_s$

(ii).$L_q = \lambda W_q$

(iii) $W_s = W_q + \frac{1}{\mu}$

(iv). $L_q = L_s - \frac{\lambda}{\mu}$

2. Customers arrive at one-man barber shop according to a poisson process with a mean interarrival of 12min. customers spend an average of 10min. in the barber's chair.

(i). what is the expected number of customers in the barber's and in the queue?

(ii). Calculate the % of time an arrival can walk straight into the barber's chair without having to wait.

(iii). How much time can customer expect to spend in the barber's shop?

(iv). Management will provide another chair and here another barber when a customer's waiting time in the shop exceeds 1.25hr. Here much must the average rate of arrivals increase to warrant a second barber?

(v). what is the average time customers spend in the queue?

(vi). what is the probability that the waiting time in the system to greater than 30 min?

(vii). Calculate the % of customers who have to wait prior to getting into barber's chair?

(viii). what is the probability that more than 3 customers are in the system?

SOLUTION:

Given: One man barber shop → single server

Customers → infinite capcity

The given problem is (M/M/1) :(∞/FIFO)

$\lambda = \frac{1}{12}$, $\quad \mu = \frac{1}{E[T]} = \frac{1}{10}$, $\quad \rho = \frac{\lambda}{\mu} = \frac{\frac{1}{12}}{\frac{1}{10}} = \frac{5}{6}$

(i). (a). The expected number of customers in the barber shop

$$L_s = \frac{\rho}{1-\rho} = \frac{5/6}{1-5/6} = 5.$$

(i). (b). The expected number of customers in the queue

$$L_q = L_s - \rho = 5 - \frac{5}{6} = 4.17$$

(ii). P [a customer straight goes to the barber's chair] = P [No customer in the system]

$$P_0 = 1 - \rho = 1 - \frac{5}{6} = 0.1666$$

The percentage of time an arrival need not wait= 0.1666X100=16.67.

(iii). Expected time a customer spends in the barber's shop $= w_s = \frac{1}{\lambda} L_s = (12)(5) = 60\ min.$

(iv). Given $w_s > 75$

$$\frac{1}{\mu-\lambda} > 75$$

$$\lambda + \frac{1}{75} > \mu$$

$$\lambda > \frac{13}{150}$$

Hence to warrant a second barber, the average arrival rate must increase by

$= \frac{13}{150} - \frac{1}{12} = \frac{1}{300}$ per min.

(v). Average waiting time per customer in the queue

$$w_q = \frac{1}{\lambda} L_q = \left(\frac{1}{\frac{1}{12}}\right)\left(\frac{25}{6}\right) = 50\ min.$$

(vi). The probability that the waiting time in the system greater than 30 min.

$$p[w > t] = e^{-(\mu-\lambda)t}$$

$$p[w > 30] = e^{-\left(\frac{1}{10}-\frac{1}{12}\right)t} = e^{-0.5} = 0.6065$$

(vii) P [a customer has to wait]= P[w>0]

= 1- P[w=0]

= 1-ρ_0 = 1-1/6 = 5/6

The percentage of customers who have to wait = 5/6 X 100 = 83.33

(viii) P (N>3) = $P_4 + P_5 + P_6 + \cdots.$

$= 1 - (P_0 + P_1 + P_2 + P_3)$

$= 1 - [1 - \left(\frac{\lambda}{\mu}\right)^4]$

= 0.4823.

3. The arrivals at the counter in a bank occur in accordance with a Poisson process at an average rate of 8 per hour .The duration of service of a customer has an exponential distribution with a mean of 6 minutes .Find the probability that an arriving customer.

(i) Has to wait

(ii) Finds 4 customers in the system

(iii) Has to spend less than 15 minutes in the bank.

Also estimate the fraction of the total time that the counter is busy. **(N/D 2005)**

SOLUTION:

$\lambda = 8/hour$

$\frac{1}{\mu} = 6\ minutes.\ \ => \ \mu = \frac{1}{6}/mins\ \ => \mu = 10/hour$

(i) Probability that customer has to wait = Probability that the system is busy = $\frac{\lambda}{\mu} = \frac{8}{10} = .8$

(ii) Probability that there are 4 customers in the system = $p_4 = \left(\frac{\lambda}{\mu}\right)^4 \left(1 - \frac{\lambda}{\mu}\right)$

$$= \left(\frac{8}{10}\right)^4 \left(1 - \frac{8}{10}\right) = 0.081922$$

(iii)Probability that a customer has to spend less than 15 minutes in the bank .

$= P[W_s < 15\ mins] = P\left[W_s < \frac{1}{4}/hour\right] = 1 - P\left[W_s > \frac{1}{4}\right] = 1 - e^{-(10-8)\frac{1}{4}}$

$= 1 - e^{-\frac{1}{2}} = 0.3935$

(iv) Fraction of the total time that the counter is busy $= \frac{\boldsymbol{\lambda}}{\boldsymbol{\mu}} = \frac{8}{10} = 0.8$

4. Customers arrive at a one-man barber shop according to a Poisson process with a mean inter arrival time of 20 minutes .Customers spend an average of 15 minutes in the barber's chair .If an hour is used as the unit of time ,then

   (i)What is the probability that a customer need not wait for a hair cut?

   (ii)What is the expected number of customers in the barber shop and in the queue?

   (iii)How much time can a customer expect to spend in the barbershop?

   (iv)Find the average time that a customer spends in the queue.

(v) What is the probability that there will be 6 or more customers waiting for service?**(N/D 2006)**

SOLUTION:

$\lambda = 3/hour$

$\frac{1}{\mu} = 15\ minutes. \quad => \quad \mu = \frac{1}{15}/mins \quad => \mu = 4/hour$

(i) $P[The\ system\ is\ empty] = 1 - \frac{\lambda}{\mu} = 1 - \frac{3}{4} = 0.25$

(ii) The expected number of customers in the barber shop $= L_s = \frac{\boldsymbol{\lambda}}{\boldsymbol{\mu} - \boldsymbol{\lambda}} = \frac{3}{\mathbf{4-3}} = 3$

(iii)The expected number of customers in the queue $= L_q = \frac{\boldsymbol{\lambda}^2}{\boldsymbol{\mu}(\boldsymbol{\mu} - \boldsymbol{\lambda})} = \frac{9}{\mathbf{4(4-3)}} = \frac{9}{4}$

(iv) The waiting time of a customer in the barbershop (or)$in\ the\ system = W_s = \frac{1}{\boldsymbol{\mu} - \boldsymbol{\lambda}} = \frac{1}{\mathbf{4-3}} = 1\ hour$

(v) The waiting time of a customer in the queue $= w_q = \frac{\boldsymbol{\lambda}}{\boldsymbol{\mu}(\boldsymbol{\mu} - \boldsymbol{\lambda})} = \frac{3}{\mathbf{4(4-3)}} = \frac{3}{4}\ hour =$ $45\ mins.$

(vi) probability that there will be 6 or more customers waiting for service

$$= P[N > k] = P[N > 6] = \left(\frac{\boldsymbol{\lambda}}{\boldsymbol{\mu}}\right)^{k+1} = \left(\frac{\boldsymbol{\lambda}}{\boldsymbol{\mu}}\right)^7 = \left(\frac{3}{\mathbf{4}}\right)^7 = 0.1334$$

5. Customers arrive at a watch repair shop according to a Poisson process at a rate of one per every 10 minutes and the service time is an exponential random variable with mean 8 minutes.

(i)Find the average number of customers Ls in the shop.

(ii)Find the average time a customer spends in the shop Ws.

(iii)Find the average number of customers in the queue Lq.

(iv) What is the probability that the server is idle? **(A/M 2005)**

SOLUTION:

$\lambda = 6/hour$

$\frac{1}{\mu} = 8\ minutes. \quad => \quad \mu = \frac{1}{8}/mins \quad => \mu = \frac{15}{2}/hour$

(i) The average number of customers in the shop = $L_s = \frac{\lambda}{\mu - \lambda} = \frac{6}{\frac{15}{2} - 6} = 4$

(ii) The average waiting time of a customer in the shop (or) $in\ the\ system = W_s$

$$= \frac{1}{\mu - \lambda} = \frac{1}{\frac{15}{2} - 3} = \frac{2}{3} hour = 40\ mins.$$

(iii)The average number of customers in the queue = $L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{36}{\frac{15}{2}\left(\frac{15}{2} - 6\right)} = \frac{16}{5}$

(iv) $P[The\ system\ is\ empty] = 1 - \frac{\lambda}{\mu} = 1 - \frac{6}{\frac{15}{2}} = \frac{1}{5}.$

6. A repairman is to be hired to repair machines which break down at an average rate of 3 per hour .The breakdowns follow Poisson distribution .Non-productive time of machine is considered to cost Rs.16/hr .Two repairmen have been interviewed . One is slow but cheap while the other is fast and expensive .The slow repairmen charges Rs.8hr and he services machines at the rate of 4 per hour .The fast repairman demands Rs.10/hr and services at the average rate of 6 per hour .Which repairmen should be hired?(assume an 8 hour working day) **(N/D 2003,M/J 2006)**

SOLUTION:

$\lambda = 3/hour$

Idle time cost of machine = Rs. $16/hour$

For slow repairman:

$\mu = 4/hour$

Charges per hour = Rs. 8.

The average waiting time of 1 machine $in\ the\ system = W_s = \frac{1}{\mu - \lambda} = \mathbf{1/hour}$

Machine hours lost in 8- hour day = (No.of machines breaking down per hour) **X** (No.of hours) **X** (W_s) = 3 **X** 8 **X** 1 = 24 hours.

Total cost per day = Cost of Idle machines + Repairman charges.

= Rs. $[(16 \times 24) + (8 \times 8)]$

= Rs.448.

For fast repairman:

$\mu = 6/hour$

Charges per hour = Rs. 10.

The average waiting time of 1 machine $in\ the\ system = W_s = \dfrac{1}{\boldsymbol{\mu - \lambda}} = \dfrac{\mathbf{1}}{\mathbf{3}}/\boldsymbol{hour}$

Machine hours lost in 8- hour day = (No.of machines breaking down per hour) **X** (No.of hours) **X** (W_s) = 3 **X** 8 **X** $\frac{\mathbf{1}}{\mathbf{3}}$ = 8 hours.

Total cost per day = Cost of Idle machines + Repairman charges.

= Rs. $[(16 \times 8) + (10 \times 8)]$

= Rs.208.

The Fast repairman should be hired since the total cost is less in his case.

EXAMPLE:7 A T.V. repairman finds that the time spent on his jobs has an exponential distribution with mean 30 minutes .If he repairs sets in the order in which they come in and if the arrival of sets is approximately Poisson ,with an average rate of 10 per 8 hour day ,what is the repairman's idle time each day? How many jobs are ahead of the average set brought in?

(N/D 2005, M/J 2006)

SOLUTION:

$\lambda = \dfrac{10}{8}/hour$

$\frac{1}{\mu} = 30\ minutes. \ => \ \mu = \frac{1}{30}/mins \ => \mu = 2/hour$

The repairman's idle time for 1 hour = $1 - \frac{\boldsymbol{\lambda}}{\boldsymbol{\mu}} = 1 - \frac{5}{8} = \frac{3}{8}$

$\therefore$ The repairman's idle time for each day = $8 \times \frac{3}{8} = 3\ hours.$

The average number of customers in the system = $L_s = \dfrac{\boldsymbol{\lambda}}{\boldsymbol{\mu - \lambda}} = \dfrac{\frac{\mathbf{5}}{\mathbf{4}}}{2 - \frac{\mathbf{5}}{\mathbf{4}}} = \dfrac{\mathbf{5}}{\mathbf{3}}$

Example:8

In a railway marshalling yard, goods trains arrive at a rate of 30 trains per day. Assuming that the inter-arrival time follows an exponential distribution and the service time (tie time taken to hump to train) distribution is also exponential with an avg. of 36 minutes. Calculate

(i) Expected queue size (line length)

(ii) Prob. that the queue size exceeds 10.

If the input of trains increase to an avg. of 33 per day, what will be the change in (i) & (ii)

$$\lambda = \frac{30}{60x24} = \frac{1}{48} \text{ trains per minute.}$$

$$\mu = \frac{1}{36} \text{ trains per minute.}$$

The traffic intensity $\rho = \lambda / \mu$

$= 0.75$

(i) Expected queue size (line length)

$$L_S = \frac{\lambda}{\lambda - \mu} \quad \text{or} \quad \frac{\rho}{1-\rho}$$

$$= \frac{0.75}{1-0.75} = 3 \text{ trains}$$

(ii) Prob. that the queue size exceeds 10

$$P[n \geq 10] = \rho^{10} = (0.75)^{10} = 0.06$$

Now, if the input increases to 33 trains per day, then we have

$$\lambda = \frac{30}{60x24} = \frac{1}{48} \text{ trains per minute.}$$

$$\mu = \frac{1}{36} \text{ trains per minute.}$$

The traffic intensity $\rho = \lambda / \mu = \frac{11}{480} x36$

$= 0.83$

Hence, recalculating the value for (i) &(ii)

(i) $L_S = \frac{\rho}{1-\rho} = 5$ trains (approx)

(ii) $P(n \geq 10) = \rho^{10} = (0.83)^{10} = 0.2$ (approx)

Hence recalculating the values for (i) & (ii)

(i) Ls = ρ/ 1-ρ= 5 trains (approx)

(ii) P(n ≥ 1010)=(0.83)10= 0.ρ2(approx)

Example:9

A super market has a single cashier. During the peak hours, customers arrive at a rate of 20 customers per hour. The average no of customers that can be processed by the cashier is 24 per hour. Find

(i) The probability that the cashier is idle.

(ii) The average no of customers in the queue system

(iii)The average time a customer spends in the system.

(iv)The average time a customer spends in queue.

(v) The any time a customer spends in the queue waiting for service

λ= 20 customers

μ= 24 customers / hour

(i) Prob. That the customer is idle = 1- λ/μ= 0.167

(ii) Average no of customers in the system.

Ls = λ/ μ-λ= 5

(iii)Average time a customer spends in the system

Ws = Ls / λ= ¼ hour = 15 minutes.

(iv) Average no of customers waiting in the queue Lq = λ2 / μ(μ-λ) = 4.167

(v) Average time a customer spends in the queue

Wq = λ/ μ(μ-λ) = 12.5 minutes

[***SINGLE SERVER WITH FINITE CAPACITY QUEUE***]

(i) Probability that there are n customers in the system = $p_n =$

$$\begin{cases} \left(\frac{\lambda}{\mu}\right)^n \frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{k+1}} & if \quad \lambda \neq \mu \\ \frac{1}{k+1} & if \quad \lambda = \mu \end{cases}$$

(ii) Probability that the system is empty $= p_0 = \begin{cases} \frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{k+1}} & if \quad \lambda \neq \mu \\ \frac{1}{k+1} & if \quad \lambda = \mu \end{cases}$

(iii) The effective arrival rate $\lambda' = \mu(1 - p_0)$

(iv) The expected number of customers in the system $= L_s$

$$= \begin{cases} \frac{\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}} - \frac{(k+1)\left(\frac{\lambda}{\mu}\right)^{k+1}}{1-\left(\frac{\lambda}{\mu}\right)^{k+1}} & if\ \lambda \neq \mu \\ \frac{k}{2} & if\ \lambda = \mu \end{cases}$$

(v) The expected number of customers in the queue $= L_q = L_s - \frac{\lambda'}{\mu}$

(vi) The average waiting time of a customer in the $system = W_s = \frac{L_s}{\lambda'}$

(vii) The average waiting time of a customer in the queue $= w_q = \frac{L_q}{\lambda'}$

EXAMPLE :10 Patients arrive at a clinic according to Poisson distribution at a rate of 30 patients per hour .The waiting room does not accommodate more than 14 patients .Examination time per patient is exponential with a mean rate of 20 per hour.

(a) Find the effective arrival rate at the clinic.

(b) What is the probability that an arriving patient does not have to wait?

(c) What is the expected waiting time until a patient is discharged from the clinic?

(May/June 2007)

SOLUTION: Given that

$\lambda = 30/hour$ and $\mu = 20/hour$

Capacity of the system =k= Waiting patient + 1 being served in the chair

⇨ K= 14+1 = 15 Here $\lambda \neq \mu$

(i) Effective arrival rate at the clinic $\lambda' = \mu(1 - p_0)$

Here $p_0 = p(no\ patient) = \begin{cases} \frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{k+1}} & if \quad \lambda \neq \mu \\ \frac{1}{k+1} & if \quad \lambda = \mu \end{cases}$

$$= \frac{1 - \frac{30}{20}}{1 - \left(\frac{30}{20}\right)^{16}} = 0.000762$$

$\therefore The\ e$ffective arrival rate at the clinic $\lambda' = \mu(1 - p_0) = 20(1 - 0.000762)$

$= 19.9848 \cong 20\ /\ hour$

(ii) The probability that an arriving patient does not have to wait =

$p(the\ system\ is\ empty)$

$$= p_0 = p(no\ patient) = \frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{k+1}} = \frac{1-\frac{30}{20}}{1-\left(\frac{30}{20}\right)^{16}} = 0.000762$$

(iii) The expected waiting time until a patient is discharged from the clinic $W_s = \frac{L_s}{\lambda'}$

$$\text{Where } L_s = \begin{cases} \frac{\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}} - \frac{(k+1)\left(\frac{\lambda}{\mu}\right)^{k+1}}{1-\left(\frac{\lambda}{\mu}\right)^{k+1}} & if\ \lambda \neq \mu \\ \frac{k}{2} & if\ \lambda = \mu \end{cases}$$

$$L_s = \frac{\frac{30}{20}}{1-\frac{30}{20}} - \frac{(16)\left(\frac{30}{20}\right)^{16}}{1-\left(\frac{30}{20}\right)^{16}} = -3 + 16.0244 = 13.024$$

$\therefore$ *From the Little's formula* $W_s = \frac{L_s}{\lambda'} = \frac{13.024}{19.9848} = 0.6517\ /hour$ = 39.102 mins

EXAMPLE:11 A one person barber shop has six chairs to accommodate people waiting for a haircut. Assume that customers who arrive when all the six chairs are full leave without entering the barber shop. Customers arrive at the average rate of 3 per hour and spend an average of 15 minutes in the barber's chair.

(i) What is the probability that a customer can get directly into the barber's chair upon arrival?
(ii) What is the expected number of customers waiting for a haircut?
(iii) How much time can a customer expect to spend time in the barber shop?
(iv) What fraction potential customers are turned away?

SOLUTION: Given that

$\lambda = 3/hour$ and $\mu = 4/hour$

Capacity of the system =k = 6+1 = 7 Here $\lambda \neq \mu$

Effective arrival rate at the clinic $\lambda' = \mu(1-p_0)$

$$\text{Here } p_0 = p(no\ patient) = \begin{cases} \frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{k+1}} & if \quad \lambda \neq \mu \\ \frac{1}{k+1} & if \quad \lambda = \mu \end{cases}$$

$$= \frac{1-\frac{3}{4}}{1-\left(\frac{3}{4}\right)^{8}} = 0.2778$$

$\therefore The$ effective arrival rate at the clinic $\lambda' = \mu(1 - p_0) = 4(1 - 0.2778) = 2.89 \,/\, hour$

(i) The probability that a customer can get directly into the barber's chair upon arrival

i.e., $p(the\ system\ is\ empty)$

$$= p_0 = p(no \text{ customer in the barber's chair }) = \frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{k+1}} = \frac{1 - \frac{3}{4}}{1 - \left(\frac{3}{4}\right)^8} = 0.2778$$

(ii) The expected number of customers waiting for a haircut $= L_q = L_s - \frac{\lambda'}{\mu}$

Where $L_s = \begin{cases} \frac{\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}} - \frac{(k+1)\left(\frac{\lambda}{\mu}\right)^{k+1}}{1-\left(\frac{\lambda}{\mu}\right)^{k+1}} & if\ \lambda \neq \mu \\ \frac{k}{2} & if\ \lambda = \mu \end{cases}$

$$L_s = \frac{\frac{3}{4}}{1 - \frac{3}{4}} - \frac{(8)\left(\frac{3}{4}\right)^8}{1 - \left(\frac{3}{4}\right)^8} = 3 - \frac{\frac{6561}{8192}}{0.8999} = 2.11$$

$\therefore From\ the\ Little's formula\ L_q = L_s - \frac{\lambda'}{\mu} = 2.11 - \frac{2.89}{4} = 1.3875$

(iii) The expected waiting time of a customer in the barber shop $= W_s = \frac{L_s}{\lambda'}$

$$= \frac{2.11}{2.89} = 0.7301/hour = 43.8062/mins$$

(iv) A fraction of potential customers are turned away

i.e., $p(the\ system\ is\ full) = p_n = \begin{cases} \left(\frac{\lambda}{\mu}\right)^n \frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{k+1}} & if \quad \lambda \neq \mu \\ \frac{1}{k+1} & if \quad \lambda = \mu \end{cases}$

$$p_7 = \left(\frac{\lambda}{\mu}\right)^n \frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{k+1}} = \left(\frac{\lambda}{\mu}\right)^n p_0 = \left(\frac{3}{4}\right)^7 (0.2778) = 0.0371$$

$= 3.71\%\ of$ potential customers are turned away

UNIT-5

MARKOV PROCESS - MARKOV CHAINS

MARKOV PROCESS

Definition

A random process {X(t)} is said to be markovian if

$$P\left[X(t_{n+1}) \le X_{n+1} / X(n) + x_n, x(t_{n-1}) = x_{n-1} \ldots x(t_0 = x_0)\right]$$
$$P\left[X(t_{n+1}) \le X_{n+1} / x(t_n) = x_n\right]$$

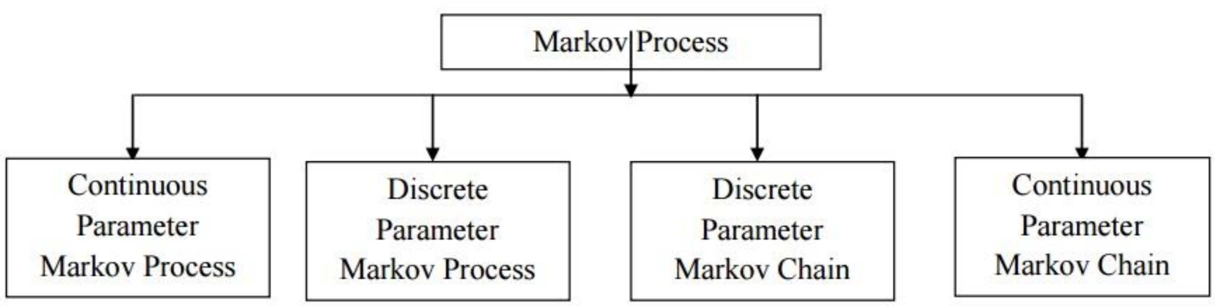
Where $t_0 \le t_1 \le t_2 \le \ldots \le t_n \le t_{n+1}$

Examples of Markov Process

1.The probability of raining today depends only on previous weather conditions existed for the last two days and not on past weather conditions.

2.A different equation is markovian.

Classification of Markov Process



MARKOV CHAIN

Definition

We define the Markov Chain as follows If P{Xn = an/Xn-1 = an-1, Xn-2 = an-2, … X0 = a0}

⇒P{Xn = an / Xn-1 = an-1} for all n. the process {Xn}, n = 0, 1, 2… is called as Markov Chains.

1.a1, a2, a3, … an are called the states of the Markov Chain.

2.The conditional probability P{X n = aj | X n −1 = ai} = Pij (n −1, n) is called the one step

transition probability from state a i to state a j at the nth step. 3.The tmp of a Markov chain is a stochastic matricx

i) Pij≥ 0

ii) ΣPij = 1   [Sum of elements of any row is 1]

EXAMPLE:1

If the tmp of a Markov Chain is $\begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, find the steady state distribution of the chain.

Given $P = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

If $\pi = (\pi_1 \pi_2)$ is the steady state distribution of the chain, then by the property of π, we have

$$\pi P = \pi \qquad ---------(1)$$

$$\pi_1 + \pi_2 = 1 \qquad ---------(2)$$

$$\Rightarrow (\pi_1\ \pi_2)\begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = (\pi_1 \pi_2)$$

$$\left[\pi_1[0] + \pi_2(\tfrac{1}{2}) + \pi_1(1) + \pi_2(\tfrac{1}{2})\right] = [\pi_1\ \pi_2]$$

$$\tfrac{1}{2}\pi_2 = \pi_1 \qquad ---------- (3)$$

$$\pi_1 + \tfrac{1}{2}\pi_2 = \pi_2 \qquad --------(4)$$

$$\pi_1 + \pi_2 = 1$$

$$\tfrac{1}{2}\pi_2 + \pi_2 = 1 \Rightarrow \tfrac{3}{2}\pi_2 = 1 \quad \text{by (3)}$$

$$\pi_2 = \tfrac{2}{3}$$

$$(3) \Rightarrow \pi_1 = \tfrac{1}{2}\pi_2 = \tfrac{1}{2}(\tfrac{2}{3}) = \tfrac{1}{3}$$

$\therefore$ The steady state distribution of the chain is $\pi = (\pi_1\ \pi_2)$

i.e. $\pi = (\tfrac{1}{3}\ \tfrac{2}{3})$

Example :2

An Engineering analysing a series of digital signals generated by a testing system observes that only 1 out of 15 highly distorted signals followed a highly distorted signal with no recognizable signal, where as 20 out of 23 recognizable signals follow recognizable signals with no highly distorted signals b/w. Given that only highly distorted signals are not recognizable. Find the fraction of signals that are highly distorted.

π_1= The fraction of signals that are recognizable [R]

π_2 = The fraction of signals that are highly distorted [D]

The tmp of the Markov Chain is

$$P = \begin{array}{c} \\ R \\ D \end{array}\begin{array}{c} \begin{array}{cc} R & D \end{array} \\ \begin{bmatrix} 20/23 & - \\ - & 1/15 \end{bmatrix} \end{array} \quad \Rightarrow P = \begin{bmatrix} 20/23 & 3/23 \\ 14/15 & 1/15 \end{bmatrix}$$

- 1 out of 15 highly distorted signals followed a highly distorted signal with no recognizable signal

$$\left[P(D \to D)\right] = 1/15$$

- 20 out of 23 recognizable signals follow recognizable signals with no highly distorted signals.
- If the tmp of a chain is a stochastic martin, then the sum of all elements of any row is equal to 1.

If $\pi = (\pi_1 \pi_2)$ is the steady state distribution of the chain, then by the property of π, we have

$$\pi P = \pi$$

$$\pi_1 + \pi_2 = 1$$

$$\Rightarrow (\pi_1\,\pi_2)\begin{bmatrix} 20/23 & 3/23 \\ 14/15 & 1/15 \end{bmatrix} = (\pi_1 \pi_2)$$

$$20/23\,\pi_1 + 14/15\,\pi_2 = \pi_1 \qquad -----(3)$$

$$3/23\,\pi_1 + 1/15\,\pi_2 = \pi_2 \qquad -----(4)$$

(3) $\pi_2 = 45/322\,\pi_1$

$\Rightarrow \pi_1 = 322/367$

(2)

$\pi_2 = 45/367$

$\therefore$ The steady state distribution of the chain is

$$\pi = (\pi_1 \pi_2)$$

i.e. $\pi = \left(\frac{322}{367} \quad \frac{45}{367}\right)$

$\therefore$ The fraction of signals that are highly distorted is $45/367$

$$\Rightarrow \frac{45}{367} x100\% = 12.26\%$$

Example :3.4.2

Transition prob and limiting distribution. A housewife buys the same cereal in successive weeks. If the buys cereal A, the next week she buys cereal B. However if she buys B or C, the next week she is 3 times as likely to buy A as the other cereal. In the long run how often she buys each of the 3 cereals.

Given : Let $\pi_1 \rightarrow$ Cereal A

$\pi_2 \rightarrow$ Cereal B

$\pi_3 \rightarrow$ Cereal C

$\therefore$ the tpm of the Markov Chain is

$$P = \begin{bmatrix} 0 & 1 & - \\ 3/4 & 0 & - \\ 3/4 & - & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 3/4 & 0 & 1/4 \\ 3/4 & 1/4 & 0 \end{bmatrix}$$

If $\pi = (\pi_1, \pi_2, \pi_3)$ is the steady - state distribution of the chain then by the property of π we have,

$$\pi P = \pi$$

$$\pi_1 + \pi_2 + \pi_3 = 1$$

$$\Rightarrow (\pi_1\ \pi_2\ \pi_3) \begin{bmatrix} 0 & 1 & 0 \\ 3/4 & 0 & 1/4 \\ 3/4 & 1/4 & 0 \end{bmatrix} = [\pi_1\ \pi_2\ \pi_3]$$

$$3/4\,\pi_2 + 3/4\,\pi_3 = \pi_1 \qquad (3)$$

$$\pi_1 + 1/4\,\pi_3 = \pi_2 \qquad (4)$$

$$1/4\,\pi_2 = \pi_3 \qquad (5)$$

$$(3) \Rightarrow 3/4\,\pi_2 + 3/4\left(1/4\,\pi_2\right) = \pi_1 \qquad \text{(by 5)}$$

$$15/16\,\pi_2 = \pi_2 \qquad (6)$$

$$(2) \Rightarrow \pi_1 + \pi_2 + \pi_3 = 1$$

$$15/16\,\pi_2 + \pi_2 + \frac{1}{4}\pi_2 = 1 \qquad \text{by (5) \& (6)}$$

$$\frac{35}{16}\pi_2 = 1$$

$$\pi_2 = \frac{16}{35}$$

$$(6) \Rightarrow \pi_1 = \frac{15}{16}\pi_2$$

$$\pi_1 = \frac{15}{35}$$

$$(5) \Rightarrow \pi_3 = \frac{1}{4}\pi_2$$

$$\pi_3 = \frac{4}{35}$$

$\therefore$ The steady state distribution of the chain is

$$\pi = (\pi_1\ \pi_2\ \pi_3)$$

$$\text{i.e., } \pi = \left(15/35\ \ 16/35\ \ 4/35\right)$$

n - step tmp P^n

Example : 3.4.1 The transition Prob. Martix of the Markov Chain $\{X_n\}$, n = 1,2,3,... having 3 states 1, 2 & 3 is $P = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$ and the initial distribution is $P^{(0)} = (0.7\ \ 0.2\ \ 0.1)$.

Find (i) $P(X_2 = 3)$ and (ii) $P[X_3 = 2, X_2 = 3, X_1 = 3, X_0 = 2]$

Solution

Given $P^{(0)} = (0.7\ \ 0.2\ \ 0.1)$.

$$\Rightarrow \quad P[X_0 = 1] = 0.7$$

$$P(X_0 = 2) = 0.2$$

$$P[X_0 = 3] = 0.1$$

$$P = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$$

$$= \begin{bmatrix} P_{11}^{(1)} & P_{12}^{(1)} & P_{13}^{(1)} \\ P_{21}^{(1)} & P_{22}^{(1)} & P_{23}^{(1)} \\ P_{31}^{(1)} & P_{32}^{(1)} & P_{33}^{(1)} \end{bmatrix}$$

$$P^2 = P\,.\,P$$

$$= \begin{bmatrix} 0.43 & 0.31 & 0.26 \\ 0.24 & 0.42 & 0.34 \\ 0.36 & 0.35 & 0.29 \end{bmatrix}$$

$$= \begin{bmatrix} P_{11}^{(2)} & P_{12}^{(2)} & P_{13}^{(2)} \\ P_{21}^{(2)} & P_{22}^{(2)} & P_{23}^{(2)} \\ P_{31}^{(2)} & P_{32}^{(2)} & P_{33}^{(2)} \end{bmatrix}$$

$$
\begin{aligned}
\text{(i) } P[X_2 = 3] &= \sum_{i=1}^{3} P[X_2 = 3 / X_0 = i] P[\lambda_0 = i] \\
&= P[X_2 = 3 / X_0 = 1]\ P[X_0 = 1] + P[X_2 = 3 / X_0 = 3] P[X_0 = 2] + \\
&\qquad P[X_2 = 3 / X_{0=3}] P[X_0 = 3] \\
&= P_{13}^{(2)} P[X_0 = 1] + P_{23}^{(2)} P[X_0 = 2] + P_{33}^{(2)} P[X_0 = 3] \\
&= (0.26)(0.7) + (0.34)(0.2) + (0.29)(0.1) \\
&= 0.279
\end{aligned}
$$

$$
\begin{aligned}
\text{(ii) } & P[X_3 = 2,\ X_2 = 3,\ X_1 = 3,\ X_0 = 2] \\
&= P_{32}^{(1)} P[X_2 = 3 / X_1 = 3, X_0 = 2]\ P[\lambda_1 = 3, X_0 = 2] \\
&= P_{32}^{(1)}\ P_{33}^{(1)}\ P_{23}^{(1)}\ P[X_0 = 2] \\
&= (0.4)\ (0.3)\ (0.2)\ (0.2) \\
&= 0.0048
\end{aligned}
$$

EXAMPLE :5 A training process is considered as two State Markov Chain. If it rain, it is considered to be state 0 & if it does not rain the chain is in stable 1. The tmp of the Markov Chain is defined as

$$
P = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}
$$

i. Find the Prob. That it will rain for 3 days from today assuming that it is raining today.

ii. Find also the unconditional prob. That it will rain after 3 days with the initial Prob. Of state) and state 1 as 0.4 & 0.6 respectively.

Solution:

Given $P=\begin{bmatrix}0.6 & 0.4\\ 0.2 & 0.8\end{bmatrix}$

$$P^{(2)} = P^2 = \begin{bmatrix}0.6 & 0.4\\ 0.2 & 0.8\end{bmatrix}\begin{bmatrix}0.6 & 0.4\\ 0.2 & 0.8\end{bmatrix}$$

$$= \begin{bmatrix}0.44 & 0.56\\ 0.28 & 0.72\end{bmatrix}$$

$$P^{(3)} = P^3 = P^2P$$

$$= \begin{bmatrix}0.376 & 0.624\\ 0.312 & 0.688\end{bmatrix}$$

(i) If it rains today, then Prob. Distribution for today is (1 0)

$$\therefore P(\text{after 2 days}) = \begin{pmatrix}1 & 0\end{pmatrix}\begin{bmatrix}0.376 & 0.624\\ 0.312 & 0.688\end{bmatrix}$$

$$= [0.376 \quad 0.624]$$

$\therefore$ P [Rain for after 3 days] $= 0.376$

(ii) Given $P^{(0)} = \begin{pmatrix}0.4 & 0.6\end{pmatrix}$

$$P[\text{after 3 days}] = \begin{pmatrix}0.4 & 0.6\end{pmatrix}\begin{bmatrix}0.376 & 0.624\\ 0.312 & 0.688\end{bmatrix}$$

$$= \begin{pmatrix}0.3376 & 0.6624\end{pmatrix}$$

$\therefore$ P[rain for after 3 days] $= 0.3376$

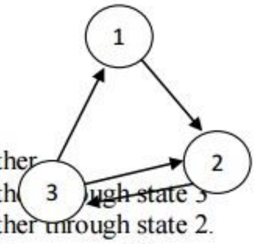
Example :3.5.1

Prove that the matrix $\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ \frac{1}{2} & \frac{1}{2} & 0\end{bmatrix}$ is the tpm of an irreducible Markov Chain?

(or)

Three boys A, B, C are throwing a ball each other. A always throws the ball to B & B always throws the ball to C but C is just as like to throw the ball to B as to A. State that the process is Markov Chain. Find the tpm and classify the status.

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$



(a) let $X_n = \{1,2,3\} \Rightarrow$ finite

State 2 & 3 are communicate with each other

State 1 & 2 are communicate with each other through state 3

State 3 & 1 are communicate with each other through state 2.

$\Rightarrow$ The Markov Chain is irreducible (3)

From (1) & (2) all the states are persistent and non-null (3)

One can get back to

State 1 in	$1 \to 2 \to 3 \to 1$	(3 steps)
State 2 in	$2 \to 3 \to 2$	(2 steps)
State 2 in	$2 \to 3 \to 1 \to 2$	(3 steps)
State 3 in	$3 \to 1 \to 2 \to 3$	(3 steps)
State 3 in	$3 \to 2 \to 2$	(2 steps)

$\Rightarrow$ The states are aperiodic (4)

[$\because$ The states are not periodc]

From (3) & (4) we get all the states are Ergodic.

TUTORIAL QUESTIONS

1.. The t.p.m of a Marko cain with three states 0,1,2 is P and the initial state distribution is Find (i)P[X2=3] ii)P[X3=2, X2=3, X1=3, X0=2]

2. Three boys A, B, C are throwing a ball each other. A always throws the ball to B and B always throws the ball to C, but C is just as likely to throw the ball to B as to A. S.T. the process is Markovian. Find the transition matrix and classify the states

3. A housewife buys 3 kinds of cereals A, B, C. She never buys the same cereal in successive weeks. If she buys cereal A, the next week she buys cereal B. However if she buys P or C the next week she is 3 times as likely to buy A as the other cereal. How often she buys each of the cereals?

4. A man either drives a car or catches a train to go to office each day. He never goes 2 days in a row by train but if he drives one day, then the next day he is just as likely to drive again as he is to travel by train. Now suppose that on the first day of week, the man tossed a fair die and drove to work if a 6 appeared. Find 1) the probability that he takes a train on the 3rd day. 2). The probability that he drives to work in the long run.