6. A five figure number is formed by the digits 0, 1, 2, 3, 4, (without repetition). Find the probability that the number formed is divisible by 4.
Ans. $\frac{3}{10}$.

7. A, B, C throw the coin alternatively in that order. One who gets Tail first wins the game. Find the probability of B winning the game if C has a start.
Ans. $\frac{1}{7}$.

8. A box contains 5 red and 4 white marbles. 2 marbles are drawn successively from the box without replacement and it is noted that the second one is white. What is the probability that the first is also white ?
Ans. $\frac{1}{6}$.

9. Box A contains 3 red and 2 blue marbles. The box B contains 2 red and 8 blue marbles. A fair coin is tossed. If the coin shows Head, a marble is chosen from box A, if it shows Tail, a marble is chosen from box B. Find the probability that a red marble is chosen.
Ans. $\frac{2}{5}$.

10. One shot is fired from each of the three guns. E_1, E_2, E_3 denote the events that the target is hit by the first, second and third guns respectively. If $P(E_1) = 0.5$, $P(E_2) = 0.6$, $P(E_3) = 0.7$ and E_1, E_2, E_3 are independent events, then find the probability that at least two hits are registered.
Ans. 0.65

11. A problem on computer mathematics is given to the three students A, B and C whose chances of solving it are $\frac{1}{2}, \frac{3}{4}$ and $\frac{1}{4}$ respectively. What is the probability that the problem will be solved ?
Ans. $\frac{29}{32}$.

12. Urn I contains 6 white and 4 black balls and urn II contains 4 white and 5 black balls. From urn I, two balls are transferred to urn II without noticing the colour. Sample of size 2 is then drawn without replacement from urn II. What is the probability that the sample contains exactly 1 white ball ?
Ans. $\frac{4}{5}$.

8.6 PROBABILITY DISTRIBUTION

In Chapter 7, we have seen that statistical data can be presented in the form of frequency distribution, giving tabulated values of variate x and corresponding frequencies. Probability distribution for a variate x can be presented in a similar manner.

8.6.1 Random Variable, Probability Density Function Sample Space

If a trial or an experiment is conducted, the set S of all possible outcomes is called **sample space.**

In an experiment of tossing a fair coin, which results in Head H or Tail T, the sample space S = {H, T}. If a coin is tossed two times successively, all possible outcomes are HH, TT, HT, TH. Sample space in this case is the set S = {HH, TT, HT, TH}.

If a die is thrown two times successively, sample space

$$S = \{(1, 1), (1, 2) \ldots (1, 6)\}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

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

Random Variable : It is a real valued function defined over the sample space of an experiment. A variable whose value is a number determined by the out come of an experiment, associated with a sample spaces is called **random variable.** It is usually denoted by capital letter X or Y etc. If out comes are o_i or x_i $i = 1, 2, 3, \ldots$ then $X(o_i)$ or $X(x_i)$ or $f(x_i)$ stands for the value x at $X = x_i$.

Probability Function : X is random variable with values x_i, $i = 0, 1, 2, \ldots$ n and associated probabilities $p(x_i)$. The set p with elements $[x_i, p(x_i)]$ is called the probability function or probability distribution function of X. It can also be called **probability density function of x.**

Illustration

Ex. 1 : A coin is tossed which results in Head or Tail. Let X be the random variable whose value for any outcome is the number of Heads obtained. Find the probability function of x and contruct a probability distribution table.

Sol. : Let H denote a head and T a tail

Sample space is $S = \{H, T\}$

$X(H) = 1,\ X(T) = 1$

x is number of Heads which takes the values 0 and 1

$$f(x) = p(X = x)$$

$$f(0) = \frac{1}{2}, f(1) = \frac{1}{2}$$

Probability distribution table is

X(x)	0	1
f(x)	$\frac{1}{2}$	$\frac{1}{2}$

Ex. 2 : *A coin is tossed two times successively X is the random variable whose value for any outcome is the number of heads obtained. Find the probability function of X and construct probability distribution table.*

Sol. : Sample space is $S = \{HH, TT, HT, TH\}$

$$X(x = 0) = 1, \; X(x = 1) = 2, \; X(x = 2) = 1$$

x is the number of heads which takes values 0, 1 and 2.

$$p(X = 0) = p(TT) = p(T)\,p(T) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$p(X = 1) = p(HT) + p(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$p(X = 2) = p(HH) = \frac{1}{4}$$

Probability distribution table is,

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

Note that, $\Sigma p(x) = 1$

Ex. 3 : *For an experiment of simultaneous throw with three coins. The results of three tosses are independent of each other. Find the probability function and construct distribution table.*

Sol.: All the outcomes can be

HHH (all the three tosses giving Heads), HHT, HTH, HTT, THH, THT, TTH and TTT.

There are in all eight outcomes. Probability of any of these outcomes viz. the event HHH is given by

$$P(HHH) = P(H)\,P(H)\,P(H)$$

$$= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

All these outcomes will have the same probabilities.

Let x denote the number of Heads appearing in each case, which takes values 0, 1, 2 and 3.

∴ (i) $P(x = 0) = P(TTT) = \frac{1}{8}$

(ii) $P(x = 1) = P(HTT) + P(TTH) + P(THT)$, mutually exclusive events

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

(iii) $P(x = 2) = P(HTH) + P(THH) + P(HHT)$, mututally exclusive events

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

(iv) $P(x = 3) = P(HHH) = \frac{1}{8}$

Probability distribution table is

x	0	1	2	3
f	1	3	3	1
P (x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Here the variate x is taking the values x = 0, 1, 2, 3 and the corresponding probabilities are $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$ respectively.

Thus, in the distribution, if the frequencies are replaced by corresponding relative frequencies, then it is called probabilities in the probability distribution.

Here, $\sum P(x) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$ which is always the case when all possible outcomes are considered. Thus, we note that

(i) $p(x) \geq 0$ for all x (ii) $\sum p(x) = 1$.

Ex. 4 : *The outcomes of a certain experiment are $x_1 = 1, x_2 = 2, x_3 = 3$. The associated probability function is*

$$\begin{aligned} p(x) &= k, & x = 1 \\ &= 2k, & x = 2 \\ &= 5k, & x = 3 \end{aligned}$$

Find $p(x < 2)$, $p(x \leq 2)$, $p(x \leq 3)$.

Sol.: $p(x(x_1)) + p(x(x_2)) + p(x(x_3)) = 1$

$$\therefore \quad k + 2k + 5k = 1, \; 8k = 1 \therefore k = \frac{1}{8}$$

$$p(x < 2) = p(x(x_1)) = \frac{1}{8}$$

$$\begin{aligned} p(x \leq 2) &= p(x = 1) + p(x = 2) \\ &= \frac{1}{8} + \frac{2}{8} = \frac{3}{8} \end{aligned}$$

$$\begin{aligned} p(x \leq 3) &= p(x = 1) + p(x = 2) + p(x = 3) \\ &= \frac{1}{8} + \frac{2}{8} + \frac{5}{8} = \frac{8}{8} = 1 \end{aligned}$$

Ex. 5 : *Given the following probability function*

x	*0*	*1*	*2*	*3*	*4*	*5*
f(x)	*0*	*c*	*2c*	*2c*	*1c*	$7c^2$

(i) Find c (ii) Find $P(x \geq 2)$ (iii) Find $P(x < 3)$.

Sol.:

(i) $c + 2c + 2c + c + 7c^2 = 1$

$$\therefore \quad 7c^2 + 6c - 1 = 0$$

$$(7c - 1)(c + 1) = 0, \; 7c - 1 = 0 \text{ or } c = \frac{1}{7}$$

c = –1 is rejected as probability cannot be negative.

(ii)
$$\begin{aligned} p(x \geq 2) &= p(x = 2) + p(x = 3) + p(x = 4) + p(x = 5) \\ &= \frac{2}{7} + \frac{2}{7} + \frac{1}{7} + \frac{7}{49} = \frac{6}{49} \end{aligned}$$

(iii)
$$\begin{aligned} p(x < 3) &= p(x = 0) + p(x = 1) + p(x = 2) \\ &= 0 + \frac{1}{7} + \frac{2}{7} = \frac{3}{7} \end{aligned}$$

8.6.2 Mathematical Expectation

If X is random variable with all possible values $x_1, x_2, x_3, \ldots x_n$ and probability functions (probabilities) $p(x_1), p(x_2), p(x_3), \ldots p(x_n)$, then the mathematical expectation of X is denoted by E(x) and is given by;

$$E(x) = \sum_{i=1}^{n} x_i \, p(x_i)$$

Since, $\Sigma \, p(x_i) = 1$.

E (x) is called ***expected value*** or ***mean value*** of X.

Note :

(1) E(x) is the arithmetic mean μ of random variable X :

We note that corresponding to the number $\bar{x}$ which is Arithmetic Mean in frequency distribution and is given by;

$$\bar{x} = \frac{\Sigma f x}{\Sigma f} = \frac{f_1}{\Sigma f} x_1 + \frac{f_2}{\Sigma f} x_2 + \ldots + \frac{f_n}{\Sigma f} x_n$$

We have,
$$\bar{x} = \sum_{i=1}^{n} p_i x_i = E(x) = \mu$$

where $p_i = \frac{f_i}{\Sigma f_i}$, i = 1, 2, … n are the relative frequencies of $x_1, x_2, \ldots, x_n$ respectively.

2. Variance Var (X) = σ^2 of a random variable X :

The expected value of X, E(X) provides a measure of central tendency of the probability distribution. However, it does not provide any idea regarding the spread of the distribution. Thus, we define variance of X.

If X be a random variable with probability distributions $[x_i, p_i]$, i = 1, 2, … n. Variance of X denoted by;

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

The mean of X is generally denoted by μ.

Thus,
$$Var(X) = \sigma^2 = E(X - \mu)^2$$

For computation simplification,

$$\begin{aligned} Var(X) &= \sigma^2 = E(x-\mu)^2 = \sum_{i=1}^{n} (x_i - \mu)^2 p_i \\ &= \sum_{i=1}^{n} x_i^2 p_i - 2\mu \sum_{i=1}^{n} x_i p_i + \mu_2^2 \Sigma p_i \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

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

3. Moments of a random variable :

The mean measures central tendency. While the variance measure spread. In order to get complete information of the probability distribution, we also have to study the shape of the probability distribution function. Thus, we need measures of skewness (lack of symmetry) and kurtosis (peakedness) of probability distribution. Moments of a random variable serves the purpose.

Let $[x_i, p_i]$, i = 1, 2, 3, … n represent a probability distribution of random variable X.

(i) Moments about any arbitrary point a :

$$\mu_r' = \mu_r'(a) = E(X-a)^r = \sum_{i=1}^{n} (x_i - a)^r p_i, \quad r = 1, 2, 3, \ldots$$

In particular,
$$\mu_1'(a) = E(X - a) = E(X) - a$$
$$\mu_2'(a) = E(X - a)^2$$

(ii) Moments about the origin (i.e. zero) or Raw moments :

$$\mu'_r = \mu'_r(0) = E(X-0)^r = \sum_{i=1}^{n} x_i^r p_i, \quad r = 1, 2, 3 \ldots$$

In particular,

$$\mu'_1 = E(X) = \text{mean}$$

$$\mu'_2 = E(X^2) = \sum_{i=1}^{n} x_i^2 p_i$$

(iii) Moments about the arithmetic mean E(x) or Central moments :

$$\mu_r = \mu'_r(E(X)) = E[X - E(X)]^r$$

$$= \sum_{i=1}^{n} [x_i - E(x)]^r, \quad r = 1, 2, 3$$

In particular,

$$\mu_1 = E[X - E(X)] = E(x) - E(x) = 0$$

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

***Ex. 6 :** A die is tossed once. Random variable x denote the digit that appears. Find the expectation of x.*

Sol.:

x	1	2	3	4	5	6
p(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(x) = \Sigma x\, p(x)$$

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

$$= \frac{(1+2+3+4+5+6)}{6} = \frac{21}{6} = \frac{7}{2}$$

***Ex. 7 :** A die is thrown twice. X denote the sum of digits in two throws. Find mathematical expectation of x.*

Sol.:

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

$$E(x) = \Sigma x\, p(x) = 2\times\frac{1}{36} + 3\times\frac{2}{36} + 4\times\frac{3}{36} + 5\times\frac{4}{36} + 6\times\frac{5}{36} + 7\times\frac{6}{36} + 8\times\frac{5}{36} + 9\times\frac{4}{36} + 10\times\frac{3}{36} + 11\times\frac{2}{36} + 12\times\frac{1}{36}$$

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

$$E(x) = 7$$

***Ex. 8 :** There are three envelopes containing ₹ 100, ₹ 400, and ₹ 700 respectively. A player selects an envelop and keep with him, what he gets. Find the expected gain of the player.*

Sol.: Each envelope has probability $\frac{1}{3}$ of getting selected probability table is;

x	100	400	700
p(x)	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$$E(x) = \Sigma\, xp(x) = 100 \times \frac{1}{3} + 400 \times \frac{1}{3} + 700 \times \frac{1}{3}$$

$$= \frac{1200}{3} = 400$$

Expected gain is ₹400.

Ex. 9 : If random variable X takes the values X = 1, 2, 3 with corresponding probabilities $\frac{1}{6}, \frac{2}{3}, \frac{1}{6}$. *Find E ($x^2$).*

Sol.: Probability distribution table is

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

$$E(x^2) = \Sigma x^2 \cdot p(x) = (1)^2 \times \frac{1}{6} + (2)^2 \times \frac{2}{3} + (3)^2 \times \frac{1}{6}$$

$$= \frac{1}{6} + \frac{8}{3} + \frac{9}{6} = \frac{1 + 16 + 9}{6} = \frac{26}{6} = \frac{13}{3}$$

$$E(x) = 1 \times \frac{1}{6} + 2 \times \frac{2}{3} + 3 \times \frac{1}{6}$$

$$= \frac{1}{6} + \frac{4}{3} + \frac{1}{2} = \frac{1 + 8 + 3}{6} = \frac{12}{6} = 2$$

$$[E(x)]^2 = 4$$

It is clear that $E(x^2) \neq [E(x)]^2$

***Ex. 10 :** Variable x takes the values 0, 1, 2, 3, 4, 5 with probability of each as* $\frac{1}{15}, \frac{2}{15}, \frac{3}{15}, \frac{3}{15}, \frac{5}{15}, \frac{1}{15}$, *find expectation of x.*

Sol.: Given that

x	0	1	2	3	4	5
p (x)	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{3}{15}$	$\frac{5}{15}$	$\frac{1}{15}$

$$E(x) = \Sigma\, xp(x) = 0 \times \frac{1}{15} + 1 \times \frac{2}{15} + 2 \times \frac{3}{15} + 3 \times \frac{3}{15} \times 4 \times \frac{5}{15} + 5 \times \frac{1}{15}$$

$$= \frac{2 + 6 + 9 + 20 + 5}{15} = \frac{42}{15}$$

8.6.3 Continuous Random Variables

In our discussion so far random variable X was taking discrete values. X can take the values in the interval $a \leq x \leq b$ or in the range [a, b], for example, temperature, height and weights take all values in the interval. Let X be a random variable which can take all the values in some interval, then x is called continuous random variable.

The ***probability density function f(x)***, called **p.d.f.** satisfies following conditions

(i) $f(x) \geq 0$ for all $x \in R_x$

(ii) $\int_{R_x} f(x)\, dx = 1$

We define probability for any interval $c < x < d$

$$p(c < x < d) = \int_c^d f(x)\, dx$$

Ex. 11 : *Given the density function*

$$f(x) = ke^{-\alpha x}, \quad x \geq 0, \alpha > 0$$
$$= 0, \quad \text{otherwise}$$

Find k.

Sol.: We have $\int_{-\infty}^{\infty} f(x, \alpha)\, dx = \int_{-\infty}^{0} f(x, \alpha)\, dx + \int_{0}^{\infty} f(x, \alpha)\, dx = 1 = 0 + \int_{0}^{\infty} k\, e^{-\alpha x}\, dx = 1$

$$\therefore \quad k\left[\frac{e^{-\alpha x}}{-\alpha}\right]_0^{\infty} = 1 \text{ or } k\left[\frac{1}{\alpha}\right] = 1$$

$$\therefore \quad k = \alpha$$

Ex. 12 : *A continuous random variable x has the p.d.f. defined as*

$$p(x) = \frac{1}{2}x, \quad 0 < x \leq 1$$
$$= \frac{1}{4}(3 - x), \quad 1 < x \leq 2$$
$$= \frac{1}{4}, \quad 2 < x \leq 3$$
$$= \frac{1}{4}(4 - x), \quad 3 < x \leq 4$$

(i) Compute p (3 < x ≤ 4) (ii) Compute p (1 < x ≤ 4)

Sol.:

(i)
$$p(x \geq 3) = \int_3^4 p(x)\, dx = \int_3^4 \frac{1}{4}(4 - x)dx = \frac{1}{4}\left[4x - \frac{x^2}{2}\right]_3^4 = \frac{1}{4}\left[(16 - 8) - \left(12 - \frac{9}{2}\right)\right]$$
$$= \frac{1}{4}\left[8 - 12 + \frac{9}{2}\right] = \frac{1}{4}\left[-4 + \frac{9}{2}\right] = \frac{1}{8}$$

(ii)
$$p(1 < x \leq 4) = \int_1^2 \frac{1}{4}(3 - x)\, dx + \int_2^3 \frac{1}{4}\, dx + \int_3^4 \frac{1}{4}(4 - x)\, dx$$
$$= \frac{1}{4}\left[3x - \frac{x^2}{2}\right]_1^2 + \frac{1}{4}[x]_2^3 + \frac{1}{4}\left[4x - \frac{x^2}{2}\right]_3^4$$
$$= \frac{1}{4}\left[(6 - 2) - \left(3 - \frac{1}{2}\right) + \frac{1}{4}(3 - 2) + \frac{1}{8}\right] = \frac{1}{4}\left[4 - \frac{5}{2} + 1 + \frac{1}{2}\right]$$
$$= \frac{3}{4}$$

Ex. 13 : *The probability density function f(x) of a continuous random variable x is defined by* $f(x) = ke^{-|x|}$, $x \in (-\infty, \infty)$. *Find the value of k.*

Sol.: We have

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

$$\therefore \quad k\left[(e^x)_{-\infty}^0 + (-e^{-x})_0^{\infty}\right] = 1$$
$$k[1 + 1] = 1$$
$$\text{or} \quad k = \frac{1}{2}$$

8.7 BINOMIAL PROBABILITY DISTRIBUTIONS

Consider the experiment or a trial which has only two outcomes, a success or failure with p as the probability of success and q as the probability of failure. Since there are only two outcomes, $p + q = 1$.

Let us consider series of n such independent trials each of which either results in success or failure.

To find the probability of r successes in n trials, consider one run of outcomes.

$$\underbrace{SSS \ldots\ldots S}_{r} \quad \underbrace{FFF \ldots\ldots F}_{n-r}$$

In which there are r consecutive successes and $n - r$ failures.

Probability of this event P(r success in n trial) is given by

$$\begin{aligned} P(SSS \ldots S\ FFF \ldots F) &= P(S)\,P(S) \ldots (r \text{ times}) \times P(F)\,P(F) \ldots \{(n-r) \text{ times}\} \\ &= pp \ldots p\ (r \text{ times}) \times qq \ldots q\ (n - r \text{ times}) \\ &= p^r q^{n-r} \end{aligned}$$

r success and $n - r$ failures can occur in ${}^nC_r$ mutually exclusive cases each of which has the probability $p^r q^{n-r}$.

$\therefore$ Probability of r success in n trials is ${}^nC_r \cdot p^r q^{n-r}$. This formula gives probability of $r = 0, 1, 2, 3, \ldots n$ success in n trials.

Putting it in tabular form,

r	0	1	2	3	......	n
p (r)	${}^nC_0 p^0 q^n$	${}^nC_1 p^1 q^{n-1}$	${}^nC_2 p^2 q^{n-2}$	${}^nC_3 p^3 q^{n-3}$	......	${}^nC_n p^n q^{n-n}$

$${}^nC_0 = 1, \quad {}^nC_n = 1$$

Consider now the Binomial expansion of

$$(q + p)^n = q^n + {}^nC_1 q^{n-1} p + {}^nC_2 q^{n-2} p^2 + \ldots + p^n$$

Terms on R.H.S. of this expansion give probability of $r = 0, 1, 2, \ldots, n$ success. This is the reason for above probability distribution to be called Binomial probability distribution. It is denoted by B (n, p, r).

Thus, $$B(n, p, r) = {}^nC_r p^r q^{n-r}$$

If n independent trials constitute one experiment and the experiment is repeated N times then r successes would be expected to occur $N \times {}^nC_r p^r q^{n-r}$ times. This is called the expected frequency of r success in N experiments.

8.7.1 Mean and Variance of Binomial Distribution :

We shall first obtain moments of the binomial distribution $X \rightarrow B(n, p, r)$ about $r = 0$ (about origin) (refer article 8.6.2).

$$\begin{aligned} \mu_1' &= \mu_1 = E(x - 0) = \text{Mean} = \sum_{r=0}^{n} r\,p(r) \\ &= \sum_{r=0}^{n} r \cdot {}^nC_r\, p^r q^{n-r} \\ &= 0 \cdot q^n + 1 \cdot {}^nC_1 p^1 q^{n-1} + 2 \cdot {}^nC_2\, p^2 q^{n-2} + 3 \cdot {}^nC_3 p^3 q^{n-3} \ldots np^n \end{aligned}$$

$$= npq^{n-1} + 2 \cdot \frac{n(n-1)}{2!} p^2 q^{n-2} + 3 \cdot \frac{n(n-1)(n-2)}{2!} p^3 q^{n-3} \ldots np^n$$

$$= np \left[q^{n-1} + (n-1) pq^{n-2} + \frac{(n-1)(n-2)}{2!} p^2 q^{n-3} \ldots p^{n-1} \right]$$

$$\mu_1' = np\,[(q+p)^{n-1}] = np$$

$$\therefore \quad \boxed{\mu_1' = \mu_1 = E(X) = \text{Mean} = np}$$

Hence, Mean of the Binomial distribution (which is also the expectation of variable r) is np.

Next, consider second moment about origin (refer article 8.6.2),

$$\mu_2' = E[X^2] = E[X(X-1)] + E(X)$$

$$\mu_2' = \sum_{r=0}^{n} r^2\, {}^nC_r\, p^r q^{n-r} = \sum_{r=0}^{n} \{r(r-1) + r\}\, {}^nC_r\, p^r q^{n-r}$$

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

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

$$= 1.2\, {}^nC_2\, p^2 q^{n-2} + 2.3\, {}^nC_3\, p^3 q^{n-3} + 3.4\, {}^nC_4\, p^4 q^{n-4} \ldots n(n-1)\, p^n + np$$

$$= 1.2 \frac{n(n-1)}{2!} p^2 q^{n-2} + 2.3 \frac{n(n-1)(n-2)}{3!} p^3 q^{n-3} + 3.4 \frac{n(n-1)(n-2)(n-3)}{4!} p^4 q^{n-4} \ldots n(n-1)\, p^n + np$$

$$= n(n-1) \times p^2 \left[q^{n-2} + (n-2)\, p\, q^{n-3} + \frac{(n-2)(n-3)}{2!} p^2 q^{n-4} \ldots p^{n-2} \right] + np$$

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

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

$$= n^2 p^2 + npq$$

$$\therefore \quad \mu_2 = \text{Variance} = \mu_2' - \mu_1'^2 = E(X^2) - [E(X)]^2$$

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

$$\therefore \quad \boxed{\text{Var}(X) = \sigma^2 = npq}$$

and $\boxed{\text{S.D.} = \sigma = \sqrt{npq}}$

ILLUSTRATIONS

***Ex. 1 :** An unbiased coin is thrown 10 times. Find the probability of getting exactly 6 Heads, at least 6 Heads.*

Sol. : Here $p = q = \frac{1}{2}$ and $n = 10$. Here occurrence of Head is treated as success.

Probability of getting exactly 6 Heads is

$$P(\text{exactly 6 heads}) = {}^{10}C_6 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^4$$

Events of at least six Heads occur when coin shows up Head 6, 7, 8, 9 or 10 times the probabilities for these events are

$$P(7) = {}^{10}C_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 \qquad P(8) = {}^{10}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^2$$

$$P(9) = {}^{10}C_9 \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right)^1 \qquad P(10) = {}^{10}C_{10} \left(\frac{1}{2}\right)^{10} \left(\frac{1}{2}\right)^0 = \left(\frac{1}{2}\right)^{10}$$

$$\begin{aligned} p(\text{at least 6 Heads}) &= p(6) + p(7) + p(8) + p(9) + p(10) \\ &= \left(\frac{1}{2}\right)^{10} [{}^{10}C_6 + {}^{10}C_7 + {}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10}] \\ &= \frac{386}{1024} = 0.37695 \end{aligned}$$

[The events of 6, 7, ... etc. Heads are mutually exclusive].

Ex. 2 : *Probability of Man aged 60 years will live for 70 years is $\frac{1}{10}$. Find the probability of 5 mean selected at random 2 will live for 70 years.*

Sol. : Here $p = \frac{1}{10}$, $q = \frac{9}{10}$, $r = 2$, $n = 5$.

$$P(\text{2 men living for 70 years}) = {}^5C_2 \left(\frac{1}{10}\right)^2 \left(\frac{9}{10}\right)^3 = 0.0729$$

Ex. 3 : *On an average a box containing 10 articles is likely to have 2 defectives. If we consider a consignment of 100 boxes, how many of them are expected to have three or less defectives ?* ***(May 2009)***

p = *Probability of box containing defective articles* $= \frac{2}{10} = \frac{1}{5}$

q = *Probability of non-defective items* $= \frac{4}{5}$

Sol. : Probability of box containing three or less defective articles

$$P(r \le 3) = p(r = 0) + p(r = 1) + p(r = 2) + p(r = 3)$$ [r denotes the number of defective items.]

$$P(r = 0) = {}^{10}C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^{10} = 0.1074$$

$$P(r = 1) = {}^{10}C_1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^9 = 0.2684$$

$$P(r = 2) = {}^{10}C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^8 = 0.302$$

$$P(r = 3) = {}^{10}C_3 \left(\frac{1}{5}\right)^3 \left(\frac{4}{5}\right)^7 = 0.2013$$

$$P(r \le 3) = 0.1074 + 0.2684 + 0.302 + 0.2013 = 0.8791$$

$$\therefore \quad NP(r \le 3) = 100 \times 0.8791 = 87.91$$

88 boxes are expected to contain three or less defectives.

Ex. 4 : *The probability of a man hitting a target is $\frac{1}{3}$. If he fires 5 times, what is the probability of his hitting the target at least twice ?*

Sol. : Probability of hitting target $= p = \frac{1}{3}$

Probability of no hit (failure) $= q = \frac{2}{3}$

r denotes the number of hits (successes) and n = 5.

Probability of man hitting the target at least twice.

$$P(r \geq 2) = P(r = 2) + P(r = 3) + P(r = 4) + P(r = 5)$$
$$= 1 - P(r < 2)$$
$$= 1 - P(r = 0) - P(r = 1)$$
$$= 1 - {}^5C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^5 - {}^5C_1 \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^4$$
$$= \frac{131}{243} = 0.539$$

Ex. 5 : *A coin is so biased that appearance of head is twice likely as that of tail. If a throw is made 6 times, find the probability that atleast 2 heads will appear.*

Sol.: Here $P(H) = \frac{2}{3}, q = P(T) = \frac{1}{3}$

A is the event of appearance of atleast two heads

$$P(A \geq 2) = P(2) + P(3) + P(4) + P(5) + P(6)$$
$$= 1 - P(0) - P(1)$$
$$= 1 - {}^6C_0 \left(\frac{2}{3}\right)^0 \left(\frac{1}{3}\right)^6 - {}^6C_1 \left(\frac{2}{3}\right)^1 \left(\frac{1}{3}\right)^5$$
$$= 1 - \left\{1 \times \left(\frac{1}{3}\right)^6 + 6 \times \frac{2}{3} \left(\frac{1}{3}\right)^5\right\}$$
$$= 0.98213$$

Ex. 6 : *Assume that an average one telephone number out of 15 called between 12 p.m. to 3 p.m. on week days is busy. What is the probability that if 6 randomly selected telephone numbers called (i) not more than 3 (ii) at least 3 of them be busy.*

Sol. : The probability that the telephone number, called beween 12 p.m. to 3 p.m., is busy is

$$p = \frac{1}{15} \quad \therefore q = 1 - \frac{1}{15} = \frac{14}{15} \qquad (\because p + q = 1)$$

Hence, probability that r numbers called out of 6 called (by binomial distribution is)

$$P(r) = {}^6C_r\, p^r q^{n-r} = {}^6C_r \left(\frac{1}{15}\right)^r \left(\frac{14}{15}\right)^{6-r} \qquad \dots (1)$$

(i) For not more than 3 calls be busy,

$$P(r \leq 3) = P(r = 0) + P(r = 1) + P(r = 2) + P(r = 3)$$
$$= {}^6C_0 \left(\frac{1}{15}\right)^0 \left(\frac{14}{15}\right)^6 + {}^6C_1 \left(\frac{1}{15}\right)^1 \left(\frac{14}{15}\right)^5 + {}^6C_2 \left(\frac{1}{15}\right)^2 \left(\frac{14}{15}\right)^4 + {}^6C_3 \left(\frac{1}{15}\right)^3 \left(\frac{14}{15}\right)^3$$
$$= \frac{(14)^3}{(15)^6} \left[(14)^3 + 6(14)^2 + \frac{6 \cdot 5}{2 \cdot 1}(14) + \frac{6 \cdot 5 \cdot 4}{3 \cdot 2 \cdot 1}(1)\right]$$
$$= (0.002409)\ [2744 + 1176 + 210 + 20]$$
$$= 0.997$$

(ii) For at least 3 calls to be busy of 6 balls, we have the probability as

$$P(r \geq 3) = 1 - P(r < 3)$$
$$= 1 - [P(r = 0) + P(r = 1) + P(r = 2)]$$
$$= 1 - \left[{}^6C_0 \left(\frac{1}{15}\right)^0 \left(\frac{14}{15}\right)^6 + {}^6C_1 \left(\frac{1}{15}\right)^1 \left(\frac{14}{15}\right)^5 + {}^6C_2 \left(\frac{1}{15}\right)^2 \left(\frac{14}{15}\right)^4\right]$$

$$= 1 - \frac{(14)^4}{(15)^6}\left[(14)^2 + 6(14) + \frac{6\cdot 5}{2\cdot 1}(1)\right]$$

$$= 1 - (0.00034)\,[196 + 84 + 15]$$

$$= 1 - 0.9949$$

$$= 0.0051$$

***Ex. 7 :** 20% of bolts produced by a machine are defective. Determine the probability that out of 4 bolts chosen at random (i) 1 is defective (ii) zero are defective (iii) at most 2 bolts are defective.*

Sol. : The probability of defective bolts is

$$p = \frac{20}{100} = 0.2 \quad \therefore \quad q = 0.8 \qquad (\because p + q = 1)$$

(i) The probability of having 1 defective bolts out of 4 is

$$P(r = 1) = {}^4C_1\,(0.2)^1\,(0.8)^3 = 4(0.2)\,(0.8)^3 = 0.4096$$

(ii) The probability of having zero bolts defective is

$$P(r = 0) = {}^4C_0\,(0.2)^0\,(0.8)^4 = 4(0.8)^4 = 0.4096$$

(iii) The probability of having at most 2 bolts out of 4 is

$$P(r \le 2) = P(r = 0) + P(r = 1) + P(r = 2)$$

$$= {}^4C_0\,(0.2)^0\,(0.8)^4 + {}^4C_1\,(0.2)^1\,(0.8)^3 + {}^4C_2\,(0.2)^2\,(0.8)^2$$

$$= (0.8)^2\left[(0.8)^2 + 4(0.2)\,(0.8) + \frac{4.3}{2.1}(0.2)^2\right]$$

$$= 0.9728$$

***Ex. 8 :** Out of 2000 families with 4 children each, how many would you expect to have (i) at least a boy (ii) 2 boys, (iii) 1 or 2 girls, (iv) no girls ?*

Sol. :

$$p = \text{probability of having a boy} = \frac{1}{2}$$

$$q = \text{probability of having a girl} = 1 - \frac{1}{2} = \frac{1}{2} \qquad (\because p + q = 1)$$

(i)
$$p(\text{at least a body}) = P(r \ge 1)$$

$$= 1 - P(r < 1) = 1 - P(r = 0)$$

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

Hence, expected number of families having at least a boy,

$$2000\,P(r \ge 1) = 2000 \times \frac{15}{16} = 1875$$

(ii)
$$P(\text{having 2 boys}) = P(r = 2)$$

$$= {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{4.3}{2.1} \times \frac{1}{16} = \frac{3}{8}$$

Expected number of families having 2 boys

$$2000\,P(r = 2) = 2000 \times \frac{3}{8} = 750$$

(iii)
$$P(\text{having 1 or 2 girls}) = P(\text{having 3 boys or 2 boys})$$

$$= P(r = 3) + P(r = 2)$$

$$= {}^4C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 + {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2$$

$$= \frac{1}{(2)^4}\left[4 + \frac{4.3}{2.1}\right] = \frac{10}{16} = \frac{5}{8}$$

Expected number of families having 1 or 2 girls

$$2000\,[P(r = 3) + p(r = 2)] = 2000 \times \frac{5}{8} = 1250$$

(iv) $$P(\text{having no girls}) = P(\text{having 4 boys}) = P(r = 4)$$

$$= {}^4C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^0 = \frac{1}{16}$$

Expected number of families having no girls

$$2000\,[p(r = 4)] = 2000 \times \frac{1}{16} = 125$$

Ex. 9 : *In a quality control department of a rubber tube manufacturing factory, 10 rubber tubes are randomly selected from each day's production for inspection. If not more than 1 of the 10 tubes is found to be defective, the production lot is approved. Otherwise it is rejected. Find the probability of the rejection of a day's production lot if the true proportion of defectives in the lot is 0.3.*

Sol. : Suppose X denotes the number of defective tubes in the 10 randomly selected tubes.

$\therefore \quad X \rightarrow B\,(n = 10,\ p = 0.3)$

The production lot is accepted if not more than one tube (i.e. at the most one tube) is found defective.

$$\therefore \quad P\,(\text{Accepting the lot}) = P\,(X = 0) + P\,(X = 1)$$

$$= q^n + npq^{n-1}$$

$$= (0.7)^{10} + 10\,(0.3)\,(0.7)^9$$

$$= (0.7)^9\,[3.7] = 0.1493$$

$$\therefore \quad P\,(\text{Rejection of the lot}) = 1 - 0.1493 = 0.8507$$

Ex. 10 : *A department in a works has 10 machines which may need adjustment from time to time during the day. Three of these machines are old, each having a probability of $\frac{1}{10}$ of needing adjustment during the day and 7 are new, having corresponding probability $\frac{1}{20}$. Assuming that no machine needs adjustment twice on the same day, determine the probabilities that on a particular day.*

(i) just two old and no new machines need adjustment.

(ii) just two machines need adjustment which are of the same type.

Sol. : Out of 3 old machines, if 2 need adjustment then this combination of 2 needing adjustment and one needing cannot occur in ${}^3C_2$ ways.

$$p_1 = \text{Probability of old machine needing adjustment} = \frac{1}{10}$$

$$q_1 = \text{Probability of old machine not needing an adjustment} = \frac{9}{10}$$

$$p_2 = \text{Probability of new machine needing an adjustment} = \frac{1}{20}$$

$$q_2 = \text{Probability of new machine not needing an adjustment} = \frac{19}{20}$$

We are looking for an event A where 2 old machines needing adjustment alongwith one not needing and remaining 7 new machines not needing an adjustment. Hence

$$P\,(A) = {}^3C_2 \left[\left(\frac{1}{10}\right)^2 \left(\frac{9}{10}\right)\right]\left(\frac{19}{20}\right)^7 = \frac{3 \times 9 \times (19)^7}{10^3\,(20)^7} = 0.0189$$

Consider the event B in which case, no old machine out of 3 need adjustment and 2 out of 7 new machines need adjustment.

Proceeding in the same way

$$P(B) = \left(\frac{9}{10}\right)^3 \left[{}^7C_2 \left(\frac{1}{20}\right)^2 \left(\frac{19}{20}\right)^5\right] = \frac{7\cdot 6}{1\cdot 2}\left(\frac{9}{10}\right)^3 \left(\frac{1}{20}\right)^2 \left(\frac{19}{20}\right)^5$$

$$= 21\left(\frac{9}{10}\right)^3 \left(\frac{1}{20}\right)^2 \left(\frac{19}{20}\right)^5 = 0.0296$$

Event, two machines needing an adjustment which are of the same type A + B.

$$P(A + B) = P(A) + P(B) \text{ [A, B are mutually exclusive].}$$

$\therefore$

$$\text{Required probability} = 0.0189 + 0.0296 = 0.0485$$

***Ex. 11 :** A r.v. $X \to B\ (n = 6, p)$. Find p if 9 P (R = 4) = P (R = 2).*

Sol. : $P(R = r) = {}^nC_r\, p^r q^{n-r} \quad r = 0, 1, \ldots n$

Here, $n = 6$

$\therefore$ $9P(4) = P(2)$

$\Rightarrow$ $9\cdot\binom{6}{4} p^4 q^2 = \binom{6}{2} p^2 q^4$

$\therefore$ $9p^2 = q^2 \quad \because \binom{6}{4} = \binom{6}{2}$

$\therefore$ $9p^2 = (1-p)^2 = 1 - 2p + p^2$

$\therefore$ $8p^2 + 2p - 1 = 0$

$\therefore$ $(4p-1)(2p+1) = 0$

$\Rightarrow$ $p = \frac{1}{4}$ or $p = -\frac{1}{2}$

The value $p = -\frac{1}{2}$ is inadmissible. Hence, the answer is $p = \frac{1}{4}$.

***Ex. 12 :** Point out the fallacy of the statement 'The Mean of Binomial distribution is 3 and variance 5'.*

Sol. : Given Mean $= np = 3$, Variance $= npq = 5$

$\therefore$ $q = \frac{npq}{np} = \frac{5}{3} > 1$

which is not possible since probability cannot exceed unity.

***Ex. 13 :** The Mean and Variance of Binomial distribution are 4 and 2 respectively. Find p (r ≥ 1).*

Sol. : Here r denotes the number of successes in n trials. Given that

$$\text{Mean} = np = 6 \quad \text{and} \quad \text{Variance} = npq = 2$$

$\therefore$ $q = \frac{npq}{np} = \frac{2}{6} = \frac{1}{3}$

$$p = 1 - q = 1 - \frac{1}{3} = \frac{2}{3}$$

$$np = 6 \quad \therefore n = \frac{6}{2/3} = 9$$

$$P(r \geq 1) = 1 - P(r = 0) = 1 - q^n = 1 - \left(\frac{1}{3}\right)^9 = 0.999949$$

8.8 HYPERGEOMETRIC DISTRIBUTION

We know that binomial distribution is applied whenever we draw a random sample *with* replacement. This is because, in sampling with replacement, the probability of getting 'success' p, remains same at every draw. Also, the successive draws remain independent. Thus, the assumptions of binomial experiment are satisfied. Now, consider the following situation.

A bag contains 4 red and 5 black balls. Suppose 3 balls are drawn at random from this bag *without* replacement and we are interested in the number of red balls drawn. Clearly at the first draw, probability of getting a red ball is $\frac{4}{9}$. Now, suppose a red ball is selected at the first draw. Because, it would be kept aside, the probability of getting a red ball at the second draw would be $\frac{3}{8}$. Thus 'p' does not remain constant. Also, the successive draws are not independent. Probability of getting red balls in the second draw is dependent on which ball you have drawn at the first draw. Thus, in case of sampling *without* replacement, the binomial distribution cannot be applied.

In such situations the hypergeometric distribution is used. Consider the following situation.

Suppose a bag contains N balls of which M are red and N – M are black. A sample of 'n' balls is drawn *without replacement* from the N balls. Let X denote the number of red balls in the *sample*. Hence, the possible values of X are 0, 1, 2, ..., n (assuming $n \leq M$). The p.m.f. is obtained in the following manner.

We want to get P [X = x].

$$
\begin{array}{ccc}
 & N & \\
 & \downarrow & \\
M & & N-M \\
\downarrow & & \downarrow \\
x & & n-x
\end{array}
$$

If the sample of 'n' balls contains 'x' red balls, then it will contain 'n – x' black balls. Hence, number of ways in which x red balls can be selected from M red balls is $\binom{M}{x}$ and number of ways in which n – x black balls can be selected from N – M black balls is $\binom{N-M}{n-x}$. The sample contains both red and black balls. Therefore, the total number of ways in which the above event can occur is $\binom{M}{x}\binom{N-M}{n-x}$. In all 'n' balls are selected from N balls. Therefore, the total number of possible selections is $\binom{N}{n}$. Using the definition of probability of an event, we get,

$$
\begin{aligned}
P(x) = P[X = x] &= \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \quad ; \quad x = 0, 1, \ldots, \min(n, M) \\
&= 0 \quad ; \quad \text{otherwise}
\end{aligned}
$$

The above P (x) is called as the p.m.f. of hypergeometric distribution with parameters N, M and n.

Notation : $X \to H(N, M, n)$.

If we don't assume $n \leq M$, then the range X is 0, 1, 2, ..., min (n, M). This is because at the most M red balls can be there in the sample.

Remark : Applicability of Hypergeometric Distribution

Hypergeometric distribution is applied whenever a random sample is taken *without* replacement from a population consisting of two classes. Following are some such situations.

(i) In quality control department, a random sample of items is inspected from a consignment containing defective and non-defective items.

(ii) A lake contains N fish. A sample of fish is taken from the lake, marked and released back in the lake. Next time, another sample of fish is selected and number of marked fish are counted.

(iii) A committee of n persons is to be formed from N persons of whom M are ladies and N – M are gentlemen. The number of ladies on the committee follows hypergeometric distribution.

(iv) In opinion surveys, where the persons have to give answers of 'yes', 'no' type.

The following conditions should be satisfied for the application of hypergeometric distribution.

1. The population is divided into two mutually exclusive categories.
2. The successive outcomes are dependent.
3. The probability of 'success' changes from trial to trial.
4. The number of draws are fixed.

ILLUSTRATIONS

***Ex. 1 :** A room has 4 sockets. From a collection of 12 bulbs of which only 5 are good. A person selects 4 bulbs at random (without replacement) and puts them in the sockets. Find the probability that (i) the room is lighted, (ii) exactly one bulb in the selected bulbs is good.*

Sol. : Notice that N = 12, M = 5, n = 4, X = number of good bulbs in the sample.

$\therefore \quad X \to H\,(N = 12, M = 5, n = 4)$

$$\therefore \quad P(x) = \frac{\binom{5}{x}\binom{7}{4-x}}{\binom{12}{4}}\ ; \qquad x = 0, 1, \ldots, 4$$

(i) The room is lighted even if a single bulb is good. Therefore the required probability is

$$P(X \geq 1) = 1 - P(X = 0)$$

$$= 1 - \frac{\binom{5}{0}\binom{7}{4}}{\binom{12}{4}} = 0.9292$$

(ii)

$$P[X = 1] = \frac{\binom{5}{1}\binom{7}{3}}{\binom{12}{4}} = 0.707$$

***Ex. 2 :** Among the 200 employees of a company, 160 are union numbers and the others are non-union. If four employees are to be chosen to serve on the staff welfare committee, find the probability that two of them will be union members and the others non-union, using hypergeometric distribution.*

Sol. : Let X denote number of union members selected in the sample.

$\therefore \quad X \to H\,(N = 200, M = 160, n = 4).$

The required probability is

$$P[X = 2] = \frac{\binom{160}{2}\binom{40}{2}}{\binom{200}{4}} = \frac{\dfrac{160 \times 159}{2} \times \dfrac{40 \times 39}{2}}{\dfrac{200 \times 199 \times 198 \times 197}{4 \times 3 \times 2}}$$

$$= 0.1534$$

8.9 POISSON DISTRIBUTION

Poisson distribution is the discrete probability distribution of a discrete random variable X.

When 'p' be the probability of success is very small and n the number of trials is very large and np is finite then we get another distribution called Poisson distribution. It is considered as limiting case of Binomial distribution with $n \to \infty$, $p \to 0$ and np remaining finite.

Consider the Binomial distribution

$$B(n, p, r) = {}^nC_r\, p^r\, q^{n-r}$$

$$= \frac{n(n-1)(n-2)\ \ldots\ (n-(r-1))}{r!}\, p^r\,(1-p)^{n-r}$$

Let $z = np \quad \therefore \quad p = \frac{z}{n}$

Hence,
$$B(n, p, r) = \frac{np\,(np-p)\,(np-2p)\ldots(np-(r-1)\,p)}{r!} \times \frac{(1-p)^n}{(1-p)^r}$$
$$= \frac{z\left(z-\frac{z}{n}\right)\left(z-\frac{2z}{n}\right)\ldots\left[z-(r-1)\frac{z}{n}\right]}{r!} \times \frac{\left(1-\frac{z}{n}\right)^n}{\left(1-\frac{z}{n}\right)^r}$$

Now taking the limit as $n \to \infty$ and $p \to 0$ in such a way that $np = z\ (> 0)$ remains constant.

$$\lim_{n\to\infty} B(n, p, r) = \frac{e^{-z} z^r}{r!} \qquad \left[\because \lim_{n\to\infty}\left(1-\frac{z}{n}\right)^n = e^{-z} \text{ and } \lim_{n\to\infty}\left(1-\frac{z}{n}\right)^r = 1\right]$$

This is called Poisson distribution which may be denoted by p (r).

Thus, the probability of r successes in a series of large number of trials n with p the probability of success at each trial, a small number, probability mass function (p.m.f.) is given by,

$$p(r) = \frac{e^{-z} z^r}{r!}; \quad r = 0, 1, 2, \ldots$$

Here $z > 0$ is called the parameter of the distribution.

We note that (i) $p(r) \geq 0\ \forall r$, $\because e^{-z} > 0$ (ii) $\sum_{r=0}^{\infty} p(r) = \sum_{r=0}^{\infty} \frac{e^{-z} z^r}{r!} = e^{-z}\left(1 + z + \frac{z^2}{3!} + \ldots\right) = e^{-z} e^{z} = 1$. Hence p(r) is a p.m.f.

We experience number of situations where chance of occurrence of an event is a short time interval is very small. However, there are infinitely many opportunities to occur. The number of occurrences of such event follows poisson distribution. Example of such events are :

(i) Number of defectives in a production centre.

(ii) Number of accidents on a highway.

(iii) Number of printing mistakes per page.

(iv) Number of telephone calls during a particular (odd) time.

(v) Number of bad (dishonoured) cheques at a bank.

8.9.1 Mean and Variance of Poisson Distribution

We shall obtain moments of the poisson distribution $X \to P(z)$ about $r = 0$ (about origin) (refer article 8.6.2).

$$\mu_1' = \mu_1 = E(x) = \text{Mean} = \sum_{r=0}^{\infty} r\,p(r)$$
$$= \sum_{r=1}^{\infty} r\frac{e^{-z} z^r}{r!} = z e^{-z} \sum_{r=1}^{\infty} \frac{z^{r-1}}{(r-1)!} \qquad \text{(Since term corresponding to } r = 0 \text{ is zero)}$$
$$= z e^{-z}\left(1 + z + \frac{z^2}{2!} + \ldots\right)$$
$$= z e^{-z} e^{z}$$
$$\boxed{\mu_1 = \mu_1' = E(x) = \text{Mean} = z}$$

Next, consider second moment about origin (refer article 8.6.2),

$$\mu_2' = E(X^2) = E[X(X-1)] + E(X)$$

or
$$\mu_2' = \sum_{r=0}^{\infty} r^2 p(r) = \sum_{r=0}^{\infty} [r(r-1) + r]\left(\frac{e^{-z} z^r}{r!}\right)$$
$$= \sum_{r=0}^{\infty} r(r-1)\frac{e^{-z} z^r}{r!} + \sum_{r=0}^{\infty} r\frac{e^{-z} z^r}{r!}$$

$$= e^{-z} z^2 \sum_{r=2}^{\infty} \frac{z^{r-2}}{(r-2)!} + E(X)$$

$$= e^{-z} z^2 \left(1 + z + \frac{z^2}{2!} + \ldots\right) + z$$

$$= e^{-z} z^2 e^z + z$$

$$= z^2 + z$$

$$\therefore \quad \mu_2 = \text{Var}(X) = \mu_2' - \mu_1'^2 = E(X^2) - [E(X)]^2$$

$$= z^2 + z - z^2 = z$$

$$\therefore \quad \boxed{\text{Var}(X) = \sigma^2 = z}$$

and $\boxed{\text{S.D.} = \sigma = \sqrt{z}}$

Thus, we note that variance of Poisson distribution = mean of Poisson distribution.

ILLUSTRATIONS

Ex. 1 : *A manufacturer of cotter pins knows that 2% of his product is defective. If he sells cotter pins in boxes of 100 pins and guarantees that not more than 5 pins will be defective in a box, find the approximate probability that a box will fail to meet the guaranteed quality.* ***(Dec. 2011)***

Sol. : Here, n = 100.

p the probability of defective pins $= \frac{2}{100} = 0.02$

$$z = \text{mean number of defective pins in a box}$$

$$z = np = 100 \times 0.02 = 2$$

Since p is small, we can use Poisson distribution.

$$P(r) = \frac{e^{-z} z^r}{r!} = \frac{e^{-2} 2^r}{r!}$$

Probability that a box will fail to meet the guaranteed quality is

$$P(r > 5) = 1 - P(r \le 5)$$

$$= 1 - \sum_{r=0}^{5} \frac{e^{-2} 2^r}{r!} = 1 - e^{-2} \sum_{r=0}^{5} \frac{2^r}{r!} = 0.0165$$

Ex. 2 : *In a certain factory turning out razor blades, there is a small chance of 1/500 for any blade to be defective. The blades are supplied in a packet of 10. Use Poisson distribution to calculate the approximate number of packets containing no defective and two defective blades, in a consignment of 10,000 packets.* ***(May 2009, 2018)***

Sol. : Here p = 0.002, n = 10, z = np = 0.02

$$P(\text{no defective}) = P(r = 0) = \frac{e^{-0.02} (0.02)^0}{0!} = \frac{1}{e^{0.02}}$$

$$P(2 \text{ defectives}) = P(r = 2) = \frac{e^{-0.02} (0.02)^2}{2}$$

Number of packets containing no defective blades in a consignment of 10,000 packets

$$= 10{,}000 \times \frac{1}{e^{0.02}} = 9802$$

Number of packets containing 2 defective blades

$$= 10{,}000 \times \frac{(0.02)^2}{2 \times e^{0.02}} = 2$$

Ex. 3 : *The average number of misprints per page of a book is 1.5. Assuming the distribution of number of misprints to be Poisson, find*

(i) The probability that a particular book is free from misprints.

(ii) Number of pages containing more than one misprint if the book contains 900 pages. **(May 2011)**

Sol. : Let X : Number of misprints on a page in the book.

Given : $X \to P\ (z = 1.5)$ $\quad E\ (r) = z = 1.5$

Here the p.m.f. is given by,

$$P(r) = \frac{e^{-z}\, z^r}{r!}$$

$$P(r) = \frac{e^{-1.5}\,(1.5)^r}{r!}$$

(i) $$P\ (r = 0) = \frac{e^{-1.5}\,(1.5)}{0!} = e^{-1.5} = 0.223130$$

Note : The Poisson probabilities for m = 0.1, 0.2, 0.3 ... 15.0 are also given in the statistical tables.

(ii)
$$\begin{aligned} P\ (r > 1) &= 1 - P\ [r \le 1] \\ &= 1 - [P(r = 0) + P(r = 1)] \\ &= 1 - \left[e^{-1.5} + \frac{e^{-1.5}\,(1.5)^1}{1!}\right] \\ &= 1 - (0.223130 + 0.334695) \\ &= 0.442175 \end{aligned}$$

∴ Number of pages in the book containing more than one misprint.

$$\begin{aligned} &= (900)\ P\ [r > 1] = (900)\ (0.442175) \\ &= 397.9575 \approx 398 \end{aligned}$$

Ex. 4 : Number of road accidents on a high way during a month follows a Poisson distribution with mean 5. Find the probability that in a certain month number of accidents on the highway will be *(i) Less than 3 (ii) Between 3 and 5 (iii) More than 3.* ***(May 2014, Nov. 2016)***

Sol. : Let X : number of road accidents on a highway during a month.

Given : $X \to P\ (z = 5)$

∴ The p.m.f is given by,

$$P(r) = \frac{e^{-z}\, z^{-r}}{r!}\ ;\ r = 0, 1, 2 \ldots\ldots$$

$$P(r) = \frac{e^{-5}\, 5^r}{r!}$$

(i)
$$\begin{aligned} P\ (r < 3) = P\ (r \le 2) &= P\ (r = 0) + P\ (r = 1) + P\ (r = 2) \\ &= \frac{e^{-5}\ 5^0}{0!} + \frac{e^{-5}\ 5^1}{1!} + \frac{e^{-5}\ 5^2}{2!} \\ &= 0.006738 + 0.033690 + 0.084224 \\ &= 0.124652 \end{aligned}$$

(ii)
$$\begin{aligned} P\ (3 \le r \le 5) &= P(r = 3) + P(r = 4) + P(r = 5) \\ &= 0.140374 + 0.175467 + 0.175467 = 0.491308 \end{aligned}$$

(iii)
$$\begin{aligned} P\ (r > 3) = 1 - P\ (r \le 3) &= [P(r = 0) + P(r = 1) + P(r = 2) + P(r = 3)] \\ &= 0.734974 \end{aligned}$$

Ex. 5 : *A car hire firm has 2 cars which it hires out day-by-day., The number of demands for the car on each day is distributed as Poisson distribution with parameter 1.5. Calculate the probability of days on which neither car is used and for the days on which demand is refused,.*

Sol. : Let r be number demands for each day, where r takes the value 0, 1, 2. The Poisson probability distribution for demand of r cars on each day is given by

$$p(r) = \frac{e^{-z}\, z^r}{r!} = \frac{e^{-1.5}\,(1.5)^r}{r!} \qquad (\because z = 1.5)$$

(i) The probability of the days on which neither car is used, is

$$p(r = 0) = \frac{e^{-1.5}\,(1.5)^0}{0!} = 0.22$$

(ii) When $r > 2$, the demand will be refused by the firm. Hence the probability of days when the demands is refused is

$$\begin{aligned} p(r > 2) &= 1 - p(r \le 2) \\ &= 1 - [p(r = 0) + p(r = 1) + p(r = 2)] \\ &= 1 - \left[\frac{e^{-1.5}(1.5)^0}{0.!} + \frac{e^{-1.5}(1.5)^1}{1!} + \frac{e^{-1.5}(1.5)^2}{2!}\right] \\ &= 1 - e^{-1.5}\left[1 + (1.5) + \frac{(1.5)^2}{2}\right] \\ &= 0.2025 \end{aligned}$$

Ex. 6 : *A telephone switch board handles 600 calls on the average during rush hour. The board can make a maximum of 20 connections per minute. Use Poisson distribution to estimate the probability the board will be overtaxed during any given minute.*

Sol. : The mean of calls handled per minute is

$$z = \frac{600}{60} = 10$$

Thus the probability of Poisson distribution for r calls to be handled is given by

$$p(r) = \frac{e^{-10}(10)^r}{r!} \qquad (\because z = 10)$$

The board will be overtaxed during given minute, when calls are more than 20 per minute. Hence, the probability that the board will be over taxed for a given minute is

$$\begin{aligned} p(r > 20) &= 1 - p(r \le 20) \\ &= 1 - \sum_{r=0}^{20} \frac{e^{-10}(10)^r}{r!} \end{aligned}$$

Hence the answer.

Ex. 7 : *The accidents per shift in a factory are given by the table :* **(Dec. 2008, May 2002)**

Accidents x per Shift	*0*	*1*	*2*	*3*	*4*	*5*
Frequency f	*142*	*158*	*67*	*27*	*5*	*1*

Fit a Poisson distribution to the above table and calculate theoretical frequencies.

Sol. : To tit a Poisson distribution, we determine the only parameter z (mean) of the distribution from given data.

$$\begin{aligned} z = np &= \text{The mean number of accidents is A.M.} = \frac{\Sigma fx}{\Sigma f} \\ &= \frac{0 \times 142 + 1 \times 158 + 2 \times 67 + 3 \times 27 + 4 \times 5 + 5 \times 1}{142 + 158 + 67 + 27 + 5 + 1} \\ &= \frac{158 + 134 + 81 + 20 + 5}{400} = \frac{398}{400} \qquad (\because \Sigma f = 400) \\ &= 0.995 \end{aligned}$$

Thus, the Poisson distribution that fits to the given data is

$$p(r) = \frac{e^{-0.995}(0.995)^r}{r!}$$

$p(0) = e^{-0.995} = 0.3697$ $\qquad p(1) = 0.36785$ $\qquad p(2) = 0.813$

$p(3) = 0.0607$ $\qquad p(4) = 0.0151$ $\qquad p(5) = 0.003$

Theoretical (expected) frequencies are = (Total frequency) × (Probabilities)

$$\begin{aligned} 400 \times p(0) &= 400 \times 0.3697 = 148 \\ 400 \times p(1) &= 400 \times 0.36785 = 147 \\ 400 \times p(2) &= 400 \times 0.183 = 73 \\ 400 \times p(3) &= 400 \times 0.0607 = 24 \\ 400 \times p(4) &= 400 \times 0.0151 = 6 \\ 400 \times p(5) &= 400 \times 0.003 = 1 \end{aligned}$$

Theoretical frequencies are compare well with observed frequencies.

Ex. 8 : *Fit a poisson distribution to the following frequency distribution and compare the theortical frequencies with observed frequencies.*

x	0	1	2	3	4	5
f	150	154	60	35	10	1

Sol.: To fit a Poisson distribution, we determine the only parameter z (mean)

$$z = np = \text{mean} = \frac{\Sigma fx}{\Sigma f} = \frac{0\times 150 + 1\times 154 + 2\times 60 + 3\times 35 + 4\times 10 + 5\times 1}{150 + 154 + 60 + 35 + 10 + 1}$$

$$= \frac{424}{410} = 1.034 \qquad (\because \Sigma f = 410)$$

The Poisson distribution that fits to the given data is

$$p(r) = \frac{e^{-1.034}(1.034)^r}{r!}$$

$$p(0) = e^{-1.034} = 0.356, \quad p(1) = 0.368, \quad p(2) = 0.190,$$
$$p(3) = 0.0656, \quad p(4) = 0.017, \quad p(5) = 0.0035$$

Theoretical frequencies are = (Total frequency) × (Probability)

$$410 \times p(0) = 410 \times 0.356 = 145.96$$
$$410 \times p(1) = 410 \times 0.368 = 150.88$$
$$410 \times p(2) = 410 \times 0.190 = 77.9$$
$$410 \times p(3) = 410 \times 0.0656 = 26.896$$
$$410 \times p(4) = 410 \times 0.017 = 6.97$$
$$410 \times p(5) = 410 \times 0.0035 = 1.435$$

Remark : Theoretical frequencies compare well with observed frequencies.

Ex. 9 : *In a Poisson distribution if p (r = 1) = 2p (r = 2), find p (r = 3).*

Sol. :

$$p(r) = \frac{e^{-z} z^r}{r!}$$

$$p(r = 1) = \frac{e^{-z} z}{1}, \quad p(r = 2) = \frac{e^{-z} z^2}{2}$$

$$\therefore \quad ze^{-z} = 2 \times \frac{e^{-z} z^2}{2} \quad \text{which gives } z = 1$$

$$p(r = 3) = \frac{e^{-1}(1)}{3!} = e^{-1}\frac{1}{6} = \frac{1}{6e} = 0.0613$$

Ex. 10 : *Show that in a Poisson distribution with unit mean, mean deviation about mean is $\frac{2}{e}$ times the standard deviation.*

Sol. : Here z = 1, $\quad p(r) = \frac{e^{-z} z^r}{r!} = \frac{e^{-1}(1)^r}{r!} = \frac{e^{-1}}{r!}$

Mean deviation about mean 1 is

$$\sum_{r=0}^{\infty} |r - 1|\, p(r) = \sum_{r=0}^{\infty} |r - 1| \frac{e^{-1}}{r!}$$

$$= e^{-1}\left[1 + \frac{1}{2!} + \frac{2}{3!} + \frac{3}{4!} + \dots\right]$$

We have, $\quad \frac{n}{(n+1)!} = \frac{(n+1) - 1}{(n+1)!} = \frac{1}{n!} - \frac{1}{(n+1)!}$

$\therefore$ Mean deviation about mean

$$= e^{-1}\left[1 + \left(1 - \frac{1}{2!}\right) + \left(\frac{1}{2!} - \frac{1}{3!}\right) + \left(\frac{1}{3!} - \frac{1}{4!}\right) + \dots\right]$$

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

But for Poisson distribution, standard deviation $\sqrt{z} = 1$.

$\therefore$ Mean deviation about mean $= \frac{2}{e}$ (standard deviation).

8.10 NORMAL DISTRIBUTION

Normal distribution is the probability distribution of a continuous random variable X, known as normal random variable or normal variate.

Normal distribution is obtained as a limiting form of Binomial distribution when n the number of trials is very large and neither p nor q is very small. Most of the modern statistical methods have been based on this distribution.

Normal distribution of a continuous random variable X with parameters μ and σ^2 is denoted $X \to N(\mu, \sigma^2)$ and given by probability density function (p.d.f.).

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

The graph (shape) of the normal distribution curve y = f(x) is bell shaped curve with symmetry about the ordinate at $x = \mu$. The mean, median and mode coincide and therefore the normal curve is unimodal (has only one maximum point). Also, normal curve is asymptotitc to both positive x-axis and negative x-axis. (Refer Fig. 8.9).

The total area under the normal curve is unity i.e. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

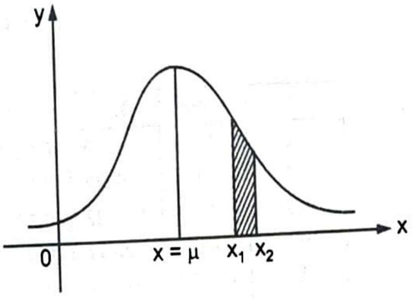

Fig. 8.9

The probability that the continuous random variable x lies between $x = x_1$ and $x = x_2$ is given by the area under the curve y = f(x) bounded by x-axis, $x = x_1$ and $x = x_2$ which is shown by shaded area in the Fig. 8.9.

$$p(x_1 \le x \le x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma^2}\right)^2} dx \qquad \text{... (2)}$$

Introducing $z = \frac{x-\mu}{\sigma}$, integral (2) becomes independent (dimensionless) of two parameters μ and σ. Here z is known as standard variable (variate). With $z = \frac{x-\mu}{\sigma}$, $dx = \sigma\, dz$ and $z_1 = \frac{x_1-\mu}{\sigma}$, $z_2 = \frac{x_2-\mu}{\sigma}$, integral (2) is

$$p(z_1 \le z \le z_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}} dz \qquad \text{... (3)}$$

If $\mu = 0$ and $\sigma = 1$, then standard variable $z = \frac{x-\mu}{\sigma}$ is called **standard normal variable** i.e. N(0, 1). The p.d.f. of standard normal random variable is given by

$$y(z) = F(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad \text{... (4)}$$

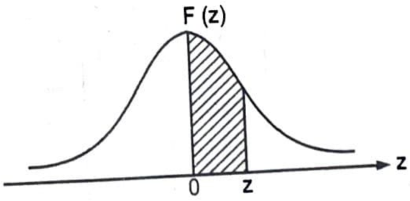

Fig. 8.10

If $z = \frac{x-\mu}{\sigma} \to N(0, 1)$ then the distribution is called **standard normal distribution** and its normal curve as standard normal cure (refer Fig. 8.10).

The probability of random variable x lying between $x = \mu$ and any value of $x = x_1$ is given by;

$$p(\mu \le x \le x_1) = \int_{\mu}^{x_1} f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma^2}\right)} dx$$

When $\frac{x-\mu}{\sigma} = z$, $dx = \sigma\, dz$. Also, $x = \mu$, $z = 0$ and $x = x_1$, $z = \frac{x_1-\mu}{\sigma} = z_1$ (say)

$$P(0 \le z \le z_1) = \frac{1}{\sqrt{2\pi}} \int_{0}^{z_1} e^{-\frac{1}{2}z^2} dz \qquad \text{... (5)}$$

If A(z) denotes the area (refer Fig. 8.10) under the normal curve $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$, from 0 to z_1, z_1 being any number then from (5), we can write

$$p(z) = A(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{1}{2}z^2} dx \qquad \text{... (6)}$$

The definite integral (6) is called normal probability integral (or error function), the value of this integral for different values of z are given in table 8.1. Thus, the entries in the normal table gives (represents) the area under the normal curve between z = 0 to z (shaded in figure 8.10). Hence the determination of normal probabilities (3) reduce to the determination of area as

$$p(x_1 \le x \le x_2) = p(z_1 \le z \le z_2) = p(z_2) - p(z_1)$$

= (Area under the normal curve from 0 to z_2) – (Area under the normal curve from 0 to z_1)

Table 8.1

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5259
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8906	0.8925	0.8943	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9494	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9871	0.9678	0.9686	0.9693	0.9699	0.9708
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9783	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9634	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.98686	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9809	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9988	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9998	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9898

In each row and each column 0.5 to be subtracted.

8.10.1 The Area under the Normal Curve

Normal distribution curve is given by

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

The area under the normal curve is

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

Put $\frac{x-\mu}{\sigma} = z \quad \therefore\ dx = \sigma\,dz$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz$$

$$= \frac{1}{\sqrt{2\pi}} \left[\int_0^{\infty} e^{-\frac{1}{2}z^2} dz + \int_0^{\infty} e^{-\frac{1}{2}(-z)^2} dz\right] \quad \text{(by definite integral theorem)}$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}z^2} dz$$

Put $\frac{z^2}{2} = t,\ z\,dz = dt$

$$= \frac{1}{\sqrt{\pi}} \int_0^{\infty} e^{-t} t^{-1/2} dt$$

$$= \frac{1}{\sqrt{\pi}} \left\lceil \frac{1}{2} = \frac{1}{\sqrt{\pi}} \sqrt{\pi} = 1\right.$$

8.10.2 Mean Deviation from the Mean

$$\text{M.D.} = \int_{-\infty}^{\infty} |x-\mu|\, p(x)\,dx$$

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

Put $\frac{x-\mu}{\sigma} = z;\ dx = \sigma\,dz$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 |z|\, e^{-z^2/2} dz = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z|\, e^{-z^2/2} dz$$

$$= \frac{\sigma}{\sqrt{2\pi}} \cdot 2 \int_0^{\infty} |z|\, e^{-z^2/2} dz \quad [f(z) = |z| = e^{-z^2/2} \text{ is even function of } z]$$

$$\text{M.D.} = \sqrt{\frac{2}{\pi}}\, \sigma \int_0^{\infty} z e^{-z^2/2} dz \quad [|z| = z \text{ for } 0 < z < \infty]$$

$$= \sqrt{\frac{2}{\pi}}\, \sigma \int_0^{\infty} \frac{1}{2} e^{-z^2/2}\, d(z^2)$$

$$= \sqrt{\frac{2}{\pi}} \cdot \frac{\sigma}{2} \left[\frac{e^{-z^2/2}}{-1.2}\right]_0^{\infty}$$

$$= \sqrt{\frac{2}{\pi}}\, \sigma\,[1] = 0.8\,\sigma = \frac{8}{10}\sigma = \frac{4}{5}\sigma \text{ (approximately)}$$

ILLUSTRATIONS

Ex. 1 : *The mean weight of 500 students is 63 kgs and the standard deviation is 8 kgs. Assuming that the weights are normally distributed, find how many students weigh 52 kgs ? The weights are recorded to the nearest kg.*

Sol. : The frequency curve for the given distribution is

$$y = 500 \frac{1}{8\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-63}{8}\right)^2} \qquad \dots (1)$$

Since the weights are recorded to the nearest kg, the students weighing 52 kgs have their actual weights between $x = 51.5$ and 52.5 kg. So the area under the curve (1) from $x = 51.5$ to $x = 52.5$ is to be obtained.

Using $z = \frac{x-\mu}{\sigma}$, we have

$$z_1 = \frac{51.5-63}{8} = -1.4375 = -1.44 \text{ (appx)}$$

$$z_2 = \frac{52.5-63}{8} = -1.3125 = -1.31 \text{ (appx)}$$

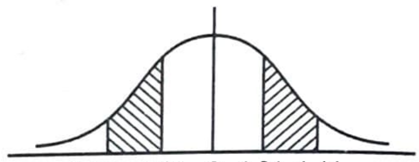

Fig. 8.11

$$P(51.5 \le x \le 52.5) = 500\, P(-1.44 \le z \le -1.31)$$

$$500 \int_{51.5}^{52.5} p(x)\, dx = \frac{500}{\sqrt{2\pi}} \int_{-1.44}^{-1.31} e^{-z^2/2}\, dz$$

The number of students weighing 52 kg

$$= 500\,(A_1 - A_2)$$

$$= 500\,(0.4251 - 0.4049) = 10 \text{ students approximately.}$$

where, $A_1 = 0.4251$ is the area for $z_1 = 1.44$,

and $A_2 = 0.4049$ is the area for $z_2 = 1.31$.

Ex. 2 : *For a normal distribution when mean $\mu = 1$, $\sigma = 3$, find the probabilities for the intervals :*

(i) $3.43 \le x \le 6.19$; (ii) $-1.43 \le x \le 6.19$

Sol. : Using $z = \frac{x-\mu}{\sigma}$, we have

(i) $z_1 = \frac{3.43-1}{3} = 0.81$, $z_2 = \frac{6.19-1}{3} = 1.73$

$$P(3.43 \le x \le 6.19) = P(0.81 \le z \le 1.73)$$

$\therefore$ Required probability $= A_2 - A_1$

$$= (0.4582 - 0.2910) = 0.1672$$

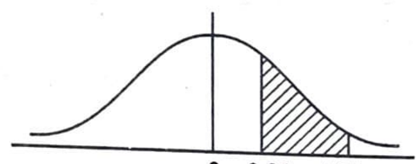

Fig. 8.12

where, $A_2 = 0.4592$ is area corresponding to $z_2 = 1.73$

$A_1 = 0.2910$ is area corresponding to $z_1 = 0.81$

(ii) $z_1 = \frac{-1.43-1}{3} = -0.81$, $z_2 = \frac{6.19-1}{3} = 1.73$

$$P(-1.43 \le x \le 6.19) = P(-0.81 \le z \le 1.73)$$

$\therefore$ Required probability $= A_1 + A_2$

$$= 0.2910 + 0.4582 = 0.7492$$

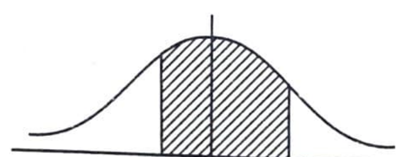

Fig. 8.13

Ex. 3 : *Assuming that the diameters of 1000 brass plugs taken consecutively from machine form a normal distribution with mean 0.7515 cm and standard deviation 0.0020 cm. How many of the plugs are likely to be approved if the acceptable diameter is 0.752 ±0.004 cm ?* **(Dec. 2008, May 2009, 2016)**

Sol. : Given,

$$\sigma = 0.0020, \quad \mu = 0.7515$$

For $x_1 = 0.752 + 0.004 = 0.756$

and $x_2 = 0.752 - 0.004 = 0.748$

Using $z = \dfrac{x-\mu}{\sigma}$, we have

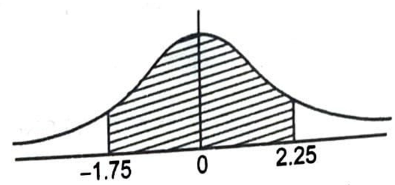

Fig. 8.14

$$z_1 = \frac{x_1-\mu}{\sigma} = \frac{0.756-0.7515}{0.0020} = 2.25$$

$$z_2 = \frac{x_2-\mu}{\sigma} = \frac{0.748-0.7515}{0.0020} = -1.75$$

(Refer table 8.1)

A_1 corresponding to $(z_1 = 2.25) = 0.4878$

A_2 corresponding to $(z_2 = 1.75) = 0.4599$

$$p(0.748 < x < 0.756) = A_1 + A_2$$
$$= 0.4878 + 0.4599 = 0.9477$$

Number of plugs likely to be approved = $1000 \times 0.9477 = 948$ approximately.

Ex. 4 : *In a certain examination test, 2000 students appeared in a subject of statistics. Average marks obtained were 50% with standard deviation 5%. How many students do you expect to obtain more than 60% of marks, supposing that marks are distributed normally ?* **(Dec. 2006, 2017)**

Sol. : Given : $\mu = 0.5$, $\sigma = 0.05$

$x_1 = 0.6$, $z_1 = \dfrac{0.6-0.5}{0.05} = 2$ $\left(\because z = \dfrac{x-\mu}{\sigma}\right)$

A corresponding to $z = 2$ is 0.4772

$\therefore$ $P(x \geq 6) = P(z \geq 2) = 0.5 - 0.4772 = 0.0228$

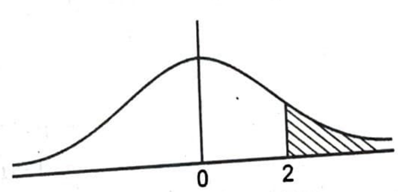

Fig. 8.15

Number of students expected to get more than 60% marks

$= 0.0228 \times 2000 = 46$ students approximately.

Ex. 5 : *In a certain city 4000 tube lights are installed. If the lamps have average life of 1500 burning hours with standard deviation 100 hours.. Assuming normal distribution*

(i) How many lamps will fail in first 1400 hours

(ii) How many lamps will last beyond 1600 hours

Sol. :

$$z = \frac{x-\mu}{\sigma}$$

(i)

$$z = \frac{1400-1500}{100} = -\frac{100}{100} = 1$$

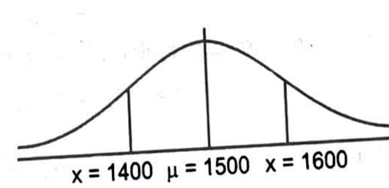

Fig. 8.16

Corresponding to $z = 1$, $A = 0.3413$

No. of tubes failing in first 1400 hours will be

$= 4000 \times 0.3413$

$= 1365.2$ or 1365

(ii)

$$z = \frac{1600-1500}{100} = 1$$

Area to the right of $x = 1600 = 0.5 - 0.3413 = 0.1587$

No. of tubes which will continue to burn after 1600 hours will be 4000×0.1587 or 635.

Ex. 6 : *In a normal distribution 10% of items are under 40 and 5% are over 80. Find the mean and standard deviation of distribution.*

Sol. : $p(x < 40) = 0.1$ and $p(x > 80) = 0.05$

$x = 40, x = 80$ are located as shown in Fig. 8.11.

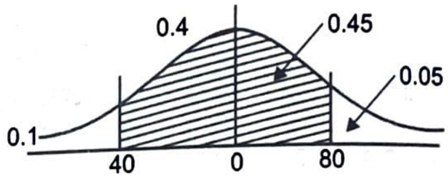

Fig. 8.17

for $x = 40$, $\quad z = \dfrac{40 - \mu}{\sigma} = -z_1$

(say – ve sign because x = 40 is to the left of $x = \mu$).

when $x = 80$, $\quad z = \dfrac{80 - \mu}{\sigma} = z_2$ (+ ve sign because x = 80 is to the right of $x = \mu$)

Corresponding to A = 0.4, z = 1.29

Corresponding to A = 0.45, z = 1.65

$$\frac{40 - \mu}{\sigma} = -1.29, \quad \frac{80 - \mu}{\sigma} = 1.65$$

$$40 - \mu = -1.29\,\sigma, \; 80 - \mu = 1.65\,\sigma$$

$$40 = (1.65 + 1.29)\sigma = 2.94\sigma$$

$$\therefore \quad \sigma = 13.6$$

$$\mu = 40 + 1.29\,\sigma = 57.54$$

Ex. 7 : *Suppose heights of students follows normal distribution with mean 190 cm and variance 80 cm^2. In a school of 1000, students how many would you expect to be above 200 cm tall.*

Sol. : Let, x = Height of students and $x \rightarrow N(190, 80)$

Proportion of students having height above 200 cm.

$$= p(x > 200) = p\left(\frac{x - \mu}{\sigma} > \frac{200 - 190}{\sqrt{80}}\right)$$

$$= p(z > 1.1180) = 0.5 - 0.3686$$

$$= 0.1314$$

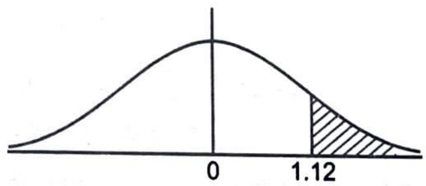

Fig. 8.18

$\therefore$ Number of students out of 1000 having height above 200 cm

= 1000 × Proportion of students having height above 200 cm

= 1000 × 0.1314 = 1.31 student

Ex. 8 : *Let $x \rightarrow N(4, 16)$. Find (i) $P(x > 5)$, (ii) $P(x < 2)$, (iii) $P(x > 0)$, (iv) $P(6 < x < 8)$, (v) $P(|x| > 6)$.*

Sol. : Let $x \rightarrow N(4, 16) = N(\mu, \sigma^2)$, hence $\mu = 4$, and $\sigma^2 = 16 \Rightarrow \sigma = 4$.

(i)
$$P(x > 5) = P\left(z = \frac{x - \mu}{\sigma} > \frac{5 - 4}{4}\right)$$

$$= P(z > 1/4)$$

$\therefore$ From normal probability integral table, we get area of shaded region as,

$$p(z > 1/4) = 0.40129$$

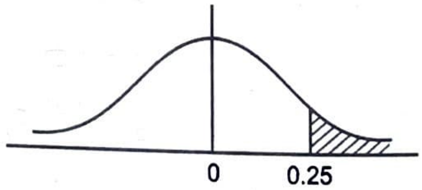

Fig. 8.19

(ii)
$$p(x < 2) = P\left(\frac{x - \mu}{\sigma} > \frac{2 - 4}{4}\right)$$

$$= P\left(z < -\frac{2}{4}\right)$$

$$= P(z < -0.5)$$

$$= P(z > 0.5) \quad \text{(Due to symmetry)}$$

$$= 0.30854 \quad \text{(From the table)}$$

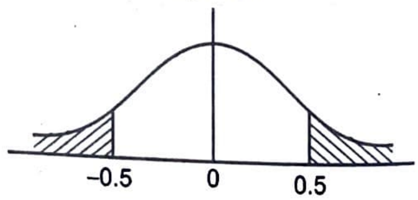

Fig. 8.20

(iii) $$p(x > 0) = P\left(\frac{x-\mu}{\sigma} > \frac{0-4}{4}\right)$$

$$= P(z > -1) = B$$

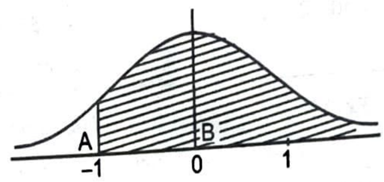

Fig. 8.21

Since, only tail area is given in the table, we use the fact that $A + B = 1$.

$\therefore$ $$p(z > -1) = 1 - A = 1 - p(z < -1)$$

$$= 1 - p(z > 1) \qquad \text{(Due to symmetry)}$$

$$= 1 - 0.15866$$

$$= 0.84134$$

(iv) $$p(6 < x < 8) = P\left(\frac{6-4}{4} < \frac{x-\mu}{\sigma} < \frac{8-4}{\sigma}\right)$$

$$= P\left(\frac{2}{4} < z < 1\right)$$

$$= P(0.5 < z < 1) = A$$

$$= (A + B) - B$$

$$= P(z > 0.5) - P(z > 1)$$

$$= 0.30854 - 0.15866$$

$$= 0.14988$$

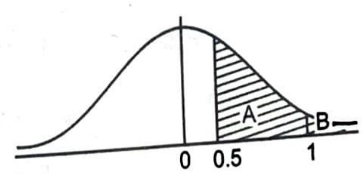

Fig. 8.22

(v) $$p(|x| > 6) = p(x > 6) + P(x < -6)$$

$$= p\left(\frac{x-\mu}{6} > \frac{6-4}{4}\right) + P\left(\frac{x-\mu}{\sigma} > \frac{-6-4}{4}\right)$$

$$= p(z > 0.5) + p(z < -2.5)$$

$$= p(z > 0.5) + p(z > 2.5)$$

$$= 0.30854 + 0.0062097)$$

$$= 0.31475$$

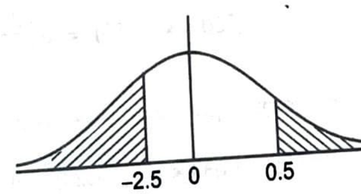

Fig. 8.23

Ex. 9 : *In a distribution, exactly normal, 7% of the items are under 35 and 89% are under 63. Find the mean and standard deviation of the distribution.* ***(May 2012)***

Sol. : From Fig. 8.24, it is clear that 7% of items are under 35 means area under 35 is 0.07. Similarly area for $x > 63$ is 0.11.

$$p(x < 35) = 0.07 \quad \text{and} \quad p(x > 63) = 0.11$$

$x = 35$, $x = 63$ are located as shown in Fig. 8.24.

When $x = 35$, $z = \dfrac{35-\mu}{\sigma} = -z_1$ (say), (– ve sign because $x = 35$ to the left of $x = \mu$)

When $x = 63$, $z = \dfrac{63-\mu}{\sigma} = z_2$ (say), (+ve sign for $x = 63$ lies to the right of $x = \mu$)

$\therefore$ From Table 8.1, we get

Area $A_1 = p(0 < z < z_1) = 0.43$ corresponds to $z_1 = 1.48$ (appx)

& Area $A_2 = p(0 < z < z_2) = 0.39$ corresponds to $z_2 = 1.23$ (appx)

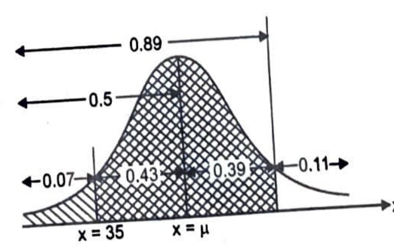

Fig. 8.24

Thus, we get two simultaneous equations

$$\frac{35-\mu}{\sigma} = -z_1 = -1.48 \qquad \text{... (1)}$$

$$\frac{63-\mu}{\sigma} = z_2 = 1.23 \qquad \text{... (2)}$$

Subtracting (1) from (2),

$$\frac{28}{\sigma} = 2.71 \quad \Rightarrow \quad \sigma = 10.33 \text{ (approximately)}$$

and (2) $\Rightarrow$

$$\mu = 63 - \sigma \times 1.23$$

$$= 63 - 10.33 \times 1.23 = 50.3 \text{ (approximately)}$$

Ex. 10 : *A fair coin is tossed 600 times. Using normal approximation find the probability of getting (i) number of heads less than 270. (ii) number of heads between 280 to 360.*

Sol. : A fair coin tossing 600 times result into head or tail each with probability p = 0.5.

Let, x = Number of heads in 600 tosses and $x \rightarrow B(600, 0.5)$

$$\mu = E(X) = np = 600 \times 0.5 = 300 \text{ and } \sigma^2 = Var(X) = npq = 600 \times 0.5 \times 0.5 = 150$$

(i) p (number of heads less than 270).

$$P(x < 270) = P\left(\frac{x-\mu}{\sigma} < \frac{270 - 30}{\sqrt{150}}\right) P\left(\frac{x-np}{\sqrt{npq}} < \frac{270-300}{\sqrt{150}}\right)$$

$$= P(z < -2.4495) \qquad \text{(Using normal approximate)}$$

$$= P(z > 2.4495) \qquad \text{(Due to symmetry)}$$

$$= 0.0071428$$

(ii) P (Number of heads are between 280 and 350)

$$P(280 < x < 350) = P\left(\frac{280-300}{\sqrt{150}} < z < \frac{350-300}{\sqrt{150}}\right)$$

Using normal approximate, we get, $z = \frac{x-\mu}{\sigma} \rightarrow N(0, 1)$

$$p \approx p(-1.633 < z < 4.0823) = B = 1 - A - C$$

$$= 1 - P(z < -1.633) - P(z > 4.0823) \qquad \text{(Due to symmetry)}$$

$$= 1 - P(z > 1.633) - P(z > 4.0823) \qquad \text{(Using standard table values)}$$

$$= 1 - 0.51551 - 0.000022518 = 0.4845$$

Ex. 11 : *If x is a random variable with p.d.f.* $f(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{18}(x-6)^2}$. *Find : (i) p (x > 5), (ii) p (2x + 3 > 10).*

Sol. : We know that $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ and given p.d.f. is $f(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{18}(x-6)^2}$

Comparing the p.d.f., we get, $\mu = 6$, $\sigma^2 = 9$ and $x \rightarrow N(6, 9)$

(i) For x = 5, $z = \frac{x-\mu}{\sigma} = \frac{5-6}{9} = -\frac{1}{9} = 0.33$

$$p(x > 5) = P(z > -0.33)$$

$$= 0.5 + p(z < 0.33) = 0.5 + 0.1293 = 0.6293$$

(ii)

$$2x + 3 \rightarrow N(\mu', \sigma^2)$$

$$\mu' = E(2x + 3) = 2\mu + 3$$

$$= 2(6) + 3 = 12 + 3 = 15$$

$$\sigma'^2 = Var(2x + 2) = 4\, Var(x) = 36$$

$$2x + 3 \rightarrow N(15, 36)$$