1

Random Variables

1.1 Introduction

In the last chapter, sample space is treated as the set of all possible outcomes of a random experiment. But the outcomes of the random experiment may be numerical or non-numerical in nature. For example the result of coin tossing experiment is non-numerical in nature while the result of dice throwing experiment is a number.

For mathematical convenience we associate one or more numbers with each possible outcome of an experiment. These numbers may correspond to the total number of defective items in a batch or to the number of heads obtained on tossing two coins or to the time of failure of a component. Hence comes the notion of random variables. The random variables provide a more compact description of an experiment and allows us to discard unimportant details in the outcome of an experiment; for instance we may be intrested only in the number of heads obtained on tossing three coins and not in the actual sequence in which head and tail are obtained.

1.2 Random Variable

A random variable X is a function $X : S \to R$ that assigns a real number X(s) to each $s \in S$ (sample space), corresponding to a random experiment E. Hence domain of the random variable is S, range is $(-\infty, \infty)$ and image may be any subset of real numbers (R). Random variables are generally denoted by capital letters X, Y, Z etc. The values taken by the random variables X, Y, Z are generally denoted by lower case letters x, y, z.

For example :-

(i) A drug is given to two sick patients. Let random variable X represent the number of cures that occur. Hence $X = 0,1,2$.

(ii) A single fair dice is rolled and the random variable X represents the number that turns up. Hence X can take values 1, 2, 3, 4, 5, 6.

(iii) Let a coin is tossed thrice and the random variable X denotes the number of tails that turn up. Hence X = 0, 1, 2, 3.

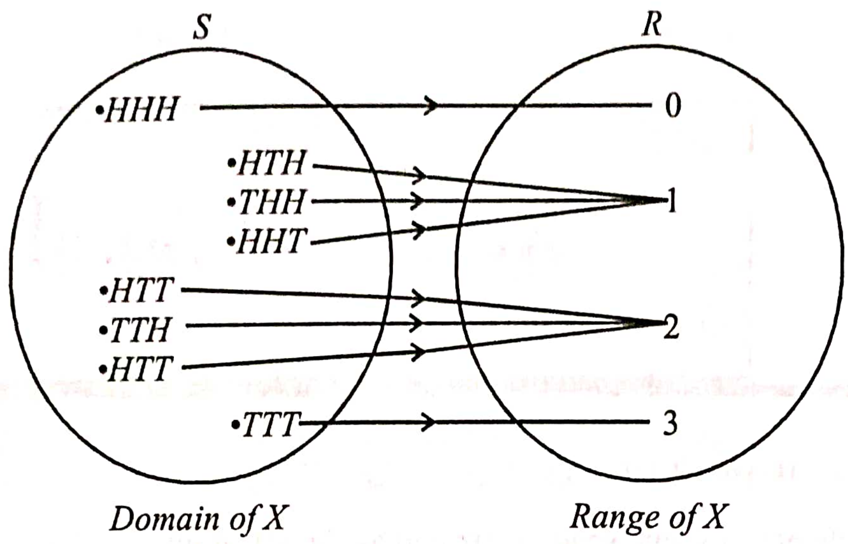

Figure 1.1

(iv) Consider a battery with maximum life of 200 hours. Let the random variable X denote the life of a randomly selected battery of such kind then $X = \{x \varepsilon R, 0 \leq x \leq 200\}$.

(v) Suppose we are interested in measuring the weight of students of Computer Science branch in a particular college. Then the random variable $X = \{x \varepsilon R, 0 < x < \infty\}$.

We now observe that a random variable is a variable whose value is determined by the outcome of a random experiment, and can be discrete or continuous in nature.

1.2.1 Discrete Random Variable

A discrete random variable has either finite or countably infinite number of values. Some examples of discrete random variables are :-

(i) The number shown when die is thrown.

(ii) The number of alpha particles emitted by a radioactive source.

(iii) The number of complaints received at the office of an airline on a given day.

(iv) The number of white balls drawn from a bag containing 6 white and 2 black balls.

(v) $X = \{x = a^2$ where $a = \ldots -3, -2, -1, 0, 1, 2, 3 \ldots\}$.

In other words we can say that a random variable defined on a discrete sample space will be discrete. But we can also define a discrete random variable on a continuous sample space. For example, let S = {s : height of individuals in a large group} and let random variable X denote height in inches rounded to nearest whole number then we can have X (s) = 59 inches.

A random variable partitions its sample space into a mutually exclusive and collectively exhaustive set of events. For example, let us consider the tossing of three coins and denote H as the event of getting head and T as the event of getting tail. Suppose the random variable X denotes the number of heads, then the sample space for this experiment is

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

Hence $X\{TTT\} = 0 \quad (\because \text{no heads})$

$$X(HTT) = X(THT) = X(TTH) = 1 \quad (\because \text{ 1 head is obtained})$$

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

Here we see that two sample points give the same value of X i.e., X may also be a many one function. Let us define a set $A_i = \{s\varepsilon S \mid X(s) = i\}$ i.e., the set consisting of all those events which have the same value of random variable X.

Here, in this experiment

$$A_0 = \{TTT\} \qquad A_1 = \{HTT,\ THT,\ TTH\}$$

and $A_2 = \{HHT,\ THH,\ HTH\}. \qquad A_3 = \{HHH\}$

It is clear that,

$$A_0 \cup A_1 \cup A_2 \cup A_3 = S$$

and $A_0 \cap A_1 = \phi,\ A_1 \cap A_2 = \phi,\ A_0 \cap A_2 = \phi.\ A_0 \cap A_3 = \phi.\ A_1 \cap A_3 = \phi.\ A_2 \cap A_3 = \phi.$

Hence the collection of events $A_i\ \forall\ i = 0, 1, 2$ partitions the sample space and thus defines an **event space.**

1.2.2 Continuous Random Variable

A random variable which can take infinite number of values in an interval is known as continuous random variable. Some examples of continuous random variables are :-

(i) The length of time during which a vaccum tube installed in a circuit functions.

(ii) The height of a person.

(iii) The weight of a fish.

(iv) The price of a house.

(v) $X = \{x\varepsilon R : 0 < x < 1\}.$

Most random variables we consider will either be discrete or continuous, but mixed random variables also occur. For example, there may be a non zero probability p_0, of initial failure of a component at time $t = 0$ due to manufacturing defects. In this case, the random variable X denoting the time to failure of a component is discrete for $t = 0$ and continuous for $0 < t < \infty$.

Such random variable is neither a discrete nor a continuous random variable and hence is known as mixed random variable.

1.3 Probability Distribution of A Discrete Random Variable

The probability distribution of a discrete random variable lists all the possible values that the random variable can assume and their corresponding probabilities. If a random variable assumes values $x_0, x_1, x_2, \ldots$ with probabilities $p_0, p_1, p_2, \ldots$ then the probability distribution is :-

$X = x$	x_0	x_1	x_2	$\cdots$
$P(x)$	p_0	p_1	p_2	$\cdots$

For example :-

(i) When rolling a single fair dice, the probability distribution is :-

$X = x$	1	2	3	4	5	6
$P(x)$	1/6	1/6	1/6	1/6	1/6	1/6

(ii) When tossing a coin and denoting random variable X as the number of heads obtained, the probability distribution is :-

$X = x$	0	1
$P(x)$	1/2	1/2

1.3.1 Probability Mass Function

Let X be a discrete random variable such that $P(X = x_i) = p_i$, then p_i is said to be probability mass function (pmf) if it satisfies the following conditions :

(i) $p_i \geq 0 \ \forall i$

(ii) $\sum\limits_i p_i = 1.$

Obviously the collection of pairs (x_i, p_i) is the probability distribution of the random variable X.

For example consider the probability distribution of the number of vehicles owned by families as :-

Number of vehicles owned (x)	0	1	2	3	4
Probability P(x)	.015	.235	.425	.245	.080

Then P(x) is a probability mass function.

The probability distribution of a discrete random variable can be presented in the form of a mathematical formula, a table, or a graph. Graphical representation of the probability distribution of above example in the form of a line or bar graph is :-

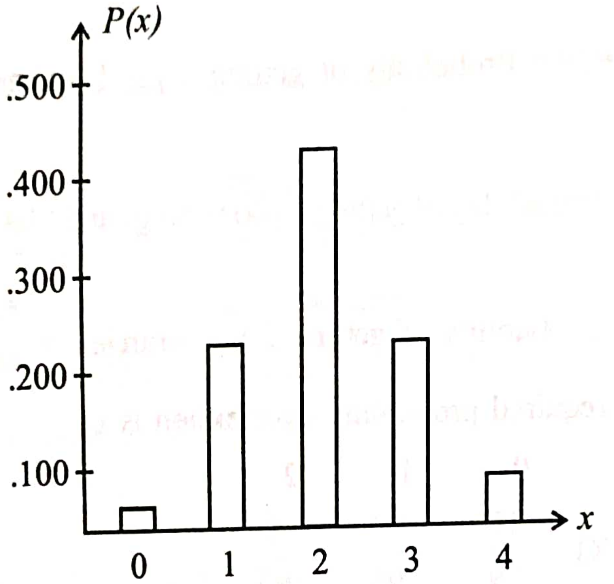

Figure 1.2

***Example* 1.** Check whether the following functions serve as probability mass function

(i) $P(X=x)=\frac{x-2}{2}\ \forall\ x=1,2,3,4$

(ii) $P(X=x)=\frac{x^2}{25}\ \forall\ x=1,2,3,4.$

Solution : (i)

x :	1	2	3	4
P(X=x) :	$-\frac{1}{2}$	0	$\frac{1}{2}$	1

As $P(X=1)=-\frac{1}{2}<0$, hence P(x) is not a probability mass function.

(ii)

x :	1	2	3	4
P(X=x) :	$\frac{1}{25}$	$\frac{4}{25}$	$\frac{9}{25}$	$\frac{16}{25}$

Though $P(X=x)>0 \;\; \forall \;\; x=1,2,3,4$ yet $\sum_{x=1}^{4} P(X=x)=\frac{30}{25}=\frac{6}{5}>1$

Hence it also does not serve as a probability mass function.

Example 2. Four bad oranges are mixed accidently with 16 good oranges. Find the probability distribution of the number of bad oranges in a draw of two oranges.

Solution : Let the random variable X denote the number of bad oranges in a draw of two oranges. Hence $X=0,\ 1,\ 2$.

Now $P(X=0)=$ Probability of getting 2 good oranges $=\frac{{}^{16}C_2}{{}^{20}C_2}=\frac{12}{19}$

$P(X=1)=$ Probability of getting 1 good orange and 1 bad orange $=\frac{{}^{4}C_1 \times {}^{16}C_1}{{}^{20}C_2}=\frac{32}{95}$

$P(X=2)=$ Probability of getting 2 bad oranges $=\frac{{}^{4}C_2}{{}^{20}C_2}=\frac{3}{95}$

Hence the required probability distribution is :

X :	0	1	2
P(X) :	$\frac{12}{19}$	$\frac{32}{95}$	$\frac{3}{95}$

1.3.2 Distribution Function

Let X be a discrete random variable, then its discrete distribution function or cumulative distribution function (cdf) is defined as

$$F(x)=\sum_{\substack{i \\ X_i \le x}} p_i$$

$$\text{i.e., } F(x_j)=P(X \le x_j)=\sum_{i \le j} p_i \,.$$

The graph of distribution function is a step function. For example if x_i is the integer i then F(X) has a jump p_i at each i and is constant between each pair of integers.

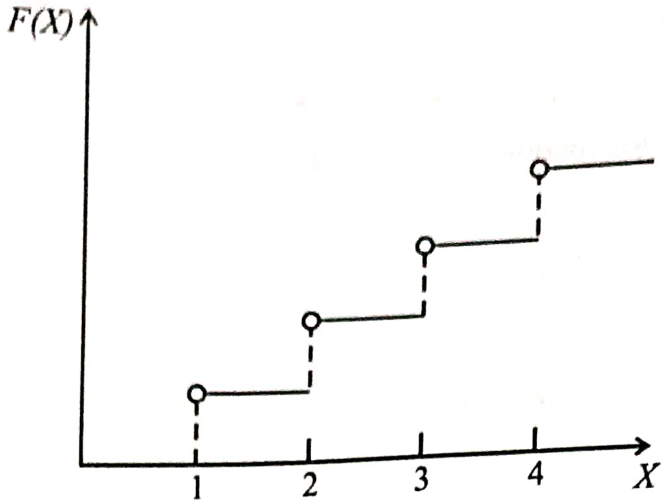

Figure 1.3

The distribution function F(X) satisfies the following properties :-

(i) $0 \le F(x) \le 1$, as F(x) is also a probability.

(ii) F(x) is a monotonic non decreasing function of x i.e., $x_1 \le x_2 \Rightarrow (-\infty, x_1] \le (-\infty, x_2]$.

Hence $P(-\infty < X \le x_1) \le P(-\infty < X \le x_2)$

i.e., $F(x_1) \le F(x_2)$.

(iii) $\lim_{x \to -\infty} F(x) = 0$, i.e., F(x) = 0 $\forall$ x sufficiently small

$\lim_{x \to \infty} F(x) = 1$, i.e., F(x) = 1 $\forall$ x sufficiently large.

(iv) F(x) has a positive jump at i = 1, 2, ... equal to $p(x_i)$ and F(x) is constant $\forall\ x_i < x \le x_{i+1}$.

$\therefore$ $F(x) = F(x_i)\ \forall\ x_i \le x < x_{i+1}$.

and $F(x_{i+1}) = F(x_i) + p(x_{i+1})$

Hereafter cdf and pmf will substitute the elementary things-sample space, event space, probability etc. as an easy alternative and important entity.

Example 3. Find the probability distribution of boys and girls in families with 3 children assuming equal probabilities of boys and girls. Also give the distribution function.

Solution : Let the random variable X denote the number of boys in the family.

Hence X = 0, 1, 2, 3. Also P(boy) = P(girl) = 1/2

and $P(X = x) = {}^3C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{3-x} = {}^3C_x \left(\frac{1}{2}\right)^3$

Hence the required probability distribution is :-

X	: 0	1	2	3
P(X)	: 1/8	3/8	3/8	1/8 .

Also the distribution function is given by

$$F(x) = \begin{cases} \dfrac{1}{8} & \text{when } x = 0 \\[2ex] \dfrac{4}{8} & \text{when } x = 1 \\[2ex] \dfrac{7}{8} & \text{when } x = 2 \\[2ex] 1 & \text{when } x = 3 \end{cases}$$

Graphically :

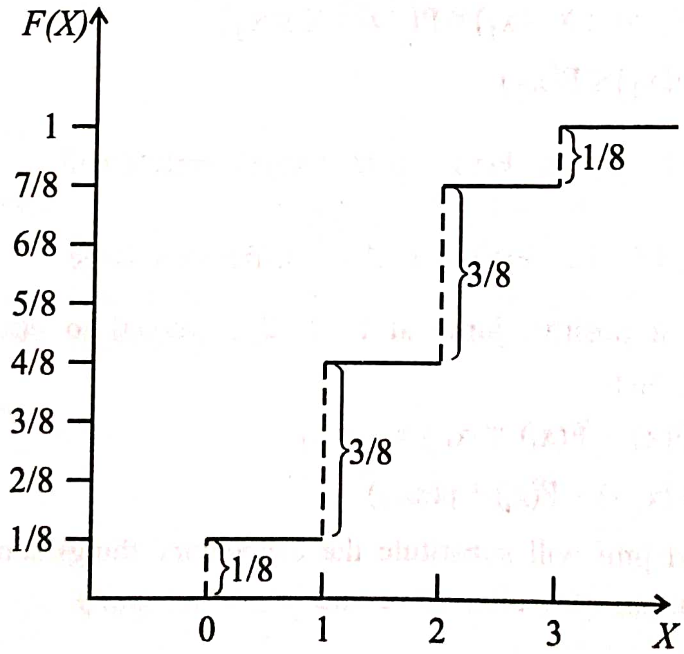

Figure 1.4

Example 4. From a lot of 10 items containing 3 defectives, a sample of 4 items is drawn at random. If the sample is drawn without replacement and the random variable X denotes the number of defective items in the sample, find :

(i) the probability distribution of X.

(ii) $P(X \leq 1)$

(iii) $P(X < 1)$

(iv) $P(0 < X < 2)$

Solution : (i) It is clear that $X = 0, 1, 2, 3$ such that

$$P(X=0)=\frac{{}^{7}C_{4}}{{}^{10}C_{4}}=\frac{1}{6}; \qquad P(X=1)=\frac{{}^{3}C_{1}\times{}^{7}C_{3}}{{}^{10}C_{4}}=\frac{1}{2}$$

$$P(X=2)=\frac{{}^{3}C_{2}\times{}^{7}C_{2}}{{}^{10}C_{4}}=\frac{3}{10}; \qquad P(X=3)=\frac{{}^{3}C_{3}\times{}^{7}C_{1}}{{}^{10}C_{4}}=\frac{1}{30}$$

Hence the required probability distribution is :-

X :	0	1	2	3
P(X) :	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

(ii) $P(X \le 1) = P(X=0) + P(X=1) = \frac{1}{6}+\frac{1}{2}=\frac{2}{3}$

(iii) $P(X<1)=P(X=0)=\frac{1}{6}$

(iv) $P(0<X<2)=P(X=1)=\frac{1}{2}$.

Example 5. A random variable X has the following probability distribution.

x :	0	1	2	3	4	5	6	7
p(x) :	0	k	2k	2k	3k	k^2	$2k^2$	$7k^2+k$

(i) Find k.

(ii) Evaluate $P(X < 6)$, $P(X \ge 6)$, $P(0 < X < 5)$.

(iii) Determine distribution function of X.

(iv) If $P(X \le c) > \frac{1}{2}$ find the minimum value of c.

(v) Find $P\left(\frac{1.5 < X < 4.5}{X > 2}\right)$ **(Raj. IV Sem IT-2003, RTU 2007, 2009)**

Solution : (i) As above given is a probability distribution

Hence $\sum_{x=0}^{7} p(x) = 1 \Rightarrow 10k^2 + 9k - 1 = 0$

$\Rightarrow \quad (k+1)(10k-1) = 0 \Rightarrow k = -1, \frac{1}{10}$

$k = -1$ is not possible as it makes $p(x) < 0$ which is impossible, as above given is a probability distribution .

Hence $k = \frac{1}{10}$. $\left[\because \sum p(x) = 1\right]$

(ii) $P(X < 6) = 1 - P(X \geq 6)$

$$= 1 - [P(X = 6) + P(X = 7)]$$

$$= 1 - (9k^2 + k) = 1 - \frac{1}{10} - \frac{9}{100} = \frac{81}{100} .$$

$$P(X \geq 6) = 1 - P(X < 6) = 1 - \frac{81}{100} = \frac{19}{100} .$$

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

$$= 8k = \frac{8}{10} = \frac{4}{5} .$$

(iii) The required distribution function is :-

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

(iv) From the distribution function it is clear that

$$F(3) = P(X \leq 3) = \frac{5}{10} = 0.5$$

$$F(4) = P(X \leq 4) = \frac{8}{10} = 0.8 > \frac{1}{2}$$

$$F(5) = P(X \leq 5) = \frac{81}{100} = 0.81 > \frac{1}{2},$$ and so on.

Hence the minimum value of c for which $P(x \leq c) > \frac{1}{2}$ is 4.

Therefore $c = 4$.

(v) $P\left(\frac{1.5<X<4.5}{X>2}\right)=\frac{P[(1.5<X<4.5)\cap(X>2]}{P(X>2)}$

$$=\frac{P(2<X<4.5)}{1-P(X\le 2)}=\frac{P(3)+P(4)}{1-[P(X=0)+P(X=1)+P(X=2)]}$$

$$=\frac{\frac{2}{10}+\frac{3}{10}}{1-\frac{3}{10}}=\frac{\frac{5}{10}}{\frac{7}{10}}=\frac{5}{7}.$$

1.4 Probability Distribution of a Continuous Random Variable

A continuous random variable can assume any value over an interval or intervals. Because the number of values contained in any interval is infinite, the possible number of values that a continuous random variable can assume is also infinite as well as uncountable. Hence a continuous random variable gives rise to continuous probability distribution.

As the number of events is infinitely large, the probability that a particular event will occur is practically zero. Here instead of finding the probability at a particular value of x we find the probability of x falling in a small interval. We thus define the continuous probability distribution of x by the function f(x) such that the probability of x lying in the small interval $\left(x-\frac{dx}{2},x+\frac{dx}{2}\right)$ is f(x) dx, i.e., $P\left(x-\frac{dx}{2}<x<x+\frac{dx}{2}\right)=f(x)dx$.

The continuous curve y = f(x) is known as probability curve.

1.4.1 Probability Density Function

The function f(x) for a continuous random variable X is said to be probability density function (p.d.f.) provided it satisfies the following conditions :-

(i) $f(x)\ge 0;\ -\infty<x<\infty$

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

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

Remark 1 When X is a continuous random variable $P(a\le X\le b)=P(a\le X<b)=P(a<X\le b)$ $=P(a<X<b)$, because $P(X=a)=P(a\le X\le a)=\int_a^a f(x)dx=0$.

Example 6. The diameter of an electric cable, say X, is assumed to be a continuous random variable with pdf = $f(x) = 6x(1-x),\ 0 \le x \le 1$.

(i) Check that above given is a pdf.

(ii) Determine a number 'b' such that $P(X < b) = P(X > b)$.

Solution : (i) $f(x) \ge 0\ \forall\ 0 \le x \le 1$

Also

$$\int_0^1 f(x)dx = 6\int_0^1 x(1-x)dx = 6\int_0^1 (x-x^2)dx = 6\left[\frac{x^2}{2}-\frac{x^3}{3}\right]_0^1 = 6\cdot\frac{1}{6} = 1$$

As $f(x) \ge 0$ and $\int_0^1 f(x)dx = 1,\ 0 \le x \le 1$.

Hence above given is a p.d.f.

(ii) $$P(X<b) = \int_0^b f(x)dx = 6\int_0^b \left(x-x^2\right)dx = 6\left[\frac{x^2}{2}-\frac{x^3}{3}\right]_0^b = 3b^2-2b^3$$

and $P(X > b) = 1 - P(X < b)$

Now $P(X < b) = P(X > b)$

$\Rightarrow$ $2P(X < b) = 1$ $\Rightarrow$ $2(3b^2 - 2b^3) = 1$

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

$\Rightarrow$ $b = \frac{1}{2}$ or $b = \frac{2 \pm \sqrt{4+8}}{2.2} = \frac{1 \pm \sqrt{3}}{2}$

Now $b = \frac{1+\sqrt{3}}{2} > 1$, hence not acceptable as $0 < x < 1$.

and $b = \frac{1-\sqrt{3}}{2} < 0$, hence not acceptable.

Hence $b = \frac{1}{2}$.

1.4.2 Distribution Function

Let X be a continuous random variable with pdf f(x), then the function $F_X(x)$ is called the continuous distribution function or cumulative distribution function of the random variable X where

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

Properties of continuous distribution function :-

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

(ii) $F(x) = \int_{-\infty}^{x} f(x)dx \Rightarrow F'(x) = f(x)$

i.e., $\frac{d}{dx}F(x) = f(x) \ge 0.$

Hence F(x) is a non decreasing function of x.

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

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

(iv) F(x) is a continuous function of x on the right having at most countable discontinuities.

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

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

i.e., $P(a \le X \le b) = F(b) - F(a).$

Remark 2 Since $F'(x) = f(x)$ we have $dF(x) = f(x)\,dx$. This is known as probability differential of X. Hence we can find pdf with the help of cdf as $f(x) = d/dx\,F(x)$ at all points where F(x) is differentiable.

We observe that the distribution functions of discrete random variables grow only by jumps, whereas the distribution functions of continuous random variables are continuous functions and have no jumps. A random variable X is said to be of mixed type if its distribution function has both jumps and continuous growth.

Example 7. Let X be a continuous random variable with p.d.f.

$$f(x) = \begin{cases} ax & 0 \le x \le 1 \\ a & 1 \le x \le 2 \\ -ax + 3a & 2 \le x \le 3 \\ 0 & \text{elsewhere}. \end{cases}$$

(i) Determine the constant a.

(ii) Find $P(X \le 1.5)$.

(iii) Determine the cdf and hence find $P(X \le 2.5)$.

Solution : (i) As f(x) is given to be a pdf, hence

$$\int_0^\infty f(x)dx = 1 \Rightarrow \int_0^1 f(x)dx + \int_1^2 f(x)dx + \int_2^3 f(x)dx = 1$$

$$\Rightarrow \quad \int_0^1 ax\,dx + \int_1^2 a\,dx + \int_2^3 (-ax+3a)dx = 1$$

$$\Rightarrow \quad a\left(\frac{x^2}{2}\right)_0^1 + a(x)_1^2 + \left(\frac{-ax^2}{2} + 3ax\right)_2^3 = 1$$

$$\Rightarrow \quad a\left(\frac{1}{2}\right) + a(1) + \left(\frac{-a}{2}\right)(5) + 3a(1) = 1$$

$$\Rightarrow \quad 2a = 1 \quad \Rightarrow \quad a = \frac{1}{2}.$$

(ii) $$P(X \le 1.5) = \int_0^1 f(x)dx + \int_1^{1.5} f(x)\,dx = \int_0^1 ax\,dx$$

$$+\int_1^{1.5} adx = a\left(\frac{x^2}{2}\right)_0^1 + (ax)_1^{1.5} = \frac{a}{2} + (0.5)a = a = \frac{1}{2}.$$

(iii) For $x < 0$ $\quad F(x) = 0.$

For $0 \le x \le 1$, $$F(x) = \int_0^x f(x)dx = \int_0^x ax\,dx = a\left(\frac{x^2}{2}\right)_0^x$$

$$= a\frac{x^2}{2} = \frac{x^2}{4} \quad \left(\because a = \frac{1}{2}\right)$$

For $1 \le x \le 2$, $$F(x) = \int_0^1 f(x)dx + \int_1^x f(x)dx = \int_0^1 ax\,dx + \int_1^x adx$$

$$= a\left(\frac{x^2}{2}\right)_0^1 + a(x)_1^x = \frac{a}{2} + a(x-1)$$

$$= \frac{1}{4} + \frac{1}{2}(x-1) = \frac{x}{2} - \frac{1}{4} \quad \left(\because a = \frac{1}{2}\right)$$

For $2 \le x \le 3$, $F(x) = \int_0^1 f(x)dx + \int_1^2 f(x)dx + \int_2^x f(x)dx$

$$= \int_0^1 ax\,dx + \int_1^2 a\,dx + \int_2^x (-ax+3a)\,dx$$

$$= \frac{a}{2} + a(1) + \left(-a\frac{x^2}{2} + 3ax\right)_2^x$$

$$= \frac{3a}{2} + \left(\frac{-a}{2}\right)\left[x^2 - 4\right] + 3a(x-2)$$

$$= \frac{3a}{2} - \frac{ax^2}{2} + 2a + 3ax - 6a$$

$$= -\frac{5a}{2} + 3ax - \frac{ax^2}{2}$$

$$= \frac{-5}{4} + \frac{3x}{2} - \frac{x^2}{4}$$

For $x \ge 3$, $F(x) = \int_0^\infty f(x)dx = 1$

Hence the required cdf is :-

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{4} & 0 \le x \le 1 \\ \frac{x}{2} - \frac{1}{4} & 1 \le x \le 2 \\ \frac{-x^2}{4} + \frac{3x}{2} - \frac{5}{4} & 2 \le x \le 3 \\ 1 & x \ge 3 \end{cases}$$

Now $P(X \le 2.5) = F(2.5) = -\frac{(2.5)^2}{4} + \frac{3}{2}(2.5) - \frac{5}{4}$

$$= -1.5625 + 3.75 - 1.25$$

$$= 0.9375.$$

***Example* 8.** A continuous random variable X has the following distribution function :-

$$F(X) = \begin{cases} 0 & \text{if} \quad x \le 1 \\ k(x-1)^4 & \text{if} \quad 1 < x \le 3 \\ 1 & \text{if} \quad x > 3 \end{cases}$$

Find (i) k (ii) pdf f(x).

Solution : The p.d.f. $f(x) = \frac{d}{dx}F(x)$ provided F(x) is differentiable.

For $1 \le x \le 3$, $f(x) = \frac{d}{dx}\left[k(x-1)^4\right] = 4k(x-1)^3$

As F(x) is differentiable for x = 1 and x = 3.

For $x < 1$, $f(x) = \frac{d}{dx}(0) = 0$

For $x > 3$, $f(x) = \frac{d}{dx}(1) = 0$

$\therefore$ The pdf is $f(x) = \begin{bmatrix} 4k(x-1)^3, & 1 \le x \le 3 \\ 0, & \text{elsewhere}. \end{bmatrix}$

Also $\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow \int_1^3 f(x)dx = 1$

$$\Rightarrow \quad \int_1^3 4k(x-1)^3 dx = 1 \Rightarrow \left[4\frac{(x-1)^4}{4}k\right]_1^3 = 1$$

$$\Rightarrow \quad 16k = 1 \quad \Rightarrow \quad k = \frac{1}{16}.$$

Also the required pdf now is :-

$$f(x) = \begin{cases} \frac{1}{4}(x-1)^3, & 1 \le x \le 3 \\ 0, & \text{elsewhere}. \end{cases}$$

Illustrative Examples

***Example* 9.** Two cards are drawn without replacement from a well shuffled deck of 52 cards. Determine the probability distribution of the number of face cards.

Solution : Let X denote the number of face cards (i.e. Jack, Queen, King and Ace) obtained in a draw of 2 cards. Then X = 0, 1, 2. A deck of 52 cards contains 16 face cards and 36 other cards.

$\therefore$ $P(X=0)$ = probability that no face card is obtained

$$= \frac{{}^{36}C_2}{{}^{52}C_2} = \frac{36}{52} \times \frac{35}{51} = \frac{105}{221}$$

Similarly

$$P(X=1) = \frac{{}^{36}C_1 \times {}^{16}C_1}{{}^{52}C_2} = \frac{36 \times 16}{52 \times 51} \times 2 = \frac{96}{221}$$

and $$P(X=2) = \frac{{}^{16}C_2}{{}^{52}C_2} = \frac{16}{52} \times \frac{15}{51} = \frac{20}{221}.$$

Hence the required probability distribution is :

X :	0	1	2
P(X) :	$\frac{105}{221}$	$\frac{96}{221}$	$\frac{20}{221}$.

***Example* 10.** A shipment of 10 television sets contains 3 defective sets. A hotel makes a random purchase of 3 of the sets. If X is the number of defective sets purchased by the hotel, find the probability distribution of X.

Solution : Let random variable X denote the number of defective sets purchased by the hotel. Hence X = 0, 1, 2, 3. Shipment of 10 sets consists of 3 defective sets and 7 good sets.

Now $P(X=r)$ = P(choosing exactly r defective sets)

= P(choosing r defective and (3 – r) good sets)

$$= \frac{{}^{3}C_r \times {}^{7}C_{3-r}}{{}^{10}C_3}$$ (considering sets are not replaced)

Hence the required probability distribution is :-

X :	0	1	2	3
P(X) :	$\frac{{}^{7}C_3}{{}^{10}C_3} = \frac{7}{24}$	$\frac{{}^{3}C_1 \times {}^{7}C_2}{{}^{10}C_3} = \frac{21}{40}$	$\frac{{}^{3}C_2 \times {}^{7}C_1}{{}^{10}C_3} = \frac{7}{40}$	$\frac{{}^{3}C_3 \times {}^{7}C_0}{{}^{10}C_3} = \frac{1}{120}$.

***Example* 11.** A random variable X can take all non negative integral values and the probability that X takes the value r is proportional to $\alpha^r (0 < \alpha < 1)$. Find $P(X = 0)$.

Solution : Given $P(X = r) \propto \alpha^r$, $r = 0, 1, 2$ ....

Hence let, $P(X = r) = A\,\alpha^r$, $r = 0, 1, 2$ ....

As $$\sum_{r=0}^{\infty} P(X = r) = 1 \Rightarrow \sum_{r=0}^{\infty} A\,\alpha^r = 1$$

$$\Rightarrow \quad A[1 + \alpha + \alpha^2 + \ldots.] = 1 \quad \Rightarrow \quad A(1-\alpha)^{-1} = 1$$

$$\Rightarrow \quad A = 1 - \alpha.$$

Therefore, $P(X = r) = (1 - \alpha)\alpha^r$

and $P(X = 0) = 1 - \alpha$.

***Example* 12.** The probability distribution of a random variable X is given by :-

x_i :	0	1	2
p_i :	$3C^3$	$4C - 10C^2$	$5C - 1$

where $C > 0$

Find (i) C, (ii) $P(X < 2)$, (iii) $P(1 < X \le 2)$.

Solution : (i) As p_i is a probability distribution therefore $\Sigma p_i = 1$.

$$\Rightarrow \quad 3C^3 + 4C - 10C^2 + 5C - 1 = 1$$

$$\Rightarrow \quad 3C^3 - 10C^2 + 9C - 2 = 0$$

$$\Rightarrow \quad (C-2)(3C^2 - 4C + 1) = 0$$

$$\Rightarrow \quad (C-2)\ (3C-1)(C-1) = 0$$

$$\Rightarrow \quad C = 1,\ 2,\ 1/3.$$

But p_i is probability hence $0 \le p_i \le 1$, therefore the values $C = 1$ and 2 are not acceptable. Hence $C = 1/3$.

(ii) $$P(X < 2) = 1 - P(X = 2) = 1 - (5C - 1) = 2 - \frac{5}{3} = \frac{1}{3}$$

(iii) $$P(1 < X \le 2) = P(X = 2) = 5C - 1 = \frac{5}{3} - 1 = \frac{2}{3}.$$

***Example* 13.** A random variable X may assume 4 values with probabilities $(1 + 3x)/4, (1-x)/4, (1+2x)/4$ and $(1-4x)/4$. Find the condition on x so that these values represent the probability function of X.

Solution : If the given values represent the probability function then :-

$$\frac{1+3x}{4}+\frac{1-x}{4}+\frac{1+2x}{4}+\frac{1-4x}{4}=1,$$

which is obviously true $\forall$ x. Also each probability should be greater than or equal to zero.

i.e., $\frac{1+3x}{4}\geq 0 \Rightarrow x\geq -\frac{1}{3}$

and $\frac{1-x}{4}\geq 0 \Rightarrow x\leq 1$

and $\frac{1+2x}{4}\geq 0 \Rightarrow x\geq \frac{-1}{2}$

and $\frac{1-4x}{4}\geq 0 \Rightarrow x\leq \frac{1}{4}.$

Therefore, the values of x, for which a probability function is defined, lie in the range $\frac{-1}{3}\leq x\leq \frac{1}{4}.$

***Example* 14.** The probability distribution of a random variable X is given below :-

x	: 0	1	2	3
P(x)	: 0.1	0.3	0.5	0.1 .

If $Y = X^2 + 2X$, find the probability distribution of Y.

Solution :

x :	0	1	2	3
$y = x^2 + 2x$:	0	3	8	15
$p(x) = p(y)$:	0.1	0.3	0.5	0.1 .

Hence the required probability distribution of Y is :

y	: 0	3	8	15
P(y)	: 0.1	0.3	0.5	0.1 .

***Example* 15.** Let a pair of fair dice be tossed and let the random variable X denote the sum of points. Obtain the probability mass function of X.

Solution : The minimum and maximum numbers obtained on throwing a dice are 1 and 6 respectively.

Hence X = 2, 3, 4, . . ., 12

Now $P(X=2) = P(1 \text{ on 1st dice}) \times P(1 \text{ on 2nd dice})$

$$= \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} = P(X=12)$$

$P(X=3) = 2 \times P(1 \text{ on 1st dice}) \times P(2 \text{ on 2nd dice})$

$$= 2 \times \frac{1}{6} \times \frac{1}{6} = \frac{2}{36} = P(X=11)$$

$P(X=4) = 2 \times P(1 \text{ on 1st dice}) \times P(3 \text{ on 2nd dice})$ or

$P(2 \text{ on 1st dice}) \times P(2 \text{ on 2nd dice})$

$$= 2 \times \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} = \frac{3}{36} = P(X=10)$$

Similarly

$$P(X=5) = \frac{4}{36} = P(X=9)$$

$$P(X=6) = \frac{5}{36} = P(X=8)$$

$$P(X=7) = \frac{6}{36}.$$

Hence required p.m.f. is :-

$$P(X=x) = \frac{6-|7-r|}{36}, x = 2, 3, \ldots, 12.$$

***Example* 16.** Find the value of k so that $f(x) = kx(2-x)$, may be a probability density function of a random variable x for $0 \le x \le 2$.

Solution : As $\int_0^2 f(x)dx = 1$

$$\Rightarrow \int_0^2 kx(2-x)dx = 1$$

$$\Rightarrow k\left(2\frac{x^2}{2} - \frac{x^3}{3}\right)_0^2 = 1 \Rightarrow k\left(4 - \frac{8}{3}\right) = 1$$

$$\Rightarrow \frac{4}{3}k = 1$$

or $k = \frac{3}{4}$.

Example 17. The pdf of the random variable X is given by $f(x)=\begin{cases}\dfrac{k}{\sqrt{x}}, & \text{for } 0<x<4\\ 0, & \text{elsewhere}\end{cases}$

Find (i) k

(ii) $P\left(X<\frac{1}{4}\right)$ and $P(X>1)$.

Solution : (i) We have $\int_0^4 f(x)dx=1 \Rightarrow k\int_0^4 x^{-1/2}dx=1$

$$\Rightarrow \quad k\left(\frac{x^{1/2}}{1/2}\right)_0^4=1 \Rightarrow 2k\times 2=1 \Rightarrow k=\frac{1}{4}.$$

(ii) $$P\left(X<\frac{1}{4}\right)=\int_0^{1/4} f(x)dx=\int_0^{1/4}\frac{1}{4}x^{-1/2}dx$$

$$=\frac{1}{4}\left(\frac{x^{1/2}}{1/2}\right)_0^{1/4}=\frac{1}{2}\left(\frac{1}{2}\right)=\frac{1}{4}$$

and $$P(X>1)=1-P(X<1)$$

$$=1-\int_0^1 f(x)dx=1-\frac{1}{4}\left(\frac{x^{1/2}}{1/2}\right)_0^1$$

$$=1-\frac{1}{2}=\frac{1}{2}.$$

Example 18. If $p(x)=\begin{cases} xe^{-x^2/2} & x\geq 0\\ 0 & x<0\end{cases}$

(i) Show that p(x) is a pdf.

(ii) Find its distribution function P(x).

Solution : (i) $p(x)\geq 0 \quad \forall \quad x\geq 0$

and $$\int_0^\infty p(x)dx=\int_0^\infty xe^{-x^2/2}dx=\int_0^\infty e^{-t}dt=-\left(e^{-t}\right)_0^\infty=1 \qquad \left(t=\frac{x^2}{2}\right)$$

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

(ii) $\forall \quad x<0 \quad P(x)=0$

$$0 \le x < \infty, \quad P(x)=\int_0^x f(x)dx=\int_0^x 1\cdot xe^{-x^2/2}dx$$

$$=\left(-1\cdot e^{-x^2/2}\right)_0^x-0=1-e^{-x^2/2}$$

Hence $P(x)=\begin{cases}0 & x<0\\ 1-e^{-x^2/2} & x\ge 0.\end{cases}$

***Example* 19.** A continuous random variable X can assume any value between $x=2$ and $x=5$ and has a density function given by $f(x)=k(1+x)$. Find $P(X<4)$.

Solution : As f(x) is a p.d.f hence

$$\int_2^5 f(x)dx=1 \Rightarrow \left(k\frac{(1+x)^2}{2}\right)_2^5=1 \Rightarrow \frac{27}{2}k=1 \Rightarrow k=\frac{2}{27}$$

Therefore $P(X<4)=\int_2^4 f(x)dx=\int_2^4 \frac{2}{27}(1+x)dx$

$$=\frac{2}{27}\left[\frac{(1+x)^2}{2}\right]_2^4=\frac{2}{27}\cdot\frac{1}{2}(25-9)=\frac{16}{27}.$$

***Example* 20.** The cdf of a continuous random variable X is given by

$$F(x)=\begin{cases}0, & x<0\\ x^2, & 0\le x<\frac{1}{2}\\ 1-\frac{3}{25}(3-x)^2, & \frac{1}{2}\le x<3\\ 1, & x\ge 3\end{cases}$$

Find pdf of X, $P(|X|\le 1)$ and $P\left(\frac{1}{3}\le X<4\right)$.

Solution :

$$f(x)=\frac{d}{dx}F(x)=\begin{cases}0, & x<0\\ 2x, & 0\le x<\frac{1}{2}\\ \frac{6}{25}(3-x), & \frac{1}{2}\le x<3\\ 0, & x\ge 3\end{cases}$$

where we have assumed that $f\left(\frac{1}{2}\right)=\frac{3}{5}$ and $f(3)=0$ as these are the points of discontinuity.

$$P(|X|\le 1)=P(-1<X<1)=F(1)-F(-1)=\frac{13}{25}-0=\frac{13}{25}$$

$$P\left(\frac{1}{3}\le X<4\right)=F(4)-F\left(\frac{1}{3}\right)=1-\frac{1}{9}=\frac{8}{9}.$$

***Example* 21.** The daily consumption of electric power (in millions of kw-hours) is a random variable having pdf as $f(x)=\begin{cases}\frac{1}{9}xe^{-x/3}, & x>0\\ 0, & x\le 0.\end{cases}$

If the total production is 12 million kw-hours, determine the probability that there is power cut (shortage) on any given day.

Solution : There will be a power cut if the demand increases beyond 12 millon kw-hours. Hence required probability $P(X>12)=1-P(0<X<12)$

$$=1-\int_0^{12} f(x)dx=1-\int_0^{12}\frac{1}{9}xe^{-x/3}dx=1-\left(-\frac{x}{3}e^{-x/3}-e^{-x/3}\right)_0^{12}=1-\left(1-5e^{-4}\right)$$

$$=5\,e^{-4}=5\times 0.018316=0.09158$$

***Example* 22.** Let X be the random variable denoting the time to failure of a component. Suppose the distribution function of X is F(x). Use this distribution function to express the probability of following events :

(i) $9<X<90$ (ii) $X<90$

(iii) $X>90$, given that $X>9$.

Solution : (i) $P(9<X<90)=F(90)-F(9)$

(ii) $P(X<90)=F(90)$

(iii) $$P\left(\frac{X>90}{X>9}\right)=\frac{P(9<X<90)}{P(X>9)}=\frac{F(90)-F(9)}{1-P(X<9)}$$

$$=\frac{F(90)-F(9)}{1-F(9)}$$

***Example* 23.** Find the distribution function of the random variable X whose p.d.f. is given by

$$f(x)=\begin{cases} x & \text{for } 0<x<1\\ 2-x & \text{for } 1\le x<2\\ 0 & \text{elsewhere}\end{cases}.$$

Solution : For $x < 0$ $F(x) = 0$

for $0 < x < 1$ $F(x) = \int_0^x f(x)dx = \int_0^x x\,dx = \frac{x^2}{2}$

for $1 \le x < 2$ $F(x) = \int_0^1 f(x)dx + \int_1^x f(x)dx$

$$= \frac{1}{2} + \int_1^x (2-x)dx = \frac{1}{2} - \left(\frac{(2-x)^2}{2}\right)_1^x$$

$$= \frac{1}{2} - \frac{(2-x)^2}{2} + \frac{1}{2} = -\frac{x^2}{2} + 2x - 1$$

for $2 \le x$ $F(x) = \int_0^1 f(x)dx + \int_1^2 f(x)dx + 0 = \frac{1}{2} + \frac{1}{2} = 1$

$$\text{Hence } F(x) = \begin{cases} 0 & x \le 0 \\ \frac{x^2}{2} & 0 < x < 1 \\ 2x - \frac{x^2}{2} - 1 & 1 \le x < 2 \\ 1 & x \ge 2. \end{cases}$$

***Example* 24.** The time one has to wait for a bus at a downtown bus stop is observed to be random phenomenon (X) with the following p.d.f.

$f(x) = 0$ for $x<0$

$= \frac{1}{9}(x+1)$ for $0 \le x < 1$

$= \frac{4}{9}\left(x - \frac{1}{2}\right)$ for $1 \le x < \frac{3}{2}$

$= \frac{4}{9}\left(\frac{5}{2} - x\right)$ for $\frac{3}{2} \le x < 2$

$= \frac{1}{9}(4-x)$ for $2 \le x < 3$

$= \frac{1}{9}$ for $3 \le x < 6$

$= 0$ for $x \le 6$.

Let A and B be the events that one waits between 0 to 2 minutes and 1 to 3 minutes, respectively. Show that

(i) $P\left(\frac{B}{A}\right)=\frac{2}{3}$ (ii) $P\left(\overline{A}\cap\overline{B}\right)=\frac{1}{3}$.

Solution : (i) $P\left(\frac{B}{A}\right)=\frac{P(A\cap B)}{P(A)}=\frac{P(1\le X\le 2)}{P(0\le X\le 2)}$ .....(2.1)

Now $P(1\le X\le 2)=\int_1^2 f(x)dx=\int_1^{3/2} f(x)dx+\int_{3/2}^2 f(x)dx$

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

$$=\frac{4}{9}\left(\frac{x^2}{2}-\frac{1}{2}x\right)_1^{3/2}+\frac{4}{9}\left(\frac{5}{2}x-\frac{x^2}{2}\right)_{3/2}^2$$

$$=\frac{4}{2.9}\left[\left(\frac{9}{4}-1\right)-\left(\frac{3}{2}-1\right)\right]+\frac{4}{2.9}\left[5\left(2-\frac{3}{2}\right)-\left(4-\frac{9}{4}\right)\right]$$

$$=\frac{4}{18}\left[\frac{5}{4}-\frac{1}{2}+\frac{5}{2}-\frac{7}{4}\right]=\frac{4}{18}\times\frac{3}{2}=\frac{1}{3}$$

and $P(0\le X\le 2)=\int_0^2 f(x)dx=\int_0^1\frac{1}{9}(x+1)dx+P(1\le X\le 2)$

$$=\frac{1}{9}\left[\frac{(x+1)^2}{2}\right]_0^1+\frac{1}{3}=\frac{1}{18}[4-1]+\frac{1}{3}=\frac{1}{6}+\frac{1}{3}=\frac{3}{6}=\frac{1}{2}$$

Using these values in equation (2.1) we get :-

$$P\left(\frac{B}{A}\right)=\frac{\frac{1}{3}}{\frac{1}{2}}=\frac{2}{3}.$$

(ii) $P\left(\overline{A}\cap\overline{B}\right)=P(\text{waiting time is more than 3 minutes})$

$$=P(X\ge 3)=\int_3^{\infty} f(x)dx$$

$$=\frac{1}{9}\int_3^6 dx+0=\frac{1}{9}(6-3)=\frac{1}{3}.$$

***Example* 25.** The amount of bread (in hundreds of pounds) X that a certain bakery is able to sell in a day is found to be a numerical valued random phenomenon, with pdf as

$$f(x) = kx \qquad \text{for } 0 \le x < 5$$
$$= k(10-x) \qquad \text{for } 5 \le x < 10$$
$$= 0 \qquad \text{otherwise}.$$

(i) Find k

(ii) If A, B, C be the events that the number of pounds of bread sold is more than 500 pounds, less than 500 pounds and between 250 and 750 pounds respectively. Find P(A), P(B), P(C).

(iii) Find P(A/B), P(A/C) and find whether

(a) A and B are independent.

(b) A and C are independent.

Solution : (i) In order that f(x) should be a pdf $\int_{-\infty}^{\infty} f(x)dx = 1$

i.e., $$\int_0^5 f(x)dx + \int_5^{10} f(x)dx = 1$$

$$\Rightarrow \quad k\int_0^5 x dx + k\int_5^{10}(10-x)dx = 1$$

$$\Rightarrow \quad k\left(\frac{x^2}{2}\right)_0^5 - k\left[\frac{(10-x)^2}{2}\right]_5^{10} = 1$$

$$\Rightarrow \quad \frac{25}{2}k + \frac{25}{2}k = 1 \Rightarrow k = \frac{1}{25}$$

(ii) P(A) = P(X > 5)
(as unit of x is hundreds of pounds)

$$= \int_5^{\infty} f(x)dx = k\int_5^{10}(10-x)dx = \frac{-1}{25}\left[\frac{(10-x)^2}{2}\right]_5^{10}$$

$$= \frac{1}{25} \times \frac{25}{2} = \frac{1}{2} = 0.5$$

$$P(B) = P(X < 5) = 1 - P(X > 5) = 1 - \frac{1}{2} = 0.5$$

and $$P(C) = P(2.5 \le X \le 7.5) = \int_{2.5}^{7.5} f(x)dx$$

$$= \int_{2.5}^{5} f(x)dx + \int_{5}^{7.5} f(x)dx = \int_{2.5}^{5} \frac{1}{25} x dx + \int_{5}^{7.5} \frac{1}{25}(10 - x)dx$$

$$= \frac{1}{50} \times (25 - 6.25) - \frac{1}{50}(6.25 - 25) = \frac{2}{50} \times 18.75 = \frac{3}{4}$$

Also $$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)} = \frac{0}{0.5} = 0 \quad [\because A \cap B = \phi]$$

and $$P\left(\frac{A}{C}\right) = \frac{P(A \cap C)}{P(C)} = \frac{P(5 < X < 7.5)}{P(C)}$$

$$= \frac{\int_{5}^{7.5} f(x)dx}{3/4} = \frac{-1}{25}\left(\frac{(10 - x)^2}{2}\right)_5^{7.5} \times \frac{4}{3}$$

$$= \frac{3}{8} \times \frac{4}{3} = \frac{1}{2}$$

Also $$P(A \cap B) = 0 \neq P(A) \cdot P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

and $$P(A \cap C) = \frac{3}{8} = P(A) \cdot P(C) = \frac{1}{2} \times \frac{3}{4} = \frac{3}{8}$$

Hence events A and B are not independent while the events A and C are independent.

***Example* 26.** Find pdf of a random variable X whose cdf is given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 \le x \le 1 \\ 1 & \text{for } x > 1 \end{cases}.$$

Solution : $f(x) = \frac{d}{dx}F(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$

As F(x) is not differentiable at $x = 0$ and $x = 1$ hence we can define $f(x) = 0$ for $x = 0$ and $x = 1$.

***Example* 27.** The time (measured in years), X, required to complete a software project has pdf of the form $f(x) = \begin{cases} kx(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise}. \end{cases}$

Find the probability that the project will be completed in less than four months.

Solution : As f(x) ia s pdf, hence $\int_{-\infty}^{\infty} f(x)dx = 1$

$$\Rightarrow \quad k\int_0^1 x(1-x)dx = 1 \Rightarrow k\left(\frac{x^2}{2} - \frac{x^3}{3}\right)_0^1 = 1 \Rightarrow k = 6.$$

Required probability $= P\left(X < \frac{4}{12}\right)$

$$\left[\because 4 \text{ months} = \frac{4}{12}\text{years}\right]$$

$$= P\left(X < \frac{1}{3}\right) = \int_0^{1/3} f(x)dx$$

$$= 6\left(\frac{x^2}{2} - \frac{x^3}{3}\right)_0^{1/3} = 6\left[\frac{1}{18} - \frac{1}{81}\right] = \frac{1}{3} - \frac{2}{27} = \frac{7}{27}.$$

***Example* 28.** Verify that the following is a distribution function ;-

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

Solution : Obviously $0 \le F(x) \le 1$.

Also F(x) is continuous at $x = \pm a$

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

Now if F(x) is a distribution function then f(x) should be pdf.

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

Hence F(x) is a cdf.

***Example* 29.** A petrol pump is supplied with petrol once a day. Its daily volume X of sales in thousands of litres is distributed by $f(x) = 5(1-x)^4 \; 0 \le x \le 1$. What must be the capacity of its tank in order that the probability that its supply will be exhausted in a given day shall be 0.01 ?

Solution : Let the capacity of the tank be 'r' thousand litres. Its supply is exhausted, means that its sales are more than its capacity.

Hence given probability is $P(X \ge r) = 0.01$.

$$\Rightarrow \quad \int_{r}^{1} f(x)dx = 0.01 \quad \Rightarrow \quad \int_{r}^{1} 5(1-x)^4 dx = 0.01$$

$$\Rightarrow \quad (1-r)^5 = \frac{1}{100} \Rightarrow 1-r = \frac{1}{(100)^{1/5}} \Rightarrow r = 1 - \frac{1}{(100)^{1/5}} = 0.6019$$

Hence capacity of tank = 601.9 litres.

***Example* 30.** Suppose the life in hours of a certain kind of radio tube has the pdf

$$f(x) = \begin{cases} \frac{100}{x^2} & \text{when } x \ge 100 \\ 0 & \text{when } x < 100 \end{cases}$$

What is the probability that (i) none of three such tubes in a given radio set will have to be replaced during the first 150 hours of operations ? (ii) all three of the original tubes will have to be replaced during the first 150 hours ? (iii) what is the maximum number of tubes that may be inserted into a set so that there is a probability 0.1 that after 150hr of service all of them are still functioning ?

Solution : Probability that a tube does not last for first 150 hours is given by :-

$$P(X \le 150) = P(0 < X < 100) + P(100 \le X \le 150)$$

$$= 0 + \int_{100}^{150} f(x)dx = -100\left(x^{-1}\right)_{100}^{150}$$

$$= -100\left[\frac{1}{150}-\frac{1}{100}\right] = 100\times\frac{50}{100\times150} = \frac{1}{3}$$

Also the probability that a tube lasts for first 150 hours is $= P(X > 150)$

$$= 1 - P(X < 150) = 1 - \frac{1}{3} = \frac{2}{3}$$

(i) Hence for first 150 hours P(none of the three tubes is replaced)

$$= \frac{2}{3}\times\frac{2}{3}\times\frac{2}{3} = \frac{8}{27}$$

(ii) Hence for first 150 hours P(all three tubes are replaced) $= \frac{1}{3}\times\frac{1}{3}\times\frac{1}{3} = \frac{1}{27}$

(iii) Let n tubes are inserted, given P(none of n tubes are replaced) = 0.1.

$$\Rightarrow \quad \left(\frac{2}{3}\right)^n = 0.1 \Rightarrow n\log\frac{2}{3} = \log 0.1$$

$$\Rightarrow \quad n = \frac{-2.30\,259}{-0.40\,547} = 5.6788$$

Hence $n \cong 5$.

Exercise 1(A)

Q.1 Determine which of the following can be probability distributions of a random variable X.

(i)

X	0	1	2
P(X)	0.4	0.4	0.2

(ii)

X	0	1	2
P(X)	0.6	0.1	0.2

Ans. (i) Yes (ii) No

Q.2 An urn contains 4 white and 6 red balls. Four balls are drawn at random from the urn. Find the probability distribution of the number of white balls.

Ans.

X	0	1	2	3	4
P(X)	1/14	8/21	6/14	4/35	1/210

Q.3 An unbiased dice is thrown twice. Find the probability distribution of the number of sixes.

Ans. X : 0 1 2

P(X) :	$\frac{25}{36}$	$\frac{5}{18}$	$\frac{1}{36}$

Q.4 A random variable X has the following probability distribution.

X :	–2	–1	0	1	2	3
P(X) :	0.1	k	0.2	2k	0.3	3k

Find the value of

(i) k

(ii) $P(X < 2)$ and $P(-2 < X < 2)$

(iii) cdf of X.

Ans. (i) $k = \frac{1}{15}$ (ii) $\frac{1}{2}, \frac{2}{5}$

(iii)
$$F(x) = \begin{cases} \frac{1}{10} & \text{when } -2 \le x < -1 \\ \frac{1}{6} & \text{when } -1 \le x < 0 \\ \frac{11}{30} & \text{when } 0 \le x < 1 \\ \frac{1}{2} & \text{when } 1 \le x < 2 \\ \frac{4}{5} & \text{when } 2 \le x < 3 \\ 1 & \text{when } x \ge 3 \end{cases}$$

Q.5 A random variable X has the following probability distribution :-

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

Determine (i) the value of a

(ii) $P(X < 3)$, $P(X \ge 3)$, $P(0 < X < 5)$.

Ans. (i) $a = \frac{1}{81}$ (ii) $\frac{1}{9}, \frac{8}{9}, \frac{8}{27}$

Q.6 A random variable X takes the values –3, –2, –1, 0, 1, 2 and 3 such that

$$P(X = 0) = P(X > 0) = P(X < 0)$$
$$P(X = -3) = P(X = -2) = P(X = -1)$$
$$P(X = 1) = P(X = 2) = P(X = 3) \ .$$

Obtain the probability distribution of X. Further find the pmf of $Y = 2X^2 + 3X + 4$.

Ans.

X :	–3	–2	–1	0	1	2	3
Y :	13	6	3	4	9	18	31
P(X) :	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

Q.7 Find whether the function given by $f(x) = \frac{(x+2)}{25}$ for x = 1, 2, 3, 4, 5 can serve as the probability distribution of discrete random variable.

Ans. Yes

Q.8 From a lot of 25 items, 5 of which are defective, 4 items are chosen at random. If X is the number of defectives found, obtain the probability distribution of X, when items are chosen

(i) without replacement (ii) with replacement .

Ans. (i) $P(X=r) = \frac{{}^5C_r \times {}^{20}C_{4-r}}{{}^{25}C_4}$ $r - 0, 1, 2, 3, 4$

(ii) $P(X=r) = {}^4C_2\left(\frac{5}{25}\right)^2\left(\frac{20}{25}\right)^{4-r}$, $r = 0, 1, 2, 3, 4$

Q.9 If the random variable X takes the values 1, 2, 3 and 4 such that $2P(X=1) = 3P(X=2) = P(X=3) = 5P(X=4)$, find the probability distribution and cumulative distribution function of X.

Ans. pdf is

x_i :	1	2	3	4
p_i :	$\frac{15}{61}$	$\frac{10}{61}$	$\frac{30}{61}$	$\frac{6}{61}$

cdf is

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{15}{61} & 1 \le x < 2 \\ \frac{25}{61} & 2 \le x < 3 \\ \frac{55}{61} & 3 \le x < 4 \\ 1 & x \ge 4 \end{cases}.$$

Q.10 A random variable can assume the values –1, 0, 1 with probability 1/3, 1/2, 1/6 respectively determine the distribution function of X.

Ans. $P(X < -1) = 0$, $P(X < 0) = 1/3$, $P(X < 1) = 5/6$, $P(X < \infty) = 1$

Q.11 The probability function of an infinite discrete distribution is given by $P(x = j) = 1/2^j$ ($j = 1, 2, \ldots, \infty$). Verify that the total probability is 1. Also find P(X is even), $P(X \geq 5)$ and P(X is divisible by 3).

Ans. $\frac{1}{3}, \frac{1}{16}, \frac{1}{7}$

Q.12 Find the values of a for which $P(x = j) = (1 - a)a^j, j = 0, 1, 2, \ldots$ represents a probability mass function. Show also that for any two positive integers m and n, $P(X > m + n / X > m) = P(X \geq n)$.

Q.13 The probability mass function of a random variable X is defined as $P(X = 0) = 3C^2$, $P(X = 1) = 4C - 10C^2$, $P(X = 2) = 5C - 1$, where $C > 0$ and $P(X = r) = 0$, if $r \neq 0, 1, 2$.

Find (i) the value of C.

(ii) $P[0 < X < 2 / X > 0]$

(iii) the distribution function of X.

(iv) the largest value of X for which $F(x) < \frac{1}{2}$.

(v) the smallest value of X for which $F(x) > \frac{1}{2}$.

Ans. (i) $C = \frac{2}{7}$ (ii) $\frac{16}{37}$

(iii) $F(x) = 0$ when $x < 0$,

$F(x) = \frac{12}{49}$, when $0 \leq x < 1$, $F(x) = \frac{28}{49}$ when $1 \leq x < 2$

$F(x) = 1$ when $x \geq 2$

(iv) $x = 0$ (v) $x = 1$.

Q.14 A computer system has five identical tape drives. Each tape drive is in one of the two states-busy, denoted by '0' and available, denoted by '1'. If X is a random variable which denotes the number of available tape drive, find the event space E_X of X and its probability mass function.

Ans. Let A_i = event that number of available drives = i and busy drives is $5 - i$.

$\therefore$ $E_X = \{A_0, A_1, A_2, A_3, A_4, A_5\}$ and $P(A_i) = \frac{{}^5C_i}{32}, i = 0, 1, 2, \ldots, 5$

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

Find (i) P(X = 1 or 2)

(ii) $P\left\{\frac{1}{2} < x < \frac{5}{2} \middle| x > 1\right\}$

Ans. (i) $\frac{1}{5}$ (ii) $\frac{1}{7}$

Q.16 Find the distribution function of the total number of heads obtained in four tosses of a fair coin.

Ans. $$F(x) = \begin{cases} 1/16, & x \le 0 \\ 5/16, & x \le 1 \\ 11/16, & x \le 2 \\ 15/16, & x \le 3 \\ 1, & x \le 4 \end{cases}$$

Q.17 The distribution function of a random variable X is given as follows ;-

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/8 & \text{if } 0 \le x < 1 \\ 3/8 & \text{if } 1 \le x < 2 \\ 6/8 & \text{if } 2 \le x < 3 \\ 7/8 & \text{if } 3 \le x < 4 \\ 1 & \text{if } x \ge 4 \end{cases}$$

(i) Determine the probability function f(x).

(ii) Evaluate $P(1 < X \le 4)$.

Ans. (i) $f(0) = \frac{1}{8}, f(1) = \frac{2}{8}, f(2) = \frac{3}{8}, f(3) = \frac{1}{8}, f(4) = \frac{1}{8}$ and $f(x) = 0$ elsewhere.

(ii) $\frac{5}{8}$

Q.18 The probability that a person will die in the time interval (t_1, t_2) is given by

$$P(t_1 \le t \le t_2) = \int_{t_1}^{t_2} f(t)dt$$

where $f(t) = \begin{bmatrix} 3 \times 10^{-9} t^2 (100-t)^2 & 0 \le t \le 100 \\ 0 & \text{eleswhere} \end{bmatrix}$

Determine :-

(i) The probability that a person will die between the ages 60 and 70.

(ii) the probability that he will die between those ages, assuming he lived up to 60.

Ans. (i) 0.1544 (ii) 0.4863

Q.19 A continuous random variable has a pdf $f(x) = 3x^2,\ 0 \le x \le 1$. Find a and b such that

(i) $P(X \le a) = P(X > a)$ (ii) $P(X > b) = 0.05$

Ans. (i) $a = 0.7937$ (ii) 0.9830

Q.20 The distribution function of a random variable X is given by $F(x) = 1 - (1 + x)\, e^{-x},\ x \ge 0$. Find the density function.

Ans. $f(x) = xe^{-x}, \quad x \ge 0$

Q.21 A continuous random variable has the cdf as :-

$$F(x) = \begin{cases} \dfrac{2x^2}{5} & \text{for } 0 < x \le 1 \\ \dfrac{-3}{5} + \dfrac{2}{5}\left(3x - \dfrac{x^2}{2}\right) & \text{for } 1 < x \le 2 \\ 1 & \text{for } x > 2 \end{cases}$$

Find the pdf.

Ans. $$f(x) = \begin{cases} \dfrac{4x}{5} & \text{for } 0 < x \le 1 \\ \dfrac{2}{5}(3 - x) & \text{for } 1 < x < 2 \\ 0 & \text{elsewhere .} \end{cases}$$

Q.22 The probability densiy of a random variable X is :-

$$f(x) = \begin{cases} 0, & x \le -a \\ \dfrac{a + x}{a^2}, & -a \le x \le 0 \\ \dfrac{a - x}{a^2}, & 0 \le x \le a \\ 0, & x \ge a\ . \end{cases}$$

Verify that $\int_{-\infty}^{\infty} f(x)dx = 1$. Also find cdf $F(x)$.

Ans. $F(x) = \begin{cases} 0, & x \le -a \\ \dfrac{ax + \frac{x^2}{2} + \frac{a^2}{2}}{a^2}, & -a < x < 0 \\ \dfrac{1}{2} + \dfrac{ax - \frac{x^2}{2}}{a^2}, & 0 < x < a \\ 1, & a \le x \end{cases}$.

Q.23 Let the phase error in a tracking device have probability density $f(x) = \begin{cases} \cos x & 0 < x < \pi/2 \\ 0 & \text{eleswhere} \end{cases}$.

Find the probability that the phase error is

(i) beween 0 and $\pi/4$

(ii) greater than $\pi/3$.

Ans. (i) 0.707 (ii) 0.1339

Q.24 Find the value of the constant k so that

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

is a pdf of continuous random variable.

Ans. $k = 6$

Q.25 The number of minutes that a flight from Delhi to Mumbai is early or late is a random variable whose pdf is given by :-

$$f(x) = \begin{cases} \frac{1}{288}(36 - x^2), & \text{for} -6 < x < 6 \\ 0, & \text{elsewhere} \end{cases},$$

where negative values indicate flight being early and positive values indicate flight being late. Find the probability that one of these flights will be

(i) at least 2 minutes early

(ii) at least 1 minute late

(iii) early from 1 to 3 minutes

(iv) exactly 5 minutes late.

Ans. (i) $\frac{7}{27}$ (ii) $\frac{325}{864}$ (iii) $\frac{95}{432}$ (iv) 0.04

Q.26 The mileage C in thousands of miles which car owners get with a certain kind of tyre is a random variable having probability density function as .

$$f(x) = \frac{1}{20}e^{-x/20} \quad \text{for } x > 0$$
$$= 0 \quad \text{for } x \le 0.$$

Find the probabilities that one of these tyres will last (i) at most 10,000 miles (ii) anywhere from 16,000 to 24,000 miles (iii) atleast 30,000 miles.

Ans. (i) 0.3935 (ii) 0.1481 (iii) 0.2231

Q.27 Suppose that the time in minutes that a person has to wait at a certain station for a train is a random phenomenon with the distribution function defined as

$$F(x) = \begin{cases} 0 & \text{for } x \le 0 \\ \frac{x}{2} & \text{for } 0 \le x < 1 \\ \frac{1}{2} & \text{for } 1 \le x < 2 \\ \frac{x}{4} & \text{for } 2 \le x < 4 \\ 1 & \text{for } x \ge 4 \end{cases}.$$

(a) Find its pdf.

(b) What is the probability that a person will have to wait

(i) more than 3 minutes (ii) less than 3 minutes and (iii) between 1 and 3 minutes ?

(c) What is the conditional probability that the person will have to wait for a train for

(i) more than 3 minutes, given that to he waits more than 1 minute

(ii) less than 3 minutes given that it he waits more than 1 minute.

Ans. (a) $f(x) = 0$ for $x < 0$, $f(x) = \frac{1}{2}$, for $0 \le x < 1$

$f(x) = 0$ for $1 \le x < 2$, $f(x) = \frac{1}{4}$ for $2 \le x < 4$, $f(x) = 0$ for $x \ge 4$

(b) (i) $\frac{1}{4}$ (ii) $\frac{3}{4}$ (iii) $\frac{1}{4}$

(c) (i) $\frac{1}{2}$ (ii) $\frac{1}{2}$

1.5 Two Dimensional Random Variables

In the previous sections of this chapter we have considered only one dimensional random variables. It often happens that we are interested in the relationships between two or more random variables defined on the same sample space. For example, both voltage and current might be of interest or we may be interested in a program consisting of two modules with execution times X and Y. Hence in this section we discuss two dimensional random variables.

Definition :-

Let S be the sample space associated with the random experiment E. Then the function $f : S \rightarrow R^2$ where $f(s) = (X, Y)$, where $s \in S$ is said to be a two dimensional random variable.

If the values of (X, Y) are finite or countably infinite, (X, Y) is called two dimensional discrete random variable and we can represent it as (X_i, Y_j) $\forall i = 1, 2, \ldots n \ldots$, $j = 1, 2, \ldots n \ldots$

If (X, Y) can assume all values in range R^2 in xy-plane then it is called a two dimensional continuous random variable.

1.6 Probability Mass Function of (x, y)

Consider two dimensional discrete random variable such that $P(X = x_i, Y = y_j) = p_{ij}$ then p_{ij} is known as probability mass function if

(i) $p_{ij} \geq 0 \ \forall \ i, j$ (ii) $\sum_j \sum_i p_{ij} = 1$.

The set $\{x_i, y_j, p_{ij}\}$ $i = 1, 2 \ldots m \ldots$; $j = 1, 2, \ldots n \ldots$ is called the joint probability distribution of (X, Y).

For example consider a program with two modules with module execution time being X and Y respectively, whose joint probability function is given as :

	Y = 1	Y = 2	Y = 3	Y = 4
X = 1	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$
X = 2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

Here (i) $p_{ij} \geq 0 \ \forall \ i = 1, 2$ and $j = 1, 2, 3, 4$

and (ii)
$$\sum_j \sum_i p_{ij} = \sum_j (p_{1j} + p_{2j})$$
$$= (p_{11} + p_{12} + p_{13} + p_{14}) + (p_{21} + p_{22} + p_{23} + p_{24})$$
$$= \left(\frac{1}{4} + \frac{1}{16} + \frac{1}{16} + \frac{1}{8}\right) + \left(\frac{1}{16} + \frac{1}{8} + \frac{1}{4} + \frac{1}{16}\right) = 2 \times \frac{1}{2} = 1$$

Hence above given is a probability mass function.

It is clear that each possible event $(X = x, Y = y)$ can be pictured as an event point on (x, y) coordinate system.

Example 31. Three balls are drawn at random without replacement from a box containing 2 white, 3 red and 4 black balls. If X denotes the number of white balls drawn and Y denotes the number of red balls drawn, find the joint probability distribution of (X, Y).

Solution : It is obvious from the question that $X = 0, 1, 2$ and $Y = 0, 1, 2, 3$

Now $P(X = 0, Y = 0) =$ P(3 balls are drawn out of which none is white or red)

$$= \text{P(all 3 black balls are drawn)} = \frac{{}^4C_3}{{}^9C_3} = \frac{1}{21}$$

and $P(X = 0, Y = 1) =$ P(1 red ball and 2 black balls are drawn)

$$= \frac{{}^4C_2 \times {}^3C_1}{{}^9C_3} = \frac{3}{14}$$

Similarly,

$$P(X=0, Y=2) = \frac{{}^3C_2 \times {}^4C_1}{{}^9C_3} = \frac{1}{7};\quad P(X=0, Y=0) = \frac{{}^3C_3}{{}^9C_3} = \frac{1}{84}$$

and

$$P(X=1, Y=0) = \frac{{}^2C_1 \times {}^4C_2}{{}^9C_3} = \frac{1}{7};\quad P(X=1, Y=1) = \frac{{}^2C_1 \times {}^3C_1 \times {}^4C_1}{{}^9C_3} = \frac{2}{7};$$

$$P(X=1, Y=2) = \frac{{}^2C_1 \times {}^3C_2}{{}^9C_3} = \frac{1}{14}$$

$P(X = 1, Y = 3) = 0$ (as only 3 balls are drawn)

and

$$P(X=2, Y=0) = \frac{{}^2C_2 \times {}^4C_1}{{}^9C_3} = \frac{1}{21};\quad P(X=2, Y=1) = \frac{{}^2C_2 \times {}^3C_1}{{}^9C_3} = \frac{1}{28}$$

$P(X = 2, Y = 2) = P(X = 2, Y = 3) = 0$ (as only 3 balls are drawn)

Hence the joint probability distribution of (X, Y) may be represented in the form of a table as given below :-

X \ Y	0	1	2	3
0	1/21	3/14	1/7	1/84
1	1/7	2/7	1/14	0
2	1/21	1/28	0	0

1.6.1 Cumulative Distribution Function

If (X, Y) is a a two dimensional discrete random variable, then

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

is called the cumulative distribution function (cdf) of (X, Y)

i.e., $$F(x,y) = \sum_{\substack{j \\ y_j \le y}} \sum_{\substack{i \\ x_i \le x}} p_{ij}$$

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

1.6.2 Marginal Probability Distribution

In situations where we are concerned with more than one random variable, the pmf/ probability distribution of a single variable is referred to as marginal pmf/ probability distribution. Hence if we consider two-dimensional random variable (x_i, y_j), i = 1, 2, . . ., m . . ., j = 1, 2, . . . n . . . then, the marginal probability function of X is defined as :-

$$P(X = x_i) = \sum_j p_{ij} = p_{i1} + p_{i2} + \ldots p_{in} + \ldots = p_{i*}$$

and the collection of pairs $\{x_i, p_{i*}\}$ i = 1, 2, ... m ... is called the marginal probability distribution of X.

Similarly, the marginal probability function of Y is defined as :-

$$P(Y = y_j) = \sum_i p_{ij} = p_{1j} + p_{2j} + \ldots + p_{mj} + \ldots = p_{*j}$$

and the collection of pairs $\{y_j, p_{*j}\}$, j = 1, 2, . . . n . . . is called the marginal probability distribution of Y.

1.6.3 Conditional Probability Distribution

Consider two dimensional discrete random variable (X, Y). The conditional probability function of X, given $Y = y_j$ is given by :-

$$P\left\{\frac{X = x_i}{Y = y_j}\right\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{p_{ij}}{p_{*j}}$$

and the collection of pairs $\{x_j, p_{ij}/p_{*j}\}$ i = 1, 2, 3 ... m ... is called the conditional probability distribution of X, given $Y = y_j$. Similarly, the conditional probability function of Y, given $X = x_i$ is given by :-

$$P\left\{\frac{Y = y_j}{X = x_i}\right\} = \frac{P\{X = x_i, Y = y_j\}}{p\{X = x_i\}} = \frac{p_{ij}}{p_{i*}}$$ and the collection of pairs $\left\{y_i, \frac{p_{ij}}{p_{i*}}\right\}$

$j = 1, 2, \ldots n \ldots$ is called the conditional probability distribution of Y given $X = x_i$.

1.6.4 Independent Random Variables

Let (X, Y) be a two dimonsional random variable such that

$$P\left\{\frac{X = x_i}{Y = y_j}\right\} = P(X = x_i), \quad \text{i.e.} \frac{p_{ij}}{p_{*j}} = p_{i*}$$

i.e., $p_{ij} = p_{i*} \cdot p_{*j} \quad \forall \; i,j$ then X and Y are said to be independent random variables.

Example 32. The joint probability mass function of (X, Y) is given by $p(x, y) = k(2x + 3y)$, $x = 0, 1, 2; \quad y = 1, 2, 3$. Find :-

(i) k

(ii) Marginal probability distribution of X

(iii) Marginal probability distribution of Y.

(iv) Conditional distribution of X given $Y = 1$.

(v) Conditional distribution of Y given $X = 2$.

(vi) The probability distribution of $(X + Y)$.

Solution : The joint probability distribution of (X, Y) can be represented in tabular form as :-

X \ Y	1	2	3
0	3k	6k	9k
1	5k	8k	11k
2	7k	10k	13k

(i) As above given is a pmf, hence

$$\sum_{j=1}^{3}\sum_{i=0}^{2} p(x_i, y_j) = 1 \Rightarrow 72k = 1 \Rightarrow k = \frac{1}{72}$$

(ii) $$P(X = 0) = p_{0*} = \sum_{j=1}^{3} p_{0j} = p_{01} + p_{02} + p_{03} = 18k = \frac{18}{72}$$

$$P(X = 1) = p_{1*} = \sum_{j=1}^{3} p_{1j} = p_{11} + p_{12} + p_{13} = 24k = \frac{24}{72}$$

$$P(X=2)=p_{2*}=\sum_{j=1}^{3} p_{2j}=p_{21}+p_{22}+p_{23}=30k=\frac{30}{72}$$

Hence marginal probability distribution of X is :-

$X=i$	p_i
0	18/72
1	24/72
2	30/72
Total	= 1

(iii) Similarly marginal probability distribution of Y is given by :-

Y=j	$p_{*j}=\sum_{i=1}^{2} p_{ij}$
1	$p_{01}+p_{11}+p_{21}=15k=\frac{15}{72}$
2	$p_{02}+p_{12}+p_{22}=24k=\frac{24}{72}$
3	$p_{03}+p_{13}+p_{23}=33k=\frac{33}{72}$
	Total = 1

(iv) Conditional distribution of X given Y =1 is $P\left\{\frac{X=i}{Y=1}\right\}=\frac{p_{i1}}{p_{*1}}=\frac{p_{i1}}{15k}$ (using iii). It can be represented in tabular form as :-

X=i	$\frac{p_{i1}}{p_{*1}} = \frac{p_{i1}}{15k}$
0	$\frac{p_{01}}{15k} = \frac{3k}{15k} = \frac{1}{5}$
1	$\frac{p_{11}}{15k} = \frac{5k}{15k} = \frac{1}{3}$
2	$\frac{p_{21}}{15k} = \frac{7k}{15k} = \frac{7}{15}$
	Total = 1

(v) Conditional distribution of Y given X=2 is :-

$$P\left\{\frac{Y=j}{X=2}\right\} = \frac{p_{2j}}{p_{2*}} = \frac{p_{2j}}{30k} \quad (\text{using (ii)})$$

It can be represented in tabular form as :-

Y=j	$\frac{p_{2j}}{p_{2*}} = \frac{p_{2j}}{30k}$
1	$\frac{p_{21}}{30k} = \frac{7k}{30k} = \frac{7}{30}$
2	$\frac{p_{22}}{30k} = \frac{10k}{30k} = \frac{10}{30}$
3	$\frac{p_{23}}{30k} = \frac{13k}{30k} = \frac{13}{30}$
	Total = 1

(vi) Let $Z = X + Y$

$X + Y$ can have values 1, 2, 3, 4, 5

Z can have values as shown below :-

X	0	0	1	0	1	2	1	2	2
Y	1	2	1	3	2	1	3	2	3
Z	1	2		3			4		5

hence $P(Z=1)=P(X=0,Y=1)=p_{01}=3k=\frac{3}{72}$

$$P(Z=2)=P(X=0,Y=2)+P(X=1,Y=1)=p_{02}+p_{11}=11k=\frac{11}{72}$$

Similarly :-

$$P(Z=3)=p_{03}+p_{12}+p_{21}=24k=\frac{24}{72}$$

$$P(Z=4)=p_{13}+p_{22}=21k=\frac{21}{72}$$

$$P(Z=5)=p_{23}=13k=\frac{13}{72}$$

Hence the required probability distribution is :-

$Z=X+Y$	1	2	3	4	5	Total
$p(X+Y=Z)$	$\frac{3}{72}$	$\frac{11}{72}$	$\frac{24}{72}$	$\frac{21}{72}$	$\frac{13}{72}$	1

1.7 Probability Density Function of (x, y)

If (X, Y) is a two dimensional continuous random variable such that

$$P\left\{x-\frac{dx}{2}\le X\le x+\frac{dx}{2} \text{ and } y-\frac{dy}{2}\le Y\le y+\frac{dy}{2}\right\}=f(x,y)dx\,dy$$

then f(x, y) is called the joint pdf of (X, Y) if

(i) $f(x,y)\ge 0 \ \forall \ -\infty<x<\infty, -\infty<y<\infty$

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

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

For example let the joint pdf of a two dimensional random variable be given by $f(x,y)=xy^2+\frac{x^2}{8}$, $0\le x\le 2$, $0\le y\le 1$. Then above given is a pdf as

(i) $f(x, y) \geq 0$

(ii) $$\int_0^1\int_0^2 f(x,y)dx\,dy = \int_0^1\int_0^2\left(xy^2 + \frac{x^2}{8}\right)dx\,dy$$

$$= \int_0^1\left[\frac{x^2y^2}{2} + \frac{x^3}{24}\right]_0^2 dy = \int_0^1\left(2y^2 + \frac{1}{3}\right)dy$$

$$= \left(2\frac{y^3}{3} + \frac{y}{3}\right)_0^1 = \frac{2}{3} + \frac{1}{3} = 1$$

1.7.1 Joint Distribution Function

If (X, Y) is a two dimensional continuous random variable then the joint distribution function $F(x,y) = P\{X \leq x \text{ and } Y \leq y\} = \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)dx\,dy$

Also at points of continuity of f(x, y) $\frac{\partial^2 F}{\partial x\,\partial y} = f(x,y)$.

***Example* 33.** Assume that the lifetime X and the brightness Y of a light bulb are being modeled as continuous random variables with joint pdf given by

$$f(x,y) = \lambda_1\lambda_2 e^{-(\lambda_1 x + \lambda_2 y)} \quad 0 < x < \infty, 0 < y < \infty.$$

Find the joint distribution function.

Solution : The joint distribution function is given by :-

$$F(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)dx\,dy = \int_0^y\left[\int_0^x \lambda_1\lambda_2 e^{-(\lambda_1 x + \lambda_2 y)}dx\right]dy$$

$$= \lambda_1\lambda_2\int_0^y e^{-\lambda_2 y}\left(\frac{e^{-\lambda_1 x}}{-\lambda_1}\right)_0^x dy$$

$$= \lambda_2\int_0^y e^{-\lambda_2 y}\left(1 - e^{-\lambda_1 x}\right)dy$$

$$= \lambda_2\left(1 - e^{-\lambda_1 x}\right)\left[\frac{e^{-\lambda_2 y}}{-\lambda_2}\right]_0^y$$

$$= \left(1 - e^{-\lambda_1 x}\right)\left(1 - e^{-\lambda_2 y}\right)$$

Hence $F(x,y) = \left(1 - e^{-\lambda_1 x}\right)\left(1 - e^{-\lambda_2 y}\right), 0 < x < \infty, 0 < y < \infty$

1.7.2 Marginal Density

Let (X, Y) be a two dimensional continuous random variable. Then the marginal density of X, $f_X(x)$ is defined as :-

$$f_X(x)dx = P\left\{x-\frac{dx}{2} \le X \le x+\frac{dx}{2}, -\infty < Y < \infty\right\}$$

$$= \int_{-\infty}^{\infty}\int_{x-\frac{dx}{2}}^{x+\frac{dx}{2}} f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} f(x,y)(x)_{x-\frac{dx}{2}}^{x+\frac{dx}{2}}\,dy$$

because f(x, y) may be treated as constant in the interval $\left[x-\frac{dx}{2}, x+\frac{dx}{2}\right]$

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

i.e., $$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

Also similarly marginal density of Y is :-

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

Remark 3 $$P(a \le X \le b) = P(a \le X \le b, -\infty < Y < \infty)$$

$$= \int_{-\infty}^{\infty}\int_{a}^{b} f(x,y)dx\,dy = \int_{a}^{b}\left[\int_{-\infty}^{\infty} f(x,y)dy\right]dx$$

$$= \int_{a}^{b} f_X(x)dx$$

Similarly $P(c \le Y \le d) = \int_{c}^{d} f_Y(y)dy$.

1.7.3 Conditional Density

Let (X, Y) be a two dimensional continuous random variable. Then the conditional density of X given Y denoted by f(x/y) is given by

$$f\left(\frac{x}{y}\right)dx = P\left\{\frac{x-\frac{dx}{2} \le X \le x+\frac{dx}{2}}{Y=y}\right\}$$

$$= \frac{P\left\{x-\frac{dx}{2} \le X \le x+\frac{dx}{2} \text{ and } y-\frac{dy}{2} \le Y \le y+\frac{dy}{2}\right\}}{P\left(y-\frac{dy}{2} \le Y \le y+\frac{dy}{2}\right)}$$

$$= \frac{f(x,y)dx\,dy}{P\left(-\infty < X < \infty, y-\frac{dy}{2} \le Y \le y+\frac{dy}{2}\right)} = \frac{f(x,y)dx\,dy}{f_Y(y)dy}$$

$$= \frac{f(x,y)}{f_Y(y)}dx$$

Hence $f\left(\frac{x}{y}\right) = \frac{f(x,y)}{f_Y(y)}$.

Similarly, the conditional density of Y given X denoted by f(y/x) , is given by :-

$$f\left(\frac{y}{x}\right) = \frac{f(x,y)}{f_X(x)}.$$

1.7.4 Independent Continuous Random Variables

Let (X, Y) be a two-dimensional continuous random variable then X and Y are said to be independent random variables if :-

$$f(x,y) = f_X(x) \times f_Y(y).$$

***Example* 34.** The joint pdf of the random variable (X, Y) is given by

$$f(x,y) = kxye^{-(x^2+y^2)}, \quad x>0, y>0\cdot$$

Find 'k' and prove also that X and Y are independent.

Solution : As above given is a pdf, hence

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dx\,dy = k\int_0^{\infty}\int_0^{\infty} xye^{-(x^2+y^2)}dx\,dy = 1$$

$$\Rightarrow \quad k\int_0^{\infty} ye^{-y^2}\left[\int_0^{\infty} xe^{-x^2}dx\right]dy = 1$$

$$\Rightarrow \quad k\int_0^{\infty} ye^{-y^2}\left[\frac{-e^{-x^2}}{2}\right]_0^{\infty} dy = 1 \Rightarrow \frac{k}{2}\int_0^{\infty} ye^{-y^2}dy = 1$$

$$\Rightarrow \quad \frac{k}{4}\left(-e^{-y^2}\right)_0^{\infty}=1 \Rightarrow \frac{k}{4}=1 \Rightarrow k=4$$

Marginal density of $X = f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$

$$= \int_0^{\infty} 4xye^{-(x^2+y^2)}dy = 4xe^{-x^2}\left[-\frac{e^{-y^2}}{2}\right]_0^{\infty}$$

$$= 2xe^{-x^2}, x>0.$$

Marginal density of $Y = f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$

$$= \int_0^{\infty} 4xye^{-(x^2+y^2)}dx = 4ye^{-y^2}\left(-\frac{e^{-x^2}}{2}\right)_0^{\infty} = 2ye^{-y^2}, y>0.$$

$$f_X(x) \times f_Y(y) = 4xye^{-(x^2+y^2)} = f(x,y), x>0, y>0$$

Hence X and Y are independent random variables.

Illustrative Examples

***Example* 35.** Given the joint probability density $f(x,y)=\begin{cases}\frac{2}{3}(x+2y), & 0<x<1, 0<y<1\\ 0, & \text{elsewhere}\end{cases}$. Find

(i) Marginal density of X and Y. (ii) Conditional density of X given Y = y

and use it to evaluate $P\left\{\frac{X\le 1/2}{Y=1/2}\right\}$ **(Raj. IV Sem CP-2006, RTU 2008)**

Solution : (i) Marginal densities of X and Y respectively are :-

$$f_X(x) = \int_0^{\infty} f(x,y)dy = \int_0^1 \frac{2}{3}(x+2y)dy = \frac{2}{2.3}\left(\frac{(x+2y)^2}{2}\right)_0^1$$

$$= \frac{1}{6}\left[(x+2)^2 - x^2\right] = \frac{1}{6}[4x+4] = \frac{2}{3}(x+1), \quad 0<x<1$$

and $$f_Y(y) = \int_0^\infty f(x,y)dx = \int_0^1 \frac{2}{3}(x+2y)dx = \frac{2}{3}\left[\frac{(x+2y)^2}{2}\right]_0^1$$

$$= \frac{1}{3}\left[(1+2y)^2 - 4y^2\right] = \frac{1}{3}(1+4y), \quad 0<y<1$$

(ii) Conditional density of X given Y = y is :

$$f\left(\frac{X}{Y}\right) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{2}{3}(x+2y)}{\frac{1}{3}(1+4y)} = \frac{2(x+2y)}{(1+4y)}, \quad 0<x<1$$

Hence, $$P\left[\frac{X \le 1/2}{Y = 1/2}\right] = \int_0^1 f\left(\frac{x}{y=1/2}\right)dx$$

$$= \int_0^{1/2} \frac{2}{3}(x+1)dx = \frac{2}{3}\left(\frac{(x+1)^2}{2}\right)_0^{1/2}$$

$$= \frac{1}{3}\left[\frac{9}{4} - 1\right] = \frac{1}{3} \times \frac{5}{4} = \frac{5}{12}.$$

Example 36. The joint probability mass function of two discrete random variates is given by

$$P(x,y) = \begin{cases} \frac{xy}{27}, & \text{for } x = 1, 2;\ y = 2, 3, 4 \\ 0, & \text{otherwise} \end{cases}$$

Prove that X and Y are independent.

Solution : $$p_{i*} = \sum_{j=2}^{4} p_{ij} = p_{i2} + p_{i3} + p_{i4}$$

$$\therefore \quad p_{1*} = p_{12} + p_{13} + p_{14} = \frac{9}{27};\ p_{2*} = p_{22} + p_{23} + p_{24} = \frac{18}{27}$$

Also $$p_{*j} = \sum_{x=1}^{2} p_{ij} = p_{1j} + p_{2j}$$

$$\therefore \quad p_{*2} = p_{12} + p_{22} = \frac{6}{27};\ p_{*3} = p_{13} + p_{23} = \frac{9}{27};$$

$$p_{*4} = p_{14} + p_{24} = \frac{12}{27}$$

Now $\quad p_{1*} \times p_{*2} = \frac{9}{27} \times \frac{6}{27} = \frac{2}{27} = p_{12}$

$$p_{1*} \times p_{*3} = \frac{9}{27} \times \frac{9}{27} = \frac{1}{9} = p_{13}$$

$$p_{1*} \times p_{*4} = \frac{9}{27} \times \frac{12}{27} = \frac{4}{27} = p_{14}$$

and $\quad p_{2*} \times p_{*2} = \frac{18}{27} \times \frac{6}{27} = \frac{4}{27} = p_{22}$

$$p_{2*} \times p_{*3} = \frac{18}{27} \times \frac{9}{27} = \frac{2}{9} = p_{23}$$

$$p_{2*} \times p_{*4} = \frac{18}{27} \times \frac{12}{27} = \frac{8}{27} = p_{24}$$

Hence X and Y are independent.

Exercise 1(B)

Q.1 1k RAM IC chips are purchased from two different semiconductor houses. Let X and Y denote the times to failure of the chips purchased from the two suppliers. The joint probability density of X and Y is estimated by :-

$$f_{X,Y}(x,y) = \begin{cases} \lambda\mu e^{-(\lambda x+\mu y)} & x>0, y>0 \\ 0 & \text{otherwise} \end{cases}.$$

Assume $\lambda = 10^{-5}$ per hour and $\mu = 10^{-6}$ per hour. Determine the probability that the time to failure is greater for chips characterized by X than it is for chips characterized by Y.

Q.2 Suppose that the two dimensional continuous random variable (X, Y) has joint pdf given by :-

$$f(x,y) = x^2 + \frac{1}{2}xy, \quad 0 \le x \le 1, 0 \le y \le 2$$
$$0, \qquad \text{elsewhere}$$

(i) Check that above given is pdf. Compute, (ii) $P(X+Y>1)$, (iii) $P\left(X>\frac{1}{2}\right)$,

(iv) $P(Y<X)$, (v) $P\left(\frac{Y<\frac{1}{2}}{X<\frac{1}{2}}\right)$

Ans. (ii) $\frac{65}{12}$ (iii) $\frac{5}{6}$ (iv) $\frac{7}{24}$ (v) $\frac{5}{3}$

Q.3 A computer centre owner wants to purchase 3 computers from a sale in which there are 3 new computers, 4 used but still working and 5 defective but can be made in working condition after minor repair. If X and Y denote new and used but still working computers in the lot of 3 purchased computers.

Find (i) the joint probability mass function of X and Y.

(ii) the marginal probability of X and Y

(iii) the conditional probability of X given Y =1.

Ans. (i) $P(X=x, Y=y) = \frac{{}^3C_x\, {}^4C_y\, {}^5C_{3-x-y}}{220}$, X, Y = 0, 1, 2, 3 and $0 \le X+Y \le 3$

(ii) $P(X=x) = \sum_{y=0}^{3} \frac{{}^3C_x\, {}^4C_y\, {}^5C_{3-x-y}}{220}$

$P(Y=y) = \sum_{x=0}^{3} \frac{{}^3C_x\, {}^4C_y\, {}^5C_{3-x-y}}{220}$

(iii) $P\left(\frac{X=x}{Y=1}\right) = \frac{{}^3C_x\, {}^5C_{2-x}}{\sum_{x=0}^{3} {}^3C_x\, {}^5C_{x-2}}$

Q.4 The joint probability density function of two continuous random variates (X, Y) is given by :-

$$f(x,y) = \begin{cases} 24xy & \text{for } x \ge 0, y \ge 0, x+y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Are X and Y independent ?

Q.5 Let the joint pmf of X and Y be given as P(0, 0) = 0.4, P(0, 1) = 0.2, P(1, 0) = 0.1, P(1, 1) = 0.3.

Find the conditional probability mass function of X given that Y = 1.

Also find the marginal probability function of X and Y of the given random variables.

Ans. $P\left(\frac{X}{Y=1}\right) = \frac{(2+x)}{5}, x = 0, 1; P(X) = \begin{cases} 1/7, & x=0 \\ 1/3, & x=1 \\ 11/21 & x=2 \end{cases}$ and $P(Y) = \begin{cases} 1/7, & \text{for } y=0 \\ 3/14, & \text{for } y=1 \\ 2/7, & \text{for } y=2 \\ 5/14, & \text{for } y=3 \end{cases}$

Q.6 Suppose that 2 cards are dealt from a pack of 52 without replacement. Let the random variables X and Y be the number of aces and queens that occur. Are X and Y independent ?

Ans. No

Q.7 If the joint pdf of (X, Y) is given by :

$$f(x, y) = 2, \quad 0 < x < 1, \quad 0 < y < x$$
$$= 0, \quad \text{otherwise} .$$

Find the marginal density function of X and Y.

Ans. $f_X(x) = 2x,\ 0 < x < 1;\ f_Y(y) = 2(1 - y),\ 0 < y < 1.$

Q.8 If the joint pdf of (X, Y) is given by $f(x, y) = k,\ 0 \le x < y \le 2$, find k and also the marginal and conditional density functions.

Ans. $k = \frac{1}{2},\ f_x(x) = \frac{1}{2}(2 - x),\ 0 \le x \le 2;$

$f_y(y) = \frac{1}{2}y,\ 0 \le y \le 2;\ f(x,y) = \frac{1}{y}\quad 0 < x < y;$

$f\left(\frac{y}{x}\right) = \frac{1}{2 - x}\quad x < y < 2$

Q.9 Given the joint probability density

$$f(x,y) = \begin{cases} kxy, & \text{for}, 0 < x < 1, 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases},$$

find k. Also find the marginal densities of X and Y and the conditional density of X given Y = y.

Ans. $k = 4,\ f_X(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases},\ f_Y(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$

and $f\left(\frac{x}{y}\right) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$

Q.10 The joint probability density function of (X, Y) is given by $f(x, y) = e^{-(x+y)}$ for $x \ge 0$; $y \ge 0$. Are X and Y independent ? Give reasons.

Ans. Yes

1.8 Expectation

The expectation of a random variable X is defined as :-

$$\overline{X} = E(X) = \begin{cases} \sum_i x_i p_i & \text{if X is discrete RV with pmf } p_i \\ \int_{-\infty}^{\infty} x f(x)dx & \text{if X is continuous RV with pdf } f(x) \end{cases}$$

provided the relevant sum or integral is absolutely convergent. If X is a random variable and g(X) is any function of X then.

$$E[g(X)] = \begin{cases} \sum_i g(x_i) p_i & \text{if X is discrete RV with } P(X = x_i) = p_i \\ \int_{-\infty}^{\infty} g(x) f(x)dx & \text{if X is continuous RV.} \end{cases}$$

Remark 4 If we take $g(X) = X^r$, r is a positive integer, then

$$E[X^r] = \begin{cases} \sum_i x_i^r p_i & \text{if X is discrete RV} \\ \int_{-\infty}^{\infty} x^r f(x)dx & \text{if X is continuous RV.} \end{cases}$$

Theorem 1 If X is a random variable and a and b are constants then

$$E(aX + b) = aE(X) + b,$$

provided all the expectations exist.

Proof :- By definition, we have

$$E(aX+b) = \int_{-\infty}^{\infty} (ax+b)f(x)dx \quad \text{(considering X as continuous RV)}$$

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

$$= aE(X) + b \qquad \left(\text{as} \int_{-\infty}^{\infty} f(x)dx = 1\right)$$

Remark 2 If $b = 0$ we get $E(aX) = aE(X)$.

Remark 3 If $a = 1$, $b = -\overline{X}$ then $E(X - \overline{X}) = \overline{X} - \overline{X} = 0$.

Remark 5 Let $g(X) = aX + b$

then $g[E(X)] = aE(X) + b$

Also $E[g(X)] = E(aX + b) = aE(X) + b$

Equating the above two relations we get :-

$$E[g(X)] = g[E(X)]$$

This implies that expectation of a linear function is the same as linear function of the expectation.

***Example* 37.** Let X be a random variable with the following probability distribution

x	:	–3	6	9
p(X=x)	:	1/6	1/2	1/3

Find E(X), $E(X^2)$, $E(2X + 1)^2$.

Solution :

$$E(X) = \sum xP(x) = -3\times\frac{1}{6}+6\times\frac{1}{2}+9\times\frac{1}{3}=\frac{11}{12}$$

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

Also $$E(2X+1)^2 = E\left[4X^2+4X+1\right] = 4E(X^2)+4E(X)+1$$

$$=4\times\frac{93}{2}+4\times\frac{11}{2}+1=209$$

***Example* 38.** Find the expectation of the number on a dice when thrown.

Solution : Let X be the random variable which represents the number on dice when thrown, then its probability distribution is :-

X	1	2	3	4	5	6
P(X)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\therefore \quad E(X) = 1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=\frac{7}{2}=3.5$$

***Example* 39.** Two unbiased dice are thrown. Find the expected value of the sum of number of points on them.

Solution : Let X denotes the sum of number of points obtained on throwing two dice. It's probability distribution is :-

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

Hence $E(X) = \sum_i x_i p_i$

$$= 2\times\frac{1}{36} + 3\times\frac{2}{36} + 4\times\frac{3}{36} + 5\times\frac{4}{36} + 6\times\frac{5}{36} + 7\times\frac{6}{36}$$

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

$$= \frac{1}{36}\times 252 = 7$$

This can also be solved by using addition theorem of expectation.

Addition Theorem of Expectation

The mathematical expectation of the sum of n random variables is equal to the sum of their expectations.

Symbolically if $X_1, X_2, \ldots X_n$ are random variables then

$$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$$

or
$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i),$$

provided all expectations exist.

In particular if $n = 2$ we have

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

Multiplication Theorem of Expectation

The mathematical expectation of the product of a number of independent random variables is equal to the product of their expectations. Symbolically if $X_1, X_2, \ldots, X_n$ are n independent random variables, then

$$E(X_1 X_2 \ldots X_n) = E(X_1)\, E(X_2) \ldots E(X_n)$$

or
$$E\left(\prod_{i=1}^{n} E(X_i)\right) = \prod_{i=1}^{n} E(X_i)$$

provided all expectations exist.

In particular if $n = 2$, then

$$E(X_1 X_2) = E(X_1)\, E(X_2).$$

Example **40.** What is the expectation of the (i) sum of points (ii) product of points obtained in a random throw of n dice ?

Solution : Let X_i denote the number obtained on throw of ith dice then, from Example 2 of this chapter it is clear that $E(X_i) = \frac{7}{2}$

(i) Addition theorem for expectation gives :-

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{7}{2} = \frac{7}{2}\sum_{i=1}^{n} = \frac{7n}{2}$$

(ii) Multiplication theorem for expectation gives :-

$$E(X_1 X_2 \ldots X_n) = E(X_1)\, E(X_2) \ldots E(X_n)$$

$$= \frac{7}{2} \times \frac{7}{2} \ldots \text{n times}$$

$$= \left(\frac{7}{2}\right)^n.$$

***Example* 41.** A coin is tossed until a head appears, what is the expectation of the number of tosses required ?

Solution : Let X denote the number of tosses required to get the first head. P(head) = P(tail) = 1/2. Then its probability distribution is :-

Event :	H	TH	TTH	TTTH	-----
X :	1	2	3	4	-----
P(X=x) :	$\frac{1}{2}$	$\left(\frac{1}{2}\right)^2$	$\left(\frac{1}{2}\right)^3$	$\left(\frac{1}{2}\right)^4$	-----

Hence $E(X) = 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 4 \times \frac{1}{16} + \ldots$

Let $E(X) = S = 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 4 \times \frac{1}{16} + \ldots$

then $\frac{1}{2}S = 1 \times \frac{1}{4} + 2 \times \frac{4}{8} + 3 \times \frac{1}{16} \ldots$

Subtracting both the series we get :-

$$\left(1-\frac{1}{2}\right)S = 1\times\frac{1}{2}+1\times\frac{1}{4}+1\times\frac{1}{8}+1\times\frac{1}{16}\cdots$$

$$\Rightarrow \quad \frac{1}{2}S = \frac{1}{2}\left(1+\frac{1}{2}+\frac{1}{4}\cdots\right) = \frac{1}{2}\times\frac{1}{1-\frac{1}{2}}$$

$$\Rightarrow \quad S = \frac{1}{1-\frac{1}{2}} = \frac{1}{\frac{1}{2}} = 2$$

Hence $E(X) = 2$.

Example 42. A doctor recommends a patient to go on a particular diet for two weeks and there is equal likelihood for the patient to lose his weight between 2kgs and 4kgs. What is the average amount the patient is expected to lose on this diet ?

Solution : Let random variable X denote the loss in weight in kg. with pdf

$$f(x) = \begin{cases} a, & 2<x<4 \\ 0, & \text{otherwise} \end{cases}$$

As $f(x)$ is a pdf, hence $\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow \int_{2}^{4} a\,dx = 1 \Rightarrow a(x)_2^4 = 1 \Rightarrow a = \frac{1}{2}$

Hence $f(x) = \begin{bmatrix} 1/2, & 2<x<4 \\ 0, & \text{otherwise} \end{bmatrix}$

and $E(x) = \int_{-\infty}^{\infty} xf(x)dx = \int_{2}^{4} x\frac{1}{2}dx = \left(\frac{x^2}{4}\right)_2^4 = 3\text{kg}$

Hence the patient is expected to loose 3kg on this diet.

1.9 Variance

Let X be a discrete random variable with pmf $P(X = x_i) = p_i$ then variance of X is defined as $Var(X) = \sigma^2 = E\left((X_i - \overline{X})^2\right)$ where $\overline{X}$ is the mean of random variable X.

Now $Var(X) = \sigma^2 = E\left((X_i - \overline{X})^2\right) = \sum_i (x_i - \overline{X})^2 p_i$

$$= \sum_i x_i^2 p_i - 2\overline{x}\sum_i x_i p_i + \overline{X}^2 \quad (\text{as } \Sigma\, p_i = 1)$$

$$= \sum_i x_i^2 p_i - 2\overline{X}^2 + \overline{X}^2 \quad \left(\text{as} \sum x_i p_i = \overline{X}\right)$$

$$= \sum_i x_i^2 p_i - \overline{X}^2$$

$$\Rightarrow \quad \sigma^2 = E\left(X^2\right) - [E(X)]^2 \qquad \left[\because E\left(X^2\right) = \sum_i x_i^2\, p_i \text{ and } E(X) = \overline{X}\right]$$

This simplified formula holds equally good for continuous random variables. Hence if X is a continuous random variable

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \overline{X})^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \left[\int_{-\infty}^{\infty} xf(x)dx\right]^2$$

Hence, for a random variable X :

$$\text{Var}(aX) = E\left(a^2X^2\right) - [E(aX)]^2$$

$$= a^2E\left(X^2\right) - [aE(X)]^2 \qquad [\because E(aX) = aE(X)]$$

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

Hence $\text{Var}(a\,X) = a^2\ \text{Var}\,(X)$.

***Example* 43.** Two cards are drawn successively with replacement from a well shuffled pack of 52 cards, find the mean and variance of number of aces.

Solution : Let random variable X denote the number of aces. Hence $X = 0, 1, 2$.

$$P(\text{drawing an ace}) = \frac{4}{52} = \frac{1}{13}; \quad P(\text{not drawing an ace}) = \frac{12}{13}$$

$$\therefore \quad P(X = 0) = \frac{12}{13} \times \frac{12}{13} = \frac{144}{169}$$

$$P(X = 1) = \frac{1}{13} \times \frac{12}{13} + \frac{12}{13} \times \frac{1}{13} = \frac{24}{169}$$

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

Hence the required distribution is :-

X :	0	1	2
P(X = x) :	$\frac{144}{169}$	$\frac{24}{169}$	$\frac{1}{169}$

Now $E(X)=\frac{24}{169}+\frac{2}{169}=\frac{26}{169}$

$$E(X^2)=\frac{24}{169}+\frac{4}{169}=\frac{28}{169}$$

$\therefore$ $Var(X)=E(X^2)-[E(X)]^2$

$$=\frac{28}{169}-\left(\frac{26}{169}\right)^2=\frac{4732-676}{169\times169}=\frac{4056}{169\times169}$$

$$=\frac{24}{169}$$

Hence $E(X)=\frac{26}{169}$ and $VarX=\frac{24}{169}$

$\Rightarrow$ Standard deviation $=\sqrt{VarX}=\frac{2\sqrt{6}}{13}$.

***Example* 44.** A random variable X has the following probability distribution :-

x_i :	−2	−1	0	1	2	3
p_i :	0.1	k	0.2	2k	0.3	k

(i) Calculate the mean of X.

(ii) Variance of X.

Solution : As above given is probability distribution hence $\sum_i p_i=1 \Rightarrow 0.6+4k=1 \Rightarrow k=0.1$

Hence we construct the following :

x_i	p_i	$p_i x_i$	$p_i x_i^2$
–2	0.1	–0.2	0.4
–1	0.1	–0.1	0.1
0	0.2	0	0
1	0.2	0.2	0.2
2	0.3	0.6	1.2
3	0.1	0.3	0.9
Total		$\sum p_i x_i = 0.8$	$\sum p_i x_i^2 = 2.8$

$$\therefore \quad E(X) = \sum p_i x_i = 0.8$$

$$E(X^2) = \sum p_i x_i^2 = 2.8$$

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= (2.8)^2 - (0.8)^2 = 2.16$$

1.10 Measures of Central Tendency and Dispersion, Skewness and Kurtosis

Before continuing, we first briefly revise some of the statistical definitions which would be useful in better understanding of the latter part of the chapter.

1.10.1 Measures of Central Tendency

They give us an idea about the concentration of the values in the central part of the distribution. Some commonly used measures of central tendency are :-

(i) Mean, (ii) Median, (iii) Mode.

(i) Mean (denoted as $\bar{x}$)

(a) Arithmetic mean or simply mean of n observations $x_1, x_2, \ldots x_n$ is defined as

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$$

(b) In case of frequency distribution

x :	x_1	x_2	...	x_i	...	x_n
f :	f_1	f_2	...	f_i	...	f_n

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i\, x_i}{\sum_{i=1}^{n} f_i}$$

(c) In case of frequency distribution as mentioned above if we take deviations d_i from any arbitrary point A, i.e. $d_i = x_i - A$

then $$\bar{x} = A + \frac{\sum_{i=1}^{n} f_i\, d_i}{\sum_{i=1}^{n} f_i}.$$

Remark 6 If instead of the frequency distribution x_i/f_i, $i = 1, 2, \ldots n$ we are given probability distribution x_i/f_i, $i = 1, 2, \ldots, n$ then $\sum f_i = 1$ and $\bar{x} = \sum f_i\, x_i$ or $= \sum p_i\, x_i$ (as here $p_i = f_i$).

(ii) Median

Median of a distribution is the value of the variable which divides it into two equal parts. Hence it is a positional average.

For example

(a) Median of 25, 20, 15, 35, 18 i.e., of 15, 18, 20, 25, 35 is 20, while

(b) Median of 8, 20, 50, 25, 15, 30 i.e., of 8, 15, 20, 25, 30, 50 is $\frac{1}{2}(20 + 25) = 22.5$

(iii) Mode

Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster densely. Hence in case of discrete frequency distribution, mode is the value of x corresponding to maximum frequency.

Measures of central tendency are inadequate to give us a complete idea of a distribution. For example, all the series (i) 7, 8, 9, 10, 11 (ii) 3, 6, 9, 12, 15 and (iii) 1, 5, 9, 13, 17 have same mean. Hence if we are given mean of a distribution then we cannot determine which of the above distribution (i), (ii) or (iii) must be taken. Hence measures of central tendency must be supported and supplemented by some other measures.

1.10.2 Measures of Dispersion

Dispersion means scatteredness. Hence it gives us an idea whether the series is less dispersed or more scattered. Some common measures of dispersion are :-

(i) Mean Deviation

(ii) Standard Deviation

(i) Mean Deviation

For a given frequency distribution x_i/f_i, $i = 1, 2, \ldots, n$ the mean deviation from average A, is $= \dfrac{1}{\sum_{i=1}^{n} f_i} \sum_{i=1}^{n} f_i |x_i - A|$.

(ii) Standard Deviation (Denoted as σ)

For a given frequency distribution x_i/f_i, $i = 1, 2, \ldots, n$.

$$\sigma = \sqrt{\frac{1}{\sum_i f_i} \sum_i f_i (x_i - \bar{x})^2} = \sqrt{\frac{\Sigma f_i x_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i x_i}{\Sigma f_i}\right)^2} = \sqrt{\frac{\Sigma f_i x_i^2}{\Sigma f_i} - (\bar{x})^2}$$

The square of the standard deviation is called as variance and is given by

$$\sigma^2 = \frac{1}{\sum_i f_i} \sum_i f_i (x_i - \bar{x})^2 .$$

1.10.3 Skewness

Skewness means lack of symmetry. We study it to have an idea about the shape of the curve. A distribution is said to be skewed if mean ≠ median ≠ mode or if the curve drawn with the help of given data is not symmetrical but stretched more to one side than the other.

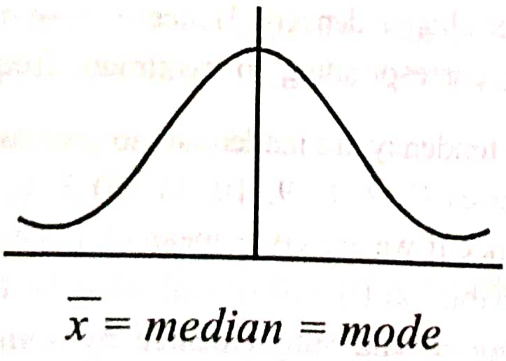

Figure 1.5 : Symmetrical Distribution

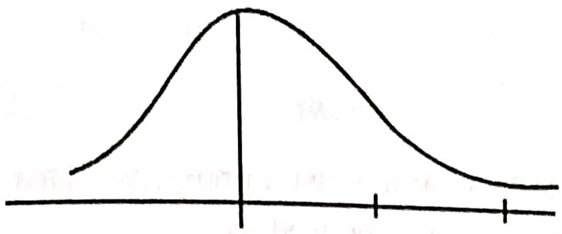

Figure 1.6 Positively Skewed Distribution

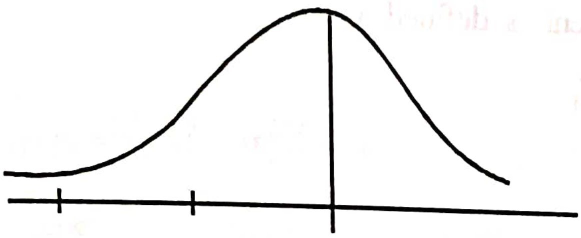

Figure 1.7 Negatively Skewed Distribution

Even after knowing all the above measures, we should know one more measure to form a complete idea about the distribution which is called by Karl Pearson as kurtosis.

1.10.4 Kurtosis

Kurtosis gives as an idea about the convexity of the curve. It enables us to determine whether the curve is a flat curve or a peaked curve.

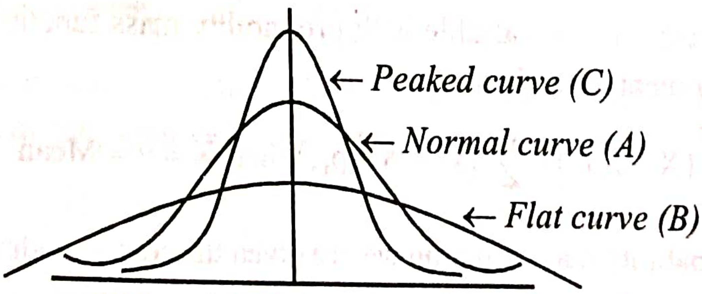

Figure 1.8

Curve (A) which is neither flat nor peaked is called the normal curve or *mesokurtic curve.* Curve (B) which is flatter than the normal curve is known as *platykurtic curve* and curve (C) which is more peaked than the normal curve is called *leptokurtic curve.*

1.11 Moments

Moment in statistics is analogous to the same term in statics according to which moment is the tendency of a force to rotate a body about a point. The concept of moment for a random variable X is also extended to the higher powers of the variable X of a given frequency distribution. In this section we give the statistical definition of the term moment.

1.11.1 Moment About Origin

Let X be a discrete random variable with probability mass function $P(X = x_i) = p_i$, then rth order moment about origin is defined as :-

$$\mu'_r = E(X^r) = \sum_i x_i^r \, p_i$$

If instead of pmf we are given the corresponding frequency distribution x_i/f_i, $i = 1, 2, \ldots, n$ then rth order moment is defined as

$$\mu'_r = \frac{\sum_i x_i^r f_i}{\sum_i f_i}$$

If X is a continuous random variable with pdf f(x) then $\mu'_r = \int_{-\infty}^{\infty} x^r f(x)dx$.

In particular for r = 1, $\mu'_1 = \sum_i x_i \, p_i = \int_{-\infty}^{\infty} xf(x)dx$, which is mean.

Hence mean is first order moment about origin.

Also $\quad r = 0 \Rightarrow \quad \mu'_0 = 1$

and $\quad \mu'_1 = \overline{X} = \mu$.

1.11.2 Moment About Mean or Central Moment

Let X be a discrete random variable with probability mass function $P(X = x_i) = p_i$ then rth order central moment is defined as

$$\mu_r = E\left[(X - \overline{X})^r\right] = \sum_i (x_i - \overline{X})^r p_i, \text{ when } \overline{X} = \mu = \text{Mean}$$

If instead of probability mass function we are given the corresponding frequency distribution x_i/f_i, $i = 1, 2, \ldots, n$ then rth order central moment is defined as :-

$$\mu_r = \frac{\sum_i (x_i - \overline{X}) f_i}{\sum_i f_i}$$

If X is a continuous random variable with pdf f(x) then $\mu_r = \int_{-\infty}^{\infty} (X - \overline{X})^r f(x)dx$.

In particular $\quad r = 0 \Rightarrow \mu_0 = 1$

$$r = 1 \Rightarrow \mu_1 = \sum_i x_i \, p_i - \overline{X} = \overline{X} - \overline{X} = 0$$

$$r = 2 \Rightarrow \mu_2 = \sum_i (x_i - \overline{X})^2 p_i = \sigma^2 = \text{var}(X).$$

We now find the interrelation between moment about origin (μ_r') and moment about mean (μ_r) considering X as discrete random variable, which would be valid in all cases.

Now $\mu_2 = \sum_i (x_i - \overline{X})^2 p_i = \sum_i x_i^2 p_i - 2\overline{X}\sum_i p_i x_i + \overline{X}^2$

$$= \sum_i x_i^2 p_i - 2\overline{X}^2 + \overline{X}^2 = \sum_i x_i^2 p_i - \overline{X}^2$$

$$= \mu_2' - \mu_1'^2 \qquad \text{..... (1.1)} \quad (\text{as } \mu_1' = \overline{X})$$

Also $\mu_3 = \sum_i (x_i - \overline{X})^3 p_i = \sum_i (x_i^3 - 3x_i^2\overline{X} + 3x_i\overline{X}^2 - \overline{X}^3)p_i$

$$= \sum_i x_i^3 p_i - 3\overline{X}\sum x_i^2 p_i + 3\overline{X}^2 \sum x_i p_i - \overline{X}^3 \sum p_i$$

$$= \mu_3' - 3\mu_1'\mu_2' + 3\mu_1'^2\mu_1' - \mu_1'^3 \quad \left(\because \sum p_i = 1 \text{ and } \overline{X} = \mu_1'\right)$$

$\Rightarrow$ $\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3$ .....(1.2)

Similarly we get

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 \qquad \text{.....(1.3)}$$

and so on.

1.11.3 Moment About An Arbitrary Point

Let X be a discrete random variable with probability mass function $P(X = x_i) = p_i$ then rth order moment about the point 'a' is defined as

$$\mu_r'' = E\left[(X-a)^r\right] = \sum_i (x_i - a)^r p_i .$$

If instead of probability distribution we are given the corresponding frequency distribution x_i/f_i, $i = 1, 2, \ldots n$ then rth order moment about the point x=a is defined as :-

$$\mu_r'' = \frac{\sum_i (x_i - a)^r f_i}{\sum_i f_i}$$

If X is a continuous random variable with pdf f(x) then

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

In particular $\quad r = 0 \quad \Rightarrow \quad \mu_0'' = 1$

$$r = 1 \quad \Rightarrow \quad \mu_1'' = \sum_i (x_i - a)p_i = \overline{X} - a$$

Now, we find the interrelation between moment about the point $x = a$ (μ_r'') and the moment about mean (μ_r).

$$\mu_2 = \sum_i (x_i - \overline{X})^2 p_i = \sum_i \left[(x_i - a) - (\overline{X} - a)\right]^2 p_i$$

$$= \sum_i ((x_i - a) - \mu_1'')^2 p_i \quad [\because \mu_1'' = \overline{X} - a]$$

$$= \sum_i (x_i - a)^2 p_i - 2\mu_1'' \sum (x_i - a)p_i + \mu_1''^2$$

$$= \mu_2'' - 2\mu_1'' \times \mu_1'' + \mu_1''^2 = \mu_2'' - \mu_1''^2$$

$$\Rightarrow \quad \mu_2 = \mu_2'' - \mu_1''^2 \qquad \text{.....(1.4)}$$

Similarly we can obtain :-

$$\mu_3 = \mu_3'' - 3\mu_2'' \mu_1'' + 2\mu_1''^3 \qquad \text{.....(1.5)}$$

$$\mu_4 = \mu_4'' - 4\mu_3'' \mu_1'' + 6\mu_2'' \mu_1''^2 - 3\mu_1''^4 \qquad \text{.....(1.6)}$$

and so on.

Also the moment about a point $x = a$ (μ_r'') can be written in terms of central moment as :-

$$\text{Also} \quad \mu_1'' = \sum_i (x_i - a)p_i = \overline{X} - a \qquad \text{.....(1.7)}$$

$$\mu_2'' = \sum_i (x_i - a)^2 p_i = \sum_i \left[(x_i - \overline{X}) + (\overline{X} - a)\right]^2 p_i$$

$$= \sum_i (x_i - \overline{X})^2 p_i + 2(\overline{X} - a) \sum_i (x_i - \overline{X})p_i + (\overline{X} - a)^2$$

$$= \mu_2 + 2\mu_1 \mu_1'' + \mu_1''^2$$

or, $\quad \mu_2'' = \mu_2 + \mu_1''^2 \quad (\text{as } \mu_1 = 0)$ .....(1.8)

Similarly

$$\mu_3'' = \sum_i (x_i - a)^3 p_i = \sum_i \left[(x_i - \bar{X}) - (\bar{X} - a)\right]^3 p_i$$

$$= \sum_i (x_i - \bar{X})^3 p_i + 3(\bar{X} - a)\sum_i (x_i - \bar{X})^2 p_i + 3(\bar{X} - a)^2 \sum_i (x_i - \bar{X}) p_i + (\bar{X} - a)^3$$

$$= \mu_3 + 3\mu_1''\mu_2 + 3\mu_1''^2 \mu_1 + \mu_1''^3$$

$\Rightarrow \quad \mu_3'' = \mu_3 + 3\mu_1''\mu_2 + \mu_1''^3$ .....(1.9)

and similarly we get :-

$$\mu_4'' = \mu_4 + 4\mu_3 \mu_1'' + 6\mu_2 \mu_1''^2 + \mu_1''^4 \quad .....(1.10)$$

and so on.

1.12 Karl Pearson β and γ Coefficients

Karl pearson defined the following four coefficients based upon the first four central moments as :-

(i) **β Coefficients**

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}; \ \beta_2 = \frac{\mu_4}{\mu_2^2}$$

(ii) **γ Coefficients**

$$\gamma_1 = +\sqrt{\beta_1}\ , \ \gamma_2 = \beta_2 - 3$$

The practical utility of these coefficients is to measure skewness (1.10.3) and Kurtosis (1.10.4).

If $\beta_1 = 0$, the curve is symmetrical. Hence β_1 can be taken as measure of skewness and β_2 is taken as the measure of kurtosis.

If $\quad \beta_2 = 3,\ \gamma_2 = 0$ curve is called mesokurtic curve.

If $\quad \beta_2 > 3,\ \gamma_2 > 0$ curve is called leptokurtic curve.

If $\quad \beta_2 < 3,\ \gamma_2 < 0$ curve is called platykurtic curve.

Example 45 The first four moments of a distribution about the value 5 are –4, 22, –117 and 560, obtain the moments about (i) mean and (ii) origin.

Solution : Given $\mu_1'' = -4, \mu_2'' = 22, \mu_3'' = -117, \mu_4'' = 560, a = 5$

From sec 3.5.3, equation (3.7)-equation (3.10) it is clear that :-

$$\mu_1'' = \overline{X} - a \Rightarrow -4 = \overline{X} - 5 \Rightarrow \overline{X} = 1$$

Hence Mean $= \mu_1' = 1$

Further moments about mean are :-

$$\mu_1 = 0$$

$$\mu_2 = \mu_2'' - \mu_1''^2 = 22 - 16 = 6$$

$$\mu_3 = \mu_3'' - 3\mu_2''\mu_1'' + 2\mu_1''^3 = -117 - 3(22)(-4) + 2(-4)^3$$

$$= -117 + 264 - 128 = 19$$

and $$\mu_4 = \mu_4'' - 4\mu_3''\mu_1'' + 6\mu_2''\mu_1''^2 - 3\mu_1''^4$$

$$= 560 - 4(-117)(-4) + 6(22)(-4)^2 - 3(-4)^2$$

$$= 560 - 1872 + 2112 - 768 = 32$$

Hence $\mu_1 = 0$, $\mu_2 = 6$, $\mu_3 = 19$, $\mu_4 = 32$.

Again using formula's in sec 3.5.2, equation (3.1)-equation (3.5) and the above central moments we get moments about mean as :-

$$\mu_2 = \mu_2' - \mu_1'^2$$

$\Rightarrow$ $$\mu_2' = \mu_2 + \mu_1'^2 = 6 + 1 = 7 \quad (\mu_1' = 1 \text{ as calculated above})$$

and $$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3$$

$\Rightarrow$ $$\mu_3' = \mu_3 + 3\mu_1'\mu_2' - 2\mu_1'^3$$

$$= 19 + 3(1)(7) - 2(1)^3$$

$$= 19 + 21 - 2 = 38$$

Similarly

$$\mu_4' = \mu_4 + 4\mu_3'\mu_1' - 6\mu_2'\mu_1'^2 + 3\mu_1'^4$$

$$= 32 + 4(38)(1) - 6(7)(1) + 3(1)$$

$$= 32 + 152 - 42 + 3 = 145$$

Hence $\mu_1' = 1, \mu_2' = 7, \mu_3' = 38, \mu_4' = 145$.

Example 46 **Calculate the first four moments about the mean of the following distribution. Also calculate β_1 and β_2.**

x :	1	2	3	4	5	6	7	8	9
y :	1	6	13	25	30	22	9	5	2

Solution : Let us first find the moment about the point x = 5.

x_i	y_i	x_i-5	$y_i(x_i-5)$	$y_i(x_i-5)^2$	$y_i(x_i-5)^3$	$y_i(x_i-5)^4$
1	1	–4	–4	16	–64	256
2	6	–3	–18	54	–162	486
3	13	–2	–26	52	–104	208
4	25	–1	–25	25	–25	25
5	30	0	0	0	0	0
6	22	1	22	22	22	22
7	9	2	18	36	72	144
8	5	3	15	45	135	405
9	2	4	8	32	128	512
Total	113	0	–10	282	2	2058

Now for the given frequency distribution :-

$$\mu_1'' = \frac{\sum y_i(x_i-5)}{\sum y_i} = \frac{-10}{113} = -0.088$$

$$\mu_2'' = \frac{\sum y_i(x_i-5)^2}{\sum y_i} = \frac{282}{113} = 2.496$$

$$\mu_3'' = \frac{\sum y_i(x_i-5)^3}{\sum y_i} = \frac{2}{113} = 0.018$$

$$\mu_4'' = \frac{\sum y_i(x_i-5)^4}{\sum y_i} = \frac{2058}{113} = 18.212$$

Now $\mu_1 = 0$

$$\mu_2 = \mu_2'' - \mu_1''^2 = 2.496 - (-0.088)^2 = 2.488$$

$$\mu_3 = \mu_3'' - 3\mu_2''\mu_1'' + 2\mu_1''^3$$

$$= 0.018 - 3(2.496)(-0.088) + 2(-0.088)^3$$

$$= 0.018 + 0.658944 - 0.001363 = 0.675581$$

$$\mu_4 = \mu_4'' - 4\mu_3''\mu_1'' + 6\mu_2''\mu_1''^2 - 3\mu_1''^4$$

$$= 18.212 - 4(0.018)(-0.088) + 6(2.496)(-0.088)^2 - 3(-0.088)^4$$

$= 18.212 + 0.006336 + 0.115974 - 0.00018$

$= 18.33413$

Hence $\mu_1 = 0,\ \mu_2 = 2.5,\ \mu_3 = 0.68,\ \mu_4 = 18.3$

Also $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{0.4564097}{15.625} = 0.029 \cong 0.03$

and $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{18.3}{6.25} = 2.928 \cong 3.$

***Example* 47** Calculate the first four moments about mean for the following distribution and also hence β_1 and β_2.

x :	0	1	2	3	4	5	6	7	8
y :	1	8	28	56	70	56	28	8	1

Solution : Here Mean $= \frac{\Sigma fx}{\Sigma f} = \frac{1024}{256} = 4$. We construct the following table :-

x_i	f_i	x_i-4	$f_i(x_i-4)$	$f_i(x_i-4)^2$	$f_i(x_i-4)^3$	$f_i(x_i-4)^4$
0	1	–4	–4	16	–64	256
1	8	–3	–24	72	–216	648
2	28	–2	–56	112	–224	448
3	56	–1	–56	56	–56	56
4	70	0	0	0	0	0
5	56	1	56	56	56	56
6	28	2	56	112	224	448
7	8	3	24	72	216	648
8	1	4	4	16	64	256
Total	Σf_i=256	0	0	512	0	2816

Hence moments about mean $x = 4$ are

$$\mu_1 = \frac{\sum f_i(x_i-4)}{\sum f_i} = 0;\quad \mu_2 = \frac{\sum f_i(x_i-4)^2}{\sum f_i} = \frac{512}{256} = 2$$

$$\mu_3 = \frac{\sum f_i(x_i-4)^3}{\sum f_i} = 0;\quad \mu_4 = \frac{\sum f_i(x_i-4)^4}{\sum f_i} = \frac{2816}{256} = 11$$

Also $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$, $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{11}{4} = 2.75$

As $\beta_1 = 0$, hence curve is symmetric about mean and $\beta_2 < 3$ hence curve is platykurtic in nature.

1.13 Moment Generating Function (mgf)

The moments of most distributions can be determined directly by evaluating the necessary integrals or sums. The moment generating function (mgf), though has some limitations, helps us to find moment without evaluating integral or sum.

The moment generating function (mgf) of a random variable X having the probability function f(x), is given by :-

$$M_X(t) = E\left(e^{tX}\right) = \sum_x e^{tx} f(x) \text{ for discrete RV}$$

$$= \int_{-\infty}^{\infty} e^{tx} f(x)dx \text{ for continuous RV}$$

provided the right hand side is absolutely convergent for some positive number h such that $-h < t < h$ where t is any real parameter.

Now $$M_X(t) = E\left(e^{tX}\right) = E\left[1 + tX + \frac{t^2X^2}{\lfloor 2} + \ldots + \frac{t^rX^r}{\lfloor r} + \ldots\right]$$

$$= 1 + tE(X) + \frac{t^2}{\lfloor 2} E\left(X^2\right) + \ldots + t^r \frac{E\left(X^r\right)}{\lfloor r} + \ldots$$

$$= 1 + \mu_1' t + \mu_2' \frac{t^2}{\lfloor 2} + \ldots + \mu_r' \frac{t^r}{\lfloor r} + \ldots$$

where μ_r' is the rth order moment about origin. Since $M_X(t)$ generates moments, hence it is known as moment generating function.

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

Similarly we can define Moment generating function about mean $(\overline{X})$ which generates central moments.

$$M_X(t)\,(\text{about } \overline{X}) = E\left(e^{t(X-\bar{x})}\right) = E\left[1 + t(X-\bar{x}) + \frac{t^2}{\lfloor 2}(X-\bar{x})^2 + \ldots + \frac{t^r}{\lfloor r}(X-\bar{x})^r + \ldots\right]$$

$$= 1 + t\,E(X-\bar{x}) + \frac{t^2}{\lfloor 2}E(X-\bar{x})^2 + \ldots + \frac{t^r}{\lfloor r}E(X-\bar{x})^r + \ldots$$

$$= 1 + \mu_1 t + \mu_2 \frac{t^2}{\lfloor 2} + \ldots + \mu_r \frac{t^r}{\lfloor r} + \ldots$$

Hence $\mu_r = \left[\frac{d^r M_X(t)}{dt^r}(\text{about } \bar{x})\right]_{t=0}$.

In general, moment generating function about the point $X = a$ is

$$M_X(\text{about } X = a) = E\left(e^{t(X-a)}\right) = 1 + \mu_1'' t + \mu_2'' \frac{t^2}{\lfloor 2} + \ldots + \mu_r'' \frac{t^r}{\lfloor r} + \ldots$$

Moment generating function suffers from some drawbacks as though all moments are present but mgf does not generate all of them and hence has restricted use in statistics. For example, Let

$$P\left(X = 2^x\right) = \frac{e^{-1}}{\lfloor x}, x = 0, 1, 2, \ldots$$

$$\mu_r' = E\left(X^r\right) = \sum_{x=0}^{\infty}(2x)^r\, P\left(X = 2^x\right) = e^{-1}\sum_{x=0}^{\infty}\frac{2^{rx}}{\lfloor x} = e^{-1}e^{2^r} = e^{2^r - 1}$$

Hence all moments exist.

But mgf of X is

$$M_x(t) = \sum_{x=0}^{\infty}\exp\left(t2^x\right)\frac{e^{-1}}{\lfloor x} = e^{-1}\sum_{x=0}^{\infty}\frac{e^{t2^x}}{\lfloor x}$$

which is convergent for $t \le 0$ and divergent for $t > 0$ by D′ Alembert's ratio test. Hence $M_X(t)$ cannot be differentiated at $t = 0$ and hence it does not generate moments. The moment generating function of a distribution, if exists, uniquely determines the distribution. Hence $M_X(t) = M_Y(t) \Leftrightarrow$ X and Y are identically distributed.

1.14 Chebyshev's Inequality

If X is a RV with $E(X) = \mu$ and $Var(X) = \sigma^2$ then $P(\ |X - \mu| \geq \varepsilon\ \} \leq \frac{\sigma^2}{\varepsilon^2}$ where $\varepsilon > 0$.

Proof : Let X be a continuous RV with pdf f(x).

$$\text{Then} \quad \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2\ f(x)dx$$

$$= \int_{-\infty}^{\mu-\varepsilon} (x-\mu)^2 f(x)dx + \int_{\mu-\varepsilon}^{\mu+\varepsilon} (x-\mu)^2 f(x)dx + \int_{\mu+\varepsilon}^{\infty} (x-\mu)^2 f(x)dx$$

$$\geq \int_{-\infty}^{\mu-\varepsilon} (x-\mu)^2 f(x)dx + \int_{\mu+\varepsilon}^{\infty} (x-\mu)^2 f(x)dx \qquad .....(1.11)$$

In the first integral $x \leq \mu - \varepsilon \Rightarrow (x - \mu)^2 \geq \varepsilon^2$

In the second integral also $x \geq \mu + c \Rightarrow (x - \mu)^2 \geq c^2$

$$\text{Hence eq. (1.11)} \Rightarrow \sigma^2 \geq \varepsilon^2 \left\{ \int_{-\infty}^{\mu-\varepsilon} f(x)dx + \int_{\mu+\varepsilon}^{\infty} f(x)dx \right\}$$

$$= \varepsilon^2 \{P(X \leq \mu - \varepsilon) + P(X \geq \mu + \varepsilon)\}$$

$$= \varepsilon^2 P\{\mu + \varepsilon \leq X \leq \mu - \varepsilon\}$$

$$= \varepsilon^2\ P\{\varepsilon \leq X - \mu \leq -\varepsilon\}$$

$$= \varepsilon^2\ P\{|X - \mu| \geq c\}$$

$$\Rightarrow \qquad P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$$

$$\text{or} \qquad P\{|X - \mu| < \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2}, \varepsilon > 0$$

$$\text{If} \qquad \mu = 0 \Rightarrow P\{|X| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$$

or $$P\{|X| < \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

This inequality though proved for continuous case holds good for discrete case also.

Remark 7 Let $\varepsilon = k\sigma$, $k > 0$ then chebyshev's inequality can be written as

$$P\left\{\left|\frac{X-\mu}{k}\right| \geq \sigma\right\} \leq \frac{1}{k^2}$$

or $$P\left\{\left|\frac{X-\mu}{k}\right| \leq \sigma\right\} \geq 1 - \frac{1}{k^2}$$

As this inequality gives $P(|X| \leq k\sigma) \geq 1 - \frac{1}{k^2}$, hence we can conclude that the standard deviation σ of a RV X is a measure of width of its pdf. Larger is σ, the wider the pdf.

For example consider the gaussian pdf with standrad deviation $\sigma = 1$ and $\sigma = 3$. Then considering $k = 3\sigma$, $P(|X| \leq 3) \geq 0.88$ and $P(|X| \leq 9) \geq 0.88$. From figure 4.27 it is clear that the pdf with $\sigma = 3$ is spread out much more than the pdf with $\sigma = 1$.

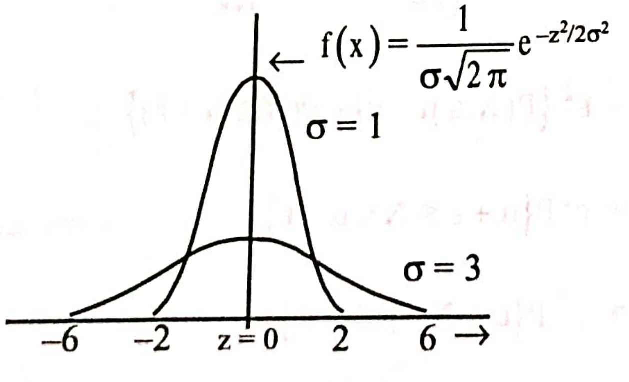

Figure 1.9

Ex.48 A RV X has mean $\mu = 12$ and variance $\sigma^2 = 9$ and an unknown probability distribution. Find P $(6 < X < 18)$

Sol. Using chebyshev's inequality

$$P\{|X-\mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}, \quad \varepsilon > 0$$

i.e, $$P\{|X-\mu| < \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2}, \quad \varepsilon > 0$$

$$\Rightarrow P\{-\varepsilon \le X-\mu \le \varepsilon\} \ge 1-\frac{\sigma^2}{\varepsilon^2}, \varepsilon > 0$$

$$\Rightarrow P\{\mu-\varepsilon \le X \le \mu+\varepsilon\} \ge 1-\frac{\sigma^2}{\varepsilon^2}$$

given $\mu = 12$, $\sigma^2 = 9$, $P\{12-\varepsilon < X < 12+\varepsilon\} \ge 1-\frac{9}{\varepsilon^2}$

Let $\varepsilon = 6 \Rightarrow P\{6 < X < 18\} \ge 1-\frac{9}{36} \Rightarrow P\{6 < X < 18\} \ge \frac{3}{4}$

Ex.49. A fair dice is tossed 720 times. Use chebyshev's inequality to find a lower bound for the probability of getting 100 to 140 sixes.

Sol. Let the RV X denote the number of sixes obtained when a fair dice is tossed 720 times.

Let $p = P\{\text{getting 6 in a single toss}\} = \frac{1}{6}$, $q = \frac{5}{6}$ and $n = 720$

then X follows Binomial distribution with mean $\mu = np = 120$ and variance $\sigma^2 = npq = 100$.

Hence by chebyshev's inequality

$$P\{|X-\mu| \le k\sigma\} \ge 1-\frac{1}{k^2}$$

$$\Rightarrow P\{|X-120| \le 10k\} \ge 1-\frac{1}{k^2} \text{ (as } \sigma = 10)$$

$$\Rightarrow P\{120-10k \le X \le 120+10k\} \ge 1-\frac{1}{k^2}$$

Let $k = 2$ $P\{100 \le X \le 140\} \ge \frac{3}{4}$

Hence required lower bound on probability = 3/4 = 0.75.

Illustrative Examples

***Example* 50.** What is the expectation of the number of failures preceeding the first success in an infinite series of independent trials with constant probability p of success in a trial ? **(Raj. IV Sem. CP-2005, 2006)**

Solution : Let random variable X denote the number of failures preceeding the first success. Then $X = 0, 1, 2 \ldots$ If p is the probability of success(S) on each trial then let q denote the probability of failure (F) in each trial, where $q = 1 - p$.

Hence the probability distribution will be :-

X	Event	P(X=xi)
0	S	p
1	FS	qp
2	FFS	q^2p
3	FFFS	q^3p
...	...	...

Hence $E(X) = 0 \times p + 1 \times qp + 2q^2p + 3q^3p + \ldots$

$$= qp(1 + 2q + 3q^2 + \ldots)$$

$$= qp(1 - q)^{-2}$$

Hence $E(X) = \dfrac{q}{p}$

***Example* 51.** A publisher earns a profit of Rs 20, on a book, if it is published at the right time, the profit earned is Rs 18.00 if the publication is delayed. The profit is reduced further to Rs 10.00 on all books, published or not published in time but which are defective. If 20% of the books are defective and there are only 70% chances of publishing the book in right time and without any defect and 10% chances are that the book is not published in right time. Find the publishers expected profit.

Solution : Let random variable X denote the profit earned by the publisher. Then its pmf is :-

X=x	20	18	10
P(x)	0.7	0.1	0.2

Hence expected profit $E(X) = \text{Rs } 17.80$.

***Example* 52.** A bag contains 2 one rupee coin and 3, 50 paise coins. A person is allowed to draw two coins indiscriminately. Find the expected value of the draw.

Solution : Let random variable X denote the amount drawn in rupees. The pmf of random variable X is :-

x	1	1.50	2
$P(x)$	$\frac{^3C_2}{^5C_2}=\frac{3}{10}$	$\frac{^2C_1 \times ^3C_1}{^5C_2}$	$\frac{^2C_2}{^5C_2}=\frac{1}{10}$

$$\text{Hence } E(X) = 1\times\frac{3}{10}+(1.50)\times\frac{6}{10}+2\times\frac{1}{10}=\frac{14}{10}=\text{Rs. } 1.40.$$

***Example* 53.** A player tosses two fair coins. He wins Rs. 2, if 2 heads occur and Rs. 1 if 1 head occurs. On the other hand he loses Rs. 3 if no head occurs. Determine the expected value of the game. Is the game favourable to the player ? If the game is fair, how much should he loose if no head occurs.

Solution : Here random variable X denotes the amount won in the game, (amount lost will be indicated by minus sign). Hence,

Event	HH	HT	TT
X	2	1	–3
P(x)	1/4	2/4	1/4

$$\text{Hence } E(X) = 2\times\frac{1}{4}+1\times\frac{2}{4}-3\times\frac{1}{4}=\frac{1}{4}= 0.25$$

As $E(X) = 0.25 > 0$. Hence the game is favourable to the player.

Now if we consider the game to be fair, then $E(X) = 0$. Here we assume that 'x' is the amount which the player looses on getting no head.

$$\therefore \text{ Now } E(X) = 2\frac{1}{4}+1\frac{2}{4}+x\frac{1}{4}= 0$$

$$\Rightarrow \quad \frac{x}{4}= -1 \Rightarrow x = -4.$$

Hence, player looses Rs 4 on getting no head, in a fair game.

***Example* 54.** In a sequence of Bernoulli trials let X be the length of the run of either successes or failures starting with the first trial. Find E(X) and Var(X).

Solution : Here X = 1, 2, 3, . . . and let 'p' denote the probability of success(S) and 'q' denote the probability of failure (F).

Now $P(X=1) = P(SF) + P(FS) = pq + qp$

$P(X=2) = P(SSF) + P(FFS) = p^2q + q^2p$

$P(X=3) = P(SSSF) + P(FFFS) = p^3q + q^3p$

- -

$$P(X=r)=P(\underbrace{SSS...SF}_{r\text{ times}})+P(\underbrace{FF...FS}_{r\text{ times}}) = p^rq+q^rp$$

- -

Hence $E(X) = 1(pq+qp) + 2(p^2q+q^2p) + 3(p^3q+q^3p) + ..$

$= pq(1+2p+3p^2+...) + qp(1+2q+3q^2+...)$

Hence, $E(X) = pq\,(1-p)^{-2} + qp(1-q)^{-2}$ $\qquad \left[\because (1-p)^{-2} = 1+2p+3p^2+....\right]$

$$= \frac{p}{q}+\frac{q}{p}$$

Also $E(X^2) = (pq+qp) + 4(p^2q+q^2p) + 9(p^3q+q^3p) + ...$

$= pq\,(1+4p+9p^2+...) + qp(1+4q+9q^2+...)$

Let $S = 1+4x+9x^2+16x^3+25x^4 ...$

$-3xS = -3x - 12x^2 - 27x^3 - 48x^4 ...$

$+3x^2S = 3x^2 + 12x^3 + 27x^4 + ...$

$-x^3S = -x^3 - 4x^4 + ...$

Adding

$\Rightarrow \quad (1-x)^3S = 1+x \qquad \Rightarrow \qquad S = (1-x)^{-3}(1+x)$

Using this in $E(X^2)$ we get :-

$E(X^2) = pq(1-p)^{-3}(1+p) + qp(1-q)^{-3}(1+q)$

$= pq(q)^{-3}(1+p) + qp(p)^{-3}(1+q)$

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

Now $Var(X) = E(X^2) - \{E(X)\}^2$

$$= \left(\frac{p}{q^2}+\frac{p^2}{q^2}+\frac{q}{p^2}+\frac{q^2}{p^2}\right)-\left(\frac{p}{q}+\frac{q}{p}\right)^2$$

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

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

***Example* 55.** From a lot of 25 items containing 5 defectives, a sample of 4 items is drawn at random (i) without replacement (ii) with replacement. Find the expected number of defectives in the sample. **(Raj. IV Sem CP 2004, RTU-2008)**

Solution : Let X denote the number of defective items. Hence X = 0, 1, 2, 3, 4 as four items are drawn. Defective items = 5 and non defective items = 20.

(i) In this case the probability distribution will be :-

X	P(X = x)
0	$\frac{{}^5C_0 \times {}^{20}C_4}{{}^{25}C_4} = 0.3830$
1	$\frac{{}^5C_1 \times {}^{20}C_3}{{}^{25}C_4} = 0.4506$
2	$\frac{{}^5C_2 \times {}^{20}C_2}{{}^{25}C_4} = 0.1502$
3	$\frac{{}^5C_3 \times {}^{20}C_1}{{}^{25}C_4} = 0.0158$
4	$\frac{{}^5C_4 \times {}^{20}C_0}{{}^{25}C_4} = 0.0004$

Hence $E(X) = 0 \times 0.3830 + 1 \times 0.4506 + 2 \times 0.1502$
$+ 3 \times 0.0158 + 4 \times 0.0004 = 0.8$

(ii) As here the items are replaced, hence the probability of being defective and non defective remains constant.

Hence P(defective) = 5/25 = 0.2 = p and P(non defective) = 20/25 = 0.8 = q

Hence the probability distribution is :-

X	P(X = x)
0	${}^4C_0\, p^0 q^4 = 0.4096$
1	${}^4C_1\, p^1 q^3 = 0.4096$
2	${}^4C_2\, p^2 q^2 = 0.1536$
3	${}^4C_3\, p^3 q = 0.0256$
4	${}^4C_4\, p^4 q^0 = 0.0016$

Hence $E(x) = 0 \times 0.4096 + 1 \times 0.4096 + 2 \times 0.1536 + 3 \times 0.0256 + 4 \times 0.0016$

$= 0.8$.

Example 56. Let X be a discrete random variate having values $x_i = (-1)^{i+1}(i+1)$, r = 1, 2, 3, . . . with probability law . Prove that E(X) does not exist.

Solution : Here $E(X) = \sum_{i=1}^{\infty} x_i p_i = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{(i+1)}{i(i+1)}$

$$= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \ldots = \log_e(1+1) = \log_e 2$$

using Libnitz test for alternating series, the series on R.H.S. is conditionally convergent (i.e. $\Sigma p_i x_i$ converges and $\Sigma|p_i x_i|$ do not converge). Hence, although $\Sigma p_i x_i$ is finite ($\log_e 2$), yet the series is not absolutely convergent; hence E(X) does not exist.

Example 57. A lot is known to contain 2 defective and 8 non-defective items. If these items are inspected at random, one after another, what is the expected number of items that must be chosen in order to remove, both the defective ones ?

Solution : Let random variable X denote the number of items chosen to remove the defective items. Hence X = 2, 3, 4, . . ., 10.

Now P(X = 2) = p(both defective items are chosen)

$$= \frac{{}^2C_2}{{}^{10}C_2} = \frac{1}{45}$$

P[X = 3] = P[1 defective and 1 non defective item is chosen uptill second draw and 1 defectective item is chosen at third draw]

$$= \left(\frac{{}^2C_1 \times {}^8C_1}{{}^{10}C_2}\right) \times \frac{1}{8} = \frac{2}{45}$$

Proceeding in the same manner :-

P[X = r] = P[1 defective and (r–2) non defective items are chosen till (r–1)th draw] × P[1 defective item is chosen at rth draw]

$$\frac{{}^2C_1 \times {}^8C_{r-2}}{{}^{10}C_{r-1}} \times \frac{1}{10-(r-1)} = \frac{2 \times 8C_{r-2}}{{}^{10}C_{r-1}(11-r)}, \quad r = 2,3,\ldots,10$$

$$\text{Hence } E(X) = \sum_{r=2}^{10} rP(X\text{-}r)$$

$$= \sum_{r=2}^{10} r.\frac{2 \times {}^8C_{r-2}}{{}^{10}C_{r-1}(11\text{-}r)}$$

$$= 2\sum_{r=2}^{10} \frac{r}{11\text{-}r}.\frac{\underline{|8}}{\underline{|r-2}\underline{|10-r}}.\frac{\underline{|r-1}\underline{|11-r}}{\underline{|10}}$$

$$= 2\sum_{r=2}^{10} \frac{\underline{|8}}{\underline{|10}} r(r-1) = \frac{1}{45}\sum_{r=2}^{10} r(r-1)$$

$$= \frac{1}{45}[21.+3.2+4.3+5.4+6.5+7.6+8.7+9.8+10.9]$$

$$= \frac{1}{45}[330] = \frac{22}{3}$$

***Example* 58.** Thirteen cards are drawn simultaneously from a deck of 52. If aces count 1, face cards 10 and others according to denomination find the expectation of the total score on 13 cards.

Solution : Let X_i be the score on the ith card.

Also

Also $P(X_i = 10) = P(\text{card is Jack or King or Queen or 10})$

$$= \frac{4}{52} + \frac{4}{52} + \frac{4}{52} + \frac{4}{52} = \frac{16}{52} = \frac{4}{13}$$

Similarly proceeding in the same manner we get the required probability distribution as :-

X_i	:1	2	3	4	5	6	7	8	9	10
$P(X=x_i)$:	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{1}{13}$	$\frac{4}{13}$

Thus,

$$E(Xi) = 1\times\frac{1}{13} + 2\times\frac{1}{13} + 3\times\frac{1}{13} + 4\times\frac{1}{13} + 5\times\frac{1}{13} + 6\times\frac{1}{13} + 7\times\frac{1}{13} + 8\times\frac{1}{13}$$

$$+ 9\times\frac{1}{13} + 10\times\frac{4}{13}$$

Hence according to addition law of expectation

$$E\left(\sum_{i=1}^{13} X_i\right) = \sum_{i=1}^{13} E(X_i) = \sum_{i=1}^{13} \frac{85}{13} \times 13 = 85$$

***Example* 59.** If t is any positive real number show that the function defined by

$$p(x) = e^{-t}(1 - e^{-t})^{x-1}$$

can represent a probability function of a random variable X assuming the values 1, 2, 3. . . . Find E(X) and Var(X) of the distribution.

Solution : $e^t > 1 \forall t > 0 \Rightarrow e^{-t} < 1 \Rightarrow 1 - e^{-t} > 0$

Also $e^t > 0 \forall t > 0 \Rightarrow e^{-t}(1 - e^{-t}) > 0 \Rightarrow p(x) > 0 \forall t > 0$, $x = 1,2,\ldots$

Also $\sum_{x=1}^{\infty} p(x) = e^{-t} \sum_{x=1}^{\infty} (1 - e^{-t})^{x-1} = e^{-t} \sum_{x=1}^{\infty} A^{x-1}$, where, $A = 1 - e^{-t}$

$$= e^{-t}\left(1+A+A^2+A^3\ldots\right)=e^{-t}\times\frac{1}{1-A}$$

$$= e^{-t}\left[1-\left(1-e^{-t}\right)\right]^{-1}=e^{-t}\left(e^{-t}\right)^{-1}=1$$

Hence p(x) represents the probability function of a random variable X.

Now $E(X) = \sum_{x=1}^{\infty} xp(x) = e^{-t}\sum_{x=1}^{\infty} x\left(1-e^{-t}\right)^{x-1} = e^{-t}\sum_{x=1}^{\infty} xA^{x-1}$, where $A = 1-e^{-t}$

$= e^{-t}[1 + 2A + 3A^2 + 4A^3 + \ldots].$

$\Rightarrow$ $E(X) = e^{-t}(1 - A)^{-2}$

$= e^{-t}(e^{-t})^{-2} = e^{t}$

Also $E(X) = \sum_{x=1}^{\infty} x^2p(x) = e^{-t}\sum_{x=1}^{\infty} x^2A^{x-1}$, $A = 1-e^{-t}$

$= e^{-t}[1 + 4A + 9A^2 + 16A^3 + \ldots]$

$E(X^2) = e^{-t}(1 + A)(1 - A)^{-3} = e^{-t}(2 - e^{-t})e^{3t}$

(as obtained in example 25 of this chapter)

$\therefore$ $Var(X) = E(X^2) - [E(X)]^2 = e^{-t}(2-e^{-t})e^{3t}-e^{2t}$

$= e^{2t}\left[\left(2-e^{-t}\right)-1\right]=e^{2t}\left(1-e^{-t}\right)$

$= e^{t}(e^{t} - 1)$

Hence $E(X) = e^{t}$; $Var(X) = e^{t}(e^{t} - 1)$.

***Example* 60.** Let the random variable X assume the value 'r' with the probability law $P(X = r) = q^{r-1}\,p$, $r = 1, 2, 3, \ldots$. Find the mgf of X and hence its mean and variance. Verify it by finding the mean from usual definition.

Solution : Mgf of $X = M_X(t) = E(e^{tX}) = \sum_{r=1}^{\infty} e^{tr}P_r$

$$= \sum_{r=1}^{\infty} e^{tr}q^{r-1}p = \frac{p}{q}\sum_{r=1}^{\infty}\left(qe^{t}\right)^{r}$$

$$= \frac{p}{q}\left[qe^t + \left(qe^t\right)^2 + \left(qe^t\right)^3 + \dots\right]$$

$$= \frac{p}{q}.qe^t\left[1 + qe^t + \left(qe^t\right)^2 + \dots\right]$$

$$= pe^t.\frac{1}{1-qe^t} = \frac{pe^t}{1-qe^t}$$

Now $\qquad \mu_1' = \text{mean} = \left[\frac{dM_x(t)}{dt}\right]_{t=0} = p\left[\frac{\left(1-qe^t\right)e^t - e^t\left(-qe^t\right)}{\left(1-qe^t\right)^2}\right]_{t=0}$

$$= \left[\frac{pe^t}{\left(1-qe^t\right)^2}\right]_{t=0} = \frac{p}{(1-q)^2} = \frac{p}{p^2} = p$$

and $\qquad \mu_2' = \left[\frac{d^2M_x(t)}{dt^2}\right]_{t=0} = \left(p\frac{\left[\left(1-qe^t\right)^2\right]e^t - 2e^t\left(1-qe^t\right)\left(-qe^t\right)}{\left(1-qe^t\right)^4}\right)_{t=0}$

$$= \left(p\left(1-qe^t\right)\frac{e^t\left[\left(1-qe^t\right)+2e^tq\right]}{\left(1-qe^t\right)^4}\right)_{t=0}$$

$$= \left[\frac{pe^t\left(1+qe^t\right)}{\left(1-qe^t\right)^3}\right]_{t=0} = p\frac{(1+q)}{(1-q)^3} = \frac{p(1+q)}{p^3} = \frac{1+q}{p^2}$$

Also $\qquad \text{Var}(X) = \mu_2' - \mu_1' = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$

By usual defination :-

$$E(X) = \sum_{i=1}^{\infty} rP(X=r) = \sum_{i=1}^{\infty} r\left(q^{r-1}p\right) = p\left[1+2q+3q^2+\dots\right]$$

$$= p(1-q)^{-2}$$

$$= \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

Example 61. The first four moments of a distribution about the value 4 of a variable are –1.5, 17, –30 and 108. Find

(i) Moments about mean, β_1 and β_2. Also give the nature of distribution.

(ii) Moments about origin.

(iii) Moment about the point $x = 2$.

Solution : Here a=4 and and about the point '4'.

Moment about mean are given by :-

$$\mu_1 = 0$$

$$\mu_2 = \mu_2'' - \mu_1''^2 = 17-(-15)2 = 14.75$$

$$\mu_3 = \mu_3'' - 3\mu_2''\mu_1'' + 2\mu_1''^3 = (-30)-3(17)(-15)+2(-1.5)^3$$

$$= -30 + 76.5 - 6.75 = 39.75$$

$$\mu_4 = \mu_4'' - 4\mu_3''\mu_1'' + 6\mu_2''\mu_1''^2 - 3\mu_1'^4$$

$$= (108) - 4(-30)(-1.5) + 6(17)(-1.5)^2 - 3(-1.5)^4$$

$$= 108 - 180 + 229.5 - 15.1875 = 142.3125$$

Also $$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(39.75)^2}{(14.75)^3} = 0.4924$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{142.3125}{(14.75)^2} = 0.6541$$

As $\beta_1 \neq 0$ hence distribution about '4' is unsymmetric.

As $\beta_1 < 3$ hence distribution about '4' is platykurtic.

Now we know that

$$\mu_1'' = \bar{X} - 4 \Rightarrow \bar{X} \Rightarrow \mu_1'' + 4$$

$$\Rightarrow \quad \bar{X} = -1.5+(4) = 2.5$$

$$\Rightarrow \quad \mu_1' = \bar{X} = 2.5$$

$$\mu_2' = \mu_2 + \mu_1'^2 = 14.75 + (2.5)^2$$
$$= 21$$

$$\mu_3' = \mu_3 + 3\mu_1'\mu_2' - 2\mu_1'^3$$
$$= 39.75 + 3(2.5)(21) - 2(2.5)^3 = 166$$

$$\mu_4' = \mu_4 + 4\mu_3'\mu_1' - 6\mu_2'\mu_1'^2 + 3\mu_1'^4$$
$$= 142.3125 + 4(166)(2.5) - 6(21)(2.5)^2 + 3(2.5)^4$$
$$= 1132$$

Now using these we find moment about point $x = 2$.

$$\mu_1'' = \bar{X} - 2 = 2.5 - 2 = 0.5$$

$$\mu_2'' = \mu_2 + \mu_1''^2 = 14.75 + (0.5)^2$$
$$= 15$$

$$\mu_3'' = \mu 3 - 3\mu_1'\mu_2 + \mu_1''^3$$
$$= 39.75 - 3(0.5)(14.75) + (0.5)^3$$
$$= 62$$

and $$\mu_4'' = \mu_4 + 4\mu_3\mu_1'' + 6\mu 2\mu_1''^2 + \mu_1'^4$$
$$= 142.3125 + 4(39.75)(0.5) + 6(14.75)(0.5)^2 + (0.5)^4 = 244.$$

Example 62. For a distribution mean is 10, variance is 16, γ_1 is 1, and β_2 is 4. Obtain the first four moments about origin. Also comment upon the nature of distribution.

Solution : Given $$\bar{X} = \mu_1' = 10, \mu_2' = 16, \gamma_1 = 1, \beta_2 = 4$$

Now $$\gamma_1 = 1 \Rightarrow \beta_1^{1/2} = 1 \Rightarrow \beta_1 \Rightarrow \frac{\mu_3}{\mu_2^{3/2}} = 1$$

$$\Rightarrow \quad \mu_3 = \mu_2^{3/4} = (16)^{3/2} = 64$$

Similarly

$$\beta_2 = 4 \Rightarrow \frac{\mu_4}{\mu_2^2} = 4 \Rightarrow \mu_4 = 4 \times \mu_2^2 = 4 \times 16^2 = 1024$$

Hence we get :-

$$\mu_2' = \mu_2 + \mu_1' = 16 + 10^2 = 116$$

$$\mu_3' = \mu_3 + 3\mu_1'\mu_2' - 2\mu_1'^3$$

$$= 64 + 3(10)(116) - 2(10)^3$$

$$= 1544$$

and $$\mu_4' = \mu 4 + 4\mu_3'\mu_1' - 6\mu_2' 2\mu_1'^2 + \mu_1'^4$$

$$= 1024 + 4(1544)(10) - 6(116)(10)^2 + 3(10)^4$$

$$= 23184$$

Nature of Distribution

Here $\beta_1 = 1 \neq 0$ Hence distribution is not symmetric

and $\beta_2 = 4 > 3$ Hence curve is a peaked curve i.e. distribution is leptokurtic in nature.

***Example* 63.** The first four moments of the distribution about the value 5 of the variable are 2, 20, 40 and 50. Find the mean, variance, β_1 and β_2. Comment upon the nature of distribution.

Solution : Here a = 5 and moments about this point are

$$\mu_1'' = 2,\ \mu_2'' = 20,\ \mu_3'' = 40,\ \mu_4'' = 50$$

we know that

$$\mu_1'' = \bar{X} - a \Rightarrow 2 = \bar{X} - 5 \Rightarrow \bar{X} = 7$$

Also $$\mu_2'' = \mu''^2 - \mu_1''^2 = 20 - 4 = 16$$

Again $$\mu_3'' = \mu_3'' - 3\mu_2''\mu_1'' + 2\mu_1''^3$$

$$= 40 - 120 + 16 = -64$$

and $$\mu_4'' = \mu_4'' - 4\mu_3''\mu_1'' + 6\mu_2''\mu_1''^2 - 3\mu_1''^4$$

$$= 50 - 320 + 480 - 48 = 162$$

Also $$\beta_1 = \frac{\mu_4}{\mu_2^2} = \frac{(-64)^2}{(16)^3} = \frac{4096}{4096} = 1$$

and $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{162}{(16)^2} = 0.6328$

As $\beta_1 \neq 0$ hence curve is not symmetric, it is skewed

and $\beta_2 < 3$ hence curve is a flat curve i.e., distribution is platykurtic.

Example 64. Let the variable X have the distribution

$P(X = 0) = P(X = 2) = p; \quad P(X = 1) = 1 - 2p$ for

For what p is Var X maximum ?

Solution : Given distribution is :-

X	: 0	1	2
P(X)	: p	1–2p	p

$\therefore$ $E(X) = 1$

$E(X^2) = 1 - 2p + 4p = 1 + 2p$

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

As and $\text{Var}(X) = 2p$ hence it is maximum for $p = 1/2$ i.e. $(\text{Var } X)_{p=1/2} = 1$.

Example 65. Find the first four moments about the mean from the following data :-

Variable :	0-10	10-20	20-30	30-40
Frequency :	1	3	4	2

Solution : Given :-

Interval :	0-10	10-20	20-30	30-40
Mid value x_i :	5	15	25	35
frequency f_i:	1	3	4	2

Here mean

We construct the following table :-

x_i	f_i	x_i-22	$f_i(x_i-22)$	$f_i(x_i-22)^2$	$f_i(x_i-22)^3$	$f_i(x_i-22)^4$
5	1	–17	–17	289	–4913	83521
15	3	–7	–21	147	–1029	7203
25	4	3	12	36	108	324
35	2	13	26	338	4394	57122
Total	10	-	0	810	–1440	148170

Hence $\mu_1 = \frac{\sum f_i (x_i - 22)}{\sum f_i} = 0$

$$\mu_2 = \frac{\sum f_i (x_i - 22)^2}{\sum f_i} = \frac{810}{10} = 81$$

$$\mu_3 = \frac{\sum f_i (x_i - 22)^3}{\sum f_i} = \frac{-1440}{10} = -144$$

and $\mu_4 = \frac{\sum f_i (x_i - 22)^4}{\sum f_i} = \frac{148170}{10} = 14817$

***Example* 66.** Calculate the first four moments of the following distribution about x = 40.5 and hence find the moments about mean of the following distribution.

Hours worked :	30.0-32.9	33.0-35.9	36.0-38.9	39.0-41.9	42.0-44.9	45.0-47.9
No. of industries	2	4	26	47	15	6

Soltion :

Hours worked	**Mid value x_i**	**d = x–40.5**	**f**	**fd**	**fd^2**	**d^3**	**fd^4**
$30.0 - 35.9 \cong 33$	31.5	–9	2	–18	162	–1458	13122
$33.0 - 35.9 \cong 36$	34.5	–6	4	–24	144	–864	5184
$36.0 - 38.9 \cong 39$	37.5	–3	26	–78	234	–702	2106
$39.0 - 41.9 \cong 42$	40.5	0	47	0	0	0	0
$42.0 - 44.9 \cong 45$	43.5	3	15	45	135	405	1215
$48.0 - 47.9 \cong 48$	46.5	6	6	36	216	1296	7776
	Total		Σf=100	Σfd= –39	Σfd^2 = 891	Σfd^3 =1323	Σfd^4 = 29403

Now Moments about the point x = 40.5 are :-

$$\mu_1'' = \frac{\Sigma fd}{\Sigma f} = \frac{-39}{100} = -0.39; \ \mu_2'' = \frac{\Sigma fd^2}{\Sigma f} = \frac{891}{100} = 8.91$$

$$\mu_3'' = \frac{\Sigma fd^3}{\Sigma f} = \frac{-1323}{100} = -13.23; \ \mu_4'' = \frac{\Sigma fd^4}{\Sigma f} = \frac{29403}{100} = -294.03$$

Again we find central moments by using equation

$$\mu_1 = 0$$

$$\mu_2'' = \mu_2'' - \mu_1''^2 = 8.91 - (-0.39)^2 \cong 8.76$$

$$\begin{aligned}\mu_3'' &= \mu_3'' - 3\mu_2''\mu_1'' + 2\mu_1''^3 \\ &= (-13.23) - 3(8.91)(-0.39) + 2(-0.39)^3 \\ &= -13.23 + 10.4247 - 0.118638 \cong -2.92\end{aligned}$$

$$\begin{aligned}\mu_4'' &= \mu_4'' - 4\mu_3''\mu_1'' + 6\mu_2''\mu_1''^2 - 3\mu_1''^4 \\ &= 294.03 - 4(-13.23)(-0.39) + 6(8.91)(-0.39)^2 - 3(-0.39)^4 \\ &= 294.03 - 20.6388 + 8.131266 - 0.06940323 \\ &\cong 281.45.\end{aligned}$$

Example 67. Calculate the first four moments about mean from the following data :

x :	6	7	8	9	10	11	12
y :	3	6	9	13	8	5	4

Solution : Here $\Sigma y = 48$ and $\Sigma xy = 432$

$$\therefore \quad \text{Mean} = \frac{\Sigma xy}{\Sigma y} = \frac{432}{48} = 9$$

Hence we find the moment about mean i.e. about point 9.

x_i	y_i	$x_i - 9$	$y_i(x_i-9)$	$y_i(x_i-9)^2$	$y_i(x_i-9)^3$	$y_i(x_i-9)^4$
6	3	–3	–9	27	–81	243
7	6	–2	–12	24	–48	96
8	9	–1	–9	9	–9	9
9	13	0	0	0	0	0
10	8	1	8	8	8	8
11	5	2	10	20	40	80
12	4	3	12	36	108	324
Total	48	0	0	124	18	760

Hence required moments are :

$$\mu_1 = \frac{\sum y_i(x_i - 9)}{\sum y} = 0$$

$$\mu_2 = \frac{\sum y_i(x_i - 9)^2}{\sum y} = \frac{124}{48} = 2.583$$

$$\mu_3 = \frac{\sum y_i(x_i - 9)^3}{\sum y} = \frac{18}{48} = 0.375$$

and $$\mu_4 = \frac{\sum y_i(x_i - 9)^4}{\sum y} = \frac{760}{48} = 15.833$$

Example 68. The standard deviation of a symmetrical distribution is 3. What must be the value of the fourth moment about the mean in order that distribution is mesokurtic ?

Soltion : Here $\sigma = 3 \Rightarrow \sigma^2 = 9 \Rightarrow \mu_2 = 9$.

Now if the distribution is symmetric then :-

$$\beta_2 = 3 \Rightarrow \frac{\mu_4}{\mu_2^2} = 3 \Rightarrow \mu_4 = 243$$

***Example* 69.** In a frequency distribution, the first four moments about the origin are 1, 4, 10 and 46 respectively. Find the first four central moments and the beta coefficients of this distribution. What is the nature of this distribution curve ?

Solution : Here given,

$$a = 0, \mu_1'' = 1, \mu_2'' = 4, \mu_3'' = 10, \mu_4'' = 46$$

Now $\mu_1 = 0$

$$\mu_2 = \mu_2'' - 3\mu_2'' = 4 - 1 = 3$$

$$\mu_3 = \mu_3'' - 3\mu_2''\mu_1'' + 2\mu_1''^3 = 0$$

$$\mu_4 = \mu_4'' - 4\mu_3''\mu_1'' + 6\mu_2''\mu_1''^2 - 3\mu_1''^4 = 27$$

Also $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$, $\beta_2 = \dfrac{\mu_4}{\mu_2^2}$

$\beta_1 = 0$ hence distribution is symmetric.

$\beta_2 = 3$ hence distribution is mesokurtic.

Exercise 1(C)

Q.1 A random variable X is associated with independent trials, each having probability of success and failure as p and q respectively, p + q = 1. The trials are made until and unless a success is obtained. Find E(x).

Ans. $\dfrac{1}{p}$

Q.2 Let a message be expected at some time past 10 A.M.. If X is a random variate which denotes the number of hours after 10 A.M. until the message arrives and its p.d.f is given by, f(x) = 1/2.5 if 0 < x < 2.5, f(x) = 0 elsewhere, find the expected amount of time past 10 A.M. until the message arrives.

Ans. 1.25 hours

Q.3 A man with n keys wants to open his door and tries the keys independently at random.

Find the mean and variance of the number of trials required to open the door.

(i) If unsuccessful keys are not eliminated from further selection and (ii) if they are eliminated.

Ans. (i) n, n (n–1)

(ii) $\frac{n+1}{2}, \frac{n^2-1}{12}$

Q.4 An urn contains balls numbered 1, 2, 3. First a ball is drawn from the urn and then a fair coin is tossed the number of times as the number shown on the drawn ball. Find the expected number of heads.

Ans. 1

Q.5 If X denotes the number of heads obtained when four coins are tossed simultaneously. Find E(X).

Ans. 2

Q.6 A box contains 2^n tickets among which ${}^nC_i$ tickets bear the number i, i = 0, 1, 2, . . . n. A group of n tickets is drawn. What is the expectation of the sum of their numbers ?

Ans. $\frac{mn}{2}$

Q.7 A player tosses three fair coins. He wins Rs. 8 if 3 heads occur, Rs 3 if 2 heads occur and Rs 1, if only 1 head occurs. On the other hand, he loses Rs x, if no head occurs. If the game is fair, find x.

Ans. Rs 20

Q.8 If X is a discrete random variate with pmf $P(X = x) = {}^{r+x-1}C_{r-1}\, p^r q^x$, x = 0, 1, 2, . . . where r is a positive integer and p + q = 1. Find its probability generati-ng function.

Ans. $\left(\frac{p}{1-qs}\right)^r$

Q.9 A bag contians a coin of value M and a number of other coins whose aggregate value is m. A person draws coins one by one till the coin of value M is drawn. Find the value of his expectation.

Ans. E(x) = M + N x/2, where N is the number of other coins.

Q.10 The first four moments of a frequency distribution about the point 5 are –0.55, 4.46, –0.43 and 68.52. Find β_1 and β_2.

Ans. $\beta_1 = 0.6055,\ \beta_2 = 4.3619$

Q.11 Show that the central moment μ_r is given by $\mu_r = \mu'_r - {}^rC_1\,\mu'_r\,\mu_x + \ldots + (-1)^{r-1}(\mu_x)^r$, where μ_x is the mean.

Q.12 The following data are given to an economist for the purpose of economic analysis $N = 100$, $\Sigma f_i d_i = 50, \Sigma f_i d_i^2 = 1967.2, \Sigma f_i d_i^3 = 2925.8$ and $\Sigma f_i d_i^4 = 86690.2$ where $d_i = x_i - 4$. Compute the moments about the arbitrary origin and the moments about the mean.

Ans. $\mu'_1 = 0.5, \mu'_2 = 19.762, \mu'_3 = 29.268, \mu'_4 = 866.502, \mu_2 = 19.4221, \mu_3 = 0, \mu_4 = 837.3065.$

Q.13 Find the moment generating function of the random variable X whose pmf is given by

$P(X = x) = \frac{1}{8}\,{}^3C_x$, x=0,1,2 and 3 then find μ'_1 and μ'_2.

Ans. $M_X(t) = \frac{1}{8}\left(1+e^t\right)^3, \mu'_1 = \frac{3}{2}, \mu'_2 = 3$

Q.14 In the two distributions A and B, the second central moments are 9 and 16 respectively and the third central moments are –8.1 and –12.8 respectively, which distribution is more skewed to the left ?

Ans. Distribution A

Q.15 A random variable assumes the values 1 and –1 with probability 1/2 each. Find (i) the moment generating function and (ii) the first four moments about origin.

Ans. (i) $M_X(t) = \frac{1}{2}\left(e^{-t} + e^t\right)$

(ii) $\mu'_1 = \mu'_3 = 0; \mu'_2 = \mu'_4 = 1$

Q.16 Show that for a random variate X which takes values $\lfloor n$ for $n = 0, 1, 2, \ldots$ with probabilities $\frac{e^{-1}}{\lfloor n}$, E(x) does not exist.

Q.17 Find the second and third central moments of the frequency distribution given below Hence find the coefficient of skewness.

Class limits	Frequency
110.0–114.9	5
115.0–119.9	15
120.0–124.9	20
125.0–129.9	35
130.0–134.9	10
135.0–139.9	10
140.0–144.9	5

Ans. $\mu_2 = 54, \mu_3 = 100.5, \gamma_1 = 0.253$

Q.18 A and B throw an ordinary dice alternately for a stake of Rs. 11, which is to be one by one who first throws 6. Find their expectation if A has the first chance.

Ans. Rs. 6, Rs 5.

Q.19 Compute the first four moments of the data 3, 5, 7, 9 about the mean. Also compute the first four moments about the point x = 4.

Ans. $\mu_1 = 0, \mu_2 = 5, \mu_3 = 0, \mu_4 = 41; \mu_1'' = 2, \mu_2'' = 9, \mu_3'' = 38.25, \mu_4'' = 177$

Q.20 Calculate the first four moments about the mean from the following data :-

Marks :	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No of sutdents :	8	12	20	30	15	10	5

Ans. $\mu_1 = 0, \mu_2 = 236.76, \mu_3 = 264.336, \mu_4 = 141290.10$

Q.21 For a frequency distribution the mean is 1.5, variance = 0.64, β_2 = 2.5 and γ_1 = 0.3. Find μ_3, μ_4 and first four moments about the origin.

Ans. $\mu_3 = 0.1536, \mu_4 = 1.024, \mu_1' = 1.5, \mu_2' = 2.89, \mu_3' = 6.4086, \mu_4' = 15.6481$

Q.26 The standard deviation of a symmetric distribution is 5. What must be the value of the fourth moment about the mean in order that the distribution be (i) leptokurtic (ii) mesokurtic (iii) platykurtic.

Ans. (i) $\mu_4 > 1875$ (ii) $\mu_4 = 1875$ (iii) $\mu_4 < 1875$

OOO

2

Theoretical Discrete Probability Distributions

2.1 Introduction

In chapter 2, we studied about the probability distributions-both discrete and continuous. In many engineering applications when probabilistic models are constructed for various observable phenomena, some probability distributions arise more frequently than the others. This chapter is devoted to some of such discrete probability distributions and derivation of some of their important characteristics.

2.2 Bernoulli Distribution

Let us consider an experiment which has two outcomes either a success(S) or failure (F). The probability of success being denoted as p and that of failure being $q = 1 - p$, i.e., here $q + p = 1$. Then the random variable X associated with the distribution, denoting the number of successes has the following probability distribution.

X	: 0	1
P(X)	: q	p

This distribution is known as Bernoulli distribution, after the Swiss mathematician James Bernoulli.

2.2.1 Mean and Variance of Bernoulli Distribution

Let X be a Bernoulli variate. Then,

Mean $\mu = E(X) = \sum x_i p(x_i) = 0 \times q + 1 \times p = p$

Also $E(X^2) = \sum x_i^2 p(x_i) = 0 \times q + 1 \times p = p$

Hence $\mathrm{Var}(X) = E\left[(X - \mu)^2\right] = E(X^2) - [E(X)]^2 = p^2 - p = p(1 - p) = pq$.

2.2.2 Moment Generating Function

The mgf of Bernoulli distribution about origin is given by :

$$M_X(t) = E\left(e^{tX}\right) = \sum_i e^{tx_i} p(x_i) = e^{t(o)} \times q + e^{t(1)} p = \left(q + pe^t\right)$$

We can obtain moments about origin from $M_X(t)$ as :

$$\mu_1' = \left(\frac{d}{dt}[M_X(t)]\right)_{t=0} = \left[\frac{d}{dt}\left(q+pe^t\right)\right]_{t=0} = p = \text{mean}$$

$$\mu_2' = \left(\frac{d^2}{dt^2}[M_X(t)]\right)_{t=0} = \left[\frac{d^2}{dt^2}\left(q+pe^t\right)\right]_{t=0} = p$$

$$\mu_3' = \left(\frac{d^3}{dt^3}[M_X(t)]\right)_{t=0} = \left[\frac{d^3}{dt^3}\left(q+pe^t\right)\right]_{t=0} = p$$

$$\mu_4' = \left(\frac{d^4}{dt^4}[M_X(t)]\right)_{t=0} = \left[\frac{d^4}{dt^4}\left(q+pe^t\right)\right]_{t=0} = p \text{ and so on.}$$

2.2.3 Probability Generating Function

The probability generating function (pgf) for the Bernoulli variate X is given as :

$$P(s) = E\left(s^x\right) = s^0 q + s^1 p = q + ps$$

2.3 Binomial Distribution

Let an experiment be repeated n times. Each of its trial are independent and each trial has two outcomes either success(S) or failure(F). The probability of success being p and the probability of failure being $q=1-p$ is constant for each trial. A random variable X (denoting number of successes) associated with this experiment is said to follow Binomial distribution if its probability mass function is given by :

$$P(X=r) = P(r) = {}^nC_r p^r q^{n-r}, r = 0, 1, 2, \ldots, n$$

It should be noted here that the word success as used here does not correspond to a good event. Selecting a defective pen may be classified as a success even though the results of such a selection may be some what less than pleasant.

The two independent constants n and p in the distribution are known as parameters of the distribution. Here X is known as a binomial variate and we use notation $X \sim B(n, p)$.

Also here $p + q = 1$. **(Raj. IV Sem. CP-2003, RTU-2008)**

Proof of the above formula for Binomial distribution

Here we are considering 'n' independent Bernoulli trials. If 'r' successes occur then there will be 'n–r' failures. One of the events may be SSFSFFFS...FSF.

Then $P(SSFSFFFS \ldots FSF) = P(S)\,P(S)\,P(F) \ldots P(F)\,P(S)\,P(F)$

$$= ppq \ldots qpq = \underbrace{p.p \ldots p}_{r\text{-times}}.\underbrace{q.q \ldots q}_{(n-r)\text{times}}$$

$$= p^r q^{n-r}.$$

But these 'r' successes in n trials can occur in ${}^nC_r$ ways.

Hence $P(r \text{ successes}) = {}^nC_r\, p^r q^{n-r}$, $r = 0, 1, 2, \ldots, n$.

Hence the formula.

To prove that above given is a pmf

(i) Obviously $P(X=r) \geq 0 \;\; \forall \;\; r = 0, 1, 2, \ldots n$ as $p, q \geq 0$

(ii) $\sum_{r=0}^{n} P(X=r) = \sum_{r=0}^{n} {}^nC_r\, p^r\, q^{n-r}$

$$= {}^nC_0\, p^0 q^{n-0} + {}^nC_1\, p^1\, q^{n-1} + \ldots + {}^nC_{n-1}\, p^{n-1}\, q + {}^nC_n\, p^n\, q^0$$

$$= (q+p)^n \qquad \text{(Binomial theorem)}$$

$$= 1^n = 1 \{\because q + p = 1\}$$

Hence Binomial distribution is a legitimate probability distribution.

Remark 1 Binomial theorem is

$$(a+b)^n = {}^nC_0\, b^0\, a^n + {}^nC_1\, b^1\, a^{n-1} + \ldots + {}^nC_n\, b^n\, a^0$$

$$= \sum_{i=0}^{n} {}^nC_i b^i a^{n-i}.$$

Remark 2 The name Binomial distribution is given since the probabilities ${}^nC_r\, p^r\, q^{n-r}$ ($r = 0, 1, 2, \ldots, n$) are the successive terms in the expansion of the Binomial expression $(q+p)^n$.

Remark 3 If we assume that n trials constitute a set and if we consider N such sets then the number of sets in which we get exactly r successes $= N \times {}^nC_r\, p^r\, q^{n-r}$, $r = 0, 1, 2, \ldots, n$.

2.3.1 Mean and Variance of Binomial Distribution

Mean $(\mu)=\mu_1'=E(X)=\sum_{r=0}^{n} x_r p(r)=\sum_{r=0}^{n} r\cdot {}^nC_r\, p^r q^{n-r}=\sum_{r=0}^{n} r\frac{\lfloor n}{\lfloor r\,\lfloor n-r} p^r q^{n-r}$

$$=\sum_{r=1}^{n}\frac{\lfloor n}{\lfloor r-1\,\lfloor n-r} p^r q^{n-r}=\sum_{r=1}^{n} n\frac{\lfloor n-1}{\lfloor r-1\,\lfloor (n-1)-(r-1)} p\cdot p^{r-1} q^{(n-1)-(r-1)}$$

$$=np\sum_{r=1}^{n} {}^{n-1}C_{r-1}\, p^{r-1} q^{(n-1)-(r-1)}$$

$$=np\left[{}^{n-1}C_0\, p^0\, q^{n-1}+{}^{n-1}C_1\, p^1\, q^{(n-1)-1}+\ldots\ldots+{}^{n-1}C_{n-1}\, p^{n-1} q^0\right]$$

$$=np\sum_{i=0}^{n-1} {}^{n-1}C_i\, p^i\, q^{(n-1)-i}=np(q+p)^{n-1}$$

$$=np\, 1^{n-1}=np. \qquad \ldots\ldots(2.\ 1)$$

Hence Mean of Binomial distribution = np.

Now $\mu_2'=E\left(X^2\right)=\sum_{r=0}^{n} x_r^2\, p(r)=\sum_{r=0}^{n} r^2\, {}^nC_r\, p^r q^{n-r}$

$$=\sum_{r=0}^{n}[r(r-1)+r]\, {}^nC_r\, p^r\, q^{n-r}$$

$$=\sum_{r=0}^{n} r(r-1)\, {}^nC_r\, p^r\, q^{n-r}+\sum_{r=0}^{n} r\, {}^nC_r\, p^r q^{n-r}$$

$$=\sum_{r=0}^{n} r(r-1)\frac{\lfloor n}{\lfloor r\,\lfloor n-r} p^r\, q^{n-r}+\sum_{r=0}^{n} r\, p(r)$$

$$=\sum_{r=2}^{n}\frac{\lfloor n}{\lfloor r-2\,\lfloor n-r} p^r\, q^{n-r}+np \qquad \text{(using (2.1))}$$

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

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

$$= n(n-1)p^2 \left[{}^{n-2}C_0 \, p^0 q^{(n-2)} + {}^{n-2}C_1 \, p^1 q^{(n-2)-1} + \ldots + {}^{n-2}C_{n-2} \, p^{n-2} q^0 \right] + np$$

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

$$= n(n-1)p^2 + np \quad (\because q+p=1)$$

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

$$= n^2p^2 + npq \quad (\because q = 1-p)$$

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

$$= n^2p^2 + npq - (np)^2 = npq \qquad \text{(using (2.1))}$$

Hence variance of Binomial distribution = npq.

***Example* 1.** If during a war, one out of 9 ships could not arrive safely, find the probability that exactly 3 out of a convoy of 6, would arrive safely.

Solution : Here X = number of ships that arrive safely,

$n = 6$ and $p = P$ (ships arrive safely) $= 1/9$, $r = 3, q = 1-p = 1-\frac{1}{9} = \frac{8}{9}$.

As $P(X=r) = {}^nC_r \, p^r \, q^{n-r}, \quad r = 0, 1, 2, \ldots, 6$

Hence $P(X=3) = {}^6C_3 \, p^3 q^3$

$$= \frac{\lfloor 6}{\lfloor 3 \lfloor 3} \left(\frac{1}{9}\right)^3 \left(\frac{8}{9}\right)^3 = \frac{2754}{531441} = 0.005.$$

Example 2. Ten coins are thrown simultaneously. Find the probability of getting at least seven heads.

Solution : Let X = number of heads obtained.

Here $n = 10$, p = P(getting head) $= 1/2$, $q = 1 - p = 1/2$

Also $P(X = r) = {}^nC_r\, p^r q^{n-r}$, $r = 0, 1, 2, \ldots, 10$

$\therefore$ P(atleast seven heads) $= P(X \geq 7)$

$= P(7) + P(8) + P(9) + P(10)$

$$= {}^{10}C_7\left(\frac{1}{2}\right)^{10} + {}^{10}C_8\left(\frac{1}{2}\right)^{10} + {}^{10}C_9\left(\frac{1}{2}\right)^{10} + {}^{10}C_{10}\left(\frac{1}{2}\right)^{10}$$

$$= \left(\frac{1}{2}\right)^{10}[120 + 45 + 10 + 1] = \frac{176}{1024} = 0.172.$$

Example 3. Write the probability distribution of the defective bulbs drawn in 3 draws when it is known that in a lot of 10, on an average 3 are defective.

Solution : Let X denote the number of defective bulbs.

Here $n = 3$, $p = P$ (defective bulbs) $= 3/10 = 0.3$

and $q = 1 - p = 0.7$

Hence the required probability distribution is :-

$X = r$	0	1	2	3
$P(X = r)$	${}^3C_0\, p^0 q^3 = (0.7)^3$ $= 0.343$	${}^3C_1\, p^1 q^{3-1}$ $= 3(0.7)^2(0.3)$ $= 0.441$	${}^3C_2\, p^2 q^1$ $= 3(0.3)^2(0.7)$ $= 0.189$	${}^3C_3\, p^3 q^0$ $= (0.3)^3$ $= 0.027$

Example 4. If m things are distributed among 'a' men and 'b' women, show that the probability that the number of things received by men is odd is $\frac{1}{2}\left[\frac{(b+a)^m - (b-a)^m}{(b+a)^m}\right]$.

Solution : Let X = number of things received by men.

Here p = P(things received by men) $= \dfrac{a}{a+b}$

and $q = 1 - p = \dfrac{b}{a+b}$.

By Binomial distribution the probability that out of 'm' things 'r' are received by men is $P(X = r) = {}^mC_r\, p^r\, q^{m-r}$, $r = 0, 1, 2, \ldots, m$. The probability that the number of things received by men is odd is given by :

$$P(\text{odd}) = P(1) + P(3) + P(5) + \ldots$$

$$= {}^mC_1\, p\, q^{m-1} + {}^mC_3\, p^3\, q^{m-3} + {}^mC_5\, p^5\, q^{m-5} + \ldots \qquad .....(2.2)$$

Now we know that :

$$(q+p)^m = q^m + {}^mC_1 q^{m-1} p + {}^mC_2 q^{m-2} p^2 + \ldots + p^m$$

and $\quad (q-p)^m = q^m - {}^mC_1\, q^{m-1}\, p + {}^mC_2\, q^{m-2}\, p^2 \ldots + (-p)^m$

Subtracting these two we get

$$(q+p)^m - (q-p)^m = 2\left[{}^mC_1 q^{m-1} p + {}^mC_3 q^{m-3} p^3 + \ldots\right] = 2P(\text{odd})$$

[using (2.2)]

$$\therefore \quad P(\text{odd}) = \frac{1}{2}\left[(q+p)^m - (q-p)^m\right]$$

Now $\quad q + p = 1$ and $q - p = \dfrac{b-a}{a+b}$

$$\therefore \quad P(\text{odd}) = \frac{1}{2}\left[1 - \frac{(b-a)^m}{(b+a)^m}\right]$$

Hence required probability $= \dfrac{1}{2}\left[\dfrac{(b+a)^m - (b-a)^m}{(b+a)^m}\right]$.

2.3.2 Moments, Moment Generating Function and Recurrence Relation for Moments

The moment generating function about origin is :-

$$M_X(t) = E(e^{tx}) = \sum_{r=0}^{n} e^{tx_r} p(r) = \sum_{r=0}^{n} e^{tr}\, {}^nC_r p^r q^{n-r}$$

$$= \sum_{r=0}^{n} {}^nC_r (pe^t)^r q^{n-r} = (q + pe^t)^n$$

We can obtain moments about origin from $M_X(t)$ as :

$$\mu_1' = \left(\frac{d}{dt}M_X(t)\right)_{t=0} = \left[n\,pe^t\left(q+pe^t\right)^{n-1}\right]_{t=0} = np$$

$$\mu_2' = \left[\frac{d^2}{dt^2}M_X(t)\right]_{t=0} = \left[np\frac{d}{dt}\left(q+pe^t\right)^{n-1}e^t\right]_{t=0}$$

$$= \left[n(n-1)p^2\left(q+pe^t\right)^{n-2}e^{2t} + npe^t\left(q+pe^t\right)^{n-1}\right]_{t=0}$$

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

We can thus derive mean and variance from the mgf as;

$$\text{mean} = \mu_1' = np$$

and variance $= \mu_2 = \mu_2' - \mu_1'^2 = n^2p^2 + npq - n^2p^2 = npq$

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

$$\mu_3' = \left[\frac{d^3M_X(t)}{dt^3}\right]_{t=0} = \left(\frac{d}{dt}\left[n(n-1)p^2e^{2t}\left(q+pe^t\right)^{n-2} + npe^t\left(q+pe^t\right)^{n-1}\right]\right)_{t=0}$$

$$= \left[2n(n-1)p^2e^{2t}\left(q+pe^t\right)^{n-2} + n(n-1)(n-2)p^3e^{3t}\left(q+pe^t\right)^{n-3}\right.$$

$$\left. + npe^t\left(q+pe^t\right)^{n-1} + np^2e^t(n-1)\left(q+pe^t\right)^{n-2}\right]_{t=0}$$

$$= 2n(n-1)p^2 + n(n-1)(n-2)p^3 + np + n(n-1)p^2$$

$$= n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np$$

Similarly we can obtain :

$$\mu_4' = \left(\frac{d^4M_X(t)}{dt^4}\right)_{t=0} = n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np$$

Remark 4 We can also obtain the moments about origin by directly using definition for moments about origin, as μ_1' and μ_2' are obtained in section 2.3.1 for Bernoulli distribution.

Central Moments of Binomial Distribution

We can use the interrelation between μ_r and μ'_r (as discussed in sec 3.2.2) to obtain μ_r.

$$\mu_1 = 0$$

$$\mu_2 = \mu'_2 - \mu'^2_1 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p) = npq$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1$$

$$= n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np - 3(n^2p^2 + npq)(np) + 2n^3p^3$$

$$= np\left[\left(n^2 - 3n + 2\right)p^2 + 3np - 3p + 1 - 3n^2p^2 - 3npq + 2n^2p^2\right]$$

$$= np\left[3np - 3np^2 + 2p^2 - 3p + 1 - 3npq\right]$$

$$= np\left[3np(1-p) + 2p^2 - 3p + 1 - 3npq\right]$$

$$= np\left[3npq + 2p^2 - 3p + 1 - 3npq\right] = np\left[2p^2 - 3p + 1\right]$$

$$= np(p-1)(2p-1) = np(1-p)(1-2p) = npq(1-2p)$$

$$= npq(q-p)$$

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1$$

$$= n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np$$

$$-4\left[n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np\right]\left[np\right] + 6\left[n^2p^2 + npq\right]\left[n^2p^2\right] - 3n^4p^4$$

$$= npq[1 + 3(n-2)pq]$$

Karl Pearson's β and γ coefficients for Binomial Distribution

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{[npq(q-p)]^2}{n^2p^2q^2} = \frac{(q-p)^2}{npq} = \frac{(1-2p)^2}{npq}$$

ans $\quad \gamma_1 = \sqrt{\beta_1} = \dfrac{1-2p}{\sqrt{npq}}$

Also $\quad \beta_2 = \dfrac{\mu_4}{\mu_2^2} = \dfrac{npq[1+3(n-2)pq]}{n^2p^2q^2} = \dfrac{[1+3(n-2)pq]}{npq}$

$$= \frac{1-6pq+3npq}{npq} = 3 + \frac{1-6pq}{npq}$$

and $\quad \gamma_2 = \beta_2 - 3 = \dfrac{1-6pq}{npq}.$

We can also obtain the central moments by mgf about mean as discussed below.

Moment generating function about $\overline{X}(=\mu = np)$

mgf about $\mu = M_X(t)$ (about μ) $= E\left(e^{t(x-\mu)}\right) = E\left(e^{(tx-\mu t)}\right) = e^{-\mu t}E\left(e^{tx}\right)$

$$= e^{-\mu t}M_X(t) \text{ (about origin)} = e^{-npt}\left(q+pe^t\right)^n \; (\because \mu = np \text{ for Binomial distribution})$$

$$= \left(qe^{-pt} + pe^te^{-pt}\right)^n$$

$$= \left(qe^{-pt} + pe^{qt}\right)^n$$

Now $\quad \mu_1 = \left[\dfrac{d}{dt}M_X(t) \text{ about mean}\right]_{t=0} = \left[n\left(qe^{-pt}+pe^{qt}\right)^{n-1}\left(-pqe^{-pt}+pqe^{qt}\right)\right]_{t=0} = 0$

$$\mu_2 = \left[\frac{d^2}{dt^2}\left[M_X(t) \text{ about mean}\right]\right]_{t=0}$$

$$= \left[n(n-1)\left(qe^{-pt}+pe^{qt}\right)^{n-2}\left(-pqe^{-pt}+pqe^{qt}\right)^2 + n\left(qe^{-pt}+pe^{qt}\right)^{n-1} \times \left(p^2qe^{-pt}+pq^2e^{qt}\right)\right]_{t=0}$$

$$= n(p^2q + pq^2) = npq(p+q) = npq$$

and so on.

Recurrence Relation for the Central Moments of Binomial Distribution

(Raj. IV Sem CP-2003)

By definition

$$\mu_r = E\left[(X-\mu)^r\right] = \sum_{x=0}^{n}(x-np)^r P(x)$$

$$\Rightarrow \quad \mu_r = \sum_{x=0}^{n}(x-np)^r \, {}^nC_x p^x q^{n-x} = \sum_{x=0}^{n} {}^nC_x (x-np)^r p^x (1-p)^{n-x}$$

Now differentiating both sides w.r.t. 'p' we get :-

$$\frac{d\mu_r}{dp} = \sum_{x=0}^{n} {}^nC_x \Big[(-n)(r)(x-np)^{r-1} p^x (1-p)^{n-x} + (x-np)^r \cdot$$

$$\left\{xp^{x-1}(1-p)^{n-x} - p^x (n-x)(1-p)^{n-x-1}\right\}\Big]$$

$$= \sum_{x=0}^{n}(-nr)(x-np)^{r-1} \, {}^nC_x p^x (1-p)^{n-x} + \sum_{x=0}^{n}(x-np)^r \, {}^nC_x p^x q^{n-x}\left\{\frac{x}{p} - \frac{n-x}{q}\right\}$$

$$= -nr\sum_{x=0}^{n}(x-np)^{r-1}P(x) + \sum_{x=0}^{n}(x-np)^r P(x)\left[\frac{xq - np + xp}{pq}\right]$$

$$= (-nr)\mu_{r-1} + \frac{1}{pq}\sum_{x=0}^{n}(x-np)^r P(x)\{x(p+q) - np\} \quad \left(\because \mu_{r-1} = E(X-np)^{r-1}\right)$$

$$= (-nr)\mu_{r-1} + \frac{1}{pq}\sum_{x=0}^{n}(x-np)^r P(x)\{x - np\}$$

$$= (-nr)\mu_{r-1} + \frac{1}{pq}\sum_{x=0}^{n}(x-np)^{r+1} P(x)$$

$$\Rightarrow \quad \frac{d\mu_r}{dp} = (-nr)\mu_{r-1} + \frac{1}{pq}\mu_{r+1}$$

$$\Rightarrow \quad \mu_{r+1} = pq\left[nr\mu_{r-1} + \frac{d\mu_r}{dp}\right]$$

which is the required recurrence relation for the central moments of Binomial distribution and gives the central moments by putting r = 1, 2, 3, ... successively in it as :-

$(\because \mu_0 = 1 \text{ and } \mu_1 = 0)$

$$r=1 \Rightarrow \quad \mu_2 = pq\left[n\mu_0 + \frac{d\mu_1}{dp}\right] = npq$$

$$r=2 \Rightarrow \quad \mu_3 = pq\left[2n\mu_1 + \frac{d\mu_2}{dp}\right] = pq\left[\frac{d}{dp}(npq)\right]$$

$$= npq\,[q-p] = npq(1-2p)$$

$$r=3 \Rightarrow \quad \mu_4 = pq\left[3n\mu_2 + \frac{d\mu_3}{dp}\right]$$

$$= pq\left[3n(npq) + \frac{d}{dp}npq(1-2p)\right]$$

$$= pq\left[3n^2pq + n\left(q(1-2p) - p(1-2p) - 2pq\right)\right]$$

$$= pq\left[3n^2pq + n\left(q - 4pq - p + 2p^2\right)\right]$$

$$= pq\left[3n^2pq + n\left((1-p) - 4pq - p + 2p^2\right)\right]$$

$$= pq\left[3n^2pq + n\left(1 - 4pq - 2p(1-p)\right)\right]$$

$$= pq\left[3n^2pq + n\left(1 - 4pq - 2pq\right)\right]$$

$$= pq\left[3n^2pq + n\left(1 - 6pq\right)\right]$$

$$= npq\left[3npq + 1 - 6pq\right]$$

$$= npq\left[1 + 3pq(n-2)\right] \text{ and so on.}$$

2.3.3 Probability Generating Function

$$P(s) = E\left(s^x\right) = \sum_{x=0}^{n} s^x P(x) = \sum_{x=0}^{n} {}^nC_x\, p^x q^{n-x} s^x$$

$$= \sum_{x=0}^{n} {}^nC_x (ps)^x q^{n-x} = (q + ps)^n$$

The coefficients of the respective powers of s, generate the Binomial probability distribution.

2.3.4 Mode of the Binomial Distribution

As it has been discussed in chapter 3, mode is the value of x for which P(x) is maximum. Let X = r be the modal value.

$\Rightarrow \quad P(X = r) > P(X = r - 1)$ and $P(X = r) > P(X = r + 1)$

Now $$\frac{P(X=r)}{P(X=r-1)} = \frac{{}^nC_r p^r q^{n-r}}{{}^nC_{r-1} p^{r-1} q^{n-r+1}} = \frac{(n-r+1)}{r}\frac{p}{q}$$

Now $$P(X=r) > P(X=r-1) \Rightarrow \frac{P(X=r)}{P(X=r-1)} > 1$$

$\Rightarrow \quad (n - r + 1)p > qr$

$\Rightarrow \quad (qr + pr) < p + np \quad \Rightarrow \quad r < (n + 1)p$ .....(2.3)

Also $\quad P(X = r) > P(X = r + 1)$

Now $$\frac{P(X=r)}{P(X=r+1)} = \frac{{}^nC_r p^r q^{n-r}}{{}^nC_{r+1} p^{r+1} q^{n-(r+1)}} = \frac{r+1}{(n-r)}\frac{q}{p}$$

Now $$P(X=r) > P(X=r+1) \Rightarrow \frac{P(X=r)}{P(X=r+1)} > 1$$

$$\Rightarrow \quad \frac{(r+1)q}{(n-r)p} > 1 \Rightarrow np - q < (rp + rq)$$

$\Rightarrow \quad r > np - q = np - 1 + p = (n + 1)p - 1$ .....(2.4)

(2.3) and (2.4) $\Rightarrow np - q < r < (n + 1)p$

or $\quad [(n+1)p-1] < r < (n+1)p$

If $(n + 1)p$ is an integer it has two modal values $m = (n + 1)p$ and $m - 1 = (n + 1)p - 1$.

If $(n + 1)p$ is not an integer there exists a unique modal value which is m = integral part of $[(n + 1)p]$.

2.3.5 Fitting of Binomial Distribution (Recurrence Relation for the Probabilities of Binomial Distribution)

We have $\dfrac{P(r+1)}{P(r)} = \dfrac{{}^nC_{r+1}p^{r+1}q^{n-(r+1)}}{{}^nC_r p^r q^{n-r}} = \dfrac{\underline{|n}}{\underline{|r+1}\,\underline{|n-(r+1)}} \cdot \dfrac{\underline{|n-r}\,\underline{|r}}{\underline{|n}} \cdot \dfrac{p}{q} = \dfrac{n-r}{r+1}\dfrac{p}{q}$

$$\Rightarrow \qquad P(r+1) = \frac{(n-r)p}{(r+1)q}P(r)$$

which is the required recurrence formula for the probabilities of Binomial distribution. This relation is helpful in fitting of Binomial distribution. Fitting a Binomial distribution means to find the theoretical frequencies for a given frequency distribution.

Let for a given frequency distribution x_i/f_i, $r = 1, 2, \ldots, n$, and $\Sigma f_i = N$. Then the parameter, p, of the Binomial distribution can be obtained as :-

$$\text{Mean of Binomial distribution } (np) = \frac{\sum_i x_i f_i}{N}.$$

Hence we can find the respective probability distribution and get the respective theoretical frequencies from successive terms in the expansion of $N(p+q)^n$.

***Example* 5.** The following data gives the number of seeds germinating out of 10 on damp filter paper for 80 sets of seeds. Fit a Binomial distribution to this data,

No of seeds (x) :	0	1	2	3	4	5	6 and above
No of sets (f) :	6	20	28	12	8	6	0

Solution : Here $n = 10$, $\Sigma f_i = 80$

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i} = \frac{174}{80} = 2.175 = np \text{ (mean of Binomial distribution)}$$

$$n = 10 \Rightarrow p = \frac{2.175}{10} = 0.2175, q = 1 - p = 0.7825.$$

Hence the Binomial distribution to be approximated for this data $= N(q+p)^n$.

$= 80(0.7825 + 0.2175)^{10}$.

Writing it in tabular form we get :

r	$P(r) = {}^nC_r\, p^r q^{n-r}$	$f(r) = N\,P(r) = 80P(r)$
0	$P(0) = q^{10} = (0.7825)^{10} = 0.08607$	$80\times0.08607 \cong 6.9$
1	$P(1) = {}^{10}C_1\,(0.2175)(0.7825)^9 = 0.2392$	$80\times0.2392 \cong 19.1$
2	$P(2) = {}^{10}C_2\,(0.2175)^2(0.7825)^8 = 0.2992$	$80\times0.2992 \cong 23.9$
3	$P(3) = {}^{10}C_3\,(0.2175)^3(0.7825)^7 = 0.2218$	$80\times0.2218 \cong 17.7$
4	$P(4) = {}^{10}C_4\,(0.2175)^4(0.7825)^6 = 0.1079$	$80\times0.1079 \cong 8.6$
5	$P(5) = {}^{10}C_5\,(0.2175)^5(0.7825)^5 = 0.0359$	$80\times0.0359 \cong 2.9$
6	$P(6) = {}^{10}C_6\,(0.2175)^6(0.7825)^4 = 0.0083$	$80\times0.0083 \cong 0.8$
7	$P(7) = {}^{10}C_7\,(0.2175)^7(0.7825)^3 = 0.0013$	$80\times0.0013 \cong 0.1$
8 and above	negligible	$\cong 0$

Hence the theoretical frequency distribution as obtained by fitting of Binomial distribution is :-

x	0	1	2	3	4	5	6	4	8 and above
f	6.9	19.1	23.9	17.7	8.6	2.9	0.8	0.1	0

***Example* 6.** Comment upon the statement that "Mean of Binomial Distribution is 3 and variance is 4".

Solution : Mean = np = 3 and Var = npq = 4

Dividing we get $q = 4/3 > 1$, which is not possible.

Hence given statement is wrong.

***Example* 7.** The mean and variance of a Binomial distribution are 4 and 4/3 respectively. Find $P(X \geq 1)$.

Solution : Given np = 4 and $npq = 4/3$

Dividing them we get, $q = \frac{1}{3}$ and $p = 1 - q = \frac{2}{3}$

$$\therefore \quad np = 4 \Rightarrow n = \frac{4}{p} = \frac{4\times3}{2} = 6$$

$$\therefore \quad P(X \geq 1) = 1 - P(X = 0) = 1 - q^6 = 1 - \left(\frac{2}{3}\right)^6 = 1 - \frac{1}{729} = 0.9986$$

***Example* 8.** Find the most probable number of heads (successes) obtained in 100 tosses .

Solution : Here n =100, $p = \frac{1}{2} \Rightarrow (n+1)p$ is not an integer.

$\therefore$ Most probable number of heads = Mode = integral part of $[(n+1)p]$

= integral part of (50.5)

= 50.

Illustrative Examples

Example **9.** If 10% of the pens manufactured by the company are defective, find the probability that a box of 12 pens contain

(i) Exactly two defective pens

(ii) Atleast two defective pens

(iii) No defective pen.

Solution : Let X denote the number of defective pens.

Here $n = 12,\ p = \dfrac{10}{100} = 0.1$ and $q = 1 - p = 0.9$

and $P(X = r) = {}^nC_r\, p^r\, q^{n-r}$.

(i) Required probability $= P(X = 2) = {}^{12}C_2\,(0.1)^2(0.9)^{10} = 0.2301$

(ii) $P(X \geq 2) = 1 - [P(0) + P(1)]$

$$= 1 - \left[{}^{12}C_0(0.1)^0(0.9)^{12} + {}^{12}C_1(0.1)^1(0.9)^{11}\right]$$

$$= 1 - [0.2824 + 0.3766]$$

$$= 1 - 0.659$$

$$= 0.341$$

(iii) $P(X = 0) = {}^{12}C_0\,(0.1)^0(0.9)^{12} = 0.2824.$

***Example* 10.** A box contains 'a' white and 'b' black balls. If balls be drawn from the box, with replacement, find the expected number of white balls in n draws.

Solution : Here X denotes number of white balls,

and number of draws = n

$$p = P(\text{white ball drawn}) = \frac{a}{a+b} \text{ and } q = 1 - p = \frac{b}{a+b}.$$

Also $P(X = r) = {}^nC_r\, p^r\, q^{n-r}$.

Therefore expected number of white balls in n draws $= n \times p = \dfrac{na}{a+b}$.

***Example* 11.** An irregular six faced dice is thrown and the probability that it gives five even numbers in 10 throws is twice the probability that it gives four even numbers in 10 throws. How many times in 10,000 sets of 10 throws each, would you expect to get no even number ?

Solution : Let X denotes the number of times an even number is obtained.

Let $p = P(\text{get an even number})$. Here, $n = 10$.

Also $P(X = r) = {}^nC_r\, p^r\, q^{n-r}, r = 0, 1, 2, \ldots n.$

Given $P(X = 5) = 2P(X = 4)$

$\Rightarrow$ ${}^{10}C_5\, p^5\, q^5 = 2 \times {}^{10}C_4\, p^4\, q^6.$

$\Rightarrow$ $252p = (210)(2q)$ $\Rightarrow$ $3p = 5q = 5(1 - p)$

$\Rightarrow$ $8p = 5$ $\Rightarrow$ $p = 5/8$

$\Rightarrow$ $q = 1 - p = \dfrac{3}{8}$

Hence $P(X = r) = {}^{10}C_r \left(\dfrac{5}{8}\right)^r \left(\dfrac{3}{8}\right)^{10-r}.$

Hence the probability of getting no even number $= P(X = 0) = \left(\dfrac{3}{8}\right)^{10} = 0.00005$

Hence the required number of times that in 10,000 sets of 10 throws each, we get no even number $= 10{,}000 \times P(X = 0) = 0.549 \cong 1$

***Example* 12.** Probability that a man aged 60 would be alive till the 70 years of age is 0.65. Find the probability that at least 7 out of 10 such men would be alive till 70 years of age. **(Raj. IV Sem. IT-2003)**

Solution : Here X denotes the number of men now 60 and would be alive till 70 years of age.

Given $n = 10$, $p = 0.65$, $q = 1 - p = 0.35$.

Also $P(X = r) = {}^nC_r\, p^r\, q^{n-r}.$

Hence required probability $= P(X \geq 7) = P(7) + P(8) + P(9) + P(10)$

$$= {}^{10}C_7(0.65)^7(0.35)^3 + {}^{10}C_8(0.65)^8(0.35)^2 + {}^{10}C_9(0.65)^9(0.35) + (0.65)^{10}$$

$$= 120 \times 0.00210 + 45 \times 0.00390 + 10 \times 0.00725 + 0.01346$$

$$= 0.252 + 0.1755 + 0.0725 + 0.01346 = 0.513$$

***Example* 13.** Find the parameters of the Binomial distribution whose mean is 10 and variance 6.

Solution : Let $X \sim B(n, p)$.

Mean $= np = 10$ and Variance $= npq = 6$.

Dividing these we get $q = \frac{6}{10} = \frac{3}{5} \Rightarrow p = 1 - q = \frac{2}{5}$

$\therefore$ $np = 10 \Rightarrow n = \frac{10 \times 5}{2} = 25$

$\therefore$ Here $n = 25$, $p = 0.4$ and $q = 0.6$.

***Example* 14.** Out of 800 families with 4 children each, how many families would be expected to have (i) 2 boys and 2 girls (ii) atleast 1 boy (iii) at most 2 girls and (iv) children of both sex. Assume equal probabilities for boys and girls.

Solution : Let random variable X denote the number of girls.

Here $n = 4$, $N = 800$ and $p = q = 1/2$ and $P(X = r) = {}^nC_r\, p^r q^{n-r}$.

(i) $P(\text{2 boys and 2 girls}) = P(X = 2) = {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = 6 \times \left(\frac{1}{2}\right)^4 = \frac{3}{8}$

$\therefore$ No of families having 2 boys and 2 girls $= 800 \times \frac{3}{8} = 300$

(ii) No. of boys $= 1, 2, 3, 4$

$\therefore$ $X = 0, 1, 2, 3$

Hence $P(X \le 3) = 1 - P(X = 4) = 1 - {}^4C_4\left(\frac{1}{2}\right)^4 = 1 - \frac{1}{16} = \frac{15}{16}$

Hence number of families having at least 1 boy $= 800 \times \frac{15}{16} = 750$

(iii) $P(\text{at most 2 girls}) = P(X \le 2)$

$$= P(X = 0) + P(X = 1) + P(X = 2)$$

$$= {}^4C_0\left(\frac{1}{2}\right)^4 + {}^4C_1\left(\frac{1}{2}\right)^4 + {}^4C_2\left(\frac{1}{2}\right)^4$$

$$=\left(\frac{1}{2}\right)^4[1+4+6]=\frac{11}{16}$$

Hence number of families having at most two girls $=800\times\frac{11}{16}=550$.

(iv) P (children of both sexes) = 1 –P(all children are of same sex)

$= 1-\{P(\text{all children are boys})+P(\text{all children are girls})\}$

$$=1-\{P(X=0)+P(X=4)\}=1-\left[\left(\frac{1}{2}\right)^4+\left(\frac{1}{2}\right)^4\right]$$

$$=1-\frac{2}{16}=1-\frac{1}{8}=\frac{7}{8}$$

Hence number of families having children of both sexes $=800\times\frac{7}{8}=700$.

***Example* 15.** The probability of a successful rocket launching is p. If the launching attempts are made until 3 successful launchings have occurred,

(i) What is the probability that exactly 5 attempts will be necessary ? What is the probability that fewer than 5 attempts will be necessary ?

(ii) If launching attempts are made until 3 consecutive successful launching have occurred, what are the respective probabilities ?

Solution : Let random variable X denote number of successful rocket launching attempts.

(i) Now 3 successes will be obtained in 5 attempts if 2 successes occur anytime in first four attempts while the third success is obtained in fifth attempt.

Hence P (exactly 5 attempts are required)

$= P(2 \text{ successes in } 4 \text{ attempts}) \times P(\text{success in 5th attempt})$

$= {}^4C_2\, p^2 q^2 \times p$ $\quad [\because X \sim B(4, p)]$

$= 6p^3 q^2$

Now P{fewer than 5 attempts are required}

$= P(\text{success in 3 attempts}) + P(\text{success in 4 attempts})$

$= P(X = 3) + P(X = 2) \times p$

$= {}^3C_3\, p^3q^0 + {}^3C_2\, p^2 q \times p$ $\quad$ [Here $X \sim B(3,p)$]

$$= p^3 + 3p^3 q = p^3(1 + 3q)$$

(ii) Now five attempts will be required to get 3 consecutive successes if first two attempts result in failure and last 3 attempts in success.

Hence P (exactly 5 attempts required) $= q\,q\,p\,p\,p = q^2 p^3$

Also P (less than 5 attempts required)

$$= \text{P (exactly 3 attempts)} + \text{P (exactly 4 attempts)}$$

$$= p.p.p + q.p.p.p$$

$$= p^3 + q\,p^3 = p^3(1 + q).$$

***Example* 16.** A communication system consists of n components, each of which will independently function with probability p. The total system will be able to operate effectively if at least one half of it's components function. For what values of p is a 5 component system more likely to operate effectively than a 3 component system.

Solution : Here the random variable X, denoting the number of components functioning is a Binomial variate $X \sim B(n, p)$.

and $P(X = r) = {}^nC_r\, p^r\, q^{n-r},\ r = 0, 1, 2, \ldots n.$

P(5 component system functions effectively)

$$= P(X = 3) + P(X = 4) + P(X = 5) = \sum_{r=3}^{5} {}^5C_r p^r q^{5-r} \qquad (\because \text{ here } n = 5)$$

Also P(3 component system functions effectively)

$$= P(X = 2) + P(X = 3) = \sum_{r=2}^{3} {}^3C_r p^r q^{3-r} \qquad (\because \text{ here } n = 3)$$

It is required to find p, such that

$$\sum_{r=3}^{5} {}^5C_r p^5 q^{5-r} \geq \sum_{r=2}^{3} {}^3C_r p^r q^{3-r}$$

$$\Rightarrow \quad 10p^3 q^2 + 5p^4 q + p^5 \geq 3p^2 q + p^3$$

$$\Rightarrow \quad 10p^3 (1 - p)^2 + 5p^4 (1 - p) + p^5 \geq 3p^2 (1 - p) + p^3$$

$$\Rightarrow \quad 3p^2 (2p^3 - 5p^2 + 4p - 1) \geq 0$$

$$\Rightarrow \quad (3p^2)(p - 1)^2 (2p - 1) \geq 0$$

As $3p^2 (p - 1)^2 \geq 0 \quad \Rightarrow \quad 2p-1 \geq 0 \quad \Rightarrow \quad p \geq 1/2.$

***Example* 17.** In how many throws of a dice the probability of throwing 6 atleast once is just greater than 0.5.

Solution : Here random variable X denotes the number of times 6 is obtained

Also $P(\text{getting } 6) = \frac{1}{6} = p \quad \Rightarrow q = 1 - \frac{1}{6} = \frac{5}{6}$

Let $X \sim B(n, p)$

$\therefore \quad P(X = r) = {}^nC_r\, p^r q^{n-r}, \; r = 0, 1, 2, \ldots n$

Given $P(X \geq 1) > 0.5$

$\Rightarrow \quad 1 - P(X = 0) > 0.5 \quad \Rightarrow \quad P(X = 0) < 0.5$

$\Rightarrow \quad {}^nC_o\, p^o q^n < 0.5$

$\Rightarrow \quad \left(\frac{5}{6}\right)^n < 0.5$

$\Rightarrow \quad n \log \frac{5}{6} < \log 0.5$

$\Rightarrow \quad n(-0.0792) < (-0.30103)$

$\Rightarrow \quad n(0.0792) > (0.30103)$

$\Rightarrow \quad n > \frac{0.30103}{0.0792} = 3.8$

$\Rightarrow \quad n = 4, 5, 6 \ldots$

Hence minimum number of required throws = 4.

Example 18. Eight coins are tossed simultaneously 256 times. Number of heads observed at each throw are recorded and the results are as given below. Find the expected frequency and fit a Binomial distribution. What are the theoretical values of the mean and standard deviation ? Also calculate the mean and standard deviation of the observed frequencies.

No. of Heads (X) :	0	1	2	3	4	5	6	7	8	Total
No. of times (f) :	2	6	30	52	67	56	32	10	1	256

Solution : Observed Mean $= \mu_{obs} = \frac{\sum f_i x_i}{\sum f_i} = \frac{1040}{256} = 4.0625$

Also $E(X^2) = \frac{\sum f_i x_i^2}{\sum f_i}$

$$= \frac{(1\times 6+4\times 30+9\times 52+16\times 67+25\times 56+36\times 32+49\times 10+64\times 1)}{256}$$

$$= \frac{4772}{256} = 18.6406$$

$\therefore$ Observed Variance $\sigma^2_{obs} = E(X^2) - [E(X)]^2 = 18.6406 - 16.5039 = 2.1367$

$\therefore$ Observed standard deviation = 1.4617

Here $n = 8$ (number of coins), $p = P$ (head) = 1/2 and $q = 1 - p = 1/2$

$\therefore$ Theoretical Mean $= np = 8 \times \frac{1}{2} = 4$

and Theoretical standard deviation $= \sqrt{npq} = \sqrt{8 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{2} = 1.414$

To find the expected frequencies (i.e. to fit the Binomial distribution) :-

$$\text{Mean} = np = 4.0625 \Rightarrow p = \frac{4.0625}{8} \cong 0.5078 \Rightarrow q = 1 - p = 0.4922$$

$\therefore$ Expected frequencies are given by respective terms in expansion of $= 256(0.4922 + 0.5078)^8$.

Example 19. For a special security in a certain protected area, it was decided to put three lighting bulbs on each pole. If each bulb has a probability p of burning out in the first 100 hours of service, calculate the probability that at least one of them is still good after 100 hours..

If $p = 0.3$, how many bulbs will be needed on each pole to ensure 99% safety so that at least one is good after 100 hours.

Solution : Here let random variable X denote the number of bulbs that do not burn out in the first 100 hours. Hence P(success) $= 1 - p$ and P(failure) $= p$

Here $n = 3$ (three lightining bulb on each pole)

Hence Required probability $= P(X \geq 1) = 1 - P(X = 0)$

$$= 1 - {}^3C_0 (1-p)^0 (p)^3$$

$$= 1 - p^3$$

$$= 1 - (0.3)^3 = 0.973$$

Now if $p = 0.3$, let number of bulbs on each pole be n to ensure 99% safety so that at least one is good after 100 hours.

Then, $P(X \geq 1) = 1 - P(X = 0) = 1 - {}^nC_0 (1-p)^0 p^n = 0.99$

$$\Rightarrow 1-(0.3)^n = 0.99$$

$$\Rightarrow 1-0.99 = (0.3)^n \Rightarrow n \log 0.3 = \log 0.01 \Rightarrow n = \frac{2}{0.5229} = 3.8$$

$$\Rightarrow n \cong 4$$

***Example* 20.** The headless matches per box of 50 matches in a total of 100 boxes is given in the following table.

No of headless matches per box :	0	1	2	3	4	5	6	7
No of boxes :	12	27	29	19	8	4	1	0

Assuming the Binomial distribution, calculate the expected frequencies.

Solution : The mean of the given distribution $= \frac{\sum f_i x_i}{\sum f_i} = \frac{200}{100} = 2 = np$ (mean of (Binomial distribution)

Here $n = 50$ and $N = 100$

$\therefore$ $np = 2 \Rightarrow p = \frac{2}{50} = 0.04 \quad \Rightarrow q = 1-0.04 = 0.96$

The theoretical frequencies are given by successive terms in the expansion of $N(q+p)^n$ i.e. in the expansion of $N \cdot {}^nC_r\, p^r q^{n-r}$. $\quad r = 0, 1, 2, \ldots 50.$

$= 100 \times {}^{50}C_r (0.04)^r (0.96)^{50-r}.$ $\quad r = 0, 1, 2, \ldots 50.$

***Example* 21.** The chance that one of the 10 telephone lines is busy at any instant is 0.2,

(i) What is the probability that at a certain instant 5 lines are busy ?

(ii) Find the most probable number of busy lines and the probability of this number.

(iii) What is the probability that all lines are busy ?

Solution : Here $X \sim B(10, 0.2)$, i.e. $n = 10$, $p = 0.2$, $q = 0.8$

and $P(X = r) = {}^nC_r\, p^r q^{n-r}$, $\quad r = 0, 1 \ldots 10.$

(i) $P(X = 5) = {}^{10}C_5 (0.2)^5 (0.8)^5 = 0.0264$

(ii) If 'r' is the most probable number then $[(n+1)p-1] \le r \le (n+1)p$

As $(n+1)p = (10+1)0.2 = 11 \times 0.2 = 2.2$ (not an integer)

$\therefore$ r = integral part of $(2.2) = 2$

Also probability of this number is given by

$$P(X = r) = P(X = 2) = {}^{10}C_2\, p^2 q^8 = {}^{10}C_2 (0.2)^2 (0.8)^8 = 0.30199$$

(iii) $P(X=10) = {}^{10}C_{10}(0.2)^{10}(0.8)^0 = 1.02 \times 10^{-7}$.

Example 22. Fit a Binomial distribution to the following data :-

x :	0	1	2	3	4	5
f :	2	14	20	34	22	8

Solution : Mean of given frequency distribution

$$= \frac{\sum fx}{\sum f} = \frac{284}{100} = 2.84 = np(\text{mean of Binomial distribution})$$

Here $n = 5$, hence $np = 2.84 \Rightarrow p = \frac{2.84}{5} = 0.568, \Rightarrow q = 1 - p = 0.432$

Hence, the expected frequencies are given by the successive terms in the expansion of $100 \times (0.432 + 0.568)^5$ as indicated in the following table :-

x	$\mathbf{P(x) = {}^5C_x p^x q^{5-x}}$	$\mathbf{f(x) = 100 \times P(x)}$
0	${}^5C_0 (0.568)^0 (0.432)^5 = 0.01505$	1.5
1	${}^5C_1 (0.568)^1 (0.432)^4 = 0.09891$	9.9
2	${}^5C_2 (0.568)^2 (0.432)^3 = 0.26010$	26.0
3	${}^5C_3 (0.568)^3 (0.432)^2 = 0.34199$	34.2
4	${}^5C_4 (0.568)^4 (0.432) = 0.2248$	22.5
5	${}^5C_5 (0.568)^5 (0.432)^0 = 0.05912$	5.9

Example 23. The probability of a man hitting a target is 1/2. How many times must he fire so that the probability of hitting the target at least once is more than 90% ?

Solution : Let X be a Binomial Variate with parameters n and $p = 1/2 = 0.5 \Rightarrow q = 1 - p = 0.5$

Given $P(X \geq 1) > \frac{90}{100} \Rightarrow 1 - P(X=0) > 0.9$

$\Rightarrow \quad P(X = 0) < 0.1$

$\Rightarrow \quad {}^nC_0 (0.5)^0 (0.5)^n < 0.1$

$\Rightarrow \quad (0.5)^n < 0.1 \quad \Rightarrow \quad n \log (0.5) < \log(0.1)$

$\Rightarrow \quad -n \times 0.3010 < -1$

$\Rightarrow \quad n > \frac{1}{0.3010} = 3.32$

$\Rightarrow \quad n = 4, 5, 6.....$

$\therefore$ Minimum number of times he has to fire to meet above said requirement = 4.

Example 24. The probability is 0.02 that an item produced by a factory is defective. A shipment of 10,000 items is sent to its warehouse. Find the expected number of defective items and standard deviation.

Solution : Here X can be assumed to be a Binomial variate with parameters n and p.

Here, $n = 10{,}000$

$\therefore$ Expected number of defective items $= 100000 \times 0.02 = 200$

Standard deviation $= \sqrt{npq} = \sqrt{196} = 14$.

Example **25.** In sampling a large number of parts manufactured by a machine, the mean number of defectives in a sample of 20 is 2. Out of 1000 such samples, how many would be expected to contain at least 3 defective parts.

Solution : Here $X \sim (20, p)$.

Mean number of defectives $= np = 2 \Rightarrow p = \dfrac{2}{20} = 0.1 \Rightarrow q = 1 - p = 0.9$

Therefore, the probability of at least three defectives $= P(X \geq 3) = 1 - P(X \leq 2)$

$$= 1 - [P(0) + P(1) + P(2)]$$

$$= 1 - {}^{20}C_0 (0.1)^0 (0.9)^{20} + {}^{20}C_1 (0.1)^1 (0.9)^{19} + {}^{20}C_2 (0.1)^2 (0.9)^{18}]$$

$$= 1 - [(0.9)^{20} + 2(0.9)^{19} + (1.9)(0.9)^{18}]$$

$$= 1 - (0.9)^{18} [(0.9)^2 + 1.8 + 1.9]$$

$$= 1 - (0.9)^{18} [4.51] = 0.3230$$

Hence the number of samples having at least three defective parts

$$= 1000 \times 0.3230 = 323$$

Example **26.** A department has 10 machines which many need adjustment from time to time during the day. Three of these machines are old, each having a probability of 1/11 of needing adjustment during the day and 7 are new, having the corresponding, probability of 1/21. Assuming that no machine needs adjustment twice on the same day, find the probabilities that on a particular day

(i) just 2 old and no new machine need adjustment.

(ii) if just 2 machines need adjustment, they are of same type.

Solution : Let X and Y be the Binomial random variates denoting the number of old and new machines respectively, which need adjustment.

Hence $X \sim B\left(3, \dfrac{1}{11}\right)$ and $Y \sim B\left(7, \dfrac{1}{21}\right)$

(i) Required probability $= P(X = 2) \times P(Y = 0)$

$$= {}^3C_2\left(\frac{1}{11}\right)^2 \times \left(\frac{10}{11}\right) \times {}^7C_0 \times \left(\frac{1}{21}\right)^0\left(\frac{20}{21}\right)^7$$

$$= 3 \times \frac{10}{11^3} \times \frac{(20)^7}{(21)^7} = 0.016$$

(ii) Required probability $= P(X = 2) \times P(Y = 0) + P(X = 0) \times P(Y = 2)$

$$= 0.016 + {}^3C_0\left(\frac{1}{11}\right)^0\left(\frac{10}{11}\right)^3 \times {}^7C_2\left(\frac{1}{21}\right)^2\left(\frac{20}{21}\right)^5$$

$$= 0.016 + (0.7513) \times (0.373)$$

$$= 0.016 + 0.028$$

$$= 0.044$$

***Example* 27.** Assuming that 20% of the population of a city are literate, so that the chance of an individual being literate is 1/5 and assuming that 100 investigators each take 10 individuals to see whether they are literate, how many investigators would you expect to report that 3 or less were literate ?

Solution : Here, $X \sim B(10, .0.2)$, where $n = 10, p = \frac{20}{100} = 0.2$

Probability that 3 or less people are literate $= P(X \le 3) = P(0) + P(1) + P(2) + P(3)$

$$= {}^{10}C_0 (0.2)^0 (0.8)^{10} + {}^{10}C_1 (0.2)^1 (0.8)^9 + {}^{10}C_2 (0.2)^2 (0.8)^8 + {}^{10}C_3 (0.2)^3 (0.8)^7$$

$$= 0.8791$$

Hence number of investigators which report that 3 or less people are literate

$$= 100 \times 0.8791 = 87.91 \cong 88$$

***Example* 28.** An experiment succeeds twice as often as it fails. Find the chance that in the next 6 trials, there will be at least four successes.

Solution : Here X denotes the success of an experiment.

Given $p = 2q \Rightarrow p = \frac{2}{3}$ and $q = \frac{1}{3}$

Also $n = 6$

Hence required probability $= P(X \ge 4) = P(4) + P(5) + P(6)$

$$= {}^6C_4\, p^4 q^2 + {}^6C_5\, p^5 q + {}^6C_6\, p^6 q^0$$

$$= 15 \times \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^2 + 6 \times \left(\frac{2}{3}\right)^5 \times \frac{1}{3} + \left(\frac{2}{3}\right)^6$$

$$= \frac{16}{729} \times 31 = 0.68.$$

Example 29. The sum of mean and variance of a Binomial distribution is 15 and the sum of their squares is 117. Determine the distribution.

Solution : Let n and p be the parameters of the distribution.

Then Mean = np and variance = npq

Now $np + npq = 15 \Rightarrow np(1 + q) = 15 \Rightarrow n^2p^2(1 + q)^2 = 225$

and $n^2p^2 + n^2p^2q^2 = 117 \Rightarrow n^2p^2(1 + q^2) = 117$

Dividing them we get :-

$$\frac{(1+q)^2}{1+q^2} = \frac{225}{117} \Rightarrow \frac{q^2+2q+1}{1+q^2} = \frac{225}{117}$$

$$\Rightarrow \quad 1 + \frac{2q}{1+q^2} = \frac{225}{117} \Rightarrow \frac{2q}{1+q^2} = \frac{12}{13}$$

$$\Rightarrow \quad \frac{1+q^2}{2q} = \frac{13}{12} \Rightarrow \frac{1+q^2+2q}{1+q^2-2q} = \frac{13+12}{13-12}$$

(by C & D rule)

$$\Rightarrow \quad \left(\frac{1+q}{1-q}\right)^2 = 25 \Rightarrow \frac{1+q}{1-q} = 5 \Rightarrow 6q = 4 \Rightarrow q = \frac{2}{3}$$

$$\Rightarrow \quad p = 1 - q = 1 - \frac{2}{3} = \frac{1}{3}$$

Also $np + npq = 15 \Rightarrow \frac{5n}{9} = 15$

$$\left(\because p = \frac{1}{3}, q = \frac{2}{3}\right)$$

$$\Rightarrow \quad n = 27$$

Hence the required distribution is :

$$P(X=r) = {}^{27}C_r\left(\frac{1}{3}\right)^r\left(\frac{2}{3}\right)^{27-r}, \quad r = 0, 1, 2, \ldots 27.$$

***Example* 30.** A manufacturer of fax machine claims that only 10% of his machines require repairs within the warranty period of one year. If 5 of 20 of his machines require repairs within one year, does this tend to support or refute the claim.

Solution ; Here $X \sim B(20,\ 0.10)$

Now probability that 5 or more machines require repairs is :-

$$P(X \geq 5) = 1 - P(X < 5)$$

$$= 1 - [P(0) + P(1) + P(2) + P(3) + P(4)]$$

$$= 1 - 0.9568 = 0.0432$$

Since this value is very small, the FAX machine manufacturer's claim should be rejected.

***Example* 31.** A Binomial variate satisfies the condition $9P(X=4) = P(X=2)$. If $n = 6$ find p, $\bar{x}$ and σ.

Solution :

$$9P(X=4) = P(X=2)$$

$$\Rightarrow \quad 9 \times {}^6C_4\, p^4(1-p)^2 = {}^6C_2 \times p^2(1-p)^4.$$

$$\Rightarrow \quad 9 \times p^2 = (1-p)^2$$

$$\Rightarrow \quad 9p^2 = 1 + p^2 - 2p$$

$$\Rightarrow \quad 8p^2 + 2p - 1 = 0$$

$$\Rightarrow \quad (4p-1)(2p+1) = 0 \Rightarrow p = \frac{1}{4}, -\frac{1}{2}$$

But $p = -1/2$ is not possible $\Rightarrow p = 1/4$

Hence $\text{Mean} = np = 6 \times \frac{1}{4} = \frac{3}{2}$

and standard deviation $= \sqrt{npq} = \sqrt{6 \times \frac{1}{4} \times \frac{3}{4}} = \sqrt{1.125} = 1.0607.$

Exercise 2(A)

Q.1 Find the total number of trials when the mean number of trials is 9 and variance is 6.

Ans. $n = 27$

Q.2 Can there be a Binomial distribution with mean 5 and standard deviation 3 ?

Ans. No.

Q.3 Find the mean and variance of the number of heads in 5 tosses of an unbiased coin.

Ans. 2.5, 1.25

Q.4 An irregular 6 faced dice is such that the probability that it gives 3 even numbers in 5 throws is twice the probability that it gives 2 even numbers in 5 throws. How many sets of exactly 5 trials can be expected to give no even number out of 2500 sets ?

Ans. nearly 10

Q.5 If the probability of functioning of a component in a system consisting of n components, each of which work independently is p. Find the expected number of components which are not functioning in the system, given that there is at least one component that is not functioning.

Ans. $\dfrac{nq}{1-p^n}$

Q.6 In a hurdle race, a player has to cross 10 hurdles. The probability of clearing a hurdle is 5/6. What is the probability that a player will knock down fewer than 2 hurdles ?

Ans. $\dfrac{25}{12} \times \left(\dfrac{5}{6}\right)^8$

Q.7 Find the moment generating function of the Binomial distribution and hence find its mean and variance.

Ans. $(q + p\,e^t)^n$, np, npq

Q.8 Out of 800 families with 5 children each, how many families would be expected to have :-

(i) 3 boys

(ii) 5 girls

(iii) either 2 or 3 boys

(iv) at most 2 girls

(v) no girl

Assume equal probabilities for boys and girls.

Ans. (i) 250 (ii) 25 (iii) 500 (iv) 150 (v) 25

Q.9 Six dice are thrown 729 times. How many times do you expect atleast three dice to show a five or a six ?

Ans. 233

Q.10 Define Binomial distribution. Find its mean and variance.

Ans. np, npq

Q.11 Fit a Binomial distribution for the following data and hence find the theoretical frequencies :-

x :	0	1	2	3	4
f :	5	29	36	25	5

Ans. 7, 26, 37, 34, 6.

Q.12 10% screws produced by a machine are defective. Find the probability of the following when they are checked at random by examining a sample of 5 screws (i) none is defective (ii) one is defective (iii) atmost one is defective.

Ans. (i) 0.5905 (ii) 0.3281 (iii) 0.9186

Q.13 A coin is tossed 4 times. What is the probability of getting (i) two heads (ii) atleast two heads ?

Ans. (i) 0.375 (ii) 0.6875

Q.14 The overall percentage of failures in a certain examination is 20. If six candidates appear in the examination, what is the probability that atleast 5 pass in the examination ?

Ans. 0.655

Q.15 Assuming that half the population of a town consumes choclates and that 100 investigators, each take 10 individuals to see whether they are consumers; how many investigators would you expect to report that three people or less were consumers ?

Ans. 17

Q.16 A man takes a step forward with probability 0.4 and backward with probability 0.6. Find the probability that at the end of eleven steps he is just one step away from the starting point.

Ans. 0.3679

Q.17 In a precision bombing attack there is a 50% chance that any bomb will strike the target. Two direct hits are required to destroy the target completely. How many bombs must be dropped to give a 99% chance or better, of completely destroying the target ?

Ans. 11

Q.18 The probability of a man hitting a target is 0.25. How many times must he fire so that the probability of hitting the target at least once is greater than 2/3 ?

Ans. 4

Q.19 An underground mine has 5 pumps installed for pumping out storm water, the probability of any one of the pumps failing during the storm is 1/8. What is the probability that (i) atleast 2 pumps will be working (ii) all the pumps will be working during a particular storm ?

Ans. (i) 0.9989 (ii) 0.5129

Q.20 Show that for p = 0.50, the Binomial distribution has a maximum probability at X = n/2 if n is even and at $X = \frac{1}{2}(n-1)$ as well as $X = \frac{1}{2}(n+1)$ if n is odd.

Q.21 If the probability that a contractor's bid will be low is 1/3 on each of the three independent jobs. Find the probability distribution of the number of successful bids and what is the probability that the contractor gets at least one bid ?

Ans. Probability distribution is $\frac{8}{27}, \frac{12}{27}, \frac{6}{27}, \frac{1}{27}; \frac{19}{27}$

Q.22 Seven coins are tossed and number of heads are noted. The experiment is repeated 128 times and the following distribution is obtained.

No. of heads :	0	1	2	3	4	5	6	7
Frequencies :	7	6	19	35	30	23	7	1

Fit a Binomial distribution assuming

(i) The coin is unbiased

(ii) The nature of the coin is not known

(iii) Probability of a head for four coins is 0.5 and for remaining three coins is 0.45.

Ans. (i) 1, 7, 21, 35, 35, 21, 7, 1, (ii) 1, 8, 23, 36, 34, 19, 6, 1

(iii) 1, 9, 36, 23, 33, 19, 6, 1

Q.23 A dealer considers a car to be successful if it does not require repairs covered by the warranty. If the probability of a car being successful is 0.8, find the mean, variance and standard deviation for the number of successful cars among the three cars sold in one day.

Ans. 24, 0.48, 0.693

Q.24 The sum and product of the mean and variance of a Binomial distribution are 24 and 128 respectively. Find the distribution

Ans. n = 32, p = q = 1/2

Q.25 The probability that a bomb dropped from a plane will strike the target is 1/5. If six such bombs are dropped find the probability that atleast two will strike the target.

Ans. 0.345

Q.26 Fit a Binomial distribution to the following data :

x :	0	1	2	3	4	5
f :	38	144	342	287	164	25

Ans. 33.2, 16.19, 316.2, 308.7 , 150.7, 29.4.

Q.27 In analyzing the employment structure of house holds in an urban area, in connection with a trip generation study, it has been found that there are 2500 households of 4 members each. Find the number of houses having no, one, two, three and four employed members, the average number of employed person in a family being one.

Ans. 790, 1055, 528, 117 and 10.

Q.28 Fit a Binomial distribution to the following data :

x :	0	1	2	3	4
f :	30	62	46	10	2

Ans.

x :	0	1	2	3	4
f :	32	60	43	13	2

Q.29 The incidence of occupational disease in any industry is such that the workers have 20% chance of suffering from it. What is the probability that out of six workers 4 or more will catch the disease ?

Ans. $\frac{53}{3125}$

Q.30 Fit a Binomial distribution to the following data ;-

x :	0	1	2	3	4	5	6	Total
f :	5	18	28	12	7	6	4	80

Ans. 4, 15, 25, 22, 11, 3, 0

Q.31 At an uncontrolled T junction, past experience indicates that the probability of a vehicle arriving on the side road during a 15 second interval and turning right into main road is 1/5. Find the probability that in a period of 1 minute there will be 0, 1, 2, 3 or 4 vehicles arriving and turning right.

Ans. $\frac{256}{625}, \frac{256}{625}, \frac{96}{625}, \frac{16}{625}, \frac{1}{625}$.

2.4 Poisson Distribution

Poisson distribution, discovered by French mathematician S D Poisson in 1837, is a distribution related to the probabilities of events which are extremely rare, but which have a large number of independent opportunities for occurrence. For example, the number of telephone calls per minute at some switchboard, number of deaths from a disease such as heart attack or the number of α particles emitted by a radioactive source.

A discrete random variable X which can take random values 0, 1, 2, . . . is said to follow Poisson distribution if its probability mass function is given by :-

$$P(X=r)=e^{-m}\frac{m^r}{\lfloor r}\ ,\ r=0,1,2,\ldots$$

$m > 0$ is said to be the parameter of the distribution and we say, $X \sim P(m)$ is Poisson variate. **(Raj. IV Sem CP-2006), (RTU-2009)**

The above given is a probability mass function as :-

(i) $$P(X=r)=e^{-m}\frac{m^r}{\lfloor r}\geq 0\ \forall\, r=0,1,2,\ldots \text{where } m>0, \text{ and}$$

(ii) $$\sum_{r=o}^{\infty}P(X=r)=\sum_{r=o}^{\infty}e^{-m}\frac{m^r}{\lfloor r}=e^{-m}\sum_{r=o}^{\infty}\frac{m^r}{\lfloor r}=e^{-m}\left[1+m+\frac{m^2}{\lfloor 2}+\ldots\right]$$

$$=e^{-m}\times e^{m}=1.$$

Proof of the formula given above

Poisson distribution can be derived as a limiting case of Binomial distribution under the conditions $n\to\infty$, $p\to 0$ and $np=m$ (finite). **(Raj. IV Sem CP-2004)**

According to the Binomial distribution the probability of r successes in 'n' independent trials is given by :-

$$P(X_B=r)={}^nC_rp^rq^{n-r}\qquad r=0,1,2,\ldots,n$$

$$=\frac{\lfloor n}{\lfloor r\,\lfloor n-r}p^r(1-p)^{n-r},\ \ r=0,1,2,\ldots,n$$

$$= \frac{n(n-1)(n-2)\ldots[n-(r-1)]\,\lfloor n-r}{\lfloor r \,\lfloor n-r} p^r(1-p)^{n-r}, \quad r = 0, 1, 2, \ldots n$$

$$\Rightarrow \quad P(X_B = r) = \frac{\overbrace{np(np-p)(np-2p)\ldots[np-(r-1)p]}^{r \text{ terms}}}{\lfloor r}(1-p)^{n-r}, \quad r = 0, 1, 2, \ldots, n$$

Now applying the limits $n \to \infty$, $p \to 0$ and $np = m$, we get :-

$$\text{Now} \quad P(X_P = r) = \frac{\overbrace{m(m-0)(m-0)\ldots(m-0)}^{r \text{ terms}}}{\lfloor r} \operatorname*{Lt}_{\substack{n\to\infty \\ p\to 0}} \frac{(1-p)^n}{(1-p)^r}, \quad r = 0, 1, 2, \ldots, n$$

$$= \frac{m^r}{\lfloor r} \operatorname*{Lt}_{\substack{n\to\infty \\ p\to 0}} \left(1 - \frac{m}{n}\right)^n \quad [\because np = m] \qquad \ldots..(2.5)$$

Now let

$$u = \left(1 - \frac{m}{n}\right)^n$$

$$\Rightarrow \quad \log u = n \log\left(1 - \frac{m}{n}\right) = n\left(-\frac{m}{n} - \frac{m^2}{2n^2} - \frac{m^3}{3n^3} \ldots\right) = -m - \frac{m^2}{2n} - \frac{m^3}{3n^2} \ldots$$

$$\Rightarrow \quad u = e^{-m - \frac{m^2}{2n} - \frac{m^3}{3n^2} \ldots..}$$

$$\Rightarrow \quad \operatorname*{Lt}_{\substack{n\to\infty \\ p\to 0}} u = e^{-m}$$

Substituting this in (2.5) we get the probability mass function for Poisson variate as :-

$$P(X_P = r) = \frac{m^r}{\lfloor r} e^{-m}, \quad r = 0, 1, 2, \ldots$$

or $$P(X=r)=\frac{m^r}{\underline{|r}}e^{-m}, \quad r=0,1,2,\ldots$$

2.4.1 Mean and Variance of Poisson Distribution

$$\text{Mean}=E(X)=\sum_r x_r p(r)=\sum_{r=0}^{\infty} r.\frac{e^{-m}m^r}{\underline{|r}}$$

$$=e^{-m}=\sum_{r=1}^{\infty}\frac{m^r}{\underline{|r-1}}=e^{-m}\left[m+\frac{m^2}{\underline{|1}}+\frac{m^3}{\underline{|2}}+\ldots\right]$$

$$=me^{-m}\left[1+m+\frac{m^2}{\underline{|2}}\ldots\right]=me^{-m}e^{m}=m.$$

Also $$E\left(X^2\right)=\sum_r x_r^2 p(r)=\sum_{r=0}^{\infty} r^2 e^{-m}\frac{m^r}{\underline{|r}}$$

$$=\sum_{r=0}^{\infty}[r(r-1)+r]e^{-m}\frac{m^r}{\underline{|r}}$$

$$=\sum_{r=0}^{\infty} r(r-1)e^{-m}\frac{m^r}{\underline{|r}}+\sum_{r=0}^{\infty} r\,e^{-m}\frac{m^r}{\underline{|r}}$$

$$=\sum_{r=2}^{\infty} e^{-m}\frac{m^r}{\underline{|r-2}}+\sum_{r=1}^{\infty} e^{-m}\frac{m^r}{\underline{|r-1}}$$

$$=e^{-m}\left[m^2+\frac{m^3}{\underline{|1}}+\frac{m^4}{\underline{|2}}+\ldots.\right]+e^{-m}\left[m+\frac{m^2}{\underline{|1}}+\frac{m^3}{\underline{|2}}+\ldots\right]$$

$$=m^2e^{-m}\left[1+m+\frac{m^2}{\underline{|2}}\ldots.\right]+e^{-m}m\left[1+m+\frac{m^2}{\underline{|2}}+\ldots\right]$$

$$= m^2 e^{-m} e^m + m.e^{-m}.e^m$$
$$= m^2 + m$$

Hence Variance (X) $= E(X^2) - [E(X)]^2$.

$$= m^2 + m - (m)^2$$
$$= m.$$

***Example* 32.** Find the probability that at the most 5 defective fuses will be found in a box of 200 fuses if experience shows that 2% of such fuses are defective.

Solution : Here the number of independent trials is 200 which is very large, hence X should be considerd as a Poisson variate denoting the number of defective fuses.

Here $n = 200, p = \frac{2}{100} = 0.02 \Rightarrow m = np = 4$

Also here $P(X = r) = e^{-m} \frac{m^r}{\lfloor r}$, $r = 0, 1, \ldots$

Hence required probability is :-

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

$$= e^{-m}\left[1 + m + \frac{m^2}{\lfloor 2} + \frac{m^3}{\lfloor 3} + \frac{m^4}{\lfloor 4} + \frac{m^5}{\lfloor 5}\right]$$

$$= e^{-4}\left[1 + 4 + \frac{4^2}{\lfloor 2} + \frac{4^3}{\lfloor 3} + \frac{4^4}{\lfloor 4} + \frac{4^5}{\lfloor 5}\right]$$

$$= 0.01832 \times 42.8667$$

$$= 0.7851.$$

***Example* 33.** A car hire firm has two cars which it hires out day by day. The number of demands for a car on each day is distributed as a Poisson variate with mean 1.5. Calculate the proportion of days on which (i) neither car is used and (ii) some demand is refused.

Solution : Given m=1.5 and $P(X = r) = \frac{e^{-m} m^r}{\lfloor r}$, $r = 0, 1, 2, \ldots$

(i) Proportion of days on which neither car is used

$= P(X = 0) = e^{-1.5} = 0.2231.$

(ii) Proportion of days on which some demand is refused

$P(X > 2) = 1 - P(X \le 2)$

$$= 1 - [P(0) + P(1) + P(2)] = 1 - e^{-1.5}\left(1 + 1.5 + \frac{(1.5)^2}{\lfloor 2}\right)$$

$$= 1 - 0.2231 \times 3.625 = 0.19126.$$

Example 34. In a book of 520 pages, 390 typographical errors occur. Assuming Poisson law for the number of errors per page, find the probability that a random sample of 5 pages will contain no error.

Solution : The average number of typographical error per page in the book is

$$= m = \frac{390}{520} = 0.75.$$

Let X denote number of errors per page and $X \sim P(0.75)$

Therefore, probability that a page will contain no error $= P(X=0) = \dfrac{e^{-m} m^0}{\lfloor 0} = e^{-0.75}$

Thus the required probability that all the 5 pages are error free

$$= [P(X=0)]^5 = \left(e^{-0.75}\right)^5 = e^{-3.75} = 0.0235.$$

2.4.2 Moments, Moment Generating Function and Recurrence Relation for Moments

The moment generating function about origin for the Poisson variate is given by :-

$$M_X(t) = E\left(e^{tX}\right) = \sum_r e^{tx_r} P(r) = \sum_{r=0}^{\infty} e^{tr} e^{-m} \frac{m^r}{\lfloor r}$$

$$= \sum_{r=0}^{\infty} e^{-m} \frac{(me^t)^r}{\lfloor r} = e^{-m}\left[1 + me^t + \frac{(me^t)^2}{\lfloor 2} + \ldots\right]$$

$$= e^{-m} e^{me^t} = e^{m(e^t - 1)}$$

Also $\quad \mu_r' = \left[\dfrac{d^r M_X(t)}{dt^r}\right]_{t=0}$

$$\Rightarrow \quad \mu_1' = \text{mean} = \left[\frac{de^{m(e^t-1)}}{dt}\right]_{t=0} = \left[me^t e^{m(e^t-1)}\right]_{t=0} = m$$

$$\mu_2' = \left(\frac{d^2 e^{m(e^t-1)}}{dt^2}\right)_{t=0} = \left(m\left[e^t e^{m(e^t-1)} + me^{2t} e^{m(e^t-1)}\right]\right)_{t=0} = m^2 + m$$

$$\mu_3' = \left(\frac{d^3 e^{m(e^t-1)}}{dt^3}\right)_{t=0} = \left(\frac{d}{dt}\left[me^t\left(1+me^t\right)e^{m(e^t-1)}\right]\right)_{t=0}$$

$$= \left(m\left[e^t\left(1+me^t\right)e^{m(e^t-1)} + me^{2t}e^{m(e^t-1)} + me^{2t}\left(1+me^t\right)e^{m(e^t-1)}\right]\right)_{t=0}$$

$$= m[(1+m)+m+m(1+m)] = m\left(m^2+3m+1\right)$$

$$= m^3 + 3m^2 + m.$$

Similarly we can get :-

$$\mu_4' = \left[\frac{d^4 e^{m(e^t-1)}}{dt^4}\right]_{t=0} = m^4 + 6m^3 + 7m^2 + m.$$

Remark 5 We can also obtain the moments about origin by using definition $\mu_r' = \sum_i x_i^r p_i$ as mean and variance have been obtained in section 2.4.1.

Central Moments of Poisson Distribution

Using interrelations between moments (refer sec. 3.5.2) we get :

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \mu_1'^2 = m^2 + m - m^2 = m$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3 = \left(m^3+3m^2+m\right) - 3(m)\left(m^2+m\right) + 2m^3$$

$$= m^3 + 3m^2 + m - 3m^3 - 3m^2 + 2m^3 = m$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

$$= \left(m^4+6m^3+7m^2+m\right) - 4\left(m^3+3m^2+m\right)m + 6\left(m^2+m\right)m^2 - 3m^4$$

$$= 2m^2 + m$$

Karl Pearson coefficient of skewness and kurtosis can be obtained as :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{1}{m} \text{ and } \gamma_1 = \sqrt{\beta_1} = \frac{1}{\sqrt{m}}$$

Also $$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3m^2 + m}{m^2} = 3 + \frac{1}{m} \text{ and } \gamma_2 = \beta_2 - 3 = \frac{1}{m}$$

Moment generating function about mean :

$$M_X(t) \text{ about mean} = E\left(e^{t(X-\overline{X})}\right) = E\left(e^{t(X-m)}\right) \qquad [\because \overline{X} = m]$$

$$= e^{-mt}E\left(e^{tX}\right) = e^{-mt}M_X(t) \text{ about origin}$$

$$= e^{-mt}e^{m(e^t-1)} = e^{m(e^t-t-1)}$$

The central moments can also be obtained with the help of moment generating function about mean as :-

$$\mu_1 = \left(\frac{d}{dt}M_X(t) \text{ about mean}\right)_{t=0} = \left(m(e^t-1)e^{m(e^t-t-1)}\right)_{t=0} = 0$$

$$\mu_2 = \left(\frac{d^2M_X(t) \text{ about mean}}{dt^2}\right)_{t=0} = \left(\left[m^2(e^t-1)^2 + me^t\right]e^{m(e^t-t-1)}\right)_{t=0} = m$$

$$\mu_3 = \left(\frac{d^3M_X(t) \text{ about mean}}{dt^3}\right)_{t=0} = \left(\left[m^3(e^t-1)^3 + 2m^2(e^t-1)e^t +\right.\right.$$

$$\left.\left. + me^t\right]e^{m(e^t-t-1)}\right)_{t=0} = m$$

and $$\mu_4 = \left(\frac{d^4M_4(t) \text{ about mean}}{dt^4}\right)_{t=0} = 3m^2 + m$$

Recurrence Relation for the Central Moments of the Poisson Distribution

By definition we have :

(Raj. IV Sem CP-2006, 02)

$$\mu_r = E\{(X-\bar{X})^r\} = E\{(X-m)^r\}$$

$$= \sum_{x=0}^{\infty}(x-m)^r P(x) = \sum_{x=0}^{\infty}(x-m)^r \frac{e^{-m}m^x}{\lfloor x}$$

Differentiating both sides with respect to 'm' we get :-

$$\frac{d\mu_r}{dm} = \sum_{x=0}^{\infty}(-r)(x-m)^{r-1}\frac{e^{-m}m^x}{\lfloor x} + \sum_{x=0}^{\infty}\frac{(x-m)^r}{\lfloor x}\left[-e^{-m}m^x + xm^{x-1}e^{-m}\right]$$

$$= (-r)\sum_{x=0}^{\infty}(x-m)^{r-1}\frac{e^{-m}m^x}{\lfloor x} + \sum_{x=0}^{\infty}\frac{(x-m)^r}{\lfloor x}e^{-m}m^x\left[-1+\frac{x}{m}\right]$$

$$= (-r)\sum_{x=0}^{\infty}(x-m)^{r-1}P(x) + \frac{1}{m}\sum_{x=0}^{\infty}(x-m)^{r+1}\frac{e^{-m}m^x}{\lfloor x}$$

$$= (-r)\sum_{x=0}^{\infty}(x-m)^{r-1}P(x) + \frac{1}{m}\sum_{x=0}^{\infty}(x-m)^{r+1}P(x)$$

$$= (-r)\mu_{r-1} + \frac{1}{m}\mu_{r+1}$$

as $\mu_{r-1} = \sum_{x=0}^{\infty}(x-m)^{r-1}P(x)$ and $\mu_{r+1} = \sum_{x=0}^{\infty}(x-m)^{r+1}P(x)$

$$\Rightarrow \quad \mu_{r+1} = rm\mu_{r-1} + \frac{md\mu_r}{dm}$$

which is the required recurrence relation.

$r = 1$ gives $\mu_2 = m\mu_0 + \frac{md\mu_1}{dm} = m$ $\quad [\because \mu_0 = 1 \text{ and } \mu_1 = 0]$

$r = 2$ gives $\mu_3 = 2m\mu_1 + m\frac{d\mu_2}{dm} = m$

$r = 3$ gives $\mu_4 = 3m\mu_2 + m\frac{d\mu_3}{dm} = 3m^2 + m$ and so on.

Example 35. If X is a Poisson variate such that $P(X = 2) = 9P(X = 4) + 90P(X = 6)$. Find (i) the mean of X (ii) β_1, the coefficient of skewness.

Solution : We know that

$$P(X = r) = \frac{e^{-m} m^r}{\lfloor r}, \quad r = 0, 1, 2, \ldots$$

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

$$\Rightarrow \quad e^{-m}\frac{m^2}{\lfloor 2} = 9e^{-m}\frac{m^4}{\lfloor 4} + 90e^{-m}\frac{m^6}{\lfloor 6}$$

$$\Rightarrow \quad \frac{m^2}{2} = \frac{9}{24}m^4 + \frac{90}{720}m^6$$

$$\Rightarrow \quad \frac{m^2}{2} = \frac{m^2}{8}(3m^2 + m^4)$$

$$\Rightarrow \quad m^4 + 3m^2 - 4 = 0$$

$$\Rightarrow \quad m^2 = \frac{-3 \pm \sqrt{9+16}}{2} = \frac{-3 \pm 5}{2} = 1, -4$$

$m^2 = -4$ is not possible. Hence $m^2 = 1 \Rightarrow m = \pm 1$

As $m > 0$ therefore $m = 1$

(i) Mean of X = m = 1

(ii) $\beta_1 = \frac{1}{m} = 1$

Example 36. Deduce the first four moments about the mean of the Poisson distribution from those of Binomial distribution.

Solution : For Binomial distribution :-

$$\mu_1 = 0$$

$$\mu_2 = npq$$

$$\mu_3 = npq\,(q - p)$$

$$\mu_4 = npq\,(1 - 6pq) + 3n^2p^2q^2$$

The central moments for Poisson distribution can be derived from these, by applying the limits $n \to \infty$, $p \to 0$, $np \to m$ as follows :-

$$\mu_1 = 0$$

$$\mu_2 = \lim_{\substack{n\to\infty \\ p\to 0 \\ np\to m}} (np)(1-p) = m$$

$$\mu_3 = \lim_{\substack{n\to\infty \\ p\to 0 \\ np\to m}} (np)(1-p)(1-2p) = m$$

$$\mu_4 = \lim_{\substack{n\to\infty \\ p\to 0 \\ np\to m}} (np)(1-p)(1-6pq)+3(np)^2(1-p)^2 = 3m^2+m$$

2.4.3 Probability Generating Function For Poisson Distribution

The probability generating function for Poisson distribution is denoted by P(s) and defined as :-

$$P(s) = E\left(s^X\right) = \sum_{x=0}^{\infty} s^x P(x) = \sum_{x=0}^{\infty} s^x \frac{e^{-m} m^x}{\underline{|x}}$$

$$= \sum_{x=0}^{\infty} e^{-m} \frac{(ms)^x}{\underline{|x}} = e^{-m}\left[1+ms+\frac{(ms)^2}{\underline{|2}}\ldots\right]$$

$$= e^{-m} \cdot e^{ms} = e^{m(s-1)}$$

2.4.4 Mode of Poisson Distribution

Let $X \sim P(m)$ with mode r.

$\Rightarrow$ $P(r) > P(r+1)$ and $P(r) > P(r-1)$

$$P(r) > P(r+1) \Rightarrow e^{-m}\frac{m^r}{\underline{|r}} > e^{-m}\frac{m^{r+1}}{\underline{|r+1}} \Rightarrow r+1 > m$$

$$\Rightarrow r > m-1$$

Also, $$P(r) > P(r-1) \Rightarrow \frac{e^{-m}m^r}{\underline{|r}} > \frac{e^{-m}m^{r-1}}{\underline{|r-1}} \Rightarrow m > r$$

Hence $m < r < m-1$

Now if m is not an integer, modal value = r = integral part of m and if m is an integer we have two modal values = m and m − 1.

2.4.5 Recurrence Relation for Probabilities of Poisson Distribution or Fitting of Poisson Distribution

Let $X \sim P(m)$

Then $P(X=r) = e^{-m}\frac{m^r}{\lfloor r}, \quad r = 0, 1, 2, \ldots$

Now $\frac{P(r+1)}{P(r)} = \frac{e^{-m}m^{r+1}}{\lfloor r+1} \cdot \frac{\lfloor r}{e^{-m}m^r} = \frac{m}{r+1}$

$\Rightarrow \quad P(r+1) = \frac{m}{r+1}P(r).$

This formula is quite useful in the fitting of Poisson distribution, i.e., finding expected frequencies with the help of given Poisson distribution as will be clear from following example :-

***Example* 37.** Fit a Poisson distribution to the following data which gives the number of doddens in a sample of clover seeds.

No. of Doddens (x) :	0	1	2	3	4	5	6	7	8
Observed frequency (f):	56	156	132	92	37	22	4	0	1

Solution : Mean of the above frequency distribution

$$= \frac{\Sigma fx}{\Sigma f} = \frac{986}{500} = 1.972 = m \text{ (mean of Poisson distribution)}$$

Now expected frequencies can be tabulated as follows :

x	$P(x)=\frac{m}{x}P(x-1)$	$f(x)=500\times P(x)$
0	$P(0)=e^{-m}=0.13918$	$500\times 0.13918=69.59\cong 70$
1	$P(1)=\frac{m}{1}P(0)=0.27446$	$500\times 0.27446=137.23\cong 137$
2	$P(2)=\frac{m}{2}P(1)=0.27062$	$500\times 0.27062=135.3\cong 135$
3	$P(3)=\frac{m}{3}P(2)=0.17789$	$500\times P(3)=88.945\cong 89$
4	$P(4)=\frac{m}{4}P(3)=0.08769$	$500\times P(4)=43.845\cong 44$
5	$P(5)=\frac{m}{5}P(4)=0.03459$	$500\times P(5)=17.295\cong 17$
6	$P(6)=\frac{m}{6}P(5)=0.01137$	$500\times P(6)=5.685\cong 6$
7	$P(7)=\frac{m}{7}P(6)=0.003203$	$500\times P(7)=1.615\cong 2$
8	$P(8)=\frac{m}{8}P(7)=0.00079$	$500\times P(8)=0.39\cong 0$

Hence the expected frequencies on fitting Poisson distribution are :-

x:	0	1	2	3	4	5	6	7	8
f:	70	137	135	89	44	17	6	21	0

2.4.6 Reproductive Property of Poisson Variate

If X_1 and X_2 are independent Poisson variates that follow Poisson distribution with parameters m_1 and m_2 respectively, then $X_1 + X_2$ is also a Poisson variate with parameter $m_1 + m_2$.

Illustrative Examples

***Example* 38.** Suppose on an average 1 house in 1000 in a certain district has a fire during a year. If there are 2000 houses in that district, what is the probability that exactly five houses will have a fire during the year ? (given $e^{-2} = 0.1353$).

Solution : Here $n = 2000$ and $p = \frac{1}{1000}$, $np = 2$

i.e., n is very large and p is very small. Hence we consider $X \sim P(2)$

Required probability $= P(X=5) = \frac{e^{-2}2^5}{\underline{|5}}$

$= 0.1353 \times 0.2667 = 0.03608.$

***Example* 39.** Letters are received in an office on each one of 100 days. Assuming the following data to form a random sample from a Poisson distribution find the expected frequencies correct to nearest units. (Given $e^{-4} = 0.0183$)

No. of letters (x) :	0	1	2	3	4	5	6	7	8	9	10
Frequency (f):	1	4	15	22	21	20	8	6	2	0	1

(Raj. IV Sem CP-2003)

Solution : Mean of the given frequency distribution $= \frac{\Sigma fx}{\Sigma f} = \frac{400}{100} = 4 = m$ (mean of Poisson distribution)

The expected frequencies, hence, are as shown in the following table :

x	$P(x) = e^{-m} \frac{m^x}{\lfloor x}$	f(x) = 100 × P(x)
0	$P(0) = e^{-4} = 0.01832$	$1.832 \cong 2$
1	$P(1) = e^{-4} \times 4 = 0.07326$	$7.326 \cong 7$
2	$P(2) = \frac{e^{-4} \times 4^2}{\lfloor 2} = 0.14652$	$14.652 \cong 15$
3	$P(3) = \frac{e^{-4} \times 4^3}{\lfloor 3} = 0.19536$	$19.536 \cong 20$
4	$P(4) = \frac{e^{-4} \times 4^4}{\lfloor 4} = 0.19536$	$19.536 \cong 20$
5	$P(5) = \frac{e^{-4} \times 4^5}{\lfloor 5} = 0.15628$	$15.628 \cong 16$
6	$P(6) = \frac{e^{-4} \times 4^6}{\lfloor 6} = 0.104192$	$10.419 \cong 10$
7	$P(7) = \frac{e^{-4} \times 4^7}{\lfloor 7} = 0.05954$	$5.954 \cong 6$
8	$P(8) = \frac{e^{-4} \times 4^8}{\lfloor 8} = 0.02977$	$2.977 \cong 3$
9	$P(9) = \frac{e^{-4} \times 4^9}{\lfloor 9} = 0.01323$	$1.323 \cong 1$
10	$P(10) = \frac{e^{-4} \times 4^{10}}{\lfloor 10} = 0.00529$	$0.529 \cong 0$

***Example* 40.** In a certain factory turning out razor blades there is a small chance of 0.002 for any blade to be defective. The blades are supplied in packets of 10. Using Poisson distribution, find the number of packets containing no defective, one defective and two defective blades respectively in a consignment of 10,000 packets.

(Raj. IV Sem CP-2004)

Solution : Let $X \sim P(m)$ where $m = np$

Here $n = 10,\ p = 0.002 \Rightarrow m = np = 0.02$. Also $N = 10{,}000$

The probabilities that a packet contains no defective, one defective and two defective blades respectively are given by $P(X = r) = \frac{e^{-m}m^r}{\lfloor r}$ as :

$$P(X = 0) = e^{-0.02} \times \frac{(0.02)^0}{\lfloor 0} = 0.9802$$

$$P(X = 1) = e^{-0.02} \times \frac{(0.02)}{\lfloor 1} = 0.019604$$

and $$P(X = 2) = e^{-0.02} \times \frac{(0.02)^2}{\lfloor 2} = 0.00019604$$

Hence the required number of packets containing no defective, one defective and two defective blades is given by $N \times P(X = r)$ respectively as :

$N \times P(X = 0) = 9802$ packets,

$N \times P(X = 1) = 196.04 \cong 196$ packets,

$N \times P(X = 2) = 1.9604 \cong 2$ packets.

***Example* 41.** A skilled typist, on routine work, kept a record of mistakes made per day during 300 working days as :-

Mistakes per days :	0	1	2	3	4	5	6
Number of days :	143	90	42	12	9	3	1

Compute the frequencies of the Poisson distribution which has the same mean and total frequency as the above distribution. **(Raj. IV Sem CP-2005)**

Solution : Here mean of Poisson distribution $= m = \frac{\Sigma fx}{\Sigma f} = \frac{267}{300} = 0.89$

Hence theoretical frequencies $= (300)P(X = r) = 300 \times e^{-m}\frac{m^r}{\lfloor r}$ as :

x	P(x)	$F(x) = 300 \times P(x)$
0	0.411	$123.3 \cong 123$
1	0.365	$109.5 \cong 110$
2	0.163	$48.9 \cong 49$
3	0.048	$14.4 \cong 14$
4	0.011	$3.3 \cong 3$
5	0.02	$0.6 \cong 1$
6	0.0003	$0.09 \cong 0$

***Example* 42.** A source of liquid is known to contain bacteria, with the mean number of bacteria per cubic centimeter equal to 3. Ten 1 cc test tubes are filled with liquid. Assuming that Poisson distribution is applicable, find the probability that all ten test tubes will show growth, i.e., contain at least one bacterium each.

Solution : Let random variable X denotes the number of bacterium which shows growth in one, 1cc test tube. Given m = 3

Therefore, the probability that one test tube shows growth i.e. contain at least one bacterium shows growth = $P(X \geq 1)$

$$= 1 - P(X = 0) = 1 - e^{-3} = 1 - 0.04979 = 0.95021$$

Hence the probability that all ten test tubes show growth is given by $= [P(X \geq 1)]^{10}$

$$= (0.95021)^{10} = 0.60006.$$

***Example* 43.** Fit a Poisson distribution to the following data :

x :	0	1	2	3	4
f :	192	100	24	3	1

and compare the calculated frequencies with actual observed one.

Solution : Here $m = \text{mean} = \frac{\Sigma fx}{\Sigma f} = \frac{161}{320} = 0.503$

and calculated frequencies $= 320 \times P(X = r) = 320 \times e^{-m} \frac{m^r}{\lfloor r}$, $r = 0, 1, 2, 3, 4$, i.e.,

x		0	1	2	3	4
P(x)		0.6047	0.3042	0.07649	0.01283	0.00194
F(x)=320×P(x)	193.5≅194	97.397	24.4≅24	F(x)=320×P(x)	193.5≅194	97.397

***Example* 44.** Suppose 300 misprints are distributed randomly throughout a book of 500 pages. Find the probability that a given page contains (i) exactly 2 misprints (ii) 2 or more misprints.

Solution : Mean number of misprints per page $= \frac{300}{500} = 0.6$

Let $X \sim P(0.6)$

(i) $P(X = 2) = e^{-0.6} \frac{(0.6)^2}{\lfloor 2} = 0.549 \times \frac{0.36}{2} = 0.09882$

(ii) $P(X \geq 2) = 1 - P(X < 2) = 1 - [P(0) + P(1)]$

$= 1 - [e^{-0.6} + e^{-0.6}(0.6)]$

$= 1 - e^{-0.6}(1.6) = 1 - 0.8784 = 0.1216.$

***Example* 45.** If X is a Poisson variate with mean m, show that the expectation of e^{-kx} is $\exp[-m(1 - e^{-k})]$.

Solution :

$$E\left(e^{-kX}\right) = \sum_{x=0}^{\infty} e^{-kx} P(x) = \sum_{x=0}^{\infty} e^{-kx} \frac{e^{-m} m^x}{\lfloor x}$$

$$= e^{-m} \sum_{x=0}^{\infty} \frac{\left(me^{-k}\right)^x}{\lfloor x} = e^{-m}\left[1 + me^{-k} + \frac{\left(me^{-k}\right)^2}{\lfloor 2} + \ldots\right]$$

$$= e^{-m} \cdot e^{me^{-k}} = e^{-m\left(1-e^{-k}\right)}$$

Hence Proved

***Example* 46.** A large number of observations on a given solution which contained bacteria were made by taking samples of 1 cc each and noting down the number of bacteria present in each sample. Assuming the Poisson distribution and given that 10% of the samples contained no bacteria, find the average number of bacteria per cc.

Solution : Let $X \sim P(m)$, where m is the average (mean) number of bacteria per cc.

Given $P(X = 0) = \frac{10}{100}$

$\Rightarrow \quad e^{-m} \frac{m^0}{\lfloor 0} = 0.1 \Rightarrow e^{-m} = 0.1 \Rightarrow e^m = 10$

$\Rightarrow \quad m = \log 10 = 2.3026.$

***Example* 47.** Records show that the probability is 0.00005 that a car will have a flat tyre while crossing a certain bridge. Use Poisson distribution to find probabilities that among

10,000 cars crossing this bridge,

(i) exactly two will have a flat tyre

(ii) at most two will have a flat tyre.

Solution : Let random variable X denote number of cars having flat tyres, which is a Poisson variate.

Here $n = 10{,}000, \quad p = 0.00005$

Hence Mean $= np = 0.5 = m$

(i) $$P(X=2) = e^{-m}\frac{m^2}{\lfloor 2} = e^{-0.5} \times \frac{(0.5)^2}{\lfloor 2} = \frac{0.6065 \times .25}{2} = 0.0758$$

(ii) $$P(X \le 2) = P(0) + P(1) + P(2)$$

$$= e^{-0.5}\left[1 + 0.5 + \frac{(0.5)^2}{2}\right]$$

$$= 0.6065 \times 1.625 = 0.98556 .$$

Example 48. The data in the following table shows the occupancy of parking spaces in a parking lot consisting of 50 spaces. The count was taken at 15 minutes intervals during the 4 hour duration between 11 A.M. and 3 P.M. on six week days. Find whether the number of vacant spaces during any count follows a Poisson distribution.

Occupancy of parking spaces (x):	50	49	48	47	46	45	44	43	42	41	40 & less
Frequency (f):	6	15	21	20	15	10	5	2	1	1	0

Solution : Let mean of Poisson variate $X = m = \frac{\Sigma fx}{\Sigma f} = \frac{4512}{96} = 47$

Now $$f(40) = 96 \times P(40) = 96 \times e^{-47}\frac{(47)^{40}}{\lfloor 40}$$

$$= 96 \times 3.874 \times 10^{-21} \times \frac{7.6544 \times 10^{66}}{8.15915 \times 10^{47}}$$

$$= \frac{96 \times 3.874 \times 7.6544}{8.15915} \times 10^{-2}$$

$$= 348.897 \times 10^{-2} = 3.49$$

$$f(41) = 96 \times P(41) = 96 \times e^{-47} \times \frac{(47)^{41}}{\lfloor 41} = 4.00$$

$$f(42) = 96 \times P(42) = f(41) \times \frac{47}{42} = 4.47$$

Example 49. A manufacturer of pins knows that on an average 5% of his products are defective. If he packs pins in packets of 100 and guarantees that not more than 10 pins will be defective in a packet, find the probability that some box would fail to meet his guarantee ($e^{-5} = 0.006738$).

Solution : Let X be a Poisson variate denoting the number of defective pins in a packet .

Then $\quad \text{Mean } m = np = 100 \times \frac{5}{100} = 5$

Probability that a box fails to meet his guarantee = Probability that more than 10 pins are defective in a packet = $P(X > 10)$

$$= 1 - P(X \le 10) = 1 - e^{-5}\left[1 + m + \frac{m^2}{\lfloor 2} + \ldots + \frac{m^{10}}{\lfloor 10}\right]$$

$$= 1 - 0.006738\left[1 + 5 + \frac{5^2}{\lfloor 2} + \ldots + \frac{5^{10}}{\lfloor 10}\right]$$

Example 50. X_1 and X_2 are two independent Poisson variates with means m_1 and m_2 respectively. Prove that their sum $X_1 + X_2$ is also a Poisson variate with mean $m_1 + m_2$.

Solution : Let $M_{X_1}(t)$ and $M_{X_2}(t)$ be the moment generating functions of X_1 and X_2 respectively and $M_X(t)$ of $X_1 + X_2$.

As X_1 and X_2 are independent, hence

$$M_X(t) = M_{X_1}(t) \cdot M_{X_2}(t)$$

$$= e^{m_1(e^t - 1)} \times e^{m_2(e^t - 1)} \qquad \left(\because X \sim P(m) \Rightarrow M_X(t) = e^{m(e^t-1)}\right)$$

$$= e^{(m_1 + m_2)(e^t - 1)}$$

which is the moment generating function of a Poisson distribution with mean $m_1 + m_2$.

Hence $X = X_1 + X_2$ is a Poisson variate with mean $m_1 + m_2$.

This property is known as reproductive property of Poisson variate.

Example 51. In a Poisson distribution if $3P(X = 3) = 4P(X = 4)$. Find $P(X = 7)$

Solution : $3P(X = 3) = 4P(X = 4)$

$$\Rightarrow \quad 3e^{-m}\frac{m^3}{\lfloor 3} = 4e^{-m}\frac{m^4}{\lfloor 4} \Rightarrow m = 3$$

$$\therefore \quad P(X = 7) = e^{-3}\frac{3^7}{\lfloor 7} = \frac{0.0498 \times 2187}{5040} = 0.0216$$

Example 52. 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, find the number of packets containing at least, exactly and at most 2 defective items in a consignment of 1000 packets using (i) Binomial distribution and (ii) Poisson distribution.

Solution : Here X denotes the number of defectives in the n items considered.

Given $n = 20, \; p = 0.05, \; q = 0.95, \; m = np = 1$

(i) $P(X = r) = {}^nC_r\, p^r\, q^{n-r}$.

No. of packets containing r defectives $= N(r) = 1000 \times P(r)$

Now $P(X \geq 2) = 1 - P(X < 2) = 1 - [P(0) + P(1)]$

$$= 1 - \left[{}^{20}C_0(0.05)^0(0.95)^{20} + {}^{20}C_1(0.05)^1 + (0.95)^{19}\right]$$

$$= 1 - [0.3585 + 0.3774]$$

$$= 0.2641$$

$\therefore \quad N(X \geq 2) = 1000 \times 0.2641 \cong 264$

and $\quad P(X = 2) = {}^{20}C_2\,(0.05)^2\,(0.95)^{18} = 0.1887$

$N(X = 2) = 1000 \times 0.1887 \cong 189$

and $\quad P(X \leq 2) = [P(0) + P(1) + P(2)] = 0.9246$

$\therefore \quad P(X \leq 2) = 1000 \times P(X \leq 2) = 1000 \times 0.9246 \cong 925$

(ii) $\quad P(X = r) = \dfrac{e^{-m}m^r}{\lfloor r}$

and number of packets containing r defectives is given by $N(X = r) = 1000P(X = r)$

Now $P(X \geq 2) = 1 - P(X < 2) = 1 - [P(0) + P(1)] = 1 - (e^{-1} + e^{-1}) = 0.2642$

$N(X \geq 2) = 1000 \times 0.2642 \cong 264$

and $P(X = 2) = \dfrac{e^{-1}1^2}{\underline{|2}} = 0.1839$

$N(2) = 1000 \times 0.1839 \cong 184$

and $P(X \leq 2) = P(0) + P(1) + P(2) = 0.9197$

$N(X \leq 2) = 1000 \times 0.9197 \cong 920.$

***Example* 53.** Find the probability of 5 or more telephone calls arriving in a 9 minute period in a college switch-board, if the telephone calls that are received at the rate of 2 every 3 minute follow a Poisson distribution.

Solution : Let X_1, X_2, X_3 denote the number of telephone calls received in three consecutive 3-min period. As each of X_1, X_2, X_3 follows a Poisson distribution with parameter $m = 2$, hence by reproductive property of Poisson distribution $X = X_1 + X_2 + X_3$ follows a Poisson distribution with parameter $= 2 + 2 + 2 = 6.$

Hence Required probability $= P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{r=0}^{4} e^{-6} \dfrac{6^r}{\underline{|r}}$

$= 1 - (0.0025 + 0.0149 + 0.0446 + 0.0892 + 0.1339)$

$= 1 - 0.2851 = 0.7149$

***Example* 54.** If X and Y are independent Poisson random variables, show that the conditional distribution of X, given the value of X + Y, is a Binomial distribution.

Solution : Let $X \sim P(m_1)$ and $Y \sim P(m_2)$

Now $P\left\{\dfrac{X = r}{(X+Y) = n}\right\} = \dfrac{P[(X = r) \text{ and } (X+Y = n)]}{P[(X+Y) = n]}$

$$= \frac{P(X = r; Y = n - r)}{P(X + Y = n)} = \frac{P(X = r) \cdot P(Y = n - r)}{P(X + Y = n)}$$

(As X and Y are independent)

$$= \frac{\left\{e^{-m_1} m_1^r / \underline{|r}\right\}\left\{e^{-m_2} (m_2)^{n-r} / \underline{|n-r}\right\}}{e^{-(m_1+m_2)} (m_1 + m_2)^n / \underline{|n}}$$

(reproductive property)

$$= \frac{\lfloor n}{\lfloor r \lfloor n-r} \left(\frac{m_1}{m_1+m_2}\right)^r \left(\frac{m_2}{m_1+m_2}\right)^{n-r}$$

$$= {}^nC_r\, p^r\, q^{n-r},$$

where $p = \frac{m_1}{m_1+m_2}$ and $q = \frac{m_2}{m_1+m_2}$ and $p+q=1$

which is clearly a Binomial distribution.

***Example* 55.** Between 2 & 4 pm the average number of phone calls per minute coming into the switchboard of a company is 2.5. Find the probability that during one particular minute there will be no phone call at all. (given $e^{-2} = 0.13435$ and $e^{-0.5} = 0.60650$)

Solution : Here $X \sim P(2.5)$

Required probability $= P(X=0) = e^{-2.5} = e^{-2} \times e^{-0.5} = 0.0821.$

***Example* 56.** The number of arrivals of customers during any day follows Poisson distribution with a mean of 5. What is the probability that the total number of customers on two days selected at random is less than 2 ?

Solution : Here $m = 2$

There can be less than 2 customers on two days selected as :-

No of customers on		
I day	*II day*	*Total*
0	0	0
0	1	1
1	0	1

Hence Required probability

$= P(0)\,P(0) + P(0)\,P(1) + P(1)\,P(0)$

$= P(0)\,[P(0) + 2P(1)]$

$= e^{-2}[e^{-2} + 2.2e^{-2}] = 5e^{-4}$

$= 4.994 \times 10^{-4}.$

Exercise 2(B)

Q.1 If 2% items are defective in certain items manufactured by a machine, determine the probability that out of 400 items 5 are found to be defective.

Ans. 0.093

Q.2 The number of monthly breakdowns of a computer is a random variable having a Poisson distribution with mean equal to 1.8. Find the probability that this computer will function for a month (i) without a breakdown (ii) with only one breakdown (iii) with at least one breakdown.

Ans. 0.1653, 0.2975, 0.8347

Q.3 A random variate X follows the Poisson distribution and satisfies the condition $3P(r=2) = P(r=4)$. Find (i) $P(r=0)$ (ii) $P(r=5)$.

Ans. (i) 0.0025 (ii) 0.1606

Q.4 Derive the moment generating function of Poisson variate with parameter m about origin and hence find its mean and variance.

Ans. m; m

Q.5 Fit a Poisson distribution to the following data :

x :	0	1	2	3	4
f :	21	18	7	3	1

Ans. 20.33, 18.30, 8.24, 2.47, 0.56

Q.6 After correcting 50 pages of the proof of a book the proof reader finds that there are, on an average 2 errors per 5 pages. How many pages would one expect to find with 0, 1, 2, 3 & 4 errors in 1000 pages of the first print of the book ?

Ans. 670, 268, 54, 7, 1.

Q.7 Five coins are tossed 1600 times. Using Poisson distribution find the probability of getting 5 heads 5 times.

Ans. $\dfrac{e^{-m}m^5}{\underline{|5}}, m = 50$

Q.8 Define Poisson distribution. Derive it as a limiting case of Binomial distribution. Find its mean and variance.

Ans. m; m

Q.9 If 3% of electric bulbs, manufactured by a company be defective, find the probability of two or less than 2 defectives in a lot of 100 bulbs ($e^{-3} = 0.0498$).

Ans. 0.4233

Q.10 Fit a Poisson distribution to the following frequency distribution.

x :	0	1	2	3	4	5	Total
f :	142	156	69	27	5	1	400

Ans. 147, 147, 74, 25, 6, 1.

Q.11 Fit a Poisson distribution to the following data :

x :	0	1	2	3	4
f :	122	60	15	2	1

Ans. 122, 61, 15, 2, 0

Q.12 In a town 10 accidents occur in a period of 50 days. Assuming that the number of accidents in a day is governed by the Poisson law, find the probability of 3 or more accidents on a certain day (given $e^{-2} = 0.8187$).

Ans. 0.001

Q.13 Suppose that in key punching of 80 column IBM cards, the arithmetic mean number of mistakes per card is 0.3. What percentage of cards will have (i) no mistake (ii) one mistake and (iii) two mistakes ?

Ans. (i) 74% (ii) 22% (iii) 3%

Q.14 Data was collected over a period of 10 years, showing the number of deaths from horse kicks in each of the 200 army corps. The distribution of death was as follows :

No of deaths :	0	1	2	3	4
Frequency :	109	65	22	3	1

Fit a Poisson distribution to the data and calculate the theoretical frequencies.

Ans. 109, 66, 20, 4, 1.

Q.15 If X has a discrete uniform distribution $P(X = k) = 1/n$, for $k = 0, 1, \ldots, n$, find its moment generating function and hence find mean and variance of X.

Ans. $\frac{1}{n}e^t\frac{(1-e^{nt})}{1-e^t}, \frac{n+1}{2}, \frac{n^2-1}{12}$.

Q.16 In a certain factory turning out razor blades, the probability of a blade to be defective is 0.01. The blades are sold in packets of 10. Use Poisson distribution to find the probabilities of a packet with (i) no (ii) one and (iii) 2 defective blades. Find the number of such packets in a consignment of 10,000 packets.

Ans. (i) .905, (ii) .0905, (iii) 0.004525; 9050, 905, 46

Q.17 The probability that a person will get food poisoning, on a day in certain state fair is 0.0012. Find, using Poisson distribution, the probability that among 1000 persons attending the state fair, at most two will get food poisoning.

Ans. 0.8795

Q.18 Define Poisson distribution. Derive it as a limiting case of Binomial distribution.