The *m.g.f.* of Bernoulli variate is :

$$M_X(t) = e^{0\times t} P(X=0) + e^{1\times t} P(X=1) = q + pe^t \qquad \ldots(8\cdot 2b)$$

Remark. *Degenerate Random Variable.* Sometimes we may come across a variate X which is degenerate at a point 'c', say, so that $P(X=c)=1$ and $=0$ otherwise, *i.e.*, the whole mass of the variable is concentrated at a single point 'c'.

Since $P(X=c)=1$, Var $(X)=0$. Hence, a degenerate *r.v.* X is characterised by : Var $(X)=0$.

m.g.f. of degenerate *r.v.* is : $\quad M_X(t) = E(e^{tX}) = e^{tc} P(X=c) = e^{ct}$.

8·4. BINOMIAL DISTRIBUTION

Binomial distribution was discovered by James Bernoulli (1654-1705) in the year 1700 and was first published posthumously in 1713, eight years after his death. Let a random experiment be performed repeatedly, each repitition being called a trial and let the occurrence of an event in a trial be called a success and its non-occurrence a failure. Consider a set of n independent Bernoullian trials (n being finite) in which the probability 'p' of success in any trial is constant for each trial, then $q = 1-p$, is the probability of failure in any trial.

The probability of x successes and consequently $(n-x)$ failures in n independent trials, in a specified order (say) $SSFSFFFS\ldots FSF$ (where S represents success and F represents failure) is given by the compound probability theorem by the expression :

$$\begin{aligned} P(SSFSFFFS\ldots FSF) &= P(S)P(S)\,P(F)\,P(S)\,P(F)\,P(F)\,P(F)\,P(S)\times\ldots\times P(F)\,P(S)\,P(F) \\ &= p\cdot p\cdot q\cdot p\cdot q\cdot q\cdot q\cdot p\ldots q\cdot p\cdot q \\ &= \underbrace{p.\,p.\,p\,\ldots p}_{\{x \text{ factors}\}}\cdot \underbrace{q\cdot q\cdot q\ldots\ldots q}_{\{(n-x) \text{ factors}\}} = p^x q^{n-x} \end{aligned}$$

But x successes in n trials can occur in $\binom{n}{x}$ ways and the probability for each of these ways is same, *viz.*, $p^x q^{n-x}$. Hence the probability of x successes in n trials in *any order* is given by the *addition theorem* of probability by the expression $\binom{n}{x} p^x q^{n-x}$.

The probability distribution of the number of successes, so obtained is called the *Binomial probability distribution,* for the obvious reason that the probabilities of $0, 1, 2, \ldots, n$ successes, *viz.*, $q^n, \binom{n}{1} q^{n-1} p, \binom{n}{2} q^{n-2} p^2, \ldots, p^n$, are the successive terms of the binomial expansion $(q+p)^n$.

Definition. *A random variable X is said to follow binomial distribution if it assumes only non-negative values and its probability mass function is given by :*

$$P(X=x) = p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x}\,;\ x = 0, 1, 2, \ldots, n\,;\ q = 1-p \\ 0 \qquad , \text{otherwise} \end{cases} \qquad \ldots(8\cdot 3)$$

The two independent constants n and p in the distribution are known as the *parameters* of the distribution. 'n' is also sometimes, known as the degree of the binomial distribution.

Binomial distribution is a discrete distribution as X can take only the integral values, *viz.*, $0, 1, 2, \ldots, n$. Any random variable which follows binomial distribution is known as *binomial variate.*

We shall use the notation $X \sim B(n, p)$ to denote that the random variable X follows binomial distribution with parameters n and p.

The probability $p(x)$ in (8·3) is also sometimes denoted by $b(x, n, p)$.

Remarks 1. The assignment of probabilities in (8·3) is permissible because

$$\sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (q+p)^n = 1$$

2. Let us suppose that n trials constitute an experiment. Then, if this experiment is repeated N times, the *frequency function* of the binomial distribution is given by :

$$f(x) = Np(x) = N \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x}; \; x = 0, 1, 2, \ldots, n \qquad \ldots (8\cdot3a)$$

and the expected frequencies of 0, 1, 2, ..., n successes are the successive terms of the binomial expansion, $N(q+p)^n$, $q + p = 1$.

3. *Physical conditions for Binomial Distribution.* We get the binomial distribution under the following experimental conditions :

(*i*) Each trial results in two exhaustive and mutually disjoint outcomes, termed as success and failure.

(*ii*) The number of trials 'n' is finite.

(*iii*) The trials are independent of each other.

(*iv*) The probability of success 'p' is constant for each trial.

The trials satisfying the conditions (*i*), (*iii*) and (*iv*) are also called *Bernoulli trials.*

The problems relating to tossing of a coin or throwing of dice or drawing cards from a pack of cards with replacement lead to binomial probability distribution.

4. Binomial distribution is important not only because of its wide applicability, but because it gives rise to many other probability distributions. Tables for $p(x)$ are available for various values of n and p.

Example 8·1. *Ten coins are thrown simultaneously. Find the probability of getting at least seven heads.*

Solution. p = Probability of getting a head $= \frac{1}{2}$

q = Probability of not getting a head $= \frac{1}{2}$

The probability of getting x heads in a random throw of 10 coins is :

$$p(x) = \binom{10}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x} = \binom{10}{x} \left(\frac{1}{2}\right)^{10}; \; x = 0, 1, 2, \ldots, 10$$

$\therefore$ Probability of getting at least seven heads is given by :

$$P(X \geq 7) = p(7) + p(8) + p(9) + p(10)$$

$$= \left(\frac{1}{2}\right)^{10} \left\{ \binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10} \right\} = \frac{120 + 45 + 10 + 1}{1024} = \frac{176}{1024}.$$

Example 8·2. *A and B play a game in which their chances of winning are in the ratio 3 : 2. Find A's chance of winning at least three games out of the five games played.*

Solution. Let p be the probability that 'A' wins the game. Then we are given :

$$n = 5, \; p = \frac{3}{5} \quad \Rightarrow \quad q = 1 - p = \frac{2}{5}.$$

Hence, by binomial probability law, the probability that out of 5 games played, A wins 'x' games is given by :

$$P(X = x) = p(x) = \binom{5}{x}\left(\frac{3}{5}\right)^x\left(\frac{2}{5}\right)^{5-x} \quad ; x = 0, 1, 2, \ldots, 5$$

The requird probability that '*A*' wins at least three games is given by :

$$P(X \geq 3) = \sum_{r=3}^{5}\binom{5}{r}\frac{3^r \cdot 2^{5-r}}{5^5} = \frac{3^3}{5^5}\left[\binom{5}{3}2^2 + \binom{5}{4}\cdot 3 \times 2 + 1.3^2 \times 1\right]$$

$$= \frac{27 \times (40 + 30 + 9)}{3125} = 0{\cdot}68$$

Example 8·3. *A coffee connoisseur claims that he can distinguish between a cup of instant coffee and a cup of percolator coffee 75% of the time. It is agreed that his claim will be accepted if he correctly identifies at least 5 of the 6 cups. Find his chances of having the claim (i) accepted, (ii) rejected, when he does have the ability he claims.*

Solution. If p denotes the probability of a correct distinction between a cup of instant coffee and a cup of percolator coffee, then we are given :

$$p = \frac{75}{100} = \frac{3}{4} \quad \Rightarrow \quad q = 1 - p = \frac{1}{4}, \quad \text{and} \quad n = 6$$

If the random variable X denotes the number of correct distinctions, then by the Binomial probability law, the probability of x correct identifications out of 6 cups is given by :

$$P(X = x) = p(x) = \binom{6}{x}\left(\frac{3}{4}\right)^x\left(\frac{1}{4}\right)^{6-x} \quad ; x = 0, 1, 2, \ldots, 6$$

(*i*) The probability of the claim being accepted is :

$$P(X \geq 5) = p(5) + p(6) = \binom{6}{5}\left(\frac{3}{4}\right)^5\left(\frac{1}{4}\right)^{6-5} + \binom{6}{6}\left(\frac{3}{4}\right)^6 = \frac{1458}{4096} + \frac{729}{4096} = 0{\cdot}534.$$

(*ii*) The probability of the claim being rejected is :

$$P(X \leq 4) = 1 - P(X \geq 5) = 1 - 0{\cdot}534 = 0{\cdot}466.$$

Example 8·4. *A multiple-choice test consists of 8 questions with 3 answers to each question (of which only one is correct). A student answers each question by rolling a balanced die and checking the first answer if he gets 1 or 2, the second answer if he gets 3 or 4 and the third answer if he gets 5 or 6. To get a distinction, the student must secure at least 75% correct answers. If there is no negative marking, what is the probability that the student secures a distinction ?*

Solution. Since there are three answers to each question, out of which only one is correct, the probability of getting an answer to a question correctly is given by :

$$p = \frac{1}{3}, \quad \text{so that} \quad q = 1 - p = \frac{2}{3}$$

By Binomial probability law, the probability of getting r correct answers in a 8-question test is given by :

$$P(X = x) = p(x) = \binom{8}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{8-x} \quad ; x = 0, 1, 2, \ldots, 8$$

Hence, the required probability of securing a distinction (*i.e.*, of getting correct answers to at least 6 out of the 8 questions) is given by :

$$p(6) + p(7) + p(8) = \binom{8}{6}\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^{8-6} + \binom{8}{7}\left(\frac{1}{3}\right)^7\left(\frac{2}{3}\right)^{8-7} + \binom{8}{8}\left(\frac{1}{3}\right)^8\left(\frac{2}{3}\right)^{8-8}$$

$$= \frac{1}{3^6}\left[28 \times \frac{4}{9} + 8 \times \frac{1}{3} \times \frac{2}{3} + \frac{1}{9}\right] = \frac{129}{729 \times 9} = 0{\cdot}0197.$$

Example 8·5. *An irregular six-faced die is thrown and the expectation that in 10 throws it will give five even numbers is twice the expectation that it will give four even numbers. How many items in 10,000 sets of 10 throws each, would you expect it to give no even number.*

Solution. Let p be the probability of getting an even number in a throw of a die. Then the probability of getting x even numbers in ten throws of a die is given by :

$$P(X = x) = \binom{10}{x} p^x q^{10-x} \quad ; x, 0, 1, 2, \ldots, 10 \qquad \ldots (*)$$

We are given that : $P(X = 5) = 2P(X = 4) \Rightarrow \binom{10}{5} p^5 q^5 = 2\binom{10}{4} p^4 q^6$

$$\Rightarrow \frac{10!\,p}{5!\,5!} = 2\frac{10!\,q}{4!\,6!} \Rightarrow \frac{p}{5} = \frac{2q}{6} = \frac{q}{3} \Rightarrow 3p = 5(1-p) \Rightarrow p = \frac{5}{8} \quad \text{and} \quad q = \frac{3}{8}.$$

Thus $$P(X = x) = \binom{10}{x}\left(\frac{5}{8}\right)^x\left(\frac{3}{8}\right)^{10-x} \qquad \ldots [\text{From } (*)]$$

Hence the required number of times that in 10,000 sets of 10 throws each, we get no even number $= 10{,}000 \times P(X = 0) = 10{,}000 \times \left(\frac{3}{8}\right)^{10} = 1$ (approx.).

Example 8·6. *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 1/11 of needing adjustment during the day, and 7 are new, having corresponding probabilities of 1/21.*

Assuming that no machine needs adjustment twice on the same day, determine the probabilities that on a particular day

(*i*) *just 2 old and no new machines need adjustment.*

(*ii*) *If just 2 machines need adjustment, they are of the same type.*

Solution. Let p_1 = Probability that an old machine needs adjustment $= \frac{1}{11} \Rightarrow q_1 = \frac{10}{11}$.

and p_2 = Probability that a new machine needs adjustment $= \frac{1}{21} \Rightarrow q_2 = \frac{20}{21}$.

Then $P_1(x)$ = Probability that 'x' old machines need adjustment

$$= \binom{3}{x} p_1^x q_1^{3-x} = \binom{3}{x}\left(\frac{1}{11}\right)^x\left(\frac{10}{11}\right)^{3-x} \quad ; x = 0, 1, 2, 3$$

and $P_2(x)$ = Probability that 'x' new machines need adjustment

$$= \binom{7}{x} p_2^x q_2^{7-x} = \binom{7}{x}\left(\frac{1}{21}\right)^x\left(\frac{20}{21}\right)^{7-x} \quad ; x = 0, 1, 2, \ldots, 7$$

(*i*) The probability that just two old machines and no new machine need adjustment is given (by the compound probability theorem) by the expression :

$$P_1(2) \cdot P_2(0) = \binom{3}{2}\left(\frac{1}{11}\right)^2\left(\frac{10}{11}\right)\left(\frac{20}{21}\right)^7 = 0{\cdot}016 \qquad \ldots (1)$$

(*ii*) Similarly, the probability that just 2 new machines and no old machine need adjustment is :

$$P_1(0).\ P_2(2) = \left(\frac{10}{11}\right)^3 \times \binom{7}{2}\left(\frac{1}{21}\right)^2\left(\frac{20}{21}\right)^5 = 0{\cdot}028 \qquad \ldots (2)$$

The probability that 'if just two machines need adjustment, they are of the same type' is the same as the probability that 'either just 2 old and no new or just 2 new and no old machines need adjustment'.

$\therefore$ Required probability = (1) + (2) = 0·016 + 0·028 = 0·044.

Example 8·7. *The probability of a man hitting a target is $\frac{1}{4}$:*

(i) If he fires 7 times what is the probability of his hitting the target at least twice ?

(ii) How many times must he fire so that the probability of his hitting the target at least once is greater than $\frac{2}{3}$?

Solution. p = Probability of the man hitting the target $= \frac{1}{4} \Rightarrow q = 1 - p = \frac{3}{4}$.

$p(x)$ = Probability of getting x hits in 7 shots $= \binom{7}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{7-x}$; $x = 0, 1, \ldots, 7$

(*i*) Probability of at least two hits

$$= 1 - \{p(0) + p(1)\} = 1 - \left\{\binom{7}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^{7-0} + \binom{7}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^{7-1}\right\} = \frac{4547}{8192}.$$

(*ii*) Probability of at least one hit in n shots $= 1 - p(0) = 1 - \left(\frac{3}{4}\right)^x$.

It is required to find n, so that $\quad 1 - \left(\frac{3}{4}\right)^n > \frac{2}{3} \quad \Rightarrow \quad \frac{1}{3} > \left(\frac{3}{4}\right)^n$

Taking logarithms of each side, $\log \frac{1}{3} > n \log \frac{3}{4} \Rightarrow \log 1 - \log 3 > n\,(\log 3 - \log 4)$

$$\Rightarrow \quad 0 - 0{\cdot}4771 > n\,(0{\cdot}4771 - 0{\cdot}6021) \quad \Rightarrow \quad 0{\cdot}4771 < 0{\cdot}1250\,n$$

$$\therefore \quad n > \frac{0{\cdot}4771}{0{\cdot}1250} = 3{\cdot}8$$

Since n cannot be fractional, the required number of shots is 4.

Example 8·8. *In a precision bombing attack there is a 50% chance that any one bomb will strike the target. Two direct hits are required to destroy the target completely. How many bombs must be dropped to give a 99% chance or better of completely destroying the target ?*

Solution. We have : p = Probability that the bomb strikes the target $= 50\% = \frac{1}{2}$.

Let n be the number of bombs which should be dropped to ensure 99% chance or better of completely destroying the target. This implies that 'probability that out of n bombs, at least two strike the target, is greater than 0 ·99'.

Let X be a *r.v.* representing the number of bombs striking the target. Then

$$X \sim B\left(n, p = \frac{1}{2}\right) \text{ with } P(X = x) = p(x) = \binom{n}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{n-x} = \binom{n}{x}\left(\frac{1}{2}\right)^n ; x = 0, 1, 2, \ldots$$

We should have : $\quad P(X \geq 2) \geq 0{\cdot}99 \quad \Rightarrow \quad [1 - p(X \leq 1)] \geq 0{\cdot}099$

$$\Rightarrow \quad [1 - \{p\,(0) + p(1)\}] \geq 0{\cdot}99 \quad \Rightarrow \quad 1 - \left\{\binom{n}{0} + \binom{n}{1}\right\}\left(\frac{1}{2}\right)^n \geq 0{\cdot}99$$

$$\Rightarrow \quad 0{\cdot}01 \geq \frac{1+n}{2^n} \quad \Rightarrow \quad 2^n \times (0{\cdot}01) \geq 1 + n \quad \Rightarrow \quad 2^n \geq 100 + 100n \qquad \ldots (*)$$

By trial method, we find that the inequality (*) is satisfied by $n = 11$. Hence the minimum number of bombs needed to destroy the target completely is 11.

Example 8·9. *In a binomial distribution consisting of 5 independent trials, probabilities of 1 and 2 successes are 0·4096 and 0·2048 respectively. Find the parameter 'p' of the distribution.*

Solution. Let $X \sim B(n, p)$. In usual notations, we are given :

$$n = 5,\ p(1) = 0{\cdot}4096 \quad \text{and} \quad p(2) = 0{\cdot}2048.$$

According to Binomial probability law :

$$P(X = x) = p(x) = \binom{5}{x} p^x (1-p)^{5-x},\ x = 0, 1, 2, \ldots, 5$$

Now $p(1) = \binom{5}{1} p\,(1-p)^4 = 0{\cdot}4096$ ... (*) and $p(2) = \binom{5}{2} p^2 (1-p)^3 = 0{\cdot}2048$ ... (**)

Dividing (*) by (**), we get

$$\frac{\binom{5}{1} p(1-p)^4}{\binom{5}{2} p^2 (1-p)^3} = \frac{0{\cdot}4096}{0{\cdot}2048} \quad \Rightarrow \quad \frac{5\,(1-p)}{10p} = 2 \quad \Rightarrow \quad p = \frac{1}{5} = 0{\cdot}2.$$

Example 8·10. *With the usual notations, find p for a binomial variate X, if n = 6 and 9 P(X = 4) = P(X = 2).*

Solution. For the binomial random variable X with parameters $n = 6$ and p, the probability function is :

$$P(X = r) = \binom{6}{r} p^r q^{6-r};\ r = 0, 1, 2, \ldots, 6$$

We are given : $9P(X = 4) = P(X = 2) \quad \Rightarrow \quad 9 \times \binom{6}{4} p^4 q^2 = \binom{6}{2} p^2 q^4$

$$\Rightarrow \quad 9p^2 = q^2 \quad \Rightarrow \quad 9p^2 = (1-p)^2 = 1 + p^2 - 2p$$

$$\Rightarrow \quad 8p^2 + 2p - 1 = 0 \quad \Rightarrow \quad p = \frac{-2 \pm \sqrt{4 + 32}}{2 \times 8} = \frac{-2 \pm 6}{16} = -\frac{1}{2}, \frac{1}{4}$$

Since probability cannot be negative, $p = -\frac{1}{2}$ is rejected. Hence $p = \frac{1}{4}$.

8·4·1. Moments of Binomial Distribution. The first four moments about origin of binomial distribution are obtained as follows :

$$\mu_1' = E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} = np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} q^{n-x} = np\,(q + p)^{n-1} = np$$

$$\left[\because \binom{n}{x} = \frac{n}{x}\binom{n-1}{x-1} = \frac{n}{x} \cdot \frac{n-1}{x-1} \cdot \binom{n-2}{x-2} = \frac{n}{x} \cdot \frac{n-1}{x-1} \cdot \frac{n-2}{x-2} \binom{n-3}{x-3}, \text{ and so on}\right]$$

Thus the mean of the binomial distribution is np.

$$\mu_2' = E(X^2) = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} \{x(x-1) + x\} \frac{n(n-1)}{x(x-1)} \cdot \binom{n-2}{x-2} p^x q^{n-x}$$

$$= n(n-1)p^2 \left\{ \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} q^{n-x} \right\} + np$$

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

$$\mu_3' = E(X^3) = \sum_{x=0}^{n} x^3 p(x) = \sum_{x=0}^{n} \{x\,(x-1)\,(x-2) + 3x\,(x-1) + x\} \binom{n}{x} p^x q^{n-x}$$

$$= n\,(n-1)\,(n-2)\, p^3 \sum_{x=3}^{n} \binom{n-3}{x-3} p^{x-3} q^{n-x}$$

$$+ 3n\,(n-1)\, p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} q^{n-x} + np$$

$$= n\,(n-1)\,(n-2)\, p^3 (q+p)^{n-3} + 3n\,(n-1)\, p^2 (q+p)^{n-2} + np$$

$$= n\,(n-1)\,(n-2)\, p^3 + 3n\,(n-1)\, p^2 + np$$

Similarly

$$x^4 = x\,(x-1)\,(x-2)\,(x-3) + 6x\,(x-1)\,(x-2) + 7x\,(x-1) + x$$

Let $x^4 = Ax\,(x-1)\,(x-2)\,(x-3) + Bx\,(x-1)\,(x-2) + Cx\,(x-1) + x$

(By giving to x the values 1, 2 and 3, we find the values of arbitrary constants A, B and C.)

$$\therefore \ \mu_4' = E(X^4) = \sum_{x=0}^{n} x^4 \binom{n}{x} p^x q^{n-x}$$

$$= n\,(n-1)\,(n-2)\,(n-3)\, p^4 + 6n(n-1)\,(n-2)\, p^3 + 7n\,(n-1)\, p^2 + np$$

[On simplification]

Central Moments of Binomial Distribution :

$$\mu_2 = \mu_2' - \mu_1'^2 = n^2p^2 - np^2 + np - n^2p^2 = np\,(1-p) = npq$$

$$\mu_3 = \mu_3' - 3\mu_2'\,\mu_1' + 2\mu_1'^3$$

$$= \{n\,(n-1)\,(n-2)\, p^3 + 3n\,(n-1)\, p^2 + np\} - 3\,\{n\,(n-1)\, p^2 + np\}\, np + 2\,(np)^3$$

$$= np\,(-3np^2 + 3np + 2p^2 - 3p + 1 - 3npq)$$

$$= np\,\{3np\,(1-p) + 2p^2 - 3p + 1 - 3npq\}$$

$$= np\,(2p^2 - 3p + 1) = np\,(2p^2 - 2p + q) = npq\,(1-2p)$$

$$= npq\,\{q + p - 2p\} = npq\,(q-p)$$

$$\mu_4 = \mu_4' - 4\mu_3'\,\mu_1' + 6\mu_2'\,\mu_1'^2 - 3\mu_1'^4 = npq\,\{1 + 3\,(n-2)\,pq\} \qquad \text{[On simplification]}$$

Hence

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{n^2 p^2 q^2 (q-p)^2}{n^3 p^3 q^3} = \frac{(q-p)^2}{npq} = \frac{(1-2p)^2}{npq} \qquad \ldots (8{\cdot}5)$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{npq\,\{1 + 3\,(n-2)\,pq\}}{n^2 p^2 q^2} = \frac{1 + 3\,(n-2)\,pq}{npq} = 3 + \frac{1-6pq}{npq} \qquad \ldots (8{\cdot}6)$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{q-p}{\sqrt{npq}} = \frac{1-2p}{\sqrt{npq}}, \quad \gamma_2 = \beta_2 - 3 = \frac{1-6pq}{npq} \qquad \ldots (8{\cdot}6a)$$

Remarks 1. If $X \sim B(n, p)$, then mean $= np$ and variance $= npq$

$$\mu_3 = npq\,(q-p) \quad \text{and} \quad \mu_4 = npq\,[1 + 3\,(n-2)\,pq] \qquad \ldots (8{\cdot}6b)$$

2. Variance $= npq < np =$ Mean $\qquad (\because 0 < q < 1) \qquad \ldots (8{\cdot}6c)$

Hence, for the binomial distribution, variance is less than mean

Example 8·11. *Comment on the following :*

The mean of a binomial distribution is 3 and variance is 4.

Solution. If the given binomial distribution has parameters n and p, then we are given : Mean $= np = 3$ ... (*) and Variance $= npq = 4$ ... (**)

Dividing (**) by (*), $q = \frac{4}{3}$, which is impossible, since probability cannot exceed unity. Hence the given statement is wrong.

Aliter. Since for a binomial distribution variance is always less than mean, the given statement is wrong.

Example 8·12. *The mean and variance of binomial distribution are 4 and $\frac{4}{3}$ respectively. Find $P(X \geq 1)$.*

Solution.

Let $X \sim B(n, p)$. Then we are given : Mean $= np = 4$ ... (*) and $\text{Var}(X) = npq = \frac{4}{3}$.

Dividing, we get $q = \frac{1}{3} \Rightarrow p = \frac{2}{3}$. Substituting in (*), we obtain $n = \frac{4}{p} = \frac{4 \times 3}{2} = 6$.

$$\therefore \quad P(X \geq 1) = 1 - P(X = 0) = 1 - q^n = 1 - \left(\frac{1}{3}\right)^6 = 1 - \frac{1}{729} = 0{\cdot}99863.$$

Example 8·13. *If $X \sim B(n, p)$, show that :*

$$E\left(\frac{X}{n} - p\right)^2 = \frac{pq}{n}; \qquad Cov\left(\frac{X}{n}, \frac{n-X}{n}\right) = -\frac{pq}{n}$$

Solution. Since $X \sim B(n, p)$, $E(X) = np$ and $\text{Var}(X) = npq$

$$\therefore \quad E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = p; \qquad \text{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2} \cdot \text{Var}(X) = \frac{pq}{n} \qquad \ldots (*)$$

(*i*) $$E\left(\frac{X}{n} - p\right)^2 = E\left[\frac{X}{n} - E\left(\frac{X}{n}\right)\right]^2 = \text{Var}\left(\frac{X}{n}\right) = \frac{pq}{n} \qquad \text{[From (*)]}$$

(*ii*) $$\text{Cov}\left(\frac{X}{n}, \frac{n-X}{n}\right) = E\left[\left\{\frac{X}{n} - E\left(\frac{X}{n}\right)\right\}\left\{\frac{n-X}{n} - E\left(\frac{n-X}{n}\right)\right\}\right]$$

$$= E\left[\left(\frac{X}{n} - p\right)\left\{\left(1 - \frac{X}{n}\right) - (1-p)\right\}\right] = E\left[\left(\frac{X}{n} - p\right)\left\{-\left(\frac{X}{n} - p\right)\right\}\right]$$

$$= -E\left(\frac{X}{n} - p\right)^2 = -\text{Var}\left(\frac{X}{n}\right) = -\frac{pq}{n}.$$

8·4·5. Mode of Binomial Distribution. We have

$$\frac{p(x)}{p(x-1)} = \binom{n}{x} p^x q^{n-x} \Big/ \binom{n}{x-1} p^{x-1} q^{n-x+1}$$

$$= \frac{n!}{(n-x)!\,x!} p^x q^{n-x} \Big/ \frac{n!}{(x-1)!\,(n-x+1)!} p^{x-1} q^{n-x+1}$$

$$= \frac{(n-x+1)\,p}{xq} = \frac{xq + (n-x+1)\,p - xq}{xq} = 1 + \frac{(n+1)\,p - x\,(p+q)}{xq}$$

$$= 1 + \frac{(n+1)\,p - x}{xq} \qquad \ldots (8{\cdot}10)$$

Mode is the value of x for which $p(x)$ is maximum.

We discuss the following two cases :

Case I. *When $(n+1)\,p$ is not an integer.* Let $(n+1)\,p = m + f$, where m is an integer and f is fractional such that $0 < f < 1$. Substituting in (8·10), we get

$$\frac{p(x)}{p(x-1)} = 1 + \frac{(m+f)-x}{xq} \qquad \ldots (*)$$

From (*), it is obvious that

$$\frac{p(x)}{p(x-1)} > 1 \text{ for } x = 0, 1, 2, \ldots m \quad \text{and} \quad \frac{p(x)}{p(x-1)} < 1 \text{ for } x = m+1, m+2, \ldots, n$$

$$\Rightarrow \quad \frac{p(1)}{p(0)} > 1, \frac{p(2)}{p(1)} > 1, \ldots, \frac{p(m)}{p(m-1)} > 1, \text{ and } \frac{p(m+1)}{p(m)} < 1, \frac{p(m+2)}{p(m+1)} < 1, \ldots, \frac{p(n)}{p(n-1)} < 1$$

$$\therefore p(0) < p(1) < p(2) < \ldots < p(m-1) < p(m) > p(m+1) > p(m+2) > \ldots > p(n).$$

$$\Rightarrow \qquad p(x) \text{ is maximum at } x = m.$$

Thus, in this case there exists unique modal value for binomial distribution and it is m, the integral part of $(n+1)p$.

Case II. *When $(n+1)\,p$ is an integer.* Let $(n+1)\,p = m$ *(an integer)*.

Substituting in (8·10), we get $\dfrac{p(x)}{p(x-1)} = 1 + \dfrac{m-x}{xq}$ ... (**)

From (**), it is obvious that :

$$\frac{p(x)}{p(x-1)} \begin{cases} > 1 \text{ for } x = 1, 2, \ldots, m-1 \\ = 1 \text{ for } x = m \\ < 1 \text{ for } x = m+1, m+2, \ldots, n \end{cases}$$

Now proceeding as in case 1, we have

$$p(0) < p(1) < \ldots < p(m-1) = p(m) > p(m+1) > p(m+2) > \ldots > p(n)$$

Thus, in this case the binomial distribution is *bimodal* and the two modal values are m and $m-1$.

Example 8·15. *Determine the binomial distribution for which the mean is 4 and variance 3 and find its mode.*

Solution. Let $X \sim B(n, p)$, then we are given that

$E(X) = np = 4$ ... (*) and $\text{Var}(X) = npq = 3$ ... (**)

Dividing (**) by (*), we get $q = \frac{3}{4} \Rightarrow p = 1 - q = \frac{1}{4}$

Hence from (*), we obtain $n = \frac{4}{p} = 16$

Thus the given binomial distribution has parameters $n = 16$ and $p = \frac{1}{4}$.

Mode. We have $(n+1)p = 4{\cdot}25$, which is not an integer. Hence the unique mode of the binomial distribution is 4, the integral part of $(n+1)\,p$.

Example 8·16. *Show that for $p = 0{\cdot}5$, the binomial distribution has a maximum probability at $X = \frac{1}{2}n$, if n is even, and at $X = \frac{1}{2}(n-1)$ as well as $X = \frac{1}{2}(n+1)$, if n is odd.*

Solution. Here we have to find the mode of the binomial distribution.

(*i*) Let n be even $= 2m$, (say), $m = 1, 2, \ldots$

$\therefore$ If $p = 0{\cdot}5$, then $(n+1)\,p = (2m+1)\times\frac{1}{2} = m + 0{\cdot}5$. Hence in this case, the distribution is unimodal, the unique mode being at $X = m = \frac{n}{2}$.

(*ii*) Let n be odd $= (2m+1)$, say. Then

$$(n+1)\,p = (2m+2)\times\tfrac{1}{2} = m+1 \text{ (Integer)} = \frac{n-1}{2} + 1 = \frac{n+1}{2}$$

Since $(n+1)\,p$ is an integer, the distribution is bimodal, the two modes being $\frac{1}{2}(n+1)$ and $\frac{1}{2}(n+1) - 1 = \frac{1}{2}(n-1)$.

8·4·6. Moment Generating Function of Binomial Distribution. Let $X \sim B(n, p)$, then :

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx}\binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x q^{n-x} = (q+pe^t)^n \qquad \ldots (8{\cdot}11)$$

***m.g.f.* about Mean of Binomial Distribution :**

$$E\{e^{t(X-np)}\} = e^{-tnp}\,.\,E(e^{tX}) = e^{-tnp}\,.\,M_X(t) = e^{-tnp}(q+pe^t)^n = (qe^{-pt} + pe^{tq})^n$$

$$= \left[q\left\{1 - pt + \frac{p^2t^2}{2!} - \frac{p^3t^3}{3!} + \frac{p^4t^4}{4!} - \ldots\right\} + p\left\{1 + tq + \frac{t^2q^2}{2!} + \frac{t^3q^3}{3!} - \ldots\right\}\right]^n$$

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

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

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

Now $\quad \mu_2 = \text{Coefficient of } \frac{t^2}{2!} = npq, \qquad \mu_3 = \text{Coefficient of } \frac{t^3}{3!} = npq\,(q-p)$

$$\mu_4 = \text{Coefficient of } \frac{t^4}{4!} = npq\,(1-3pq) + 3n\,(n-1)\,p^2q^2$$
$$= 3n^2p^2q^2 + npq\,(1-6pq).$$

Example 8·17. *X is binomially distributed with parameters n and p. What is distribution of $Y = n - X$?*

Solution. $X \sim B(n, p)$, represents the number of successes in n independent trials with constant probability p of success for each trial.

$\therefore$ $Y = n - X$, represents the number of failures in n independent trials with constant probability 'q' of failure of each trial. Hence $Y = (n - X) \sim B(n, q)$

Aliter. Since $X \sim B(n, p)$, $\quad M_X(t) = E(e^{tX}) = (q + pe^t)^n$

$$\therefore \quad M_Y(t) = E(e^{tY}) = E[e^{t(n-X)}] = e^{nt}.\ E(e^{-tX}) = e^{nt}\ M_X(-t)$$

$$= e^{nt}.\ (q + pe^{-t})^n = \left[\ e^t\ (q + pe^{-t})\right]^n = (p + qe^t)^n$$

Hence, by uniqueness theorem of *m.g.f.*, $Y = (n - X) \sim B(n, q)$.

Example 8·18. *The m.g.f. of a r.v. X is* $\left(\frac{2}{3} + \frac{1}{3}e^t\right)^9$. *Show that*

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = \sum_{x=1}^{5} \binom{9}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{9-x}$$

Solution. Since $M_X(t) = \left(\frac{2}{3} + \frac{1}{3}e^t\right)^9 = (q + pe^t)^n$, by uniqueness theorem of *m.g.f.*

$X \sim B\left(n = 9, p = \frac{1}{3}\right)$. Hence $E(X) = \mu_X = np = 3$; $\qquad \sigma_X^2 = npq = 9 \times \frac{1}{3} \times \frac{2}{3} = 2$

$$\mu \pm 2\sigma = 3 \pm 2 \times \sqrt{2} = 3 \pm 2 \times 1{\cdot}4 = (0{\cdot}2, 5{\cdot}8)$$

$$\therefore \quad P(\mu - 2\sigma < X < \mu + 2\sigma) = P(0{\cdot}2 < X < 5{\cdot}8) = P(1 \le X \le 5)$$

$$= \sum_{x=1}^{5} p(x) = \sum_{x=1}^{5} \binom{n}{x} p^x q^{n-x} = \sum_{x=1}^{5} \binom{9}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{9-x}$$

8·4·7. Additive Property of Binomial Distribution.

Let $X \sim B(n_1, p_1)$ and $Y \sim B(n_2, p_2)$ be independent random variables. Then

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

What is the distribution of $X + Y$?

We have $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$ $\qquad$ [$\because$ X and Y are independent]

$$= (q_1 + p_1 e^t)^{n_1}.\ (q_2 + p_2 e^t)^{n_2} \qquad \ldots (**)$$

Since (**) cannot be expressed in the form $(q + p\,e^t)^n$, from uniqueness theorem of *m.g.f's* it follows that $X + Y$ is not a binomial variate. Hence, *in general the sum of two independent binomial variates is not a binomial variate. In other words, binomial distribution does not possess the additive or reproductive property.*

However, if we take $p_1 = p_2 = p$, (say), then from (**) $\qquad M_{X+Y}(t) = (q + pe^t)^{n_1 + n_2}$ which is the *m.g.f.* of a binomial variate with parameters $(n_1 + n_2, p)$. Hence by uniqueness theorem of *m.g.f.'s* $X + Y \sim B(n_1 + n_2, p)$. *Thus the binomial distribution possesses the additive or reproductive property if* $p_1 = p_2$.

Generalisation. *If* X_i, $(i = 1, 2, \ldots, k)$ *are independent binomial variates with parameters* (n_i, p), $(i = 1, 2, \ldots k)$, *then their sum* $\sum_{i=1}^{k} X_i \sim B\left(\sum_{i=1}^{k} n_i, p\right)$.

The proof is left as an exercise to the reader.

Example 8·19. *If the independent random variables X, Y are binomially distributed respectively with* $n = 3, p = \frac{1}{3}$, *and* $n = 5, p = \frac{1}{3}$, *write down the probability that* $X + Y \ge 1$.

Solution. We are given : $X \sim B\,(3, \frac{1}{3})$ and $Y \sim B\,(5, \frac{1}{5})$.

Since X and Y are independent binomial random variables, with $p_1 = p_2 = \frac{1}{3}$, by the additive property of binomial distribution, we get

$$X + Y \sim B\,(3 + 5\,, \tfrac{1}{3}), \qquad i.e., \qquad X + Y \sim B\,(8\,, \tfrac{1}{3})$$

$$\therefore \qquad P\,(X + Y = r) = \binom{8}{r} \left(\frac{1}{3}\right)^r \left(\frac{2}{3}\right)^{8-r} \qquad \ldots (*)$$

Hence $\quad P\,(X + Y \geq 1) = 1 - P\,(X + Y < 1) = 1 - P\,(X + Y = 0) = 1 - \left(\frac{2}{3}\right)^8$. [From (*)]

8·5. POISSON DISTRIBUTION

Poisson distribution was discovered by the French mathematician and physicist Simeon Denis Poisson (1781-1840) who published it in 1837. Poisson distribution is a limiting case of the binomial distribution under the following conditions :

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

(*ii*) p, the constant probability of success for each trial is indefinitely small, *i.e.*, $p \to 0$.

(*iii*) $np = \lambda$, (say) is finite.

Thus $p = \lambda/n$, $q = 1 - \lambda/n$, where λ is a positive real number.

The probability of x successes in a series of n independent trials is :

$$b(x\,;n, p) = \binom{n}{x} p^x q^{n-x};\ x = 0, 1, 2, \ldots, n \qquad \ldots (*)$$

We want the limiting form of (*) under the above conditions. Hence

$$\lim_{n\to\infty} b(x; n, p) = \lim_{n\to\infty} \frac{n\,!}{x\,!(n-x)\,!}\left(\frac{\lambda}{n}\right)^x \cdot \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Using Stirling's approximation for $n\,!$ as $n \to \infty$, *viz.*,

$$\lim_{n\to\infty} n\,! \approx \sqrt{2\pi}\, e^{-n}\, n^{n+(1/2)}, \text{ we get}$$

$$\lim_{n\to\infty} b(x; n, p) = \lim_{n\to\infty}\left\{\frac{\sqrt{2\pi}\, e^{-n}\,.\, n^{n+(1/2)}}{x\,!\,\sqrt{2\pi}\, e^{-(n-x)}.\,(n-x)^{n-x+(1/2)}}\right\}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{e^x.\,x\,!}\cdot \lim_{n\to\infty}\frac{n^{n-x+(1/2)}}{(n-x)^{n-x+(1/2)}}\cdot\left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{e^x\, x\,!}\cdot\lim_{n\to\infty}\frac{\left(1-\frac{\lambda}{n}\right)^{n-x}}{\left(1-\frac{x}{n}\right)^{n-x+(1/2)}}$$

$$= \frac{\lambda^x}{e^x\, x\,!}\cdot\frac{\lim\limits_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n .\ \lim\limits_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x}}{\lim\limits_{n\to\infty}\left(1-\frac{x}{n}\right)^n\ \lim\limits_{n\to\infty}\left(1-\frac{x}{n}\right)^{-x+(1/2)}}$$

But we know that

$$\left.\begin{cases}\lim\limits_{n\to\infty}\left(1-\dfrac{\lambda}{n}\right)^n = e^{-\lambda}, \\ \lim\limits_{n\to\infty}\left(1-\dfrac{\lambda}{n}\right)^{\alpha} = 1,\ \alpha \text{ is not a function of } n\end{cases}\right\} \qquad \ldots (**)$$

$$\therefore \quad \lim_{n\to\infty} b(x; n, p) = \frac{\lambda^x}{e^x\, x\,!}\times\frac{e^{-\lambda}.\,1}{e^{-x}\,.\,1} = \frac{e^{-\lambda}.\,\lambda^x}{x\,!};\ x = 0, 1, 2, \ldots, \infty, \qquad \text{[Using (**)]}$$

which is the required probability function of the Poisson distribution. 'λ' is known as the *parameter* of Poisson distribution.

Aliter. Poisson distribution can also be derived without using Stirling's approximation as follows :

$$b(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} \left(\frac{p}{1-p}\right)^x (1-p)^n$$

$$= \frac{n(n-1)(n-2)\ldots(n-x+1)}{x!} \cdot \frac{\left(\frac{\lambda}{n}\right)^x}{\left(1-\frac{\lambda}{n}\right)^x} \left(1-\frac{\lambda}{n}\right)^n$$

$$= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{x-1}{n}\right)}{x!\left(1-\frac{\lambda}{n}\right)^x} \lambda^x \left(1-\frac{\lambda}{n}\right)^n$$

$$\therefore \quad \lim_{n\to\infty} b(x;n,p) = \frac{e^{-\lambda}\lambda^x}{x!}; x = 0, 1, 2, \ldots \qquad \text{[From (**)]}$$

Definition. *A random variable X is said to follow a Poisson distribution if it assumes only non-negative values and its probability mass function is given by :*

$$p(x, \lambda) = P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}; x = 0, 1, 2, \ldots ; \lambda > 0 \\ 0, \text{ otherwise} \end{cases} \qquad \ldots (8{\cdot}15)$$

Here λ is known as the parameter of the distribution. We shall use the notation $X \sim P(\lambda)$, to denote that X is a Poisson variate with parameter λ.

Remarks 1. It should be noted that $\sum_{x=0}^{\infty} P(X = x) = e^{-\lambda} \sum_{x=0}^{\infty} \lambda^x/x! = e^{-\lambda} e^{\lambda} = 1$

2. The corresponding distribution function is :

$$F(x) = P X \le x) = \sum_{r=0}^{x} p(r) = e^{-\lambda} \sum_{r=0}^{x} \lambda^r/r!; x = 0, 1, 2, \ldots$$

3. Poisson distribution occurs when there are events which do not occur as outcomes of a definite number of trials (unlike that in binomial distribution) of an experiment but which occur at random points of time and space wherein our interest lies only in the number of occurrences of the event, not in its non-occurrences.

4. Following are some instances where Poisson distribution may be successfully employed :

(*i*) Number of deaths from a disease (not in the form of an epidemic) such as heart attack or cancer or due to snake bite.

(*ii*) Number of suicides reported in a particular city.

(*iii*) The number of defective material in a packing manufactured by a good concern.

(*iv*) Number of faulty blades in a packet of 100.

(*v*) Number of air accidents in some unit of time.

(*vi*) Number of printing mistakes at each page of the book.

(*vii*) Number of telephone calls received at a particular telephone exchange in some unit of time or connections to wrong numbers in a telephone exchange.

(*viii*) Number of cars passing a crossing per minute during the busy hours of a day.

(*ix*) The number of fragments received by a surface area '*A*' from a fragment atom bomb.

(*x*) The emission of radioactive (alpha)particles.

8·5·1. The Poisson Process. The Poisson distribution may also be obtained independently (*i.e.*, without considering it as a limiting form of the binomial distribution) as follows :

Let X_t be the number of telephone calls received in time interval 't' on a telephone switch board. Consider the following experimental conditions :

(1) The probability of getting a call in small time interval $(t, t + dt)$ is $\lambda\, dt$, where λ is a positive constant and dt denotes a small increment in time 't'.

(2) The probability of getting more than one call in this time interval is very small, *i.e.*, is of the order of $(dt)^2$, *i.e.*, $0\,[(dt)^2]$ such that $\lim\limits_{dt \to 0} \dfrac{0\,(dt)^2}{dt} = 0.$

(3) The probability of any particular call in the time interval $(t, t + dt)$ is independent of the actual time t and also of all previous calls.

Under these conditions it can be shown that the probability of getting x calls in time 't', say, $P_X(t)$ is given by : $\quad P_X(t) = \dfrac{e^{-\lambda t}\,(\lambda t)^x}{x\,!}\ ; x = 0, 1, 2, \ldots \infty$

which is a Poisson distribution with parameter $\lambda\, t$.

Proof. Let $P_x(t) = P$ {of getting x *calls in a time interval of length* 't'}.

Also P {of at least one call during $(t, t + dt)$} $= \lambda\, dt + 0\,[(dt)^2]$

and P {of more than one call during $(t, t + dt)$} $= 0\,[(dt)^2]$.

The event of getting exactly x calls in time $t + dt$ can materialise in the following two mutually exclusive ways :

(*i*) x calls in $(0, t)$, and none during $(t, t + dt)$ and the probability of this event is:

$$P_x(t)\,[1 - \{(\lambda\, dt + 0\,(dt)^2\}].$$

(*ii*) exactly $(x - 1)$ calls during $(0, t)$ and one call in $(t, t + dt)$ and the probability of this event is $\quad P_{x-1}(t)\,(\lambda\, dt)$.

Hence by the addition theorem of probability, we get

$$\begin{aligned} P_x(t + dt) &= P_x(t)\,\{1 - \lambda\, dt - 0\,(dt)^2\} + P_{x-1}(t)\,\lambda\, dt. \\ &= P_x(t)\,(1 - \lambda\, dt) + P_{x-1}(t)\,\lambda\, dt - 0\,(dt)^2\,P_x(t) \qquad \ldots (1) \end{aligned}$$

$$\Rightarrow \quad \frac{P_x(t + dt) - P_x(t)}{dt} = -\lambda\, P_x(t) + \lambda P_{x-1}(t) - \frac{0\,(dt)^2}{dt}\,P_x(t)$$

Proceeding to the limit as $dt \to 0$, we get

$$\lim_{dt \to 0} \frac{P_x(t + dt) - P_x(t)}{dt} = -\lambda\, P_x(t) + \lambda P_{x-1}(t)$$

$$\Rightarrow \qquad P_x'(t) = -\lambda\, P_x(t) + \lambda\, P_{x-1}(t),\ x \geq 1 \qquad \ldots (2)$$

where (′) denotes differentiation *w.r.* to 't'.

For $x = 0$, $P_{x-1}(t) = P_{-1}(t) = P$ {(-1) calls in time 't'} $\equiv 0$

Hence from (1), we get $\qquad P_0(t + dt) = P_0(t)\,\{1 - \lambda\, dt\} - 0\,(dt)^2$

which on taking the limit $dt \to 0$, gives , $\quad P_0'(t) = -\lambda\, P_0(t) \quad \Rightarrow \quad \dfrac{P_0'(t)}{P_0(t)} = -\lambda.$

Integrating *w.r. to* 't', $\qquad \log P_0(t) = -\lambda t + C,$

where C is an arbitrary constant to be determined from the condition $P_0(0) = 1$.

Hence $\quad C = \log 1 = 0 \quad \therefore \ \log P_0(t) = -\lambda t \quad \Rightarrow \quad P_0(t) = e^{-\lambda t}$

Substituting this value of $P_0(t)$ in (2), we get, with $x = 1$,

$$P_1'(t) = -\lambda P_1(t) + \lambda e^{-\lambda t} \quad \Rightarrow \quad P_1'(t) + \lambda P_1(t) = \lambda e^{-\lambda t}$$

This is an ordinary linear differential equation whose integrating factor is $e^{\lambda t}$.

Hence its solution is : $$e^{\lambda t} P_1(t) = \lambda \int e^{\lambda t} e^{-\lambda t} dt + C_1 = \lambda t + C_1,$$

where C_1 is an arbitrary constant to be determined from $P_1(0) = 0$, which gives $C_1 = 0$.

$$\therefore \qquad P_1(t) = e^{-\lambda t} \lambda t$$

Again substituting this in (2) with $x = 2$, $\quad P'_2(t) + \lambda P_2(t) = \lambda e^{-\lambda t} \lambda t.$

Integrating factor of this equation is $e^{\lambda t}$ and its solution is :

$$P_2(t) e^{\lambda t} = \lambda^2 \int t e^{-\lambda t} e^{\lambda t} dt + C_2 = \frac{\lambda^2 t^2}{2} + C_2$$

where C_2 is an arbitrary constant to be determined from $P_2(0) = 0$, which gives $C_2 = 0$.

Hence $\quad P_2(t) = e^{-\lambda t} \dfrac{(\lambda t)^2}{2}$. $\quad$ Proceeding similarly step by step, we shall get

$$P_x(t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}; x = 0, 1, 2, \ldots, \infty \qquad \ldots (8 \cdot 15a)$$

which is the *p.m.f.* of Poisson distribution with parameter λt.

8·5·2. Moments of the Poisson Distribution.

$$\mu_1' = E(X) = \sum_{x=0}^{\infty} x\, p(x, \lambda) = \sum_{x=0}^{\infty} x. \frac{e^{-\lambda} \lambda^x}{x!} = \lambda e^{-\lambda} \left\{ \sum_{x=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \right\}$$

$$= \lambda e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots \right) = \lambda e^{-\lambda} . e^{\lambda} = \lambda$$

Hence *the mean of the Poisson distribution is* λ.

$$\mu_2' = E(X^2) = \sum_{x=0}^{\infty} x^2 p(x, \lambda) = \sum_{x=0}^{\infty} \{x(x-1) + x\} \frac{e^{-\lambda} \lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x}{x!} + \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \lambda^2 e^{-\lambda} \left[\sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} \right] + \lambda$$

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

$$\mu_3' = E(X^3) = \sum_{x=0}^{\infty} x^3 p(x, \lambda) = \sum_{x=0}^{\infty} \{x(x-1)(x-2) + 3x(x-1) + x\} \frac{e^{-\lambda} \lambda^x}{x!}$$

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

$$= e^{-\lambda} \lambda^3 \left\{ \sum_{x=3}^{\infty} \frac{\lambda^{x-3}}{(x-3)!} \right\} + 3e^{-\lambda} \lambda^2 \left\{ \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} \right\} + \lambda$$

$$= e^{-\lambda} \lambda^3 e^{\lambda} + 3e^{-\lambda} \lambda^2 e^{\lambda} + \lambda = \lambda^3 + 3\lambda^2 + \lambda$$

$$\mu_4' = E(X^4) = \sum_{x=0}^{\infty} x^4 . p(x, \lambda)$$

$$= \sum_{x=0}^{\infty} \{x(x-1)(x-2)(x-3) + 6x(x-1)(x-2) + 7x(x-1) + x\} \frac{e^{-\lambda} \lambda^x}{x!}$$

$$= e^{-\lambda} \lambda^4 \left\{ \sum_{x=4}^{\infty} \frac{\lambda^{x-4}}{(x-4)!} \right\} + 6e^{-\lambda} \lambda^3 \left\{ \sum_{x=3}^{\infty} \frac{\lambda^{x-3}}{(x-3)!} \right\} + 7e^{-\lambda} \lambda^2 \left\{ \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} \right\} + \lambda$$

$$= \lambda^4 (e^{-\lambda} e^{\lambda}) + 6\lambda^3 (e^{-\lambda} e^{\lambda}) + 7\lambda^2 (e^{-\lambda} e^{\lambda}) + \lambda$$

$$= \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$$

The four central moments are now obtained as follows :

$$\mu_2 = \mu_2' - \mu_1'^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda$$

Thus the mean and the variance of the Poisson distribution are each equal to λ.

$$\mu_3 = \mu_3' - 3\mu_2' \mu_1' + 2\mu_1'^3 = (\lambda^3 + 3\lambda^2 + \lambda) - 3\lambda(\lambda^2 + \lambda) + 2\lambda^3 = \lambda.$$

$$\mu_4 = \mu_4' - 4\mu_3' \mu_1' + 6\mu_2' \mu_1'^2 - 3\mu_1'^4$$

$$= (\lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda) - 4\lambda(\lambda^3 + 3\lambda^2 + \lambda) + 6\lambda^2(\lambda^2 + \lambda) - 3\lambda^4 = 3\lambda^2 + \lambda$$

Co-efficients of skewness and kurtosis are given by :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\lambda^2}{\lambda^3} = \frac{1}{\lambda} \quad \text{and} \quad \gamma_1 = \sqrt{\beta_1} = \frac{1}{\sqrt{\lambda}}$$

Also

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{\lambda} \quad \text{and} \quad \gamma_2 = \beta_2 - 3 = \frac{1}{\lambda} \qquad \ldots (8·16)$$

Hence the Poisson distribution is always a skewed distribution. Proceeding to the limit as $\lambda \to \infty$, $\beta_1 = 0$ and $\beta_2 = 3$.

8·5·3. Mode of the Poisson Distribution

$$\frac{p(x)}{p(x-1)} = \frac{\dfrac{e^{-\lambda} \lambda^x}{x!}}{\dfrac{e^{-\lambda} \lambda^{x-1}}{(x-1)!}} = \frac{\lambda}{x} \qquad \ldots (8·17)$$

We discuss the following cases :

Case I. *When* λ *is not an integer.* Let us suppose that S is the integral part of λ, so that $\lambda = S + f$, $0 < f < 1$. Hence from (8·17), we get :

$$\frac{p(x)}{p(x-1)} = \frac{S+f}{x} = \begin{cases} > 1 & , \text{if } x = 0, 1, \ldots, S \\ < 1 & , \text{if } x = S+1, S+2, \ldots \end{cases}$$

$$\frac{p(1)}{p(0)} > 1, \frac{p(2)}{p(1)} > 1, \ldots, \frac{p(S-1)}{p(S-2)} > 1, \frac{p(S)}{p(S-1)} > 1,$$

and

$$\frac{p(S+1)}{p(S)} < 1, \frac{p(S+2)}{p(S+1)} < 1, \ldots$$

Combining the above expressions into a single expression, we get

$$p(0) < p(1) < p(2) < \ldots < p(S-2) < p(S-1) < p(S) > p(S+1) > p(S+2) > \ldots,$$

which shows that $p(S)$ is the maximum value. Hence, in this case, the distribution is unimodal and the integral part of λ is the unique modal value.

Case II. *When $\lambda = k$ (say) is an integer.* Here, as in case I, we have

$$\frac{p(1)}{p(0)} > 1, \quad \frac{p(2)}{p(1)} > 1, \ldots, \frac{p(k-1)}{p(k-2)} > 1$$

and
$$\frac{p(k)}{p(k-1)} = 1, \quad \frac{p(k+1)}{pk} < 1, \frac{p(k+2)}{p(k+1)} < 1, \ldots$$

$$\therefore \quad p(0) < p(1) < p(2) < \ldots < p(k-2) < p(k-1) = p(k) > p(k+1) > p(k+2) \ldots$$

In this case we have two maximum values, *viz.*, $p(k-1)$ and $p(k)$ and thus the distribution is bimodal and two modes are at $(k-1)$ and k, *i.e.*, at $(\lambda - 1)$ and λ, (since $k = \lambda$).

8·5·4. Recurrence Relation for Moments of the Poisson Distribution. By *def.*,

$$\mu_r = E\{X - E(X)\}^r = \sum_{x=0}^{\infty} (x - \lambda)^r p(x, \lambda) = \sum_{x=0}^{\infty} (x-\lambda)^r \frac{e^{-\lambda}\lambda^x}{x!}$$

Differentiating *w.r.to* λ, we get

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

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

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

$$\therefore \quad = -r\,\mu_{r-1} + \frac{1}{\lambda}\mu_{r+1}$$

$$\Rightarrow \quad \mu_{r+1} = r\lambda\,\mu_{r-1} + \lambda\frac{d\mu_r}{d\lambda} \quad \ldots (8{\cdot}18)$$

Putting $r = 1, 2$ and 3 successively, we get

$$\mu_2 = \lambda\,\mu_0 + \lambda\frac{d\mu_1}{d\lambda} = \lambda, \quad \mu_3 = 2\lambda\,\mu_1 + \lambda\frac{d\mu_2}{d\lambda} = \lambda, \quad \mu_4 = 3\lambda\,\mu_2 + \lambda\frac{d\mu_3}{d\lambda} = 3\lambda^2 + \lambda.$$

8·5·5. Moment Generating Function of the Poisson Distribution.

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx}.\frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=0}^{\infty}\frac{e^{-\lambda}(\lambda e^t)^x}{x!}$$

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

$$= e^{\lambda(e^t - 1)} \quad \ldots (8{\cdot}19)$$

8·5·6. Characteristic Function of the Poisson Distribution.

$$\phi_X(t) = \sum_{x=0}^{\infty} e^{itx}.p(x,\lambda) = \sum_{x=0}^{\infty} e^{itx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^{it})^x}{x!} = e^{-\lambda}e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)} \quad \ldots (8{\cdot}20)$$

8·5·7. Cumulants of the Poisson Distribution.

$$K_X(t) = \log M_X(t) = \log[e^{\lambda(e^t-1)}] = \lambda(e^t - 1)$$

$$= \lambda\left[\left(1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \ldots + \frac{t^r}{r!} + \ldots\right) - 1\right] = \lambda\left[t + \frac{t^2}{2!} + \frac{t^3}{3!} + \ldots + \frac{t^r}{r!} + \ldots\right]$$

Hence, using Poisson probability law, the required probability of getting 6 heads r times is given by :

$$P(X = r) = \frac{e^{-\lambda} . \lambda^r}{r!} = \frac{e^{-100} . (100)^r}{r!} ; r = 0, 1, 2, \ldots$$

Example 8·37. *In a book of 520 pages, 390 typo-graphical errors occur. Assuming Poisson law for the number of errors per page, find the probability that a random sample of 5 pages will contain no error.*

Solution. The average number of typographical errors per page in the book is given by $\lambda = (390/520) = 0{\cdot}75$

Hence using Poisson probability law, the probability of x errors per page is given by :

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-0{\cdot}75} (0{\cdot}75)^x}{x!} ; x = 0, 1, 2, \ldots$$

The required probability that a random sample of 5 pages will contain no error is given by :

$$[P(X = 0)]^5 = (e^{-0{\cdot}75})^5 = e^{-3{\cdot}75}.$$

Example 8·38. *In a Poisson frequency distribution, frequency corresponding to 3 successes is 2/3 times frequency corresponding to 4 successes. Find the mean and standard deviation of the distribution.*

Solution. Let X be a random variable following Poisson distribution with parameter λ. Then the frequency function is given by :

$$f(x) = N . p(x) = NP(X = x) = N . \frac{e^{-\lambda} . \lambda^x}{x!} ; x = 0, 1, 2, \ldots \qquad \ldots (*)$$

Putting $x = 3$ and 4 in (*), $f(3) = N . \dfrac{e^{-\lambda} \lambda^3}{3!}$ and $f(4) = N . \dfrac{e^{-\lambda} \lambda^4}{4!}$

We are given : $f(3) = \frac{2}{3} f(4) \Rightarrow N . \dfrac{e^{-\lambda} \lambda^3}{3!} = \dfrac{2}{3} N . \dfrac{e^{-\lambda} \lambda^4}{4!}$

$$\Rightarrow \quad \frac{1}{3!} = \frac{2}{3} \frac{\lambda}{4!} \quad \Rightarrow \quad \lambda = \frac{1}{3!} \times \frac{3}{2} \times 4! = 6.$$

Mean of the Poisson distribution $= \lambda = 6$ and *s.d.* of the distribution $= \sqrt{\lambda} = \sqrt{6}$.

$$\Rightarrow \quad -p\,(200 \times 0{\cdot}4346) \geq -0{\cdot}0044 \quad \Rightarrow \quad p \leq \frac{0{\cdot}0044}{86{\cdot}92} = 0{\cdot}0000506.$$

(*ii*) The day's work of 20 letters of 200 words each is accepted, when there is no mistake in any of the $n = 20 \times 200 = 4{,}000$ words. Assuming Poisson distribution, probability of no mistake in the day's work $= e^{-\lambda}$, where $\lambda = np = 4{,}000p$. We want to find p such that :

$$\therefore \quad e^{-4000p} = 0{\cdot}90 \quad \Rightarrow \quad -4{,}000\,p\,(\log 2{\cdot}72) = \log 0{\cdot}90$$

$$\Rightarrow \quad -p\,(4{,}000 \times 0{\cdot}4346) = -\ 0{\cdot}0458 \quad \Rightarrow \quad p = \frac{0{\cdot}0458}{1738{\cdot}4} = 0{\cdot}0000263.$$

Example 8·40. *Suppose that the number of telephone calls coming into a telephone exchange between 10 A.M. and 11 A.M. say, X_1 is a random variable with Poisson distribution with parameter 2. Similarly the number of calls arriving between 11 A.M. and 12 noon, say, X_2 has a Poisson distribution with parameter 6. If X_1 and X_2 are independent, what is the probability that more than 5 calls come in-between 10 A.M. and 12 noon ?*

Solution. We are given : $X_1 \sim P(2)$ and $X_2 \sim P(6)$. Let $X = X_1 + X_2$. By the additive property of Poisson distribution, X is also a Poisson variate with parameter (say) $\lambda = 2 + 6 = 8$.

Hence the probability of x calls in-between 10 A.M. and 12 noon is given by :

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x\,!} = \frac{x^{-8}\,8^x}{x\,!};\ x = 0, 1, 2, \ldots$$

Probability that more than 5 calls come in-between 10 A.M. and 12 noon is :

$$P(X > 5) = 1 - P\,(X \leq 5) = 1 - \sum_{x=0}^{5} \frac{e^{-8}\,8^x}{x\,!} = 1 - 0{\cdot}1912 = 0{\cdot}8088.$$

Example 8·41. *A Poisson distribution has a double mode at $x = 1$ and $x = 2$. What is the probability that x will have one or the other of these two values ?*

Solution. We know that if the Poisson distribution is bimodal, then the two modes are at the points $x = \lambda - 1$ and $x = \lambda$, where λ is the parameter of the Poisson distribution. Therefore, since we are given that the two modes are at the points $x = 1$ and $x = 2$, we find that $\lambda = 2$.

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

$$\Rightarrow \quad P(X = 1) = e^{-2}\,2 \quad \text{and} \quad P(X = 2) = \frac{e^{-2}.2^2}{2\,!} = e^{-2}\,.\,2.$$

Required probability $= P\,(X = 1) + P\,(X = 2) = 2e^{-2} + 2e^{-2} = 0{\cdot}542$

Example 8·42. *If X is a Poisson variate such that*

$$P(X = 2) = 9P(X = 4) + 90\,P\,(X = 6) \qquad \ldots (*)$$

Find (i) λ, (ii) the mean of X, (iii) β_1, the coefficient of skewness.

Solution. If X is a Poisson variate with parameter λ, then

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

Hence (*) gives

$$\frac{e^{-\lambda}.\lambda^2}{2\,!} = e^{-\lambda}\left(9\,\frac{\lambda^4}{4\,!} + 90\,\frac{\lambda^6}{6\,!}\right) = \frac{e^{-\lambda}\lambda^2}{8}(3\lambda^2 + \lambda^4) \quad \Rightarrow \quad \lambda^4 + 3\lambda^2 - 4 = 0.$$

Solving as a quadratic in λ^2, $\lambda^2 = \dfrac{-3 \pm \sqrt{9+16}}{2} = \dfrac{-3 \pm 5}{2}$. Since $\lambda > 0$, $\lambda^2 = 1 \Rightarrow \lambda = 1$.

Hence, Mean $= \lambda = 1$, and β_1 = Coefficient of skewness $= \dfrac{1}{\lambda} = 1$.

Example 8·43. *If X and Y are independent Poisson variates such that*

$$P(X = 1) = P(X = 2) \quad \textit{and} \quad P(Y = 2) = P(Y = 3) \qquad \ldots (*)$$

Find the variance of $X - 2Y$.

Solution. Let $X \sim P(\lambda)$ and $Y \sim P(\mu)$. Then we have

$$P(X = x) = \frac{e^{-\lambda} . \lambda^x}{x!},\ x = 0, 1, 2, \ldots ; \lambda > 0 \quad \text{and} \quad P(Y = y) = \frac{e^{-\mu} . \mu^y}{y!},\ y = 0, 1, 2, \ldots ; \mu > 0$$

Using (*), $\quad \lambda e^{-\lambda} = \dfrac{\lambda^2 e^{-\lambda}}{2!} \quad$ and $\quad \dfrac{\mu^2 e^{-\mu}}{2} = \dfrac{\mu^3 e^{-\mu}}{3!} \qquad \ldots (**)$

Solving (**), we obtain

$$\lambda e^{-\lambda} (\lambda - 2) = 0 \quad \text{and} \quad \mu^2 e^{-\mu} (\mu - 3) = 0 \Rightarrow \lambda = 2 \text{ and } \mu = 3, \text{ since } \lambda > 0, \mu > 0.$$

Now $\quad \text{Var}(X) = \lambda = 2, \quad$ and $\quad \text{Var}(Y) = \mu = 3 \qquad \ldots (***)$

$\therefore \quad \text{Var}(X - 2Y) = 1^2\ \text{Var}(X) + (-2)^2 . \text{Var}\ Y,$

covariance term vanishes since X and Y are independent.

Hence, on using (***), we get $\text{Var}(X - 2Y) = 2 + 4 \times 3 = 14$.

Example 8·44. *If X and Y are independent Poisson variates with means* λ_1 *and* λ_2 *respectively, find the probability that (i)* $X + Y = k$, and *(ii)* $X = Y$.

Solution. We have : $X \sim P(\lambda_1)$ and $Y \sim P(\lambda_2)$. Hence

$$P(X = x) = \frac{e^{-\lambda_1} \lambda_1^x}{x!},\ x = 0, 1, 2, 3, \ldots ; \lambda_1 > 0 \quad \text{and} \quad P(Y = y) = \frac{e^{-\lambda_2} \lambda_2^y}{y!},\ y = 0, 1, 2, 3, \ldots ; \lambda_2 > 0$$

(i)
$$\begin{aligned}
P(X + Y = k) &= \sum_{r=0}^{k} P(X = r \cap Y = k - r) \\
&= \sum_{r=0}^{k} P(X = r)\, P(Y = k - r) \qquad [\because \text{X and Y are independent}] \\
&= \sum_{r=0}^{k} \frac{e^{-\lambda_1} \lambda_1^r}{r!} . \frac{e^{-\lambda_2} \lambda_2^{k-r}}{(k-r)!} = e^{-(\lambda_1+\lambda_2)} \sum_{r=0}^{k} \frac{\lambda_1^r \lambda_2^{k-r}}{r!\,(k-r)!} \\
&= e^{-(\lambda_1+\lambda_2)} \left[\frac{\lambda_2^k}{k!} + \frac{\lambda_1 \lambda_2^{k-1}}{1!\,(k-1)!} + \frac{\lambda_1^2 \lambda_2^{k-2}}{2!\,(k-2)!} + \ldots + \frac{\lambda_1^k}{k!} \right] \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \left[\lambda_2^k + {}^kC_1\, \lambda_2^{k-1} . \lambda_1 + {}^kC_2\, \lambda_2^{k-2} . \lambda_1^2 + \ldots + \lambda_1^k \right] \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \times (\lambda_1 + \lambda_2)^k ; k = 0, 1, 2, \ldots
\end{aligned}$$

which is the probability function of Poisson distribution with parameter $\lambda_1 + \lambda_2$.

Aliter. Since $X \sim P(\lambda_1)$ and $Y \sim P(\lambda_2)$ are independent, by the additive property of Poisson distribution $X + Y \sim P(\lambda_1 + \lambda_2)$. Hence

$$P(X+Y=k) = \frac{e^{-(\lambda_1+\lambda_2)} \times (\lambda_1+\lambda_2)^k}{k!}; k = 0, 1, 2, \ldots$$

(ii) $$P(X=Y) = \sum_{r=0}^{\infty} P(X=r \cap Y=r)$$

$$= \sum_{r=0}^{\infty} P(X=r)\, P(Y=r) = e^{-(\lambda_1+\lambda_2)} \sum_{r=0}^{\infty} \frac{(\lambda_1 \lambda_2)^r}{(r!)^2}$$

[$\because$ X and Y are independent]

Example 8·45. *Show that in Poisson distribution with unit mean, mean deviation about mean is (2/e) times the standard deviation.*

Solution. Let $X \sim P(\lambda)$. We are given : Mean $= \lambda = 1$.

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

Mean deviation about mean 1 is :

$$E(|X-1|) = \sum_{x=0}^{\infty} |x-1|\, P(X=x) = e^{-1} \sum_{x=0}^{\infty} \frac{|x-1|}{x!} = e^{-1}\left(1 + \frac{1}{2!} + \frac{2}{3!} + \ldots\right)$$

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

$\therefore$ Mean deviation about mean

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

$$= e^{-1}(1+1) = \frac{2}{e} \times 1 = \frac{2}{e} \times \text{standard deviation},$$

since for the Poisson distribution, variance = mean = 1 (given) $\Rightarrow$ *s.d.* = 1.

Example 8·46. *Let $X_1, X_2, \ldots, X_n$ be identically and independently distributed Bin (1, p) variates. Let $S_n = \sum_{j=1}^{n} X_j$, be a binomial (n, p) variate and $M_n(t)$ be the m.g.f. of S_n. Find*

$$\lim_{n\to\infty} M_n(t), \text{ using } np = \lambda \text{ (constant).}$$

Solution. Since X_i, $i = 1, 2, \ldots, n$ are *i.i.d.* binomial variates $B(1, p)$,

$$S_n = \sum_{j=1}^{n} X_j, \text{ is a binomial } B(n, p) \text{ variate.}$$

$$\therefore \qquad M_n(t) = m.g.f. \text{ of } S_n = (q + pe^t)^n = \left[1 + (e^t - 1)p\right]^n$$

If we take $np = \lambda \Rightarrow p = \lambda/n$ and let $n \to \infty$, we get

$$\lim_{n\to\infty} M_n(t) = \lim_{n\to\infty} \left[1 + \frac{(e^t-1)\lambda}{n}\right]^n = \exp[\lambda(e^t - 1)],$$

which is the *m.g.f.* of Poisson distribution with parameter λ.

Hence by uniqueness theorem of *m.g.f.*, $S_n = \sum_{j=1}^{n} X_j \to P(\lambda)$, as $n \to \infty$, with $np = \lambda$ (fixed).

9.1. INTRODUCTION

We consider some univariate continuous distributions in this chapter. The main continuous distributions like uniform distribution, normal distribution, gamma, beta, exponential, Laplace, Weibul, Logistic and Cauchy distributions will be discussed in detail in the subsequent sections.

9.2. NORMAL DISTRIBUTION

The normal distribution was first discovered in 1733 by English mathematician De-Moivre, who obtained this continuous distribution as a limiting case of the binomial distribution and applied it to problems arising in the game of chance. It was also known to Laplace, no later than 1774 but through a historical error it was credited to Gauss, who first made reference to it in the beginning of 19th century (1809), as the distribution of errors in Astronomy. Gauss used the normal curve to describe the theory of accidental errors of measurements involved in the calculation of orbits of heavenly bodies. Throughout the eighteenth and nineteenth centuries, various efforts were made to establish the normal model as the underlying law ruling all continuous random variables. Thus, the name *"normal"*. These efforts, however, failed because of false premises. The normal model has, nevertheless, become the most important probability model in statistical analysis.

Definition *A r.v. X is said to have a normal distribution with parameters μ (called 'mean') and σ^2 (called 'variance') if its p.d.f. is given by the probability law :*

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

or

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2};\ -\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma > 0 \qquad \dots (9.1)$$

Remarks 1. When a *r.v.* is normally distributed with mean μ and standard deviation σ, it is customary to write X is distributed as $N(\mu, \sigma^2)$ and is expressed by $X \sim N(\mu, \sigma^2)$.

2. If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma}$, is a standard normal variate with $E(Z) = 0$ and $Var(Z) = 1$ and we write $Z \sim N(0, 1)$.

3. The *p.d.f.* of standard normal variate Z is given by :

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2},\ -\infty < z < \infty$$

and the corresponding distribution function, denoted by $\Phi(z)$ is given by :

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \varphi(u)\,du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\,du$$

We shall prove below two important results on the distribution function $\Phi(.)$ of standard normal variate.

Result 1. $\quad \Phi(-z) = 1 - \Phi(z),\ z > 0$

Proof. $\quad \Phi(-z) = P(Z \le -z) = P(Z \ge z) = 1 - P(Z \le z) = 1 - \Phi(z)$

Result 2. $\quad P(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$, where $X \sim N(\mu, \sigma^2)$

Proof.
$$P(a \le X < b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right), \quad \left(Z = \frac{X-\mu}{\sigma}\right)$$
$$= P\left(Z \le \frac{b-\mu}{\sigma}\right) - P\left(Z \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

4. The graph of $f(x)$ is famous *'bell-shaped'* curve. The top of the bell is directly above the mean μ. For large values of σ, the curve tends to flatten out and for small values of σ, it has a sharp peak.

9·2·1. Normal Distribution as a Limiting form of Binomial Distribution.

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

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

(*ii*) neither p nor q is very small.

The *p.m.f.* of the binomial distribution with parameters n and p is given by :

$$p(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}\ ;\ x = 0, 1, 2, \ldots, n \qquad \ldots (*)$$

Let us now consider the standard binomial variate :

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{npq}}\ ;\ X = 0, 1, 2, \ldots, n \qquad \ldots (**)$$

When $X = 0$, $Z = \dfrac{-np}{\sqrt{npq}} = -\sqrt{\dfrac{np}{q}}$ and when $X = n$, $Z = \dfrac{n - np}{\sqrt{npq}} = \sqrt{\dfrac{nq}{p}}$

Thus in the limit as $n \to \infty$, Z takes the values from $-\infty$ to ∞. Hence the distribution of X will be a continuous distribution over the range $-\infty$ to ∞.

We want the limiting form of (*) under the above two conditions. Using Stirling's approximation to $r!$ for large r, *viz.*, $\lim\limits_{r\to\infty} r! \simeq \sqrt{2\pi}\, e^{-r}\, r^{r+(1/2)}$,

we have in the limit as $n \to \infty$ and consequently $x \to \infty$,

$$\lim p(x) = \lim \left[\frac{\sqrt{2\pi}\, e^{-n}\, n^{n+\frac{1}{2}}\, p^x\, q^{n-x}}{\sqrt{2\pi}\, e^{-x}\, x^{x+\frac{1}{2}}\, \sqrt{2\pi}\, e^{-(n-x)}\, (n-x)^{n-x+\frac{1}{2}}}\right]$$

$$= \lim \left[\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} \frac{(np)^{x+\frac{1}{2}}\, (nq)^{n-x+\frac{1}{2}}}{x^{x+\frac{1}{2}}\, (n-x)^{n-x+\frac{1}{2}}}\right]$$

$$= \lim \left[\frac{1}{\sqrt{2\pi}\,\sqrt{npq}} \left(\frac{np}{x}\right)^{x+\frac{1}{2}} \left(\frac{nq}{n-x}\right)^{n-x+\frac{1}{2}}\right] \qquad \ldots (***)$$

From (**), we get $\quad X = np + Z\sqrt{npq} \quad \Rightarrow \quad \dfrac{X}{np} = 1 + Z\sqrt{\dfrac{q}{np}}$

Further

$$n - X = n - np - Z\sqrt{npq} = nq - Z\sqrt{npq} \Rightarrow \frac{n-X}{nq} = 1 - Z\sqrt{\frac{p}{nq}}. \text{ Also } dz = \frac{1}{\sqrt{npq}}\, dx$$

Hence the probability differential of the distribution of Z, in the limit is :

$$dG(z) = g(z)\, dz = \lim_{n\to\infty} \left(\frac{1}{\sqrt{2\pi}} \times \frac{1}{N}\right) dz,$$

where
$$N = \left(\frac{x}{np}\right)^{x+\frac{1}{2}} \left(\frac{n-x}{nq}\right)^{n-x+\frac{1}{2}} \qquad \ldots (9\cdot2)$$

$$\log N = (x + \tfrac{1}{2}) \log (x/np) + (n - x + \tfrac{1}{2}) \log \{(n-x)/nq\}$$

$\Rightarrow$
$$= (np + z\sqrt{npq} + \tfrac{1}{2}) \log \left\{1 + z\sqrt{q/np}\right\} + (nq - z\sqrt{npq} + \tfrac{1}{2}) \log \left\{1 - z\sqrt{(p/nq)}\right\}$$

$$= (np + z\sqrt{npq} + \tfrac{1}{2}) \left\{z.\sqrt{(q/np)} - \tfrac{1}{2} z^2 (q/np) + \tfrac{1}{3} z^3 (q/np)^{3/2} - \ldots\right\}$$

$$+ (np - z\sqrt{npq} + \tfrac{1}{2}) \left\{-z\sqrt{(p/nq)} - \tfrac{1}{2} z^2 (p/nq) - \tfrac{1}{3} z^3 (p/nq)^{3/2} - \ldots\right\}$$

$$= \left[\left\{z\sqrt{npq} - \tfrac{1}{2} qz^2 + \tfrac{1}{3} z^3 \frac{q^{3/2}}{\sqrt{np}} + z^2 q - \tfrac{1}{2} z^3 \frac{q^{3/2}}{\sqrt{np}} + \frac{1}{2} z \sqrt{\frac{q}{np}} - \tfrac{1}{4} z^2 \frac{q}{np} + \ldots\right)\right.$$

$$\left. + \left\{-z\sqrt{npq} - \tfrac{1}{2} z^2 p - \frac{1}{3} z^3 \frac{p^{3/2}}{\sqrt{nq}} + z^2 p + \frac{1}{2} z^3 \frac{p^{3/2}}{\sqrt{np}} - \tfrac{1}{2} z\sqrt{p/nq} - \tfrac{1}{4} z^2 \frac{p}{np} + \ldots\right\}\right]$$

$$= \left[-\frac{1}{2} z^2 (p+q) + z^2 (p+q) + \frac{z}{2\sqrt{n}} \left(\frac{q}{p} + \frac{p}{q}\right) + 0\,(n^{-1/2})\right]$$

$$= \frac{z^2}{2} + 0\,(n^{-1/2}) \to \frac{z^2}{2} \text{ as } n \to \infty$$

$$\therefore \quad \lim_{n\to\infty} \log N = \frac{z^2}{2} \quad \Rightarrow \quad \lim_{n\to\infty} N = e^{z^2/2}$$

Substituting in (9·2), we get

$$dG(z) = g(z)\, dz = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz, -\infty < z < \infty \qquad \ldots (9\cdot2a)$$

Hence the probability function of Z is :

$$g(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < \infty \qquad \ldots (9\cdot2b)$$

This is the probability density function of the *normal distribution* with mean 0 and unit variance.

If X is normal variate with mean μ and *s.d.* σ, then $Z = (X - \mu)/\sigma$, is standard normal variate. Jacobian of transformation is $1/\sigma$. Hence substituting in {9·2 (*b*)}, the *p.d.f.* of a normal variate X with $E(X) = \mu$, $\text{Var}(X) = \sigma^2$ is given by :

$$f_X(x) = \begin{cases} \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, & -\infty < x < \infty \\ 0, \text{ otherwise} \end{cases}$$

Remark. Normal distribution can also be obtained as a limiting case of Poisson distribution with the parameter $\lambda \to \infty$.

9·2·2. Chief Characteristics of the Normal Distribution and Normal Probability Curve. The normal probability curve with mean μ and standard deviation σ is given by the equation :

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

and has the following properties :

(*i*) The curve is bell-shaped and symmetrical about the line $x = \mu$.

(*ii*) Mean, median and mode of the distribution coincide.

(*iii*) As x increases numerically, $f(x)$ decreases rapidly, the maximum probability occurring at the point $x = \mu$, and is given by : $[p(x)]_{max} = \dfrac{1}{\sigma\sqrt{2\pi}}$

(*iv*) $\beta_1 = 0$ and $\beta_2 = 3$.

(*v*) $\mu_{2r+1} = 0$, $(r = 0, 1, 2, \ldots)$, and $\mu_{2r} = 1.3.5 \ldots (2r-1)\,\sigma^{2r}$, $(r = 0, 1, 2, \ldots)$

(*vi*) Since $f(x)$ being the probability, can never be negative, no portion of the curve lies below the x-axis.

(*vii*) Linear combination of independent normal variates is also a normal variate.

(*viii*) x-axis is an asymptote to the curve.

(*ix*) The points of inflexion of the curve are : $x = \mu \pm \sigma$, $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-1/2}$

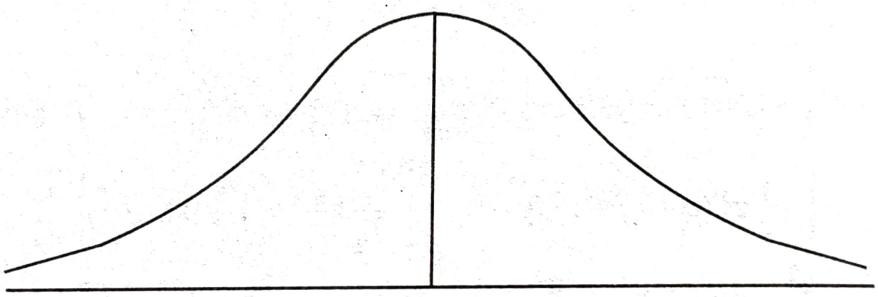


Fig. 9·1. Normal Probability Curve

(*x*) Mean deviation about mean $= \sqrt{\dfrac{2}{\pi}}\,\sigma \simeq \dfrac{4}{5}\,\sigma$ (approx.)

(*xi*) Quartiles are given by :

$$Q_1 = \mu - 0{\cdot}6745\,\sigma\,; \qquad Q_3 = \mu + 0{\cdot}6745\,\sigma$$

(*xii*) Q.D. $= \dfrac{Q_3 - Q_1}{2} \simeq \dfrac{2}{3}\,\sigma$. We have (approximately)

$$Q:D.:M.D.:S.D. :: \tfrac{2}{3}\sigma : \tfrac{4}{5}\sigma : \sigma :: \tfrac{2}{3} : \tfrac{4}{5} : 1 \Rightarrow Q.D.:M.D.:S.D. :: 10:12:15$$

(*xiii*) *Area Property* :

$$P(\mu - \sigma < X < \mu + \sigma) = 0{\cdot}6826, \qquad P(\mu - 2\sigma < X < \mu + 2\sigma) = 0{\cdot}9544,$$

and

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0{\cdot}9973$$

The adjoining table gives the area under the normal probability curve for some important values of standard normal variate Z.

Distances from the mean ordinates in terms of $\pm\sigma$	*Area under the curve*
$Z = \pm 0{\cdot}745$	$50\% = 0{\cdot}50$
$Z = \pm 1{\cdot}00$	$68{\cdot}26\% = 0{\cdot}6826$
$Z = \pm 1{\cdot}96$	$95\% = 0{\cdot}95$
$Z = \pm 2{\cdot}0$	$95{\cdot}44\% = 0{\cdot}9544$
$Z = \pm 2{\cdot}58$	$99\% = 0{\cdot}99$
$Z = \pm 3{\cdot}0$	$99{\cdot}73\% = 0{\cdot}9973$

(*xiv*) If X and Y are independent standard normal variates, then it can be easily proved that $U = X + Y$ and $V = X - Y$ are independently distributed, $U \sim N(0, 2)$ and $V \sim N(0, 2)$.

We state (without proof) the converse of this result which is due to *D. Bernstein*.

Bernstein's Theorem. *If X and Y are independent and identically distributed random variables with finite variances and if U = X + Y and V = X − Y are independent, then all r.v.'s X, Y, U and V are normally distributed.*

(*xiv*) We state below another result which characterises the normal distribution. If $X_1, X_2, \ldots, X_n$ are *i.i.d. r.v.*'s with finite variance, then the common distribution is normal if and only if :

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ or $\sum_{i=1}^{n} X_i$ and $\sum_{i=1}^{n}(X_i - \bar{X})^2$ are independent.

In the following sequences we shall establish some of these properties.

9·2·3. Mode of Normal Distribution. Mode is the value of x for which $f(x)$ is maximum, *i.e.*, mode is the solution of

$$f'(x) = 0 \text{ and } f''(x) < 0$$

For normal distribution with mean μ and standard deviation σ,

$$\log f(x) = c - \frac{1}{2\sigma^2}(x-\mu)^2,$$

where $c = \log(1/\sqrt{2\pi}\,\sigma)$, is a constant. Differentiating *w.r.* to x, we get

$$\frac{1}{f(x)} \cdot f'(x) = -\frac{1}{\sigma^2}(x-\mu) \quad \Rightarrow \quad f'(x) = -\frac{1}{\sigma^2}(x-\mu)\,f(x)$$

and $$f''(x) = -\frac{1}{\sigma^2}\left[1.f(x) + (x-\mu)\,f'(x)\right] = -\frac{f(x)}{\sigma^2}\left[1 - \frac{(x-\mu)^2}{\sigma^2}\right] \qquad \ldots (9\cdot3)$$

$f'(x) = 0 \Rightarrow x - \mu = 0 \Rightarrow x = \mu$. At the point $x = \mu$, we have from (9·3) :

$$f''(x) = -\frac{1}{\sigma^2}[f(x)]_{x=\mu} = -\frac{1}{\sigma^2} \cdot \frac{1}{\sigma\sqrt{2\pi}} < 0$$

Hence $x = \mu$, is the mode of the normal distribution.

9·2·4. Median of Normal Distribution. If M is the median of the normal distribution, we have

$$\int_{-\infty}^{M} f(x)\,dx = \frac{1}{2} \quad \Rightarrow \quad \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{M} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx = \frac{1}{2}$$

$$\Rightarrow \quad \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mu} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx + \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{M} \exp\{-(x-\mu)^2/(2\sigma^2)\}\,dx = \frac{1}{2} \qquad \ldots (9\cdot4)$$

But $$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mu} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} \exp(-z^2/2)\,dz = \frac{1}{2}$$

$\therefore$ From (9·4), we have $$\frac{1}{2} + \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{M} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx = \frac{1}{2}$$

$$\Rightarrow \quad \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{M} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx = 0, \quad i.e., \quad \mu = M.$$

Hence, for the normal distribution, Mean = Median.

Remark. From § 9·2·3. and § 9·2·4, we find that for the normal distribution mean, median and mode coincide. Hence the distribution is *symmetrical*.

9·2·5. M.G.F. of Normal Distribution. The *m.g.f.* (about origin) is given by :

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} \exp\{-(x-\mu)^2/2\sigma^2\}\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{t(\mu+\sigma z)\} \exp(-z^2/2)\, dz, \qquad \left(z = \frac{x-\mu}{\sigma}\right)$$

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

$$= e^{\mu t} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp[-\tfrac{1}{2}\{(z-\sigma t)^2 - \sigma^2 t^2\}]dz$$

$$= e^{\mu t + t^2\sigma^2/2} \times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}(z-\sigma t)^2\}\, dz$$

$$= e^{\mu t + t^2\sigma^2/2} \times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-u^2/2)\, du$$

Hence $\qquad M_X(t) = e^{\mu t + t^2\sigma^2/2}$ ...(9·5)

Remark. *M.G.F. of Standard Normal Variate.* If $X \sim N(\mu, \sigma^2)$, then standard normal variate is given by :

$$Z = (X-\mu)/\sigma.$$

$$M_Z(t) = e^{-\mu t/\sigma} M_X(t/\sigma) = \exp(-\mu t/\sigma).\exp\{(\mu t/\sigma) + (t^2/\sigma^2)(\sigma^2/2)\} = \exp(t^2/2) \quad ...(9·5a)$$

Aliter $Z \sim N(0, 1)$. Hence, taking $\mu = 0$ and $\sigma^2 = 1$ in (9·5), we get :

$$M_Z(t) = \exp(t^2/2).$$

9·2·6. Cumulant Generating Function (c.g.f.) of Normal Distribution. The *c.g.f.* of normal distribution is given by :

$$K_X(t) = \log_e M_X(t) = \log_e(e^{\mu t + t^2\sigma^2/2}) = \mu t + \frac{t^2\sigma^2}{2}$$

$\therefore$ $\quad$ Mean $= \kappa_1 =$ Coefficient of t in $K_X(t) = \mu$

Variance $= \kappa_2 =$ Coefficient of $\frac{t^2}{2!}$ in $K_X(t) = \sigma^2$

and $\quad \kappa_r =$ Coefficient of $\frac{t^r}{r!}$ in $K_X(t) = 0 \,;\, r = 3, 4 \ldots$

Thus $\quad \mu_3 = \kappa_3 = 0 \quad$ and $\quad \mu_4 = \kappa_4 + 3\kappa_2^2 = 3\sigma^4$

Hence $\quad \beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = 0 \quad$ and $\quad \beta_2 = \dfrac{\mu_4}{\mu_2^2} = 3$ ... (9·6)

9·2·7. Moments of Normal Distribution. Odd order moments about mean are given by :

$$\mu_{2n+1} = \int_{-\infty}^{\infty} (x-\mu)^{2n+1} f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^{2n+1} \exp\{-(x-\mu)^2/2\sigma^2\}dx$$

$$\therefore \quad \mu_{2n+1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1} \exp(-z^2/2)\, dz, \qquad \left(z = \frac{x-\mu}{\sigma}\right)$$

$$= \frac{\sigma^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n+1} \exp(-z^2/2)\, dz = 0, \qquad ... (9·7)$$

since the integrand $z^{2n+1} e^{-z^2/2}$ is an odd function of z.

Even order moments about mean are given by :

$$\mu_{2n} = \int_{-\infty}^{\infty} (x-\mu)^{2n} f(x)\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n} \exp(-z^2/2)\, dz$$

$$= \frac{\sigma^{2n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n} \exp(-z^2/2)\, dz = \frac{\sigma^{2n}}{\sqrt{2\pi}} . 2 \int_0^{\infty} z^{2n} \exp(-z^2/2)\, dz$$

(Since integrand is an even function of z.)

$$= \frac{2\sigma^{2n}}{\sqrt{2\pi}} \cdot \int_0^{\infty} (2t)^n e^{-t} \frac{dt}{\sqrt{2t}}, \qquad \left(t = \frac{z^2}{2}\right)$$

$$\therefore \quad \mu_{2n} = \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \int_0^{\infty} e^{-t}\, t^{(n+\frac{1}{2})-1}\, dt \quad \Rightarrow \quad \mu_{2n} = \frac{2^n \sigma^{2n}}{\sqrt{\pi}} . \Gamma\left(n + \frac{1}{2}\right)$$

Changing n to $(n-1)$, we get

$$\mu_{2n-2} = \frac{2^{n-1} . \sigma^{2n-2}}{\sqrt{\pi}} \Gamma\left(n - \frac{1}{2}\right)$$

$$\therefore \quad \frac{\mu_{2n}}{\mu_{2n-2}} = 2\sigma^2 . \frac{\Gamma\left(n + \frac{1}{2}\right)}{\left(n - \frac{1}{2}\right)} = 2\sigma^2 \left(n - \frac{1}{2}\right) \qquad [\because \Gamma(r) = (r-1)\Gamma(r-1)]$$

$$\Rightarrow \qquad \mu_{2n} = \sigma^2 (2n-1)\, \mu_{2n-2} \qquad \dots (9{\cdot}8)$$

which gives the *recurrence relation* for the moments of normal distribution.

From (9·8), we have

$$\mu_{2n} = [(2n-1)\sigma^2]\,[(2n-3)\sigma^2]\, \mu_{2n-4}$$

$$= [(2n-1)\, \sigma^2]\, [2n-3)\, \sigma^2\,]\, [2n-5)\, \sigma^2]\, \mu_{2n-6}$$

$$= [(2n-1)\sigma^2]\,[(2n-3)\sigma^2]\,[(2n-5)\sigma^2] \dots (3\sigma^2)(1\sigma^2). \mu_0$$

$$= 1.3.5. \dots (2n-1)\, \sigma^{2n} \qquad \dots (9{\cdot}9)$$

From (9·7) and (9·9), we conclude that *for the normal distribution all odd order moments about mean vanish and even order moments about mean are given by* (9·9).

Aliter. The above result can also be obtained quite conveniently as follows :

The *m.g.f.* (about mean) is given by : $E\,[e^{t(X-\mu)}] = e^{-\mu t}\, E\,(e^{tX}) = e^{-\mu t}\, M_X(t)$,

where $M_X(t)$ is the *m.g.f.* (about origin).

$\therefore$ *m.g.f.* (about mean) $= e^{-\mu t}\, e^{\mu t + t^2\sigma^2/2} = e^{t^2\sigma^2/2}$

$$= \left[1 + (t^2\sigma^2/2) + \frac{(t^2\sigma^2/2)^2}{2!} + \frac{(t^2\sigma^2/2)^3}{3!} + \dots + \frac{(t^2\sigma^2/2)^n}{n!} + \dots\right] \qquad \dots (9{\cdot}10)$$

The coefficient of $\frac{t^r}{r!}$ in (9·10) gives μ_r, the rth moment about mean. Since there is no term with odd powers of it t (9·10), all moments of odd order about mean vanish, *i.e.*,

$$\mu_{2n+1} = 0;\ n = 0, 1, 2, \dots$$

and

$$\mu_{2n} = \text{Coefficient of } \frac{t^{2n}}{(2n)!} \text{ in } (9{\cdot}10) = \frac{\sigma^{2n} \times (2n)!}{2^n\, n!}$$

$$= \frac{\sigma^{2n}}{2^n\, n!} \left[2n\,(2n-1)\,(2n-2)\,(2n-3) \dots 5.4.3.2.1\right]$$

$$= \frac{\sigma^{2n}}{2^n . n!} \left[1.3.5 \dots (2n-1)\right] \left[2.4.6 \dots (2n-2) . 2n\right]$$

$$= \frac{\sigma^{2n}}{2^n \cdot n!}\left[1.3.5. \ldots (2n-1)\right] 2^n \left[1.2.3 \ldots n\right]$$

$$= 1.3.5 \ldots (2n-1)\,\sigma^{2n}$$

Remark. In particular, from (9·7) and (9·9), $\mu_3 = 0$ and $\mu_2 = \sigma^2$, $\mu_4 = 1.3\,\sigma^4 = 3\sigma^4$

Hence $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = 0$ and $\beta_2 = \dfrac{\mu_4}{\mu_2^2} = \dfrac{3\sigma^4}{\sigma^4} = 3$, the results which have already been obtained in (9·6).

9·2·8. A linear combination of independent normal variates is also a normal variate. Let X_i, $(i = 1, 2, 3, \ldots, n)$ be n independent normal variates with mean μ_i and variance σ_i^2 respectively. Then

$$M_{X_i}(t) = \exp\{\mu_i t + (t^2 \sigma_i^2/2)\} \qquad \ldots (9·11)$$

The *m.g.f.* of their linear combination $\sum_{i=1}^{n} a_i X_i$, where $a_1, a_2, \ldots, a_n$ are constants, is given by :

$$M_{\Sigma a_i X_i}(t) = \prod_{i=1}^{n} M_{a_i X_i}(t) \qquad (\because X_i\text{'s are independent})$$

$$= M_{X_1}(a_1 t) . M_{X_2}(a_2 t) \ldots M_{X_n}(a_n t) \qquad [\because M_{cX}(t) = M_X(ct)] \qquad \ldots (9·12)$$

From (9·11), we have $\quad M_{X_i}(a_i t) = e^{\mu_i a_i t + t^2 a_i^2 \sigma_i^2/2}$

$$\therefore\; M_{\Sigma_i a_i X_i}(t) = \left[e^{\mu_1 a_1 t + t^2 a_1^2 \sigma_1^2/2} \times e^{\mu_2 a_2 t + t^2 a_2^2 \sigma_2^2/2} \times \ldots \times e^{\mu_n a_n t + t^2 a_n^2 \sigma_n^2/2}\right] \qquad [\text{From } (9·12)]$$

$$= \exp\left[\left(\sum_{i=1}^{n} a_i \mu_i\right) t + t^2 \left(\sum_{i=1}^{n} a_i^2 \sigma_i^2\right)/2\right],$$

which is the *m.g.f.* of a normal variate with mean $\sum_{i=1}^{n} a_i \mu_i$ and variance $\sum_{i=1}^{n} a_i^2 \sigma_i^2$.

Hence by uniqueness theorem of *m.g.f.*,

$$\sum_{i=1}^{n} a_i X_i \sim N\left[\sum_{i=1}^{n} a_i \mu_i,\; \sum_{i=1}^{n} a_i^2 \sigma_i^2\right] \qquad \ldots (9·12a)$$

Remarks 1. If we take $a_1 = a_2 = 1, a_3 = a_4 = \ldots = 0$, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

If we take $a_1 = 1, a_2 = -1, a_3 = a_4 = \ldots = 0$, then $X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

Thus we see that *the sum* as well as *the difference of two independent normal variates is also a normal variate.* This result provides a sharp contrast to the Poisson distribution, in which case though the sum of two independent Poisson variates is a Poisson variate, the difference is not a Poisson variate.

2. If we take $a_1 = a_2 = \ldots = a_n = 1$, then we get $\sum_{i=1}^{n} X_i \sim N\left[\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right] \qquad \ldots (9·12b)$

i.e., the sum of independent normal variates is also a normal variate, which establishes the *additive property* of the normal distribution.

3. If $X_i; i = 1, 2, \ldots, n$ are identically and independently distributed as $N(\mu, \sigma^2)$ and if we take $a_1 = a_2 = \ldots = a_n = 1/n$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\frac{1}{n}\sum_{i=1}^{n}\mu, \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2\right) \Rightarrow \bar{X} \sim N(\mu, \sigma^2/n), \text{ where } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

This leads to the following important conclusion :

If X_i $(i = 1, 2, \ldots, n)$, are identically and independently distributed normal variates with mean μ and variance σ^2, then their mean $\bar{X}$ is also $N(\mu, \sigma^2/n)$.

9·2·9. Points of Inflexion of Normal Curve. At the point of inflexion of the normal curve, we should have $f''(x) = 0$, and $f'''(x) \neq 0$.

For normal curve, we have from (9·3), $f''(x) = -\dfrac{f(x)}{\sigma^2}\left\{1 - \dfrac{(x-\mu)^2}{\sigma^2}\right\}$

$$\therefore \quad f''(x) = 0 \quad \Rightarrow \quad 1 - \frac{(x-\mu)^2}{\sigma^2} = 0 \quad \Rightarrow \quad x = \mu \pm \sigma.$$

It can be easily verified that at the points $x = \mu \pm \sigma, f'''(x) \neq 0$. Hence the points of inflexion of the normal curve are given by $x = \mu \pm \sigma$ and $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-1/2}$, *i.e.*, they are equi-distant (at a distance σ) from the mean.

9·2·10. Mean Deviation About the Mean for Normal Distribution.

$$\text{M.D. (about mean)} = \int_{-\infty}^{\infty} |x - \mu|\, f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |x - \mu|\, e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z|\, e^{-z^2/2}\, dz, \qquad \left(\frac{x-\mu}{\sigma} = z\right)$$

$$= \frac{2\sigma}{\sqrt{2\pi}} \cdot \int_0^{\infty} |z|\, e^{-z^2/2}\, dz,$$

($\because$ The integrand $|z|\, e^{-z^2/2}$ is an even function of z.)

Since in $[0, \infty]$, $|z| = z$, we have

$$\text{M.D. (about mean)} = \sqrt{2/\pi}\ \sigma \int_0^{\infty} e^{-z^2/2}\, dz\ \sqrt{2/\pi}\ \sigma \int_0^{\infty} e^{-t}\, dt, \qquad (z^2/2 = t)$$

$$= \sqrt{2/\pi}\ \sigma \left|\frac{e^{-t}}{-1}\right|_0^{\infty} = \sqrt{2/\pi}\ \sigma = \frac{4}{5}\sigma \text{ (approx.)}$$

9·2·11. Area Property (Normal Probability Integral). If $X \sim N(\mu, \sigma^2)$, then the probability that random value of X will lie between $X = \mu$ and $X = x_1$ is given by :

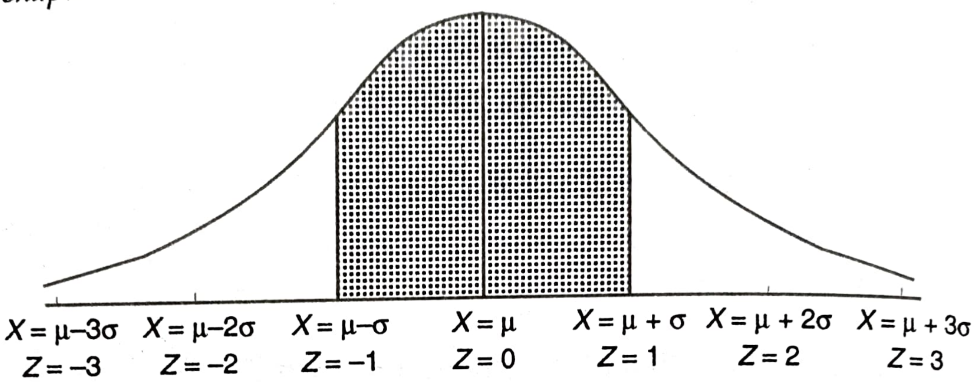
$$P(\mu < X < x_1) = \int_{\mu}^{x_1} f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{x_1} e^{-(x-\mu)^2/2\sigma^2} dx$$

Put $\qquad \dfrac{X-\mu}{\sigma} = Z \quad \Rightarrow \quad X - \mu = \sigma Z$

When $X = \mu, Z = 0$ and when $X = x_1, Z = \dfrac{x_1 - \mu}{\sigma} = z_1$, (say).

$$\therefore \quad P(\mu < X < x_1) = P(0 < Z < z_1) = \frac{1}{\sqrt{2\pi}} \int_0^{z_1} e^{-z^2/2}\, dz = \int_0^{z_1} \varphi(z)\, dz$$

where $\varphi(z) = \dfrac{1}{\sqrt{2\pi}} e^{-z^2/2}$, is the probability function of standard normal variate. The definite integral $\displaystyle\int_0^{z_1} \varphi(z)\, dz$ is known as *normal probability integral* and gives the area under standard normal curve between the ordinates at $Z = 0$ and $Z = z_1$. These areas have been tabulated for different values of z_1, at intervals of 0·01 in a table given at *the end of the chapter.*



In particular, the probability that a random value of X lies in the interval $(\mu - \sigma, \mu + \sigma)$ is given by :

$$P(\mu - \sigma < X < \mu + \sigma) = \int_{\mu - \sigma}^{\mu + \sigma} f(x)dx$$

$$\Rightarrow \qquad P(-1 < Z < 1) = \int_{-1}^{1} \varphi(z)\, dz, \qquad \left[z = \frac{x - \mu}{\sigma}\right]$$

$$= 2\int_{0}^{1} \varphi(z)\, dz \qquad \text{(By symmetry)}$$

$$= 2 \times 0{\cdot}3413 = 0{\cdot}6826 \qquad \text{(From Tables)} \quad \ldots (9{\cdot}14)$$

Similarly, $P(\mu - 2\sigma < X < \mu + 2\sigma) = P(-2 < Z < 2) = \int_{-2}^{2} \varphi(z)\, dx$

$$= 2\int_{0}^{2} \varphi(z)\, dz = 2 \times 0{\cdot}4772 = 0{\cdot}9544 \quad \ldots (9{\cdot}15)$$

and $\quad P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < Z < 3) = \int_{-3}^{3} \varphi(z)\, dz$

$$= 2\int_{0}^{3} \varphi(z)\, dz = 2 \times 0{\cdot}49865 = 0{\cdot}9973 \quad \ldots (9{\cdot}16)$$

Thus the probability that a normal variate X lies outside the range $\mu \pm 3\sigma$ is given by :

$$P(|X - \mu| > 3\sigma) = P(|Z| > 3) = 1 - P(-3 \leq Z \leq 3) = 0{\cdot}0027$$

Thus in all probability, we should expect a normal variate to lie within the range $\mu \pm 3\sigma$, though theoretically, it may range from $-\infty$ to ∞.

***Remarks* 1.** The total area under normal probability curve is unity, *i.e.*,

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

2. Since in the normal probability tables, we are given the areas under standard normal curve, in numerical problems we shall deal with the standard normal variate Z rather than the variable X itself.

3. If we want to find area under normal curve, we will somehow or other try to convert the given area to the form $P(0 < Z < z_1)$, since the areas have been given in this form in the Tables.

9·2·12. Error Function. If $X \sim N(0, \sigma^2)$, then $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}, -\infty < x < \infty$

If we take $h^2 = \dfrac{1}{2\sigma^2}$, then

$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}$$

The probability P that a random value of the variate lies in the range $\pm x$ is :

$$P = \int_{-x}^{x} f(x)\, dx = \frac{h}{\sqrt{\pi}} \int_{-x}^{x} e^{-h^2x^2}\, dx = \frac{2h}{\sqrt{\pi}} \int_0^x e^{-h^2x^2}\, dx = \frac{2}{\sqrt{\pi}} \int_0^x e^{-h^2x^2}\, (hdx) \qquad \ldots (*)$$

Taking $\psi(y) = \dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^y e^{-y^2}\, dy$, (*) may be re-written as :

$$P = \psi(hx) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-h^2x^2}\, (hdx) \qquad \ldots (**)$$

The function $\psi(y)$, known as the *error function*, is of fundamental importance in the *theory of errors* in Astronomy.

9·2·13. Importance of Normal Distribution. Normal distribution plays a very important role in statistical theory because of the following reasons :

(*i*) Most of the distributions occurring in practice, *e.g.*, Binomial, Poisson, Hypergeometric distributions, etc., can be approximated by normal distribution . Moreover, many of the sampling distributions, *e.g.*, Student's t, Snedecor's F, Chi-square distributions, etc., tend to normality for large samples.

(*ii*) Even if a variable is not normally distributed, it can sometimes be brought to normal form by simple transformation of variable. For example, if the distribution of X is skewed, the distribution of $\sqrt{X}$ might come out to be normal [*c.f.* Variate Transformations, § 9·13 at the end of this Chapter].

(*iii*) If $X \sim N(\mu, \sigma^2)$, then $P(\mu - 3\sigma \le X \le \mu + 3\sigma) = P(-3 < Z < 3) = 0{\cdot}9973$

$\therefore \qquad P(|Z| > 3) = 1 - P(|Z| \le 3) = 0{\cdot}0027$

This property of the normal distribution forms the basis of entire *Large Sample* theory.

(*iv*) Many of the distributions of sample statistics (*e.g.*, the distributions of sample mean, sample variance, etc.) tend to normality for large samples and as such they can best be studied with the help of the normal curves.

(*v*) The entire theory of small sample tests, *viz.*, t, F, χ^2 tests, etc., is based on the fundamental assumption that the parent populations from which the samples have been drawn follow normal distribution.

(*vi*) Normal distribution finds large applications in Statistical Quality Control in industry for setting control limits.

The following quotation due to Lipman rightly reveals the popularity and importance of normal distribution :

Example 9·2. *For a certain normal distribution, the first moment about 10 is 40 and the fourth moment about 50 is 48. What is the arithmetic mean and standard deviation of the distribution ?*

Solution. We know that if μ_1' is the first moment about the point $X = A$, then arithmetic mean is given by : Mean $= A + \mu_1'$

We are given : μ_1' (about the point $X = 10) = 40 \quad \Rightarrow \quad$ Mean $= 10 + 40 = 50$

Also μ_4' (about the point $X = 50) = 48$, *i.e.*, $\mu_4 = 48$ $\quad (\because$ Mean $= 50)$

But for a normal distribution with standard deviation σ,

$$\mu_4 = 3\sigma^4 \quad \Rightarrow \quad 3\sigma^4 = 48 \quad \Rightarrow \quad \sigma = 2.$$

Example 9·3. *X is normally distributed and the mean of X is 12 and S.D. is 4. (a) Find out the probability of the following :*

(a) (i) $X \geq 20$, (ii) $X \leq 20$, and (iii) $0 \leq X \leq 12$ (b) Find x', when $P(X > x') = 0{\cdot}24$.

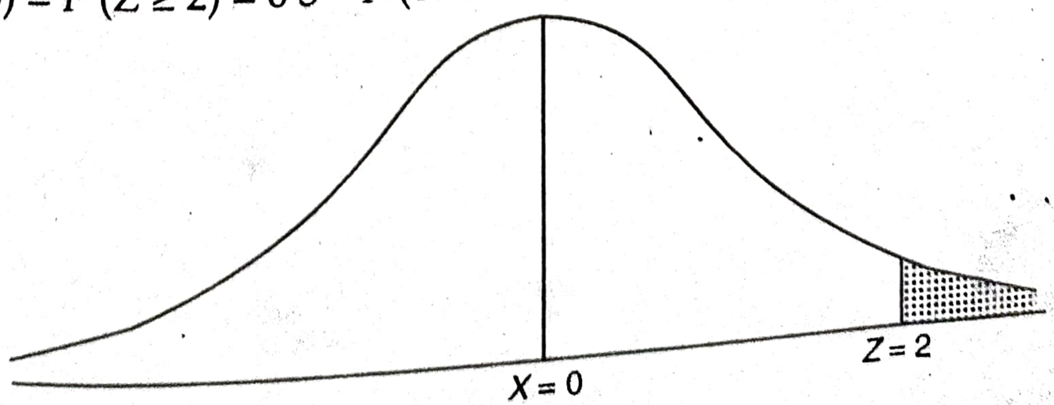
(c) Find x_0' and x_1', when $P(x_0' < X < x_1') = 0{\cdot}50$ and $P(X > x_1') = 0{\cdot}25$.

Solution. (*a*) We have $\mu = 12$, $\sigma = 4$, *i.e.*, $X \sim N(12, 16)$.

(*i*) $P(X \geq 20) = ?$

When $X = 20$, $Z = \dfrac{20 - 12}{4} = 2$

$\therefore P(X \geq 20) = P(Z \geq 2) = 0{\cdot}5 - P(0 \leq Z \leq 2) = 0{\cdot}5 - 0{\cdot}4772 = 0{\cdot}0228$



(*ii*) $P(X \le 20) = 1 - P(X \ge 20) = 1 - 0{\cdot}0228 = 0{\cdot}9722$ [From part (*i*)]

(*iii*) $P(0 \le X \le 12) = P(-3 \le Z \le 0)$, $\left(Z = \dfrac{X-12}{4}\right)$

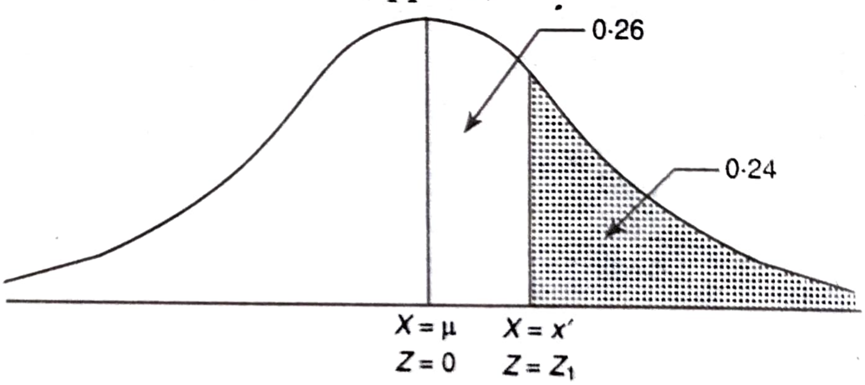
$= P(0 \le Z \le 3) = 0{\cdot}49865$ (From symmetry)

(*b*) When $X = x'$, $Z = \dfrac{x'-12}{4} = z_1$, (say).

Then, we are given :

$$P(X > x') = 0{\cdot}24 \Rightarrow P(Z > z_1) = 0{\cdot}24 \Rightarrow P(0 < Z < z_1) = 0{\cdot}26$$

$\therefore$ From Normal Tables, $z_1 = 0{\cdot}71$ (approx.)



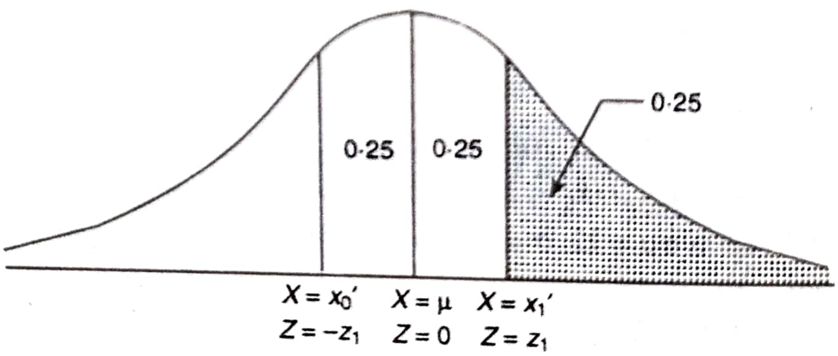
Hence $\dfrac{x'-12}{4} = 0{\cdot}71 \Rightarrow x' = 12 + 4 \times 0{\cdot}71 = 14{\cdot}84.$

(*c*) We are given : $P(x_0' < X < x_1') = 0{\cdot}50$ and $P(X > x_1') = 0{\cdot}25$ ... (*)

From (*), obviously the points x_0' and x_1' are located as shown in following adjoining.

When $X = x_1'$, $Z = \dfrac{x_1'-12}{4} = z_1$, (say),

and when $X = x_0'$, $Z = \dfrac{x_0'-12}{4} = -z_1$ (It is obvious from the figure.)



We have $P(Z > z_1) = 0{\cdot}25 \Rightarrow P(0 < Z < z_1) = 0{\cdot}25 \quad \therefore z_1 = 0{\cdot}67$ (From Tables)

Hence $\dfrac{x_1'-12}{4} = 0{\cdot}67 \Rightarrow x_1' = 12 + 4 \times 0{\cdot}67 = 14{\cdot}68$

and $\dfrac{x_0'-12}{4} = -0{\cdot}67 \Rightarrow x_0' = 12 - 4 \times 0{\cdot}67 = 9{\cdot}32.$

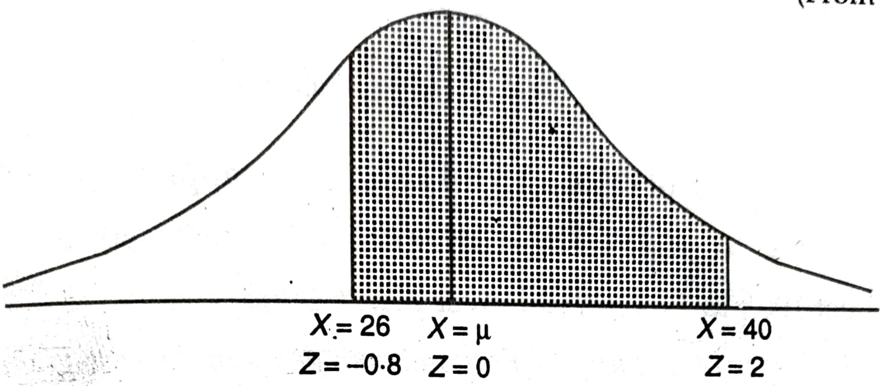
Example 9·4. *X is a normal variate with mean 30 and S.D. 5. Find the probabilities that* (*i*) $26 \le X \le 40$, (*ii*) $X \ge 45$, *and* (*iii*) $|X - 30| > 5$.

Solution. Here $\mu = 30$ and $\sigma = 5$.

(*i*) When $X = 26$, $Z = \dfrac{X-\mu}{\sigma} = \dfrac{26-30}{5} = -0{\cdot}8$

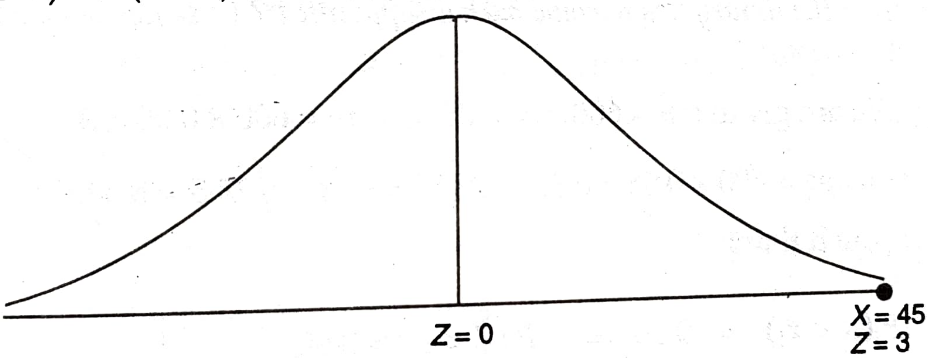
and when $X = 40$, $Z = \dfrac{40-30}{5} = 2$

$$\therefore\ P(26 \le X \le 40) = P(-0{\cdot}8 \le Z \le 2) = P(-0{\cdot}8 \le Z \le 0) + P(0 \le Z \le 2)$$
$$= P(0 \le Z \le 0{\cdot}8) + (0 \le Z \le 2) = 0{\cdot}2881 + 0{\cdot}4772 \quad \text{[By symmetry]}$$
$$= 0{\cdot}7653 \quad \text{(From Normal Tables)}$$



(ii) When $x = 45$, $Z = \dfrac{45-30}{5} = 3.$

$$\therefore\ P(X \ge 45) = P(Z \ge 3) = 0{\cdot}5 - P(0 \le Z \le 3) = 0{\cdot}5 - 0{\cdot}49865 = 0{\cdot}00135$$



(iii) $$P(|X-30| \le 5) = P(25 \le X \le 35) = P(-1 \le Z \le 1)$$
$$= 2P(0 \le Z \le 1) = 2 \times 0{\cdot}3413 = 0{\cdot}6826$$
$$\therefore\ P(|X-30| > 5) = 1 - P(|X-30| \le 5) = 1 - 0{\cdot}6826 = 0{\cdot}3174.$$

Example 9·5. *The mean yield for one-acre plot is 662 kilos with a s.d. 32 kilos. Assuming normal distribution, how many one-acre plots in a batch of 1,000 plots would you expect to have yield (i) over 700 kilos, (ii) below 650 kilos, and (iii) what is the lowest yield of the best 100 plots ?*

Solution. If the *r.v.* X denotes the yield (in kilos) for one-acre plot, the we are given that $X \sim N(\mu, \sigma^2)$, where $\mu = 662$ and $\sigma = 32$.

(i) The probability that a plot has a yield over 700 kilos is given by

$$P(X > 700) = P(Z > 1{\cdot}19), \qquad Z = \frac{700-662}{32}$$
$$= 0{\cdot}5 - P(0 \le Z \le 1{\cdot}19) = 0{\cdot}5 - 0{\cdot}3830 = 0{\cdot}1170$$

Hence in a batch of 1,000 plots, the expected number of plots with yield over 700 kilos is $1{,}000 \times 0{\cdot}117 = 117$.

(ii) Required number of plots with yield below 650 kilos is given by :

$$1000 \times P(X < 650) = 1{,}000 \times P(Z < -0{\cdot}38) \qquad \left(Z = \frac{650-662}{32}\right)$$