1.17. BINOMIAL DISTRIBUTION

Binomial distribution is a discrete probability distribution which is obtained when the probability p of the happening of an event is same in all the trials, and there are only two events in each trial. For example, the probability of getting a head, when a coin is tossed a number of times, must remain same in each toss, *i.e.*, $\frac{1}{2}$.

Let an experiment consisting of n trials be performed and let the occurrence of an event in any trial be called a success and its non-occurrence a failure. Let p be the probability of success and q be the probability of the failure in a single trial, where $q = 1\ p$, so that $p + q = 1$.

Let us assume that trials are independent and the probability of success is same in each trial. Let us claim that we have n trials, then the probability of happening of an event r times and failing $(n-r)$ times in any specified order is $p^r q^{n-r}$ (by the theorem on multiplication of probability). But the total number of ways in which the event can happen r times exactly in n trails is $C(n, r)$. These $C(n, r)$ ways are equally likely, mutually exclusive and exhaustive.

Therefore, the probability of r successes and $(n - r)$ failures in n trails in any order, whatever is, $C(n, r)\, p^r\, q^{n-r}$

It can also be expressed in the form

$$P(X = r) = P(r) = C(n, r)\, p^r\, q^{n-r};\ r = 0, 1, 2, 3, 4, \ldots..n,$$

where $P(X = r)$ or $P(r)$ is the probability distribution a random variable X of the number of succeses, Giving different values of r, *i.e.*, putting $r = 1, 2, 3, \ldots.n$, we get the corresponding probabilities ${}^nC_0\, q^n$, $C_1 q^{n-1} p$, ${}^nC_2\, q^{n-2} p^2$, ${}^nC_3\, q^{n-3} p^3$, ... p^n, which are the different terms in the Binomial expansion of $(q + p)^2$.

As a result of it, the distribution $\boldsymbol{P(r) = C(n, r)\, p^r\, q^{n-r}}$ is called Binomial probability distribution. The two independent constants. viz., n and p in the distribution are called the parameter of the distribution.

Again if the experiment (each consisting of n trials) be repeated N times, the frequency function of the Binomial distributions is given by

$$f(r) = NP(r) = NC\,(n, r)\, p^r\, q^{n-r}$$

The expected frequencies of 0, 1, 2, 3 ... n successes in the above set of experiment are the successive terms in the Binomial expansion of $N(q + p)^2$; where $p + q = 1$ which is also called the Binomial frequency distribution.

1.18. PROPERTIES OF A BINOMIAL DISTRIBUTION

1. It is a discrete distribution which gives the theoretical probabilities.
2. It depends on the parameters p or q, the probability of success or failure and n (the number of trails). The parameter n is always a positive integer.
3. The distrbution will be symmetrical if $p = q$.
4. The statistics of the Binomial distribution one mean $= np$, variance $= npq$; and standard deviation $= \sqrt{npq}$.

5. The mode of the binomial distribution is equal to that value X which has the largest frequency.
6. The shape and location of a binomial distribution changes as p changes for a given n or n changes for a given p.

1.19. MEAN OF BINOMIAL DISTRIBUTION

For a binomial distribution the probability function is

$$P(X = r) = {}^nC_r p^r q^{n-r}$$

The discrete probability distribtion for the binomial distribution can be displayed as follows:

X	0	1	2	...	r	...	n
$P(X)$	${}^nC_0q^n$	${}^nC_1pq^{n-1}$	${}^nC_2p^2q^{n-2}$	...	${}^nC_rp^rq^{n-r}$	...	${}^nCp^n$

$$\therefore \quad \text{Mean } (\mu) = E(X) = \sum_{r-0}^{n} rP(X = r)$$

$$= {}^nC_0\, q^n \times 0 + {}^nC_1\, pq^{n-1} \times 1 + {}^nC_2\, p^2\, q^{n-2} \times 2 + \ldots + {}^nC_r\, pr\, q^{n-r} \times + \ldots + {}^nC_n\, p^n \times n$$

$$= 0 + npq^{n-1} + \frac{n(n-1)}{2!} p^2 q^{n-2} \times 2 + \ldots + np^n$$

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

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

$\therefore$ **Mean = *np***

20. VARIANCE OF BINOMIAL DISTRIBUTION

Since Variance $= \Sigma px^2 - \mu^2$

Now $\Sigma px^2 = {}^nC_0\, q^n \times (0)^2 + {}^nC_1\, pq^{\,n-1} \times (1)^2 + {}^nC_2\, p^2\, q^{n-2} \times (2)^2 + {}^nC_3\, p^3\, q^{n-3} \times (3)^2 + \ldots + {}^nC_n\, p^n \times n^2$

$$= 0 + n.pq^{n-1} + \frac{n(n-1)}{2!} p^2 q^{n-2} \times 4 + \frac{n(n-1)(n-2)}{3!} p^3 q^{n-3} \times 9 + \ldots + p^n \times n^2$$

Breaking second, third and following terms into parts, we get

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

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

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

$$= np + n\,(n-1)\,p^2 = np\,[1 + (n-1)\,p] = np\,[q + np]\; npq + n^2 p^2$$

$$\therefore \quad \text{Variance} = npq + n^2p^2 - (np)^2 = npq.$$

1.22. CONDITIONS FOR APPLICATION OF BINOMIAL DISTRIBUTION

1. The variable should be discrete *i.e.*, defectives should could be 1, 2, 3, 4 or 5 etc., and never 1.5, 2.1 or 3.41 etc.
2. A dichotomy exists. In other words, the happening of events must be of two alternative. It must be either a success or failure.
3. The number of trials n should be finite and small.
4. The trials or events must be independent. The happening of one event must not affective happening of other events. In other words, statistical independence must exist.
5. The trial or events must be repeated under identical conditions.

1.23. RECURSION FORMULA OR RECURRENCE RELATION FOR BINOMIAL DISTRIBUTION

We known that for the Binomial distribution

$$P(X=r) = {}^nC_r p^r q^{n-r}$$

and

$$P(X=r+1) = {}^nC_{r+1} p^{r+1} q^{n-r-1}$$

$$\Rightarrow \quad \frac{P(X=r+1)}{P(X=r)} = \frac{{}^nC_{r+1} p^{r+1} q^{n-r-1}}{{}^nC_r p^r q^{n-1}}$$

$$= \frac{n!}{(r+1)!(n-r-1)!} \times \frac{r!(n-r)!}{n!} \times \frac{p^{r+1} q^{n-r-1}}{p^r q^{n-r}} = \frac{n-r}{r+1} \cdot \frac{p}{q}$$

$$\Rightarrow \quad P(X=r+1) = \frac{n-r}{r+1} \cdot \frac{p}{q} P(X=r); \quad r = 1, 2, 3 \ldots..$$

which is the required recurrence formula. Applying this formula successively, we can find $P(X=1)$, $P(X=2)$, $P(X=3)$ .... if $P(X=0)$ is known.

SOLVED EXAMPLES

Example 1.54. Ten coins are thrown simultaneously. Find the probability of getting at least seven heads.

Solution. When one coin is thrown,

$$\text{The probability of getting a head} = \frac{1}{2}$$

$$\therefore \quad p = \frac{1}{2}$$

$$\text{The probability of not getting a head} = 1 - \frac{1}{2} = \frac{1}{2}$$

$$\therefore \quad q = \frac{1}{2}$$

Then P (at least 7 heads) $= P$ (7 heads) $+ P$(8 heads) $+ P$(9 heads) $+ P$(10 heads)

$$= {}^{10}C_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 + {}^{10}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^2 + {}^{10}C_9 \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right) + {}^{10}C_{10} \left(\frac{1}{2}\right)^{10}$$

$$= \frac{1}{2^{10}}\left[{}^{10}C_7 + {}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10}\right]$$

$$= \frac{120+45+10+1}{1024} = \frac{176}{1024} = \frac{11}{64}$$

Example 1.55. In a lot of 200 ariticles 10 are defective, find the probability of : (*i*) no defective article, (*ii*) one defective article, (*iii*) at least one defective article, in a random sample of 20 articles.

Solution. The probability of defective article is $\frac{10}{200} = \frac{1}{20}$

$$\therefore \qquad p = \frac{1}{20}$$

The probability of non-defective article $= 1 - \frac{1}{20} = \frac{19}{20} \Rightarrow q = \frac{19}{20}$

(*i*) The probability of no defective article out of 20

$$= {}^{20}C_0 (p)^0 q^{20} = \left(\frac{19}{20}\right)^{20} \qquad \left[\because {}^{20}C_0 = 1\right]$$

(*ii*) The probability of exactly one defective article

$$= {}^{20}C_1 (p)^1 (q)^{19} = 20 \times \frac{1}{20} \times \left(\frac{19}{20}\right)^{19} = \left(\frac{19}{20}\right)^{19}$$

(*iii*) The probability of at least one will be defective

$$= 1 - [\text{probability that none will be defective}]$$

$$= 1 - {}^{20}C_{20}\left(\frac{19}{20}\right)^{20} = 1 - \left(\frac{19}{20}\right)^{20}$$

Example 1.56. If on an average, one ship out of 10 is wrecked, find the probability that out of 5 ships expected to arrive the port, at least four will arrive safely.

Solution. p be the probability of a ship arriving safely $= 1 - \frac{1}{10} = \frac{9}{10}$

$$q = 1 - \frac{9}{10} = \frac{1}{10}$$

Binomial distribution is $\left(\frac{1}{10} + \frac{9}{10}\right)^5$

Probability that at least four ships out of five arrive safely

$$= P(4) + P(5) = {}^5C_4\left(\frac{1}{10}\right)^1\left(\frac{9}{10}\right)^4 + {}^5C_5\left(\frac{1}{10}\right)^0\left(\frac{9}{10}\right)^5$$

$$= \left(\frac{9}{10}\right)^4 \frac{14}{10} = \left(\frac{9}{10}\right)^4 \frac{7}{5} = 0.91854$$

Example 1.57. The probability that a man aged 60 will live to be 70 is 0.65. What is the ·obability that out of 10 men aged 60 now, at least 7 would live to be 70?

Solution. Probability of survival upto the age of 70

$$= p = 0.65$$

Probability of non-survival upto the age of 70

$$= q = 1 - p = 1 - 0.65 = 0.35$$

Probability that out of 10 such men at least 7 would survive as desired

= Probability that exactly 7 would survive +
Probability that exactly 8 would survive +
Probability that exactly 9 would survive +
Probability that exactly 10 would survive

$$= P(7) + P(8) + P(9) + P(10)$$
$$= {}^{10}C_7 p^7 q^3 + {}^{10}C_8 p^8 q^2 + {}^{10}C_9 p^9 q + {}^{10}C_{10} p^{10}$$
$$= 120\, p^7 q^3 + 45\, p^8 q^2 + 10\, p^9 q + p^{10}$$
$$= p^7 (120\, q^3 + 45\, pq^2 + 10\, p^2 q + p^3)$$
$$= (0.65)^7 [120 \times (0.35)^3 + 45\, (0.65)\, (0.35)^2 + 10\, (0.65)^2\, (0.35) + (0.65)^3]$$

= 0.514, the required result.

Example 1.58. Six dice are thrown together at a time, the process is repeated 729 times. How many times do you expect at least three dice to have 4 to 6?

Solution. The chance of getting 4 or 6 with one dice is

$$\frac{2}{6} \text{ i.e., } p = \frac{1}{3} \quad \text{and} \quad q = 1 - \frac{1}{3} = \frac{2}{3}$$

In one throw of six dice together, we have probability of getting at least 3 dice to have 4 or 6.

$$= P(3) + P(4) + P(5) + P(6)$$
$$= {}^6C_3 p^3 q^3 + {}^6C_4 p^4 q^2 + {}^6C_5 p^5 q + {}^6C_6 p^6$$
$$= 20\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^3 + 15\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2 + 6\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right) + \left(\frac{1}{6}\right)^6$$
$$= \frac{1}{(3)^6}[160 + 60 + 12 + 1] = \frac{233}{(3)^6}$$

Now the process is repeated 729 times

∴ Required number of times at least 3 dice have 4 or 6

$$= 729 \times \frac{233}{(3)^6} = 233, \text{ the required result.}$$

Note. In the above case the binomial distribution is $N(q + p)^n$ where $N = 729$, $n = 6$

Example 1.59. If the sum of the mean and the variance of binomial distribution of 5 trials is 4.8, find the distribution.

Solution. Let the required binomial distribution be $^nC_r p^r q^{n-r}$ where n = number of trials = 5

Mean of the distribution = np

and the variance of the distribution = npq

By the given condition

$$np + npq = 4.8$$

$$\Rightarrow \quad 5p + 5pq = 4.8 \qquad [\because p = 1-q]$$

$$5p(1+q) = 4.8$$

$$\Rightarrow \quad 50(1-q^2) = 48 \quad \Rightarrow \quad 50 - 50q^2 = 48$$

$$\Rightarrow \quad 50q^2 = 2 \Rightarrow q = \frac{1}{5}$$

$$\therefore \quad p = 1 - q = 1 - \frac{1}{5} = \frac{4}{5}$$

Hence, the required binomial distribution is $^5C_r \left(\frac{4}{5}\right)^r \left(\frac{1}{5}\right)^{5-r}$

Example 1.60. The probability that a bomb dropped from a place will strike the target is $\frac{1}{5}$. If six boms are dropped, find the probability that : (*i*) exactly two will strike the target, (*ii*) at least two will strike the target.

Solution. The probabilities of 0,1, 2 ..., successes are given by the respective terms in the expansion of

$$(q+p)^n = \left(\frac{4}{5}+\frac{1}{5}\right)^6, \text{ since } p = \frac{1}{5}, q = \frac{4}{5} \text{ and } n = 6.$$

$\therefore P(2)$ = The probability that exactly two bombs will strike the target

$$= {^6C_2} p^2 q^4 = \frac{6.5}{1.2} \cdot \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^4 = 0.246$$

The probability that at least 2 bombs will strike the target

$$= 1 - [P(0) + P(1)]$$

$$= 1 - q^6 - {^6C_1} q^5 p = 1 - (0.8)^6 - 6(0.2)(0.8)^5$$

$$= 1 - 0.2621 - 0.3932 = 0.345$$

Example 1.61. Assuming that half the population are consumers of rice so that the chance of an individual being a rice consumer is $\frac{1}{2}$ and assuming that 100 investigations each take 10 individuals to see whether they are rice consumers. How many investigations would you expect to report that three people or less consumers?

Solution. Here $p = \frac{1}{2}, q = \frac{1}{2}, n = 10, N = 100$

$\therefore$ The probability that r persons out of 10 persons are consumers of rice is given by

$$P(r) = {}^{10}C_r \left(\frac{1}{2}\right)^r \left(\frac{1}{2}\right)^{10-r}$$

$\therefore$ The expected number of investigators (*i.e.*, expected frequencies) who would report that three or less people were consumers of rice.

$$= 100\,[P(0) + P(1) + P(2) + P(3)]$$

$$= 100\left[{}^{10}C_0\left(\frac{1}{2}\right)^{10} + {}^{10}C_1\left(\frac{1}{2}\right)^{10} + {}^{10}C_2\left(\frac{1}{2}\right)^{10} + {}^{10}C_3\left(\frac{1}{2}\right)^{10}\right]$$

$$= \frac{100}{2^{10}}[1 + 10 + 45 + 120] = \frac{17600}{1024} = 17 \text{ approx.}$$

Example 1.62. A die is thrown 5 times. Getting an even number greater than 2 is considered a success. Calculate $P(X = r)$ for $r = 1, 2, 3, 4, 5$ from recurrence formula.

Solution. Let p be the probability of getting an even number greater than 2 on a die.

$$\Rightarrow \qquad p = \frac{2}{6} = \frac{1}{3}$$

$$\therefore \qquad q = 1 - p \;\Rightarrow\; q = 1 - \frac{1}{3} = \frac{2}{3}$$

$$\therefore \qquad \frac{p}{q} = \frac{1}{2}. \quad \text{Also } n = 5$$

$$P(X) = 0 = \text{Probability of no success in 5 trials}$$

$$= {}^5C_0(q)^5 = \left(\frac{2}{3}\right)^5 = 0.1317$$

Recurrence formula for binomial distribution is

$$P(X = r + 1) = \frac{n - r}{r + 1} \cdot \frac{p}{q} P(X = r)$$

$$= \frac{5 - r}{r + 1} \cdot \left(\frac{1}{2}\right) P(X = r) \qquad \ldots(i)$$

Putting $r = 0$, in (*i*), $P(X = 1) = 5\left(\frac{1}{2}\right) P(X = 0) = 5\left(\frac{1}{2}\right)(0.1317) = 0.3292$

Putting $r = 1$, in (*i*), $P(X = 2) = 2\left(\frac{1}{2}\right) P(X = 1) = P(X = 1) = 0.3292$

Putting $r = 2$, in (*i*), $P(X = 3) = (1)\left(\frac{1}{2}\right) P(X = 2) = \frac{1}{2} \cdot (0.3292) = 0.1646$

Putting $r = 3$, in (*i*), $P(X = 4) = \frac{2}{4} \cdot \frac{1}{2} P(X = 3) = \frac{1}{4}(0.1646) = 0.0412$

Putting $r = 4$, in (*i*), $P(X = 5) = \frac{1}{5} \cdot \frac{1}{2} P(X = 4) = \frac{1}{10}(0.0412) = 0.0041.$

Example 1.63. Out of 800 families with 4 children each, how many families would be expected to have: (*i*) 2 boys and 2 girls, (*ii*) at least one boy, (*iii*) no girl, (*iv*) at most two girls? Assume equal probabilities for boys and girls.

Solution. Since probability for boys and girls are equal

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

$$q = \text{Probability of having a girl} = \frac{1}{2}$$

$$n = 4, \ N = 800$$

The binomial distribution is $800\left(\frac{1}{2}+\frac{1}{2}\right)^4$.

(*i*) The expected number of families having 2 boys and 2 girls

$$= 800 \,.\, {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = 800 \times 6 \times \frac{1}{16} = 300$$

(*ii*) The expected number of families having at least one boy

$$= 800\left[{}^4C_1\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) + {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + {}^4C_3\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 + {}^4C_4\left(\frac{1}{2}\right)^4\right]$$

$$= 800 \times \frac{1}{16}[4+6+4+1] = 750$$

(*iii*) The expected number of families having no girl having 4 boys

$$= 800 \times {}^4C_4\left(\frac{1}{2}\right)^4 = 50$$

(*iv*) The expected number of families having at most two i.e., having at least 2 boys

$$= 800\left[{}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + {}^4C_3\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 + {}^4C_4\left(\frac{1}{2}\right)^4\right]$$

$$= 800 \times \frac{1}{16}[6+4+1] = 550.$$

Example 1.64. A student obtained the following answer to a certain problem given to him. Mean = 2.4; variance = 3.2 for a binomial distribution. Comment on the result.

Solution. The mean of binomial distribution is np and variance npq. We are given mean $= np = 2.4$

$$\text{Variance} = npq$$

$$2.4q = 3.2$$

$$q = \frac{3.2}{2.4} = 1.333$$

Since the value of q is greater than 1, the given results are inconsistent.

Example 1.65. Ten coins are tossed 1024 times and the following frequencies are observed. Compare these frequencies with the expected frequencies :

Number of heads	0	1	2	3	4	5	6	7	8	9	10
Frequencies	2	10	38	106	188	257	226	128	59	7	3

Solution. Here $n = 10, N = 1024$

$$p = \text{The change of getting a head in one toss} = \frac{1}{2}$$

$$\therefore \quad q = 1 - p = \frac{1}{2}$$

The expected frequencies are the respective terms of the binomial $1024\left(\frac{1}{2}+\frac{1}{2}\right)^{10}$

The frequency of r heads $(0 \geq r \geq 10)$ is

$$= 1024.{}^{10}C_r\left(\frac{1}{2}\right)^{10-r} \cdot \left(\frac{1}{10}\right) = 1024 \times {}^{10}C_r\left(\frac{1}{2}\right)^{10} = {}^{10}C_r$$

Hence, we have the following comparison.

Number of heads	0	1	2	3	4	5	6	7	8	9	10
Observed frequency	2	10	38	106	188	257	226	128	59	7	3
Expected frequency	1	10	45	120	210	252	210	120	45	10	1

$({}^{10}C_0, {}^{10}C_1, {}^{10}C_2$ and so on.)

Example 1.66. Probability of man hitting a target is $\frac{1}{3}$.

(*a*) If the fires of 6 times, what is the probability of hitting: (*i*) at most 5 times, (*ii*) at least 5 times, (*iii*) exactly once

(*b*) If he fires so that the probability of his hitting target atleast once is greater than $\frac{3}{4}$, find n.

Solution. (*a*) Given $p = 1/3, q = 1 - 1/3 = 2/3$ $n = 6$

(*i*) The probability of hitting the target almost 5 times.

$$P(X \leq 5) = 1 - P(X > 5) = 1 - P(X = 6)$$

$$= 1 - \left[{}^6C_6\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^0\right] = 1 - \frac{1}{729} = \frac{728}{729}$$

(*ii*) The probability of hitting the target atlesat 5 times

$$P(X \geq 5) = P(X = 5) + P(X = 6)$$

$$= {}^6C_5\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)^1 + {}^6C_6\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^0 = \frac{13}{729}$$

(*iii*) The probability of hitting the target exactly once

$$P(X=1)\ {}^6C_1\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^5 = \frac{192}{729}$$

(*b*) If he fires so that the probability of his hitting the target atleast once is greater than 3/4 then

$$\Rightarrow \quad P(X \geq 1) = 3/4$$

$$\Rightarrow \quad 1 - P(X<1) > 3/4$$

$$\Rightarrow \quad 1 - P(X=0) > 3/4$$

$$\Rightarrow \quad 1 - (2/3)^n > 3/4$$

$$\Rightarrow \quad 1/4 > (2/3)^n$$

$$\Rightarrow \quad \frac{1}{2^2} > \left(\frac{2}{3}\right)^n$$

$$\Rightarrow \quad 3n > 2^{n+2}$$

This inequality is satisfied for $n = 4$

$\therefore$ He must fire 4 times to that probability of hitting the target atleast once is greater than 3/4.

Example 1.67. A student takes a true-false examination consisting of 8 questions. He guesses a answer. The guesses are made at random. Find the smallest value of *n* so that the probability guessing atleast *n* correct answers is less than 1/2.

Solution. For given data, we have to find

$$P(X \geq n) < 1/2$$

Now probability of guessing a correct answer is 1/2 and guessing a wrong answer is 1/2

$$\therefore \quad p = 1/2, q = 1/2, n = 8$$

Using Binomial distribution.

$$P(X \geq n) < 1/2$$

$$\Rightarrow \quad 1 - P(X<n) < 1/2$$

$$\Rightarrow \quad 1 - [P(X=0) + P(X=1) \ldots P(X=n-1)] < 1/2$$

$$\Rightarrow \quad P(X=0) + P(X=1) \ldots P(X=n-1) > 1/2$$

$$\Rightarrow \quad {}^8C_0\left(\frac{1}{2}\right)^8 + {}^8C_1\left(\frac{1}{2}\right)^8 + \ldots {}^8C_{n-1}\left(\frac{1}{2}\right)^8 > \frac{1}{2}$$

$$\Rightarrow \quad \left(\frac{1}{2}\right)^8\left[{}^8C_0 + {}^8C_1 + \ldots + {}^8C_{n-1}\right] > \frac{1}{2}$$

This inequality is satisfied it $n - 1 = 4$

$$\Rightarrow \quad n = 5$$

Example 1.68. Fit a binomial distribution to following data, when tossing 5 coins.

x	0	1	2	3	4	5
f	2	14	20	34	22	8

Solution.

$$\Sigma f = 100 = N$$
$$n = 5$$
$$\bar{x} = \text{mean}$$
$$= \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{0\times2+1\times14...+5\times8}{100}$$
$$= 2.84$$
$$\text{Mean} = \bar{x} = n.p = 2.84$$

$\Rightarrow$ $p = 0.57$ and $q = 0.43$ $[\because n = 5]$

Using binomial distribution

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

and expected frequency obtained from

$$f(x) = N.\,P(x)$$

x	f	$p(x) = {}^nC_x p^x q^{n-x}$	$f(x) = N.P(x)$
0	2	${}^5C_0\,(0.57)^0\,(0.43)^5 = (0.43)^5$	$100 \times (0.43)^5 \approx 1$
1	14	${}^5C_1\,(0.57)^1\,(0.43)^4 = 0.098$	$100 \times (0.098) \approx 10$
2	20	${}^5C_2\,(0.57)^2\,(0.43)^3 = 0.260$	$100 \times (0.260) \approx 26$
3	34	${}^5C_3\,(0.57)^3\,(0.43)^2 = 0.342$	$100 \times (0.342) \approx 34$
4	22	${}^5C_4\,(0.57)^4\,(0.43) = 0.224$	$100 \times (0.224) \approx 22$
5	8	$C_5\,(0.57)^5\,(0.43)^0 = 0.059$	$100 \times (0.059) \approx 6$

Example 1.69. In a Binomial distribution consisting of 5 independent trials, probability of 1 and 2 success are 0.4096 and 0.2048 respectively. Find the parameter 'p' of distribution.

Solution. Given $n = 5$

For Binomial distribution

$$P(X=x) = {}^nC_x p^x q^{n-x}; \ x = 0, 1, 2, 3 \ldots n$$
$$P(X=1) = {}^nC_1 p^1 q^{n-1} = 0.4096 \qquad \ldots(i)$$
$$P(X-2) = {}^nC_2 p^2 q^{n-2} = 0.2048 \qquad \ldots(ii)$$

Dividing eqn. (ii) with eqn. (i)

$$\frac{{}^5C_2 p^2 q^3}{{}^5C_1 p^1 q^4} = \frac{10p}{5q} = \frac{0.4096}{0.2048}$$

$\Rightarrow$
$$\frac{2p}{1-p} = 1/2$$

$\Rightarrow$
$$4p = 1 - \text{p}$$

$\Rightarrow$
$$p = 1/5$$

1.24. POISSON DISTRIBUTION

It is due to the French Mathematician S.D. Poisson (1781–1840) who published its derivation in 1837. It is a discrete probability distribution which has the following characteristics:

(*i*) It is the limiting form of the Binomial distribution as n becomes infinitely large *i.e.*, $n \to \infty$ and p, the constant probability of success for each trial becomes indefinitely small *i.e.*, $p \to 0$ in such a manner that $np = m$ remains a finite number.

(*ii*) It consists of a single parameter m only. The entire distribution can be obtained once m is known.

It has wide applications in physical, engineering and management sciences as well as economics, operations research and reliability technology.

Poisson distribution occurs when there are events which do no occur as outcomes of definite number of trials of an experiment but which occurs at random point of time and space wherein our interest lies only in the number of occurrences of the events. Here we are concerned with processes taking place over continuous intervals of time such as arrival of a telephone calls at a switch board, or the passing by cars of an electric checking device. That is why is has important applications is queuing theory where we are interested, for example, in the number of aircraft arriving at an air field, the number of ships and trucks arriving to be unloaded at a receiving dock etc.

Some of the situations where the probability of the event remains very small and the Poisson distribution is applicable, are: (*i*) the occurrence of accidents in a factory in a given period, (*ii*) the number of defective articles produced by some factory in a fixed time period, (*iii*) the number of twins births per year in some hospital, (*iv*) and the number of deaths due to snake bite in a city per year, (*v*) the number of printing mistakes whch page of the book, (*vi*) the number of persons blind in a large city every year etc.

1.25. BINOMIAL APPROXIMATION TO POISSON DISTRIBUTION

We will derive Poisson distribution as a limiting case of Binomial distribution when $p \to 0, n \to \infty$ such that $np = m$ (a finite quantity). We know that in a Binomial distribution, the probability of r successes is given by

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

on expanding ${}^nC_r$ and replacing q by $(1-p)$.

$$P(r) = \frac{n(n-1)(n-2)\ldots(n-r+1)}{r!}\left(1-\frac{m}{n}\right)^{n-r}\left(\frac{m}{n}\right)^r$$

since $$np = m \text{ or } p = \frac{m}{n}$$

or
$$P(r) = \frac{n(n-1)(n-2)\ldots(n-r+1)}{r!n^r}\left(1-\frac{m}{n}\right)^n\left(1-\frac{m}{n}\right)^{-r}\cdot m^r$$
$$= \frac{n}{n}\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)\ldots\left(\frac{n-r+1}{n}\right)\left(1-\frac{m}{n}\right)^n\times\left(1-\frac{m}{n}\right)^{-r}\cdot\frac{m^r}{r!}$$
$$= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{r-1}{n}\right)\left(1-\frac{m}{n}\right)^n\times\left(1-\frac{m}{n}\right)^{-r}\frac{m^r}{r!}$$

Now for a given value of r, when $n \to \infty$, we have
$$\lim_{n\to\infty}\left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{r-1}{n}\right)\right] = 1,$$
$$\lim_{n\to\infty}\left(1-\frac{m}{n}\right)^{-r} = 1 \text{ and } \lim_{n\to\infty}\left(1-\frac{m}{n}\right)^n = e^{-m}$$

As such in the limiting form
$$P(r) = \frac{e^{-m}\cdot m^r}{r!}$$

This is the probability of r successes for Poisson distribution. For $r = 0, 1, 2, 3, \ldots$, we get the probabilities of 0, 1, 2, 3, ... successes as
$$P(0) = e^{-m}, P(1) = me^{-m}, P(2) = \frac{m^2}{n!}e^{-m}$$
$$P(3) = \frac{m^3}{3!}e^{-m} \text{ ... and so on}$$

Note : 1. The sum of the probabilities $P(r)$ for $r = 0, 1, 2, 3, \ldots$ is 1.

i.e.,
$$\Sigma P(r) = P(0) + P(1) + P(2) + P(3) + \ldots$$
$$= e^{-m}\left(1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\ldots\right) = e^{-m}\times e^m = 1.$$

2. Poisson distribution possesses only one parameter 'm'. If we consider the length of interval as 'd' instead of unit length, the average number of occurrences in 'd' length of interval is 'md' Thus, the probability function in this situation is,
$$P(r) = \frac{e^{-md}(md)^r}{r!}$$

1.26. CONDITIONS UNDER WHICH POISSON DISTRIBUTION IS USED

1. The random variable x should be discrete.
2. A dichotomy exists, *i.e.*, the happening of the event must be of two alternatives such as success and failure, ocurrences and non-occuraence etc.
3. It is applicable in those case where the number of trials n is very large and the probabilities of success p is very small but the mean $np = m$ is finite.

4. p should be very small (close of zero). If $p \to 0$, then the distribution is J-shaped and unimodal.
5. Statistical independence is assumed. In other words, it is applicable to those cases when the happening of an even does not affect the happening of the other events.

1.27. MEAN OF POISSON DISTRIBUTION

Poisson distribution is $P(r) = \dfrac{e^{-m}.m^r}{r!}, r = 0, 1, 2, ...$

$$\text{Mean,} \quad (\mu) = \sum_{i=0}^{i=n} x_i p_i = p_1 x_1 + p_2 x_2 + p_3 x_3 \cdots p_n x_n$$

$$= 0e^{-m} + 1e^{-m} + 2\frac{m^2 e^{-m}}{2!} + 3\frac{m^3 e^{-m}}{3!} + 4\frac{m^4 e^{-m}}{4!} + \ldots$$

$$= me^{-m} + m^2 e^{-m} + \frac{m^3 e^{-m}}{2\times 1} + \frac{m^4 e^{-m}}{3.2.1} + \ldots$$

$$= me^{-m}\left(1 + m + \frac{m^2}{2.1} + \ldots\right)$$

$$= me^{-m} e^m \qquad \left(\because e^m = 1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \ldots +\right)$$

$$= me^{-m+n} = me^0 = m$$

$\therefore$ The mean of the Poisson distribution is m.

1.28. VARIANCE OF POISSON DISTRIBUTION

$$\text{Variance } (\sigma^2) = (\Sigma x_i^2 p_1) - \mu^2$$

Now

$$\sum_{i=1}^{i=n} p_i x_i^2 = 0.e^{-m} + 1\cdot me^{-m} + 2^2.\frac{m^2 e^{-m}}{2!} + 3^2.\frac{m^3 e^{-m}}{3!} + 4^2.\frac{m^4 e^{-m}}{4!} + .$$

$$= me^{-m} + 2m^2 e^{-m} + \frac{3m^3 e^{-m}}{2\times 1} + \frac{4m^4 e^{-m}}{3\times 2\times 1} + \ldots$$

$$= me^{-m}\left[1 + 2m + \frac{3m^2}{2\times 1} + 4\frac{m^3}{3\times 2\times 1} + \ldots\right]$$

$$= me^{-m}\left[\left(1 + m + \frac{m^2}{2\times 1} + \frac{m^3}{3\times 2\times 1} + \ldots\right) + \left(m + \frac{2m^2}{2} + \frac{3m^3}{3\times 2\times 1} + \ldots\right)\right]$$

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

$$= me^{-m}(e^m + me^m) = m(m+1)e^{-m}e^m = m(m+1)$$

$\therefore$ $$\text{Variance} = \Sigma x_i^2 p_i = \mu^2 m(m+1) - m^2 = m^2 + m - m^2 = m$$

$$\text{Standard deviation} = \sqrt{\text{Variance}} = \sqrt{m}$$

1.29. MODE OF POISSON DISTRIBUTION

Mode is a value of at which $p(r)$ has maximum value, then

$$p(r) = p\,(r + 1) \quad \text{and} \qquad p(r) = p\,(r - 1)$$

Consider

$$p(r) \geq \quad p\,(r + 1)$$

i.e.,

$$\frac{e^{-m}}{r!} \geq \frac{e^{-m} m^{r+1}}{(r+1)!} \Rightarrow \frac{r+1}{m} \geq 1$$

$\Rightarrow$

$$r + 1 \geq m \Rightarrow r \geq m - 1$$

i.e.,

$$m - 1 \leq r \qquad \text{...}(i)$$

Similarly consider

$$p(r) \geq p\,(r - 1) \text{ then we get } r \geq m \qquad \text{...}(ii)$$

From eqns. (*i*) and (*ii*)

$$m - 1 \leq r \leq m$$

Therefore the mode of the Poisson distribution lies between $(m - 1)$ and m.

Constants of Poisson Distribution

$$\mu_1 = 0,\ \mu_2 = m,\ \mu_3 = m,\ \mu_4 = m + 3m^2$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{m^2}{m^3} = \frac{1}{m}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{m + 3m^2}{m^2} = 3 + \frac{1}{m}$$

1.30. RECURRENCE FORMULA FOR THE POISSON DISTRIBUTION

We have

$$P(r) = \frac{e^{-m} \cdot m^r}{r!}$$

and

$$P(r + 1) = \frac{e^{-m} \cdot m^{r+1}}{(r+1)!}$$

Then

$$\frac{P(r+1)}{P(r)} = \frac{\dfrac{e^{-m} \cdot m^{r+1}}{(r+1)!}}{\dfrac{e^{-m} \cdot m^r}{r!}} = \frac{e^{-m} \cdot m^{r+1}}{e^{-m} \cdot m^r} \times \frac{r!}{(r+1)!} + \frac{m}{r+1}$$

$\Rightarrow$

$$P(r + 1) = \frac{m}{r+1} P(r); \ r = 0, 1, 2, \ldots$$

which is the required recurrence formula for Poisson distribution. With this formula, we can find $P(1), P(2), P(3), P(4), \ldots$ if $P(0)$ is given.

SOLVED EXAMPLES

Example 1.70. The number of telephone calls arriving on an internal switch board of an office is 90 per hour. Find the probability that at the most 1 to 3 calls in a minute on the board arrive. (Use $e^{-1.5} = 0.223$)

Solution. $m = \text{mean} = \dfrac{90}{60} = 1.5$. Obviously, X will follow Poisson distribution.

Now the probability that at the most 1 to 3 calls in one minute

$$= P(1) + P(2) + P(3)$$

where

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

$$P(1) = \frac{e^{-1.5}(1.5)^1}{1!} = 1.5 \times e^{-1.5} = 1.5 \times 0.223$$

$$P(2) = \frac{e^{-1.5}(1.5)^2}{2!} = 1.125 \times e^{-1.5} = 1.125 \times 0.223$$

$$P(3) = \frac{e^{-1.5}(1.5)^3}{3!} = 0.562 \times e^{-1.5} = 0.562 \times 0.223$$

$\therefore$ Required probability $= 0.233\ (1.5 + 1.125 + 0.562) = 0.711$

Example 1.71. Suppose a book of 585 pages contains 43 typographical errors. If these errors are randomly distributed throughout the book, what is the probability that 10 pages, selected at random, will be free from errors? (Use $e^{-0.735} = 0.4795$).

Solution. Here $p = \dfrac{43}{585} = 0.0735$ and $n = 10$.

$\therefore$

$$m = np = 10 \times 0.0735 = 0.735$$

Clearly, p is very small and n is large.

So, it is a case of Poisson distribution.

Let X denote the number of errors in 10 pages.

Then,

$$P(X = r) = \frac{e^{-m} \cdot m^r}{r!} = \frac{e^{-0.735} \times (0.735)^r}{r!}$$

$\therefore$

$$P(\text{no error}) = P(X = 0) = \frac{e^{-0.735} \times (0.735)^0}{0!} = e^{-0.735} = 0.4795$$

Hence, the required probability is 0.4795.

Example 1.72. Average number of accidents on any day on a national highway is 1.8. Determine the probability that the number of accidents are: (*i*) at least one, (*ii*) at most one. (Given $e^{-1.8} = 0.16529$).

Solution. The probability function of the Poisson distribution is

$$P(X=r) = \frac{e^{-m} \cdot m^r}{r!} \qquad (r = 0, 1, 2, 3, \ldots)$$

Given that $m = 1.8$

(*i*) $P(X \geq 1)$:
$$P(X \geq 1) = 1 - p(r < 1)$$
$$= 1 - p(r = 0)$$
$$= 1 - \frac{e^{-m} m^0}{0!} = 1 - \frac{e^{1.8}(1.8)^0}{0!}$$
$$= 1 - e^{-1.8} = 1 - 0.16529 = 0.8347$$

$\therefore$ The probability the number of accidents at least one is 0.8347.

(*ii*) $P(X \leq 1)$:
$$P(X \leq 1) = p(X = 0) + p(X = 1)$$
$$= \frac{e^{-m} m^0}{0!} + \frac{e^{-m} m^1}{1!} = \frac{e^{-1.8}(1.8)^0}{0!} + \frac{e^{-1.8}(1.8)^1}{1!}$$
$$= e^{-1.8}(1 + 1.8) = 0.4628$$

$\therefore$ The probability that the number of accidents at most one is 0.4628.

Example 1.73. 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. Calculate the approximate number of packets containing no defective, one defective and two defective blades in a consignment of 10,000 packets. (Given $e^{-0.02} = 0.9802$).

Solution. Here $N = 10000, p = 0.02, n = 10.$

$\therefore$
$$\text{Mean } (m) = np = 10 \times 0.002 = 0.02$$

Let r be the number of defective blades in a packet.

Let $P(r)$ be the number of packets containing r defective blades, then

$$P(r) = N \times \frac{e^{-m} m^r}{r!}$$

(*i*)
$$P(0) = \text{Number of packets with no defective blades}$$
$$= 10000 \times \left[0.9802 \times \frac{(0.02)^0}{0!}\right] = 10000\,(0.9802) = 9802$$

$\therefore$ Number of packets with no defective blade = 9802

(*ii*) $P(1) =$ Number of packets with 1 defective blade

$$= 1000\left[0.9802 \times \frac{(0.02)^1}{1!}\right] = 10000\,[0.9802 \times 0.02]$$
$$= 10000\,(0.019604) = 196.$$

$\therefore$ Number of packets with one defective blade = 196.04 or 196 (*iii*) $P(2)$ = Number of packets with 2 defective blades

$$= 10000\left[0.9802 \times \frac{(0.02)^2}{2!}\right] = 10000\left[0.9802 \times \frac{0.0004}{2}\right]$$

$$= 10000\,(0.00019604) = 1.96$$

$\therefore$ Number of packets with two defective blades = 1.96 or 2.

Example 1.74. If the variance of the Possion distribution is 2, find the probabilities for r, 1, 2, 3, 4 from the recurrence relation of the Possion distribution. Also find $P(x \geq 4)$. (Use $e^{-2} = 0.1353$)

Solution. Here variance $= m = 2$, $P(0) = e^{-2} = 0.1353$ (given)

We known that $$P(r+1) = \frac{m}{r+1}P(r) = \frac{2}{r+1}P(r)$$

Putting $r = 0, 1, 2, 3$ in (*i*) we get

$$P(1) = \frac{2}{0+1} \times P(0) = 2e^{-2} = 2 \times 0.1353 = 0.27706$$

$$P(2) = \frac{2}{2} \times P(1) = 0.2706$$

$$P(3) = \frac{2}{3} \times P(2) = \frac{2}{3} \times 0.2706 = 0.1804$$

$$P(4) = \frac{2}{4} \times P(3) \times = \frac{1}{2} \times P(3) = 0.0902$$

Now, $$P(X \geq 4) = 1 - [P(0) + P(1) + P(2) + P(3)]$$
$$= 1 - [0.1353 + 0.2706 + 0.2706 + 0.1804] = 0.1431$$

Example 1.75. If a random variable X follows a Poisson distribution such that $P(X=2) = 9.\ P(X=4) + 90\ P(X=6)$, find the mean and variance of X.

Solution. By Poisson distribution, we have

$$P(X=r) = \frac{e^{-m}m^r}{r!}$$

Now $$P(X=2) = 9.\ P(X=4) + 90\ P(X=6)$$

$$\Rightarrow \quad \frac{e^{-m} \cdot m^4}{r!} = \frac{9 \cdot e^{-m}m^4}{4!} + \frac{90 \cdot e^{-m} \cdot m^6}{6!}$$

$$\Rightarrow \quad \frac{3}{4}m^2 + \frac{1}{4}m^4 = 1 \Rightarrow m^4 + 3m^2 - 4 = 0$$

$\Rightarrow (m^2 + 4)(m^2 - 1) \Rightarrow m^2 = 1 \Rightarrow m = 1$ $\quad [\because m > 0 \text{ and } m^2 + 4 \neq 0]$

Hence, mean = 1 and variance = 1 $\quad [\because \text{Mean} = \text{Variance} = m]$

Example 1.76. An insurance company has discovered that only 0.1% of the population is involved in a certain type of accident every year. If its 1000 policy-holders are selected at random from the population, what is the probability that not more than 5 of its clients are involved in such accident next year? (Use $e^{-1} = 0.3668$)

Solution. Here, n = total number of policy-holders = 1000

Let p = probability of a person being involved in an accident $= \dfrac{0.1}{100} = 0.001$

$$m = np = 1000 \times 0.001 = 1$$

$\therefore$

Here, n is large and p is very small.

So, Poisson distribution can be used to find the required probability.

Let X be the Poisson variable with $m = 1$

$$\therefore \qquad P(X=r) = \frac{e^{-m}m^r}{r!} = \frac{1^r e^{-1}}{r!} = \frac{e^{-1}}{r!} \quad [[\because m=1]$$

Required probability $= P(X \le 5)$

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

$$= \left(\frac{e^{-1}}{0!} + \frac{e^{-1}}{1!} + \frac{e^{-1}}{2!} + \frac{e^{-1}}{3!} + \frac{e^{-1}}{4!} + \frac{e^{-1}}{5!}\right)$$

$$= e^{-1}\left(1+1+\frac{1}{2}+\frac{1}{6}+\frac{1}{24}+\frac{1}{120}\right)$$

$$= 0.368 \times \frac{326}{120} = \frac{3.68 \times 3.26}{1200} = 0.9997$$

Hence, the required probability is 0.9997.

Example 1.77. Fit a Poisson distribution on the following :

x	0	1	2	3	4
f	192	100	24	3	1

(Given that $e^{-0.5} = 0.6065$)

Solution.
$$P(r) = \frac{e^{-m}m^r}{r!}$$

m = mean of the distribution

$$= \frac{0\times192 + 1\times100 + 2\times24 + 3\times3 + 4\times1}{192+100+24+3+1} = \frac{161}{320} = 0.5 \text{ (approx)}.$$

$$P(0) = \frac{e^{-0.5}(0.5)^0}{0!} = 0.6065 \text{ i.e., } f = 320 \times 0.6056 = 194 \text{ (approx)}$$

$$P(1) = \frac{e^{-0.5}(0.5)^0}{1!} = 0.30325 \text{ i.e., } f = 320 \times 0.30325 \times 0.30325$$
$$= 97 \text{ (approx)}$$

$$P(2) = \frac{e^{-0.5}(0.5)^1}{2!} = 0.07581 \text{ i.e., } f = 320 \times 0.07581 = 24 \text{ (approx)}$$

$$P(3) = \frac{e^{-0.5}(0.5)^3}{3!} = 0.0126 \text{ i.e., } f = 320 \times 0.0126 = 4 \text{ (approx)}$$

$$P(4) = \frac{e^{-0.5}(0.5)^4}{4!} = 0.0016 \text{ i.e., } f = 320 \times 0.0016 = 0.512 \text{ or } 1 \text{ (approx)}$$

As total number of trials = 320

We have the approximate value as obtained by Poisson distribution as :

x	0	1	2	3	4
f	194	97	24	4	1

Example 1.78. 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 poisson distribution with mean 1.5. Calculate the proportion, of days on which (*i*) neither car is used (*ii*) some demand is refused.

Solution. For Poisson distribution

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}; \quad x = 0, 1, 2\ 3...$$

Now $\lambda = 1.5$

(*i*) The probability that there is no demand

$$P(X=0) = \frac{e^{-1.5}(1.5)^0}{0!} = e^{-1.5} = 0.223$$

The proportion of days on which neither car is used = 0.2231 = 22.31%

(*ii*) The probability that the demand is refused only when demand is more than 2 cars.

$$P(X>2) = 1 - P(X \le 2)$$
$$= 1 - [P(X=0) + P(X=1) + P(X=2)]$$
$$= 1 - \left\{\frac{e^{1.5}(1.5)^0}{0!} + \frac{e^{-1.5}(1.5)}{0!1} + \frac{e^{1.5}(1.5)^2}{2!}\right\}$$
$$= 0.1912$$

The proportion of days on which some demand is refused = 0.1912 = 19.12%

2.6. NORMAL DISTRIBUTION

Normal distribution is the most popular and commonly used distribtion. It was discovered by De Moivre in 1733 after 20 years when Bernoulli gave Binomial distribution. This distribution is a limiting case of Binomial distribution when neither $p \to q$ is too small and n, the number of trials becomes infinitely large *i.e.*, $n \to \infty$. In fact any quantity whose variation depends on random cause will be distributed according to the normal distribution whereas in Binomial and Poisson distribution X assume values like 0, 1, 2,... and thus these distribution are discrete distributions. Cases when the variables can assume any value between 0 and 1 are classified under the continuous variate, for example, in case of height, weight etc.

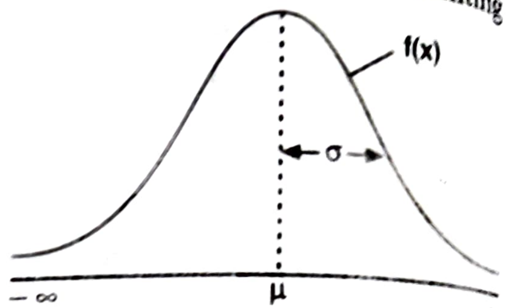

Fig. 2.2.

The continuous random varible X is said to have a normal distribution, if its probability density function is defined as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty \le X \le \infty, \ \sigma > 0$$

where μ and σ are parameters of the normal distribution.

2.6.1. Mean of the Normal Distribution

The normal distribution is

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

$$\text{Mean} = E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Put $\frac{x-\mu}{\sigma} = z$ then $dx = \sigma dz$ and $x = \mu + \sigma z$ limits are unaltered

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

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

$$= \frac{\mu}{\sqrt{2\pi}} 2\int_0^{\infty} e^{-\frac{1}{2}z^2} dz + 0$$

($\because$ $ze^{-\frac{1}{2}z^2}$ is an odd and $e^{-\frac{1}{2}z^2}$ is even function)

$$= \frac{2\mu}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{\sqrt{2}}$$

$$= \mu$$

$$E(X) = \mu$$

$\therefore$ The mean of the normal distribution is 'μ'.

2.6.2. Variance of the Normal Distribution

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

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

Put $\frac{x-\mu}{\sigma} = z$ then $dx = \sigma dz$ limits are unaltered

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

($\because$ $z^2 e^{-\frac{1}{2}z^2}$ is an even function)

Put $\frac{z^2}{2} = t$ then $zdz = dt$ and $z = \sqrt{2t}$ also limits are unaltered

$$= \frac{2\sigma^2}{\sqrt{2\pi}}\int_0^{\infty} 2te^{-t} \frac{dt}{\sqrt{2t}}$$

$$= \frac{2\sigma^2}{\sqrt{2\pi}}\int_0^{\infty} \sqrt{2t}\, e^{-t} dt = \frac{2\sigma^2}{\sqrt{\pi}}\int_0^{\infty} t^{\frac{1}{2}} e^{-t} dt = \frac{2\sigma^2}{\sqrt{\pi}}\int_0^{\infty} e^{-t} t^{\frac{3}{2}-1} dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right) \qquad \because \int_0^{\infty} e^{-t} t^{\frac{3}{2}-1} dt = \sqrt{\left(\frac{3}{2}\right)}$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{2\sigma^2}{\sqrt{\pi}} \frac{1}{2}\sqrt{\pi} \qquad \because \sqrt{\frac{1}{2}} = \sqrt{\pi}$$

$$\because \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$= \sigma^2$$

Variance of normal distribution is σ^2.

2.6.3. Median of the Normal Distribution

Suppose 'M' is median of normal distribution then

$$\int_{-\infty}^{M} f(x)\,dx = \int_{M}^{\infty} f(x)\,dx = \frac{1}{2} \qquad \ldots(i)$$

Now take, $$\int_{-\infty}^{M} f(x)\,dx = \frac{1}{2} \Rightarrow \int_{-\infty}^{\mu} f(x)\,dx + \int_{\mu}^{M} f(x)\,dx$$

$$\int_{-\infty}^{\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx + \int_{\mu}^{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2} \qquad \ldots(ii)$$

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

Put $\dfrac{x-\mu}{\sigma} = z$ then $dx = \sigma dz$

Also limits are when $x = -\infty$ then $z = -\infty$

Also when $x = \mu$ then $z = 0$

$$\int_{-\infty}^{\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_{-\infty}^{0} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} \sigma\, dx$$

$$= \frac{1}{2\pi}\int_{0}^{\infty} e^{-\frac{1}{2}(z)^2} dz \qquad \text{by symmetry property}$$

$$\therefore \quad \int_{-\infty}^{\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{\sqrt{2}} = \frac{1}{2} \qquad \ldots(iii)$$

From (*ii*) and (*iii*)

$$\frac{1}{2} + \int_{\mu}^{M} f(x)\,dx = \frac{1}{2}$$

$\therefore \displaystyle\int_{\mu}^{M} f(x)\,dx = 0$. This implies that $\mu = M$ $\qquad$ ($\because$ if $\displaystyle\int_{a}^{b} f(x)\,dx = 0$ then $a = b$)

Hence, the median of the normal distribution is 'μ'.

2.6.5. Properties of the Normal Distribution

1. The normal probability curve with mean μ and standard deviation σ is given by

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

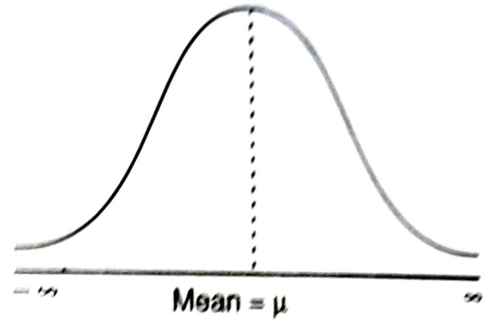

Fig. 2.2.

2. The curve is bell-shaped and symmetrical about the line $x = \mu$.
3. Mean, median and mode of the normal distribution of coincide the normal distribution is unimodal.
4. $f(x)$ decreases rapidly as x increases.
5. X-axis is an asymptote to the curve (the tangent to the curve at ∞ is X-axis)
6. The maximum probability occurs at the point $x = \mu$ and is $\dfrac{1}{\sigma\sqrt{2\pi}}$.
7. Mean deviation about mean $= \dfrac{4}{5}\sigma$.
8. Since $f(x)$, being the probability, can never the negative, so that no portion of the curve lies below the X-axis.
9. A linear function of independent normal variates is also normal variate.
10. The point of inflextion of the curve are at $x = \mu \pm \sigma$.

   (i) Area of the normal curve between $(\mu - \sigma)$ and $(\mu + \sigma)$ is 0.6826

   i.e., $P(\mu - \sigma < X < \mu + \sigma) = 0.6826$

(ii) Area of the normal curve between $\mu - 2\sigma$ and $\mu + 2\sigma$ is 0.9544

i.e., $P(\mu - 2\sigma < X \mu + 2\sigma) = 0.9544$

(iii) Area of the normal curve between $\mu - 3\sigma$ and $\mu + 3\sigma$ is 0.9973

i.e., $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$

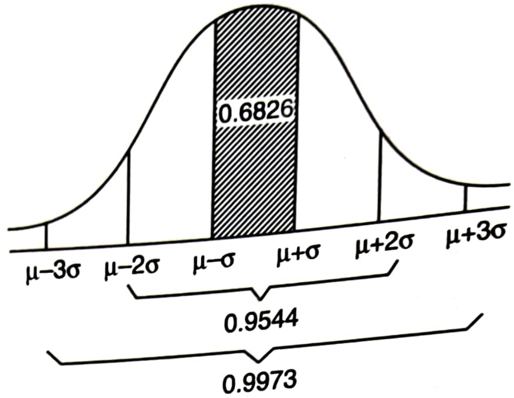

Fig. 2.3

2.6.6. Normal Distribution as a Limiting Form of Binomial Distribution

Normal distribution is another limiting form of the binomial distribution under the following conditions:

(i) The number of trials n, is indefinitely large, *i.e.*, $n \to \infty$

(ii) Neither 'p' or nor 'q' is very small.

For large n, the calculation of binomial probabilities is very difficult. In such a case we can use normal distribution and compute the required probability.

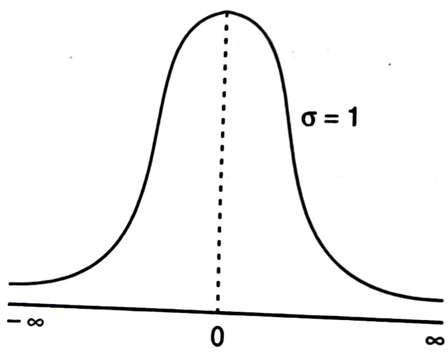

Fig. 2.4.

Suppose the number of successes ranges from x_1 to x_2 then the probability of getting x_1 to x_2 successes is given by

$$\sum_{x=x_1}^{x_2} C_x p^x q^{n-x}$$

Now consider two cases.

Case 1: $p = q = \frac{1}{2}$ and n is not large. The mean of the binomial distribution = np and standard deviation = $\sqrt{npq}$.

The corresponding mean and standard deviation of the normal distribution are μ and s respectively.

Since we know that

Let z_1 and z_2 be the values of Z, corresponding to x_1 and z_2 of X respectively. Then

$$= P(x_1 < X < x_2)\ P = (z_1 < Z < z_2) = \int_{z_1}^{z_2} \phi(z)dz$$

And the value can be computed from the standard normal distribution.

Case 2: $p \neq q$ for large n.

We can approximate the binomial to the normal curve and calculate the corresponding probabilities.

For any success x which lies in the interval $\left(x - \frac{1}{2}, x + \frac{1}{2}\right)$ then z_1 can be calculated corresponding to the lower limit of the x_1 class and z_2 to the limit of the x_2 class.

$$\therefore \qquad z_1 = \frac{\left(x_1 - \frac{1}{2}\right) - \mu}{\sigma} = \frac{x_1 - \frac{1}{2} - np}{\sqrt{npq}}$$

and

$$z_2 = \frac{\left(x_2 + \frac{1}{2}\right) - \mu}{\sigma} = \frac{x_2 + \frac{1}{2} - np}{\sqrt{npq}}$$

$\therefore$ The required probability $= \int_{z_1}^{z_2} \phi(z)dz$,

and this value can be computed from standard normal distribution.

Constant of Normal Distribution

The mean of the normal distribution is $\bar{x}$

The standard deviation of the normal distribution is σ

$$\mu_2 = \sigma^2, \quad \mu_3 = 0 \quad \text{and} \quad \mu_4 = 3\sigma^4$$

β_1 or moment coefficient of skeness, $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = 0$

β_2 or moment coefficient of kurtosis, $\beta_2 = \dfrac{\mu_4}{\mu_2^2} = \dfrac{3\sigma^4}{\sigma^4} = 3$

SOLVED EXAMPLES

Example 2.12. If $\mu = 50$ and $\sigma = 10$ find : (*i*) $P(50 \le X \le 80)$, (*ii*) $P(60 \le X \le 70)$ (*iii*) $P(30 \le X \le 40)$, (*iv*) $P(40 \le X \le 60)$. Use Table : Area under the normal curve.

Solution. Standard normal variate $z = \dfrac{X - \mu}{\sigma} = \dfrac{X - 50}{10}$

(*i*) $z = \dfrac{50-50}{10} = 0$ when $X = 50$ and $z = \dfrac{80-50}{10} = 3$, when $X = 80$

Hence, $P(50 \le X \le 80) = P(0 \le z \le 3) = 0.4987$ [From table]

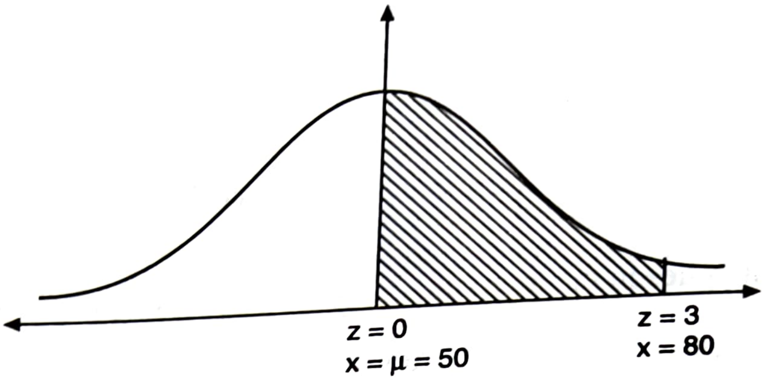

Fig. 2.5.

(*ii*) $P(60 \le X \le 70) = P(1 \le z \le 2) =$ Area from $z = 1$ to $z = 2$

$=$ (Area from $z = 0$ to $z = 2$) – (Area from $z = 0$ to $z = 1$)

$= 0.4772 - 0.3413 = 0.1359$

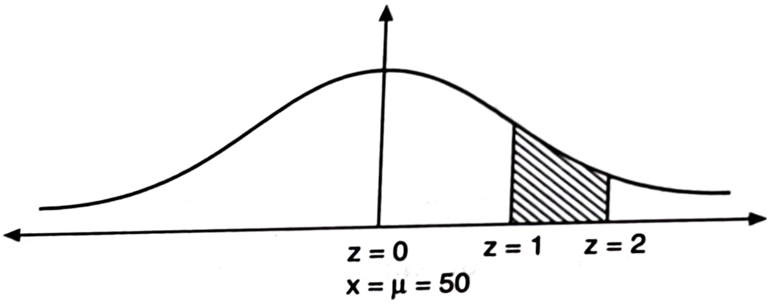

Fig. 2.6.

(*iii*) $P(30 \le X \le 4) = P(-2 \le z \le -1)$

Due to symmetry, area between $z = -1$ to $z = -2$ will be the same as between $z = -1$ to $z = -2$, which is the same as in (*ii*) *i.e.*, 0.1359.

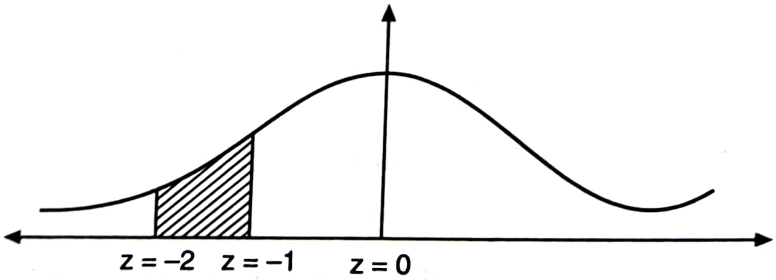

Fig. 2.7.

(*iv*) $P(40 \le X \le 60) = P(-1 \le z \le 1)$

$=$ area between $z = -1$ to $z = 1$

$=$ twice the area between $z = 0$ to $z = 1$

$= 2 \times 0.3413 = 0.6826$

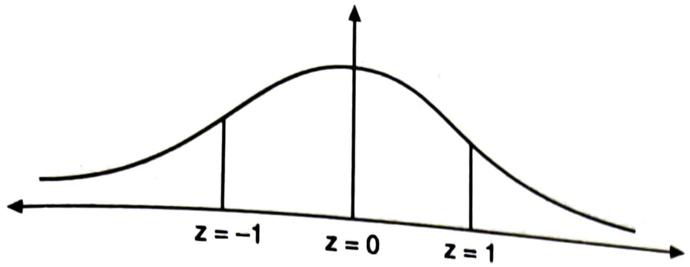

Fig. 2.8.

Example 2.13. In a normal distribution 31% of items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution.

Solution. Let μ be the mean and σ be the standard deviation of the distribution

Normal variate $z = \dfrac{x-\mu}{\sigma}$

As 31% of items are under 45 and 8% over 64

∴ At $x = 64$, we have $$z = \frac{64-\mu}{\sigma} = z_1 \text{ (say)}$$

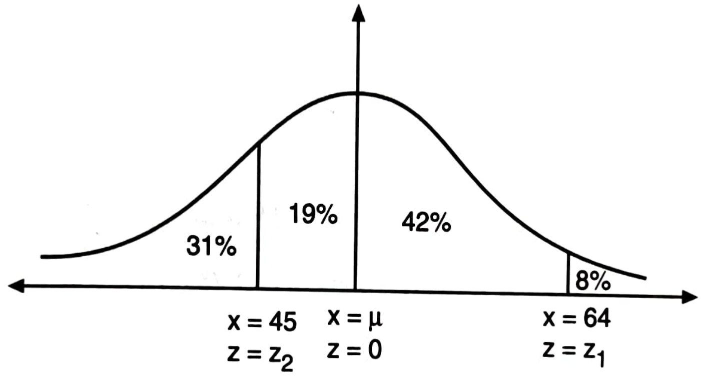

Fig. 2.9.

Area from $z = 0$ to $z = z_1$ is 42% *i.e.*, 0.42. The value of z_1 corresponding to area 0.42 from the table 1 is 1.405

$$\therefore \qquad 1.405 = \frac{64-\mu}{\sigma} \qquad ...(i)$$

Similarly at $x = 45$, $$|z| = \left|\frac{45-\mu}{\sigma}\right| = |z_2|$$

Area between $x = 45$ (*i.e.*, $z = z_2$) to $x = \mu$ ($z = 0$) is the same numerically as between $z = 0$ to $z = z_2$ which is 19%.

Now the value of normal variate z_2 corresponding to area 0.19 is 0.495

$$\therefore \qquad \left|\frac{45-\mu}{\sigma}\right| = |0.495|$$

$$\therefore \qquad \frac{\mu-45}{\sigma} = 0.495 \qquad ...(ii)$$

From (i) and (ii) $$\frac{64-\mu}{\mu-45} = \frac{1.405}{0.495} = 2.8.$$

$$\mu = 50 \text{ and } \sigma = 10.$$

Example 2.14. A sample of 100 dry battery cells tested to find the length of life produced the following results:

$$\mu = 12 \text{ hours}, \sigma = 3 \text{ hours}$$

Assuming the data to be normally distributed, what percentage of battery calls are expected to have life:

(a) more than 15 hours

(b) less than 6 hours

(c) between 10 and 14 hours.

Solution. Let X denotes the length of life of dry battery cells.

Also $$z = \frac{X-\mu}{\sigma} = \frac{X-12}{3}$$

(a) When $X = 15$, $$z = \frac{15-12}{3} = 1$$

$\therefore$ $$P(X > 15) = P(z > 1)$$

Area of the right of $z = 1$ is $0.5 - 0.3413 = 0.1587$

$\therefore$ Percentage of battery cells having life more than 15 hours

$$= 0.1586 \times 100 = 15.87\%$$

(b) When $X = 6$, $$z = \frac{6-12}{3} = -2$$

$\therefore$ $$P(X < 6) = P(z < -2)$$

$\therefore$ Percentage of battery cells have life less than 6 hours

$$= 0.0028 \times 100 = 2.28\%$$

(c) When $X = 10$, $$z = \frac{10-12}{3} = -0.67$$

When $X = 14$, $$z = \frac{14-12}{3} = 0.67$$

$\therefore$ $$P(10 < X < 14) = P(-0.67 < z < 0.67) = 2P(0 < Z < 0.67)$$

Area between $X = 10$ to $X = 14$ is twice the area of the left of $z = 0.67$

$$= 2 \times 0.2487 = 0.4974$$

$\therefore$ Percentage of battery cells having life span between 10 hours and 14 hours = 49.74%

Example 2.15. The customer accounts of certain department store have an average balance of ₹ 120 and a standard deviation of ₹ 40. Assuming that the account balances are normally distributed:

(i) What proportion of the account is over ₹ 150?

(ii) What proportion of account is between ₹ 100 and ₹ 150?

(iii) What proportation of account is between ₹ 60 and ₹ 90?

Solution. Standard normal variate $z = \dfrac{X - \mu}{\sigma}$

(i) When $X = 150$, $\quad z = \dfrac{150-120}{40} = 0.75$

To find out the proportion of accounts greater than ₹ 150, we refer to right of $z = 0.75$

The area to the right of $z = 0$ is 0.5

Less the area between $z = 0$ and $z = 0.75$ is 2735.

The area of the right to $z = 0.75$ is 0.2266.

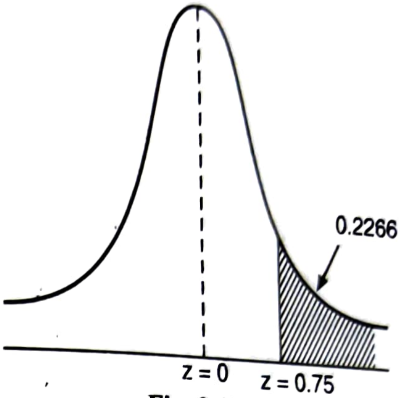

Fig. 2.10.

Therefore, 22.66% of account have balance is excess of ₹150 (shaded area)

(ii) When $X = 100$, $\quad z = \dfrac{100-120}{40} = -0.5$

When $X = 150$, $\quad z = \dfrac{150-120}{40} = 0.75$

The area between $z = -0.5$ and $z = 0$ is 0.1915.

The area between $z = 0$ and $z = 0.175$ is 0.2734

$\therefore$ The total area *i.e.*, $z = -0.5$ and $z = 0$ is 0.1915

The area between $z = 0$ and $z = 0.75$ is 0.2734.

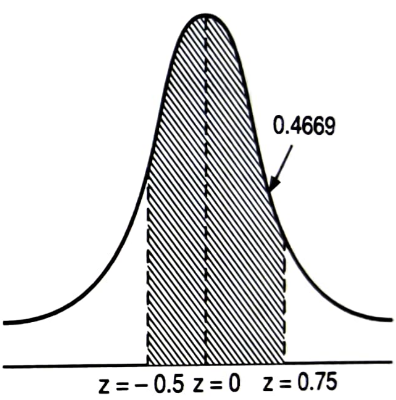

Fig. 2.11

$\therefore$ The area total area *i.e.*, $z = -0.5$ to $z = 0.75$ is $0.1915 + 0.273 = 0.4649$

$\therefore$ 46.49% of accounts have an average balance of accounts between ₹ 100 and 150.

(iii) When $X = 90$, $\quad z = \dfrac{90-120}{40} = -0.75$

When $X = 60$, $\quad z = \dfrac{60-120}{40} = -1.5$

Area between $z = 0$ to $z = -1.5$ is 0.4332

Area between $z = 0$ to $z = -0.75$ is 0.2734

$\therefore$ The area between $X = 90$ and $X = 60$

($z = -1.5$ to $z = -0.75$) is $0.4332 - 0.2734 = 0.1598$

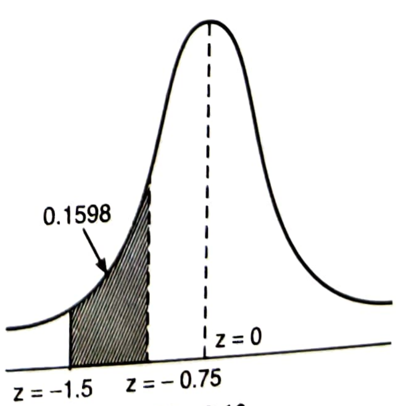

Fig. 2.12.

$\therefore$ 15.98% of accounts are between ₹ 60 and ₹ 90

Example 2.16. The average height of soliders of a country is given as 68.22 inches with variance 10.8 sq. inch. How many soldiers out of 1000 would you expect to be over 72 inches tall? Given that the area under the normal curve between $z = 0$ to $z = 0.35$ is 0.1368 and between $z = 0$ and $z = 1.15$ is 0.3746.

Solution. Standard normal variate

$$z = \frac{X-\mu}{\sigma} = \frac{72-68.22}{\sqrt{10.8}} = 1.15$$

$P(X \geq 72) = P(z \geq 1.15) =$ Total area to the right to ordinate $z = 0)$ – (area between $z = 0$ to $z = 1.15$)

$$= 0.5 - 0.3746 = 0.1254$$

$\therefore$ Total number of soldiers k above 72″ tall $= 1000 \times 0.1254 = 125$

Example 2.17. In an intelligence test administered to 1000 students the average score was 42 and standard deviation 24. Find: (*a*) the number of students exceeding a score of 50, (*b*) the number of students lying between 30 and 54, (*c*) the value of score exceeded by the top 100 students.

Solution. (*a*) Given $\mu = 42, x = 50, \sigma = 24$

$$\therefore \quad z = \frac{x-\mu}{\sigma} = \frac{50-42}{24} = 0.333$$

Area of the right to ordinate at 0.333 is $0.5 - 0.1304 = 0.3696$

$\therefore$ The expected number of children exceeding a score of 50

$$= 0.3696 \times 1000 = 369.6 \text{ or } 370$$

(*b*) Standard normal variate for score 30

$$z = \frac{x-\mu}{\sigma} = \frac{30-42}{24} = -0.5$$

Standard normal variate for score 54

$$z = \frac{x-\mu}{\sigma} = \frac{54-42}{24} = 0.5$$

Area from $z = 0$ to $z = 0.5$ is 0.1915

Area from $z = -0.5$ to $z = 0$ is 0.1915

$\therefore$ Area from $z = -0.5$ to $z = 0.5 = 0.1915 + 0.1915 = 0.3830$

Thus the number of children having score between 30 and 54 $= 0.383 \times 1000 = 383$

(*c*) Probability of getting top 100 students $= \dfrac{100}{1000} = 0.1$

Standard normal variate having 0.1 area of the right $= 1.281$

Standard normal variate for score x

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

$$1.281 = \frac{x-42}{24}$$

$$1.28 \times 24 = x - 42$$

$$x = (1.28 \times 24) + 42 = 72.72 \text{ or } 73$$

Example 2.18. The distribution of a random variable is given by

$$f(x) = Ce^{-\frac{1}{50}(9x^2 - 30x)}, -\infty \le x \le \infty$$

Find the constant C, the mean and the variance of the random variable. Find also the upper 5% value of the random variable.

Solution. $P(-\infty \le x \le \infty) = 1$ = unit area under the curve

$y = f(x)$ and x-axis

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

$$1 = Ce^{\frac{1}{2}} \int_{-\infty}^{\infty} e^{\frac{(x-\mu)^2}{2\sigma^2}} dx$$

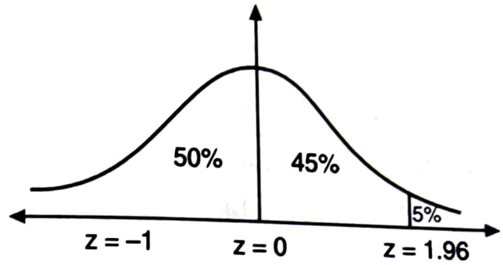

Fig. 2.13.

Comparing with normal distribution, we have

We have
$$\mu = \frac{5}{3} = 1.667$$

$$2\sigma^2 = \frac{50}{9} \text{ or } \sigma^2 = \frac{25}{9} = 2.778$$

$$\sigma = \frac{5}{3}, \; C = \frac{1}{\sqrt{e}\sqrt{2\pi\sigma}} = \frac{1}{\sqrt{2\pi \times 1.667}\sqrt{e}} = 0.145$$

Standard normal variate $t = \dfrac{x-\mu}{\sigma} = \dfrac{x - \frac{5}{3}}{\frac{5}{3}}$ corresponding to 95% area, we have $1.96 = \dfrac{x - \frac{5}{3}}{\frac{5}{3}}$

$$\therefore \quad 1.96 \times \frac{5}{3} + \frac{5}{3} = x \text{ or } x = 4.933$$

Example 2.19. Assuming that the diameters of 1000 brass plugs taken consecutively from a machine form a normal distribution with mean 0.7515 cm and standard deviation 0.002 cm. Find the number of plugs likely to be rejected if the approved diameter is 0.752 ± 0.004 cm.

Solution. $\mu = 0.7515, \sigma = 0.002$

Limits of the diameter of non-defective plugs are

$$0.752 - 0.004 = 0.748 \text{ and } 0.752 + 0.004 = 0.756 \text{ cm}$$

At $x = 0.748$ m the standard normal variate

$$z_1 = \frac{x-\mu}{\sigma} = \frac{0.7480 - 0.7515}{0.002} = -1.75$$

and at
$$x = 0.756, \; z_2 = \frac{0.756 - 0.7515}{0.002} = 2.25$$

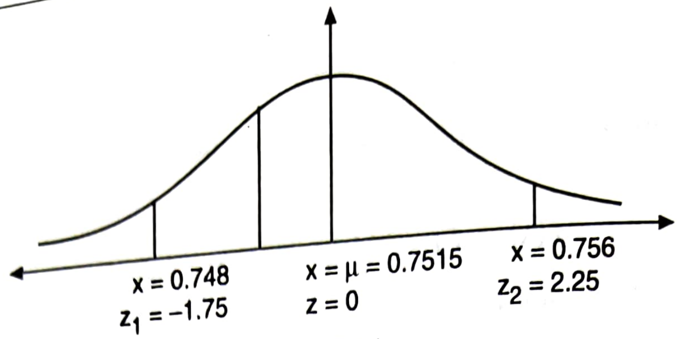

Fig. 2.14.

Area between z_1 and z_2 is numerical equal to the sum of the area between 0 to z_1 and z_2. From the table, when $z = 1.75$ area is equal to 0.4599 and when $z = 2.25$ area is equal to 0.4878.

Hence, the required are for non-defective items $= 0.4599 + 0.4878 = 0.9477$.

Non-defective plugs are $100 \times 0.9477 = 948$

Hence, defective plugs are $1000 - 948 = 52$

$\therefore$ 52 pulgs are likely to be rejected.

Example 2.20. A sales tax officer reported that the average sales of 500 business establishment in a year is ₹ 36,000/and standard deviation of ₹ 10,000/-. Assuming that the sales in these business are normally distributed, find: (*i*) the number of businesses the sales of which are more than ₹ 40,000/-, (*ii*) the percentage of businesses the sales of which are likely to range between ₹ 30,000 and ₹ 40,000.

Solution. From the given data

$$N = 500,\ \mu = 36{,}000 \quad \text{and} \quad \sigma = 10{,}000$$

The standard normal variate $z = \dfrac{X - \mu}{\sigma} \Rightarrow z = \dfrac{X - 36{,}000}{10{,}000}$

(*i*) $P(X > 40{,}000)$:

When $x = 40{,}000$ then $z = \dfrac{40{,}000 - 36{,}000}{10{,}000}$

$$= 0.4 \text{ (say } z_1)$$

$$\begin{aligned} P(X > 40{,}000) &= P(z > 0.4) \\ &= 1 - P(z \le 0.4) \\ &= 1 - (F(0.4)) = 1 - 0.6554 \\ &= 0.3446 \end{aligned}$$

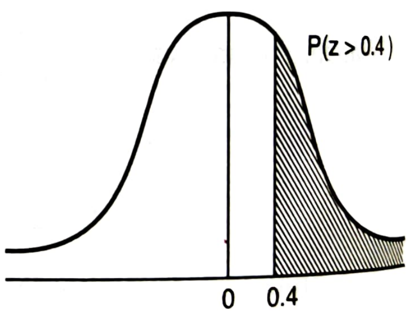

Fig. 2.15.

$\therefore$ The number of businesses the sales which are more than ₹ 40,000/- are

$$\begin{aligned} &= N.P(X > 40{,}000) \\ &= 500 \times 0.3446 = 172.3 \\ &= 172 \end{aligned}$$

(*ii*) $P(30{,}000 < X < 40{,}000)$:

When $x = 30{,}000$, then $z = \dfrac{30{,}000 - 36{,}000}{10{,}000} = -0.6$ (say z_1)

When $x = 40{,}000$ then $z = \dfrac{40{,}000 - 36{,}000}{10{,}000} = 0.4$ (say z_2)

(iii) $P(30{,}000 < X < 40{,}000) = P(-0.6 < z < 0.4)$

$$= F(0.4) - F(-0.6)$$
$$= 0.6554 - 0.2743$$
$$= 0.3811$$

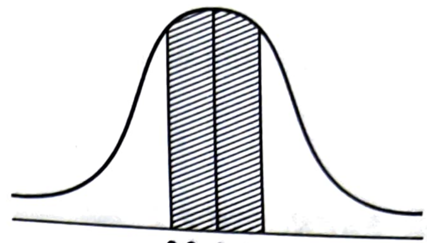

Fig. 2.16.

∴ The number of business the sales which are likely to range between 30,000 and 40,000 are

$$= N.P\,(30{,}000 < X < 40{,}000)$$
$$= 500 \times 0.3811 = 190.55 = 191$$

Out of 500 businesses 191 having sales between 30,000 and 40,000. For 100 businesses 38.2.

∴ The required percenatge is 38.2%.

Example 2.21. Suppose 10 percent of probability for a normal distribution $N(\mu, \sigma^2)$ is below 25 and 5 percent above 90, what are the value of μ and σ.

Solution. $$P(x \le 35) = \frac{10}{100} = 0.1 \qquad \text{... } (i)$$

$$P(x > 90) = \frac{5}{100} = 0.05 \qquad \text{... } (ii)$$

The standard normal variate $z = \dfrac{x - \mu}{\sigma}$

When $x = 35$ then $z = \dfrac{35 - \mu}{\sigma}$

When $x =$ then $z = \dfrac{90 - \mu}{\sigma}$

From eqn. (i) $P(x \le 35) = 0.1$

$$\Rightarrow \quad P\left(z \le \frac{35 - \mu}{\sigma}\right) = 0.1$$

$$\Rightarrow \quad \left(\frac{35 - \mu}{\sigma}\right) = -1.28 \text{ from table}$$

$$\Rightarrow \quad 35 - \mu = 1.28\,\sigma$$

$$\Rightarrow \quad \mu - 1.28\,\sigma = 35 \qquad \text{...}(iii)$$

From eqn. (ii) $P(x > 90) = 0.$

$$P\left(z > \frac{90 - \mu}{\sigma}\right) = 0.05$$

$$\Rightarrow \quad \left(\frac{90 - \mu}{\sigma}\right) = -1.645$$

$$\Rightarrow \quad 90 - \mu = -1.645$$

$$\Rightarrow \qquad \mu - 1.645\,\sigma = 90 \qquad \ldots(iv)$$

From eqns. (*iii*) and (*iv*)

$$\mu = 157.87$$
$$\sigma = 150.68$$

EXERCISE 2.2

1. Students of a class were given an aptitude test. Their marks were found to be normally distributed with mean 60 and standard 5. What percentage of students scored more than 60 marks? **[Ans.** 50]
2. Find the value of z in each of the cases
   (*i*) area between $z = 0$ and $z = 0.3770$.
   (*ii*) area of the left of z is 0.8621. **[Ans.** (*i*) ±1.6, (ii) 1.09]
3. In a sample of 1000 cases, the mean of a certain test is 14 and standard deviation is 2.5. Assuming the distribution to be normal, find:
   (*i*) how many students score between 12 and 15?
   (*ii*) how many score above 18?
   (*iii*) how many score below 8?
   (*iv*) how many score 16? **[Hint :** (*iv*) Take marks between 15.5 and 16.5]
   [Ans. (*i*) 444, (*ii*) 55, (*iii*) 8, (*iv*) 116]
4. Find the area under the normal curven in each of the following case :
   (*i*) $z = 0$ and $z = 1.2$
   (*ii*) $z = -0.68$ and $z = 0$
   (*iii*) $z = -0.46$ and $z = 2.21$
   (*iv*) $z = 0.81$ and $z = 1.94$ **[Ans.** (*i*) 0.3849, (*ii*) 0.2518, (*iii*) 0.6637, (*iv*) 0.1828)
   [Hints: Use standard curve table]
5. Students of a class were given a mechanical aptitude test. These marks were found to be normally distributed with mean 60 and standard deviation 5. What percent of students scored: (*i*) more than 60 marks, (*ii*) less than 56 marks, (*iii*) between 45 and 65 marks.
   (Ans. (*i*) 50%, (ii) 21.21%, (*iii*) 84%)
6. The mean height of 500 students's is 151 cm and the standard deviation is 15 cm. Assuming that the heights are normally distributed, find how many students's heights lie between 120 and 155 cm.
   [Ans. 294]
7. Fit a normal curve

$$P(x) = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} ; -\infty \le x \le \infty$$

Given the following data :

x	8.60	8.59	8.57	8.56	8.55	8.54	8.53	8.52
f	2	3	4	10	8	4	1	1

$$\left[\textbf{Ans. } P(x) = \frac{42}{0.176\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-8.563}{0.176}\right)^2}\right]$$

2.7. EXPONENTIAL DISTRIBUTION

A random variable X is said to have an exponential distribution with parameter $\lambda > 0$; if its probability density function is given by:

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

The cumulative distribution function $F(x)$ is given by $F(x) = P(X \leq x) = \int_{-\infty}^{x} \lambda e^{-\lambda x}\, dx$

$$\Rightarrow \qquad F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \\ 1, & x = \infty \end{cases}$$

Remark: Sometimes exponential distribution is defined by the p.d.f.

$$f(x) = \frac{1}{\beta} e^{-\frac{1}{\beta}x}, \qquad x > 0$$

$$0, \qquad \text{otherwise}$$

Graph of Exponential Probability Density Function

For $\qquad f(x) = \lambda e^{-\lambda x}$

x	0	1	2	....	∞
$f(x)$	λ	$\lambda e^{-\lambda}$	$\lambda e^{-2\lambda}$	....	0

when $\quad \lambda = 1, \quad f(x) = 1, e^{-1}, e^{-2}, \ldots\ldots 0$

$\lambda = 2, \quad f(x) = 2, 2e^{-2}, 2e^{-4}, \ldots\ldots 0$

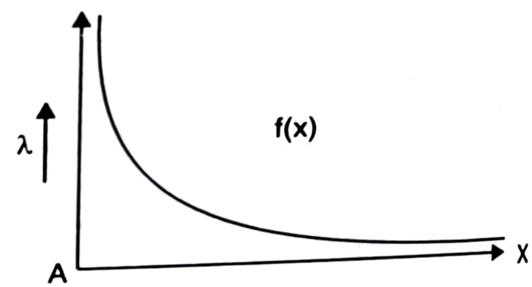

Fig. 2.17.

Graph of Distribution Function

For $\qquad F(x) = 1 - e^{-\lambda x}$

x	0	1	2	....	∞
$F(x)$	0	$1 - e^{-\lambda}$	$1 - e^{-2\lambda}$	....	1

when $\quad \lambda = 1, \quad F(x) = 0, 1 - e^{-1}, 1 - e^{-2}, \ldots\ldots 1$

$\lambda = 2, \quad F(x) = 0, 1 - e^{-2}, 1 - c^{-4}, \ldots\ldots 1$

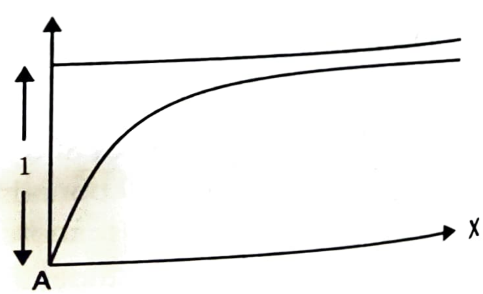

Fig. 2.18.

SOLVED EXAMPLES

Example 2.22. A power supply unit for a computer component is assumed to follow an exponential distribution with a mean life of 1200 hours. What is the probability that the component will

(i) fail in the first 300 hours?

(ii) survive more than 1500 hours?

(iii) last between 1200 hours and 1500 hours?

Solution. Given that $\lambda = 1200$

(i) The probability that the component will fail in the first 300 hours is

$$P(X \leq 300) = 1 - e^{-\frac{300}{1200}}$$
$$= 1 - e^{-0.25}$$
$$= 1 - 0.7788 = 0.2212$$

(ii) The probability that the component will servive more than 1500 hours is

$$P(X \geq 1500) = e^{-1500/1200}$$
$$= e^{-1.25} = 0.2865$$

(iii) The probability that the component will lost between 1200 hours and 1500 hours is given as $P(1200 \leq X \leq 1500)$

$$\Rightarrow \qquad P[\text{inside}] = 1 - P[\text{outside}]$$
$$= 1 - [P(X \leq 1200) + P(X \geq 1500)]$$

Now, $$P(X \leq 1200) = 1 - e^{-\frac{1200}{1200}} = 0.6321$$

and $$P(X \geq 1500) = e^{-\frac{1500}{1200}} = 0.2865$$
$$P(\text{Inside}) = 1 - [0.6321 + 0.2865] = 0.0813$$

Hence, $$P(1200 \leq X \leq 1500) = 0.0813.$$

Example 2.23. A random variable has an exponential distribution with probability density function given by

$$f(x) = \begin{cases} 3e^{-3x}, & \text{for } x > 0 \\ 0, & \text{for } x \leq 0. \end{cases}$$

What is the probability, that X is not less than 4? Find mean and standard deviation. Show that

$$\text{Coefficient of variation} = \left(\frac{\text{mean}}{\text{S.D.}}\right) = 1.$$

Solution. $P(X \nless 4) = P(X \geq 4) = \int_4^{\infty} 3e^{-3x}\, dx = [-e^{-3x}]_4^{\infty} = e^{-12}.$

$$\text{Mean} = E(X) = \int_0^{\infty} x3e^{-3x}\, dx = 3\int_0^{\infty} xe^{-3x}\, dx$$

$$= 3.\frac{\Gamma(2)}{3^2} = \frac{1}{3}.$$

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

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

$$= \int_0^{\infty} x^2 . 3e^{-3x}\, dx - \left(\frac{1}{3}\right)^2$$

$$= 3.\frac{\Gamma(3)}{3^3} - \frac{1}{9} = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$$

$\therefore$ Standard deviation, S.D. $= \sqrt{\frac{1}{9}} = \frac{1}{3}$

Hence, $\frac{\text{Mean}}{\text{S.D.}} = \frac{1/3}{1/3} = \frac{1}{3} \times \frac{3}{1} = 1.$

Example 2.24. The sales tax, X, of a shopkeeper has an exponential distribution with *p.d.f.*

$$f(x) = \frac{1}{4}e^{-x/4}, \quad x \geq 0$$

$$= 0, \quad x < 0$$

If sales tax is levied at the rate of 5% what is the probability that his sales exceed ₹ 10,000?

Solution. Sales tax on ₹ 10,000 = 10,000 × $\frac{5}{100}$ = ₹ 500

$\therefore$ If sales exceed ₹ 10,000 tax will exceed ₹ 500. Hence, the required probability is

$$P(X > 500) = \int_{500}^{\infty} \frac{1}{4}e^{-x/4}\, dx$$

$$= \left[\frac{1}{4}\frac{e^{-x/4}}{\left(-\frac{1}{4}\right)}\right]_{500}^{\infty} = e^{-\frac{500}{4}} = e^{-125}$$

Example 2.25. If families are selected at random in a certain thickly populated area and their annual income in excess of ₹ 4000 is treated as a random variable having an exponential distribution

$$f(x) = \frac{1}{2000} e^{-\frac{x}{2000}}$$

for $x > 0$, what is the probability that 3 out of 4 families selected in the area have income in excess of ₹ 5000?

Solution. ₹ 5000 – ₹ 4000 = ₹ 1000

$$\therefore \quad P(X > 1000) = \int_{1000}^{\infty} \frac{1}{2000} e^{-\frac{x}{2000}} dx = \left[e^{-\frac{x}{2000}} \right]_{1000}^{\infty}$$

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

$$\therefore \quad P(X < 1000) = 1 - \frac{1}{\sqrt{e}}.$$

Hence, required probability $= {}^4C_3 \left(\frac{1}{\sqrt{e}}\right)^3 \left(1 - \frac{1}{\sqrt{e}}\right) = \frac{4(\sqrt{e}-1)}{e^2}$

Example 2.26. The income tax of a man is exponentially distributed with the probability density function given by

$$f(x) = \frac{1}{3} e^{-\frac{1}{3}x}, \text{ for } x > 0$$

$$= 0, \qquad \text{for } x < 0.$$

What is the probability that his income will exceed ₹ 17,000 assuming that the income tax is levied at the rate of 15% on the income above ₹ 15,000?

Solution. If the income exceeds ₹ 17,000 then income tax will exceed by 15% of (17000 – 15000),

i.e., exceeds by $\frac{15}{100} \times 2000$ *i.e.*, exceeds by ₹ 300.

Hence, the required probability

$$= P(X > 300) = \int_{300}^{\infty} \frac{1}{3} e^{-\frac{1}{3}x} dx$$

$$= \frac{1}{3} \left[-3e^{-\frac{1}{3}x} \right]_{300}^{\infty} = e^{-\frac{300}{3}} = e^{-100}.$$

Example 2.27. If X has a uniform distribution in [10, 1] with probability density function

$$f_x(x) = \begin{cases} 1, & 0 < x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the distribution (probability density function) of $-2 \log x$.

Solution. Suppose $y = -2 \log X$. Now the distribution function G of y will be defined as follows:

$$G_y(y) = P(Y \le y) = P(-2 \log X \le y)$$

$$= P\left(\log X \ge -\frac{1}{2}y\right) = P(X \ge e^{-y/2}) = 1 - P(X \le e^{-y/2})$$

$$= 1 - \int_0^{e^{-y/2}} f(x)\, dx = 1 - \int_0^{e^{-y/2}} 1\, dx = 1 - e^{-y/2}.$$

$$\therefore \quad g_y(y) = (d/dy) G(y) = \frac{1}{2} e^{-y/2}$$

Example 2.28. A random variable X has p.d.f.

$$f(x) = \frac{1}{\sigma} e^{-x/\sigma}, \ 0 \le x \le \infty.$$

Find the rth cumulant.

Solution. Moment generating function of X is

$$M_x(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} \cdot \frac{1}{\sigma} e^{-x/\sigma}\, dx = \frac{1}{\sigma}\int_0^{\infty} e^{(t - 1/\sigma)^x}\, dx$$

$$= -\frac{1}{\sigma} \cdot \frac{1}{t - 1/\sigma} = \frac{1}{t\sigma - 1} = (1 - t\sigma)^{-1}$$

$$\therefore \quad K_x(t) = \log (1 - t\sigma)^{-1} = t\sigma + \frac{t^2\sigma^2}{2} + \ldots + \frac{t^r\sigma^r}{r} + \ldots$$

$\therefore$ rth cumulant = coefficient of $\frac{t^r}{r!}$ in $K_x(t) = (r - 1)!\, \sigma^r$.

2.9. GAMMA DISTRIBUTION

A continuous random variable X is said to be follow gamma distribution with parameter α and λ if its p.d.f. is

$$f(x) = \begin{cases} \dfrac{\alpha^{\lambda} e^{-\alpha x} x^{\lambda-1}}{\Gamma(\lambda)}, & x \geq 0, \alpha, \lambda > 0 \\ 0 \quad , & \text{otherwise} \end{cases}$$

Remarks:

1. The function $f(x)$ represents a probability density function since

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

$$= \int_{0}^{\infty} f(x)\,dx \qquad (\because \quad x \geq 0)$$

$$= \int_{0}^{\infty} \frac{\alpha^{\lambda} e^{-\alpha x}}{\Gamma(\lambda)} x^{\lambda-1}\,dx$$

$$= \frac{\alpha^{\lambda}}{\Gamma(\lambda)} \int_{0}^{\infty} e^{-\alpha x} . x^{\lambda-1}\,dx$$

$$= \frac{\alpha^{\lambda}}{\Gamma(\lambda)} \cdot \frac{\Gamma(\lambda)}{\alpha^{\lambda}} = 1 \quad \left[\because \int_{0}^{\infty} e^{-ax} x^{n-1}\,dx = \frac{\Gamma(\lambda)}{a^{n}} \text{(gamma function)}\right]$$

2. **Mean of Gamma Distribution**

As we know,

$$\text{Mean} = \mu_1' = E(X) = \bar{x}$$

i.e.,

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

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

$$= 0 + \int_{0}^{\infty} x\,f(x)\,dx \qquad (\because \quad x \geq 0)$$

$$\int_{0}^{\infty} x\,f(x)\,dx = \int_{0}^{\infty} \frac{x\,\alpha^{\lambda} e^{-\alpha x}}{\Gamma(\lambda)} . x^{\lambda-1}\,dx$$

$$= \frac{\alpha^{\lambda}}{\Gamma(\lambda)} \int_{0}^{\infty} e^{-\alpha x} . x^{\lambda-1+1}\,dx$$

$$= \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty e^{-\alpha x} . x^{(\lambda+1)-1}\, dx$$

$$= \frac{\alpha^\lambda}{\Gamma(\lambda)} . \frac{\Gamma(\lambda+1)}{\alpha^{\lambda+1}} \quad \left[\because \int_0^\infty e^{-ax} x^{n-1}\, dx = \frac{\Gamma(\lambda)}{a^n} \text{(gamma function)}\right]$$

$$= \frac{\cancel{\alpha^\lambda} . \Gamma(\lambda+1)}{\Gamma(\lambda)\, \cancel{\alpha^\lambda} . \alpha}$$

$$= \frac{\Gamma(\lambda+1)}{\Gamma(\lambda) . \alpha} = \frac{\lambda\, \cancel{\Gamma\lambda}}{\cancel{\Gamma\lambda} . \alpha} = \frac{\lambda}{\alpha} \qquad (\because \quad \Gamma(\lambda+1) = \lambda\Gamma(\lambda))$$

$\therefore$ $$\boxed{E(X) = \frac{\lambda}{\alpha}.}$$

3. Variance of Gamma Distribution

$$\text{Var}(X) = \int_0^\infty x^2 f(x)\, dx - [E(x)]^2$$

So,
$$\text{Var}(X) = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty x^{(\lambda+1)} e^{-\alpha x}\, dx - \left(\frac{\lambda}{\alpha}\right)^2$$

$$= \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty \left(\frac{t}{\alpha}\right)^{\lambda+1} e^{-t} \frac{dt}{\alpha} - \frac{\lambda^2}{\alpha^2} \qquad [\because \text{ substituting } t = \alpha x]$$

$$= \frac{e^\lambda}{\alpha^{\lambda+2}\Gamma(\lambda)} \int_0^\infty t^{\lambda+1} e^{-t}\, dt - \frac{\lambda^2}{\alpha^2}$$

$$= \frac{\Gamma(\lambda+2)}{\alpha^2\Gamma(\lambda)} - \frac{\lambda^2}{\alpha^2} \qquad (\because \text{ by definition of gamma function})$$

$$= \frac{\Gamma(\lambda+2) - \lambda^2\Gamma\lambda}{\alpha^2\Gamma(\lambda)}$$

$$= \frac{\lambda(\lambda+1)\,\Gamma(\lambda) - \lambda^2\Gamma(\lambda)}{\alpha^2\Gamma(\lambda)}$$

$$= \frac{\lambda\Gamma(\lambda)\,(\lambda+1-\lambda)}{\alpha^2\Gamma(\lambda)} = \frac{\lambda}{\alpha^2}$$

$\therefore$ $$\boxed{\text{Var}(X) = \frac{\lambda}{\alpha^2}}$$

SOLVED EXAMPLES

Example 2.29. The daily consumption of milk in a city, in excess of 20,000 litres, is approximately distributed as a Gamma variate with parameters $a = \dfrac{1}{10,000}$ and $\lambda = 2$. The city has a daily stock of 30,000 litres. What is the probability that the stock is insufficient on a particular day?

Solution. If the *r.v.* X denotes the daily consumption of milk (in litres) in a city, then the *r.v.* $Y = X - 20,000$ has a gamma distribution with p.d.f.;

$$g(y) = \frac{1}{(10,000)^2\, \Gamma(2)} y^{2-1} e^{-y/10,000} = y\frac{e^{-y/10,000}}{(10,000)^2};\quad 0 < y < \infty$$

Since the daily stock of the city is 30,000 litres, the required probability 'p' that the stock is insufficient on a particular day is given by:

$$p = P(X > 30,000) = P(Y > 10,000)$$

$$= \int_{10,000}^{\infty} g(y)\, dy = \int_{10,000}^{\infty} \frac{y\, e^{-y/10,000}}{(10,000)^2}\, dy$$

$$= \int_1^\infty z e^{-z}\, dz \qquad \text{[Taking } z = y/10{,}000]$$

Integrating by parts, $p = \left| -ze^{-z} \right|_1^\infty + \int_1^\infty e^{-z}\, dz = e^{-1} - \left| e^{-z} \right|_1^\infty = e^{-1} + e^{-1} = \frac{2}{e}.$

Remark. Since $\lambda = 2$, the integration is easily done. However, for general values of a and λ, the integral is evaluated by using tables of Incomplete Gamma Integral of the form:

$\int_0^\infty \frac{e^{-x} x^{n-1}}{\Gamma(n)}\, dx$, which have been tabulated for different values of α and n.

Example 2.30. If $X \sim N(\mu, \sigma^2)$, obtain the p.d.f. of $U = \frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2$.

Solution. Since $X \sim N(\mu, \sigma^2) \Rightarrow z = \frac{x-\mu}{\sigma} \sim N(0, 1)$ with p.d.f.:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right); \quad -\infty < z < \infty. \qquad ...(i)$$

The distribution function $G(\cdot)$ of $U = \frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2$ is given by:

$$G_U(u) = P(U \le u) = P\left[\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2 \le u\right]$$

$$= P(Z^2 \le 2u) = P(-\sqrt{2u} \le Z \le \sqrt{2u}), \text{ where } Z \sim N(0, 1)$$

$$= P(Z \le \sqrt{2u}) - P(Z \le \sqrt{2u})$$

$$= \Phi(\sqrt{2u}) - \Phi(-\sqrt{2u}) \qquad ...(ii)$$

where $\Phi(\cdot)$ is the distribution function of standard normal variate (SNV)Z.
Differentiating w.r.to u, the p.d.f. $g(\cdot)$ of U is given by:

$$g(u) = \phi(\sqrt{2u}) \cdot \frac{d}{du}(\sqrt{2u}) - \phi(-\sqrt{2u}) \cdot \frac{d}{du}(-\sqrt{2u})$$

[$\phi(\cdot)$ is p.d.f. of SNV Z]

$$= \frac{1}{\sqrt{2u}}\left[\phi(\sqrt{2u}) + \phi(-\sqrt{2u})\right]$$

$$= \frac{1}{\sqrt{2u}} \cdot 2\phi(\sqrt{2u}) \qquad [\because \ \phi(u) \text{ is even function of } u]$$

$$= \sqrt{\frac{2}{u}} \cdot \frac{1}{\sqrt{2\pi}} \exp.\left(-\frac{1}{2} \cdot 2u\right) \qquad \text{[From (i)]}$$

$$= \frac{1}{\Gamma(1/2)} e^{-u} \cdot u^{(1/2)-1} \quad u \geq 0$$

$$\left[\because \sqrt{\pi} = \Gamma(1/2) \text{ and } u = \frac{1}{2}\left(\frac{x-u}{\sigma}\right)^2 \geq 0\right]$$

which is the p.d.f. of gamma distribution with parameter $\frac{1}{2}$.

Hence, $$U = \frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2 \text{ is a } \gamma\left(\frac{1}{2}\right) \text{ variate.}$$

Example 2.31. Show that the mean value of positive square root of *a* $\gamma(\mu)$ variate is $\Gamma\left(\mu+\frac{1}{2}\right)/\Gamma(\mu)$. Hence, prove that the mean deviation of a normal variate from its mean is $\sqrt{2/\pi}$, where σ is the standard deviation of the distribution.

Solution. Let X be a $\gamma(\mu)$ variate. Then $f(x) = \dfrac{e^{-x} x^{\mu-1}}{\Gamma(\mu)}$; $\mu > 0, 0 < x < \infty$

$$\therefore \quad E(\sqrt{X}) = \int_0^\infty x^{1/2} f(x)\, dx = \frac{1}{\Gamma(\mu)} \int_0^\infty e^{-x} x^{\mu+(1/2)-1}\, dx = \frac{\Gamma\left(\mu+\frac{1}{2}\right)}{\Gamma(\mu)} \qquad ...(i)$$

If $X \sim N(\mu, \sigma^2)$, then $U = \dfrac{1}{2}\left(\dfrac{X-\mu}{\sigma}\right)^2$ is a $\gamma\left(\dfrac{1}{2}\right)$ variate. (refer example 2.30)

$$\therefore \quad |X-\mu| = \sqrt{2}\sigma U^{1/2}, \text{ where } U \text{ is a } \gamma\left(\frac{1}{2}\right) \text{ variate.}$$

Hence, mean deviation of X about mean is given by:

$$E|X-\mu| = E(\sqrt{2}\,\sigma\, U^{1/2}) = \sqrt{2}\,\sigma\, E(U^{1/2})$$

$$= \sqrt{2}\sigma \frac{\Gamma\left(\frac{1}{2}+\frac{1}{2}\right)}{r\left(\frac{1}{2}\right)} = \frac{\sqrt{2}\sigma}{\sqrt{\pi}} = \sigma\sqrt{2/\pi}. \quad \left[\text{Using } (i) \text{ with } \mu = \frac{1}{2}\right]$$