13. Find the mean and standard deviation of the following two samples put together :

Sample No.	Size	Mean	S.D.
1	50	158	5.1
2	60	164	4.6

14. A distribution consists of three components with frequencies 200, 250 and 300 having means 25, 10 and 15 and S.Ds. 3, 4 and 5 respectively. Show that the mean of the combined distribution is 16 and its S.D. is 7.2 approximately.

25.9 (1) MOMENTS

The rth moment about the mean $\bar{x}$ of a distribution is denoted by μ_r and is given by

$$\mu_r = \frac{1}{N}\Sigma f_i(x_i - \bar{x})^r \qquad ...(1)$$

The corresponding moment about any point a is defined as

$$\mu_r' = \frac{1}{N}\Sigma f_i(x_i - a)^r \qquad ...(2)$$

In particular, we have $\mu_0 = \mu_0' = 1$ ...(3)

$$\mu_1 = \frac{1}{N}\Sigma f_i(x_i - \bar{x}) = 0;\ \mu_1' = \frac{1}{N}\Sigma f_i(x_i - a) = \bar{x} - a = d, \text{ say} \qquad ...(4)$$

$$\mu_2 = \frac{1}{N}\Sigma f_i(x_i - \bar{x})^2 = \sigma^2. \qquad ...(5)$$

(2) Moments about the mean in terms of moments about any point.

We have

$$\mu_r = \frac{1}{N}\Sigma f_i(x_i - \bar{x})^r = \frac{1}{N}\Sigma f[(x_i - a) - (\bar{x} - a)]^r$$

$$= \frac{1}{N}\Sigma f_i(X_i - d)^r \quad \text{where} \quad X_i = x_i - a,\ d = \bar{x} - a.$$

$$= \frac{1}{N}[\Sigma f_i X_i^r - {}^rC_1 d\Sigma f_i X_i^{r-1} + {}^rC_2 d^2 \Sigma f_i X_i^{r-2} - ...]$$

$$= \mu_r' - {}^rC_1\, d\mu_{r-1}' + {}^rC_2 d^2 \mu_{r-2}'' - ...... \qquad ...(6)$$

In particular,

$$\mu_2 = \mu'_2 - \mu'^2_1 \qquad ...(7)$$

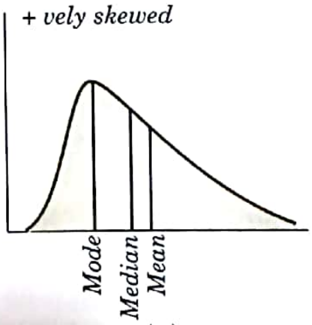
$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1 \qquad ...(8)$$

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1 \qquad ...(9)$$

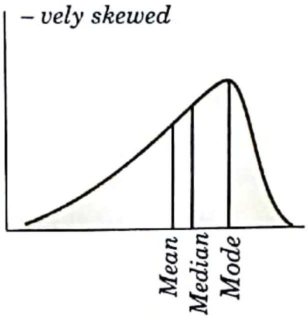
These three results should be committed to memory. It should be noted that in each of these relations, the sum of the coefficients of the various terms on the right side is zero. Also each term on the right side is of the same dimension as the term on the left.

25.10 SKEWNESS

Skewness measures the degree of asymmetry or the departure from symmetry. If the frequency curve has a longer 'tail' to the right, *i.e.*, the mean is to the right of the mode [as in Fig. 25.4 (*a*)], then the distribution is said to have *positive skewness*. If the curve is more elongated to the left, then it is said to have *negative skewness* [Fig. 25.4 (*b*)].



(*a*)



(*b*)

The following three measures of skewness deserve mention :

(*i*) *Pearson's* coefficient of skewness* $= \dfrac{\text{mean} - \text{mode}}{\sigma}$

* After the English statistician and biologist *Karl Pearson* (1857–1936) who did pioneering work and found the English school of statistics.

(ii) *Quartile coefficient of skewness* $= \dfrac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

Its value always lies between – 1 and + 1.

(iii) *Coefficient of skewness based on third moment* $\gamma_1 = \sqrt{\beta_1}$.

where $\beta_1 = \mu_3^2/\mu_2^3$

Thus $\gamma_1 = \sqrt{\beta_1}$ gives the simplest measure of *skewness.*

25.11 KURTOSIS

Kurtosis measures the degree of peakedness of a distribution and is given by $\beta_2 = \mu_4/\mu_2^2$.

(a) *Leptokurtic* (b) *Platykurtic* (c) *Mesokurtic*

Fig. 25.5

$\gamma_2 = \beta_2 - 3$ gives the excess of *Kurtosis.* The curves with $\beta_2 > 3$ are called *Leptokurtic* and those with $\beta_2 < 3$ as *Platykurtic*. The normal curve for which $\beta_2 = 3$, is called *Mesokurtic* [Fig. 25.5].

Example 25.11. *The first four moments about the working mean 28.5 of a distribution are 0.294, 7.144, 42.409 and 454.98. Calculate the moments about the mean. Also evaluate β_1, β_2 and comment upon the skewness and kurtosis of the distribution.* (V.T.U., 2003 S)

Solution. The first four moments about the arbitrary origin 28.5 are $\mu'_1 = 0.294$, $\mu'_2 = 7.144$, $\mu'_3 = 42.409$, $\mu'_4 = 454.98$.

$$\therefore \quad \mu'_1 = \frac{1}{N}\sum f_i(x_i - 28.5) = \frac{1}{N}\sum f_i x_i - 28.5 = \bar{x} - 28.5 = 0.294 \text{ or } \bar{x} = 28.794$$

$$\mu_2 = \mu'_2 - \mu'^2_1 = 7.144 - (0.294)^2 = 7.058$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1 = 42.409 - 3(7.144)(0.294) + 2(0.294)^3 = 36.151.$$

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1$$

$$= 454.98 - 4(42.409) \times (0.294) + 6(7.144)(0.294)^2 - 3(0.294)^4 = 408.738$$

Now $\beta_1 = \mu_3^2/\mu_2^3 = (36.151)^2/(7.058)^3 = 3.717$

$\beta_2 = \mu_4/\mu_2^2 = 408.738/(7.058)^2 = 8.205.$

$\therefore$ $\gamma_1 = \sqrt{\beta_1} = 1.928$, which indicates considerable skewness of the distribution.

$\gamma_2 = \beta_2 - 3 = 5.205$ which shows that the distribution is *leptokurtic.*

Example 25.12. *Calculate the median, quartiles and the quartile coefficient of skewness from the following data :*

Weight (lbs) :	70–80	80–90	90–100	100–110	110–120	120–130	130–140	140–150
No. of persons :	12	18	35	42	50	45	20	8

Solution. Here total frequency $N = \Sigma f_i = 230$.

The cummulative frequency table is

Weight (lbs) :	70–80	80–90	90–100	100–110	110–120	120–130	130–140	140–150
f :	12	18	35	42	50	45	20	8
cum. f. :	12	30	65	107	157	202	222	230

Now $N/2 = 230/2 = 115$th item which lies in 110–120 group.

$$\therefore \quad \text{median or } Q_2 = L + \frac{N/2 - C}{f} \times h = 110 + \frac{115 - 107}{50} \times 10 = 111.6$$

Also $N/4 = 230/4 = 57.5$ *i.e.* Q_1 is 57.5th or 58th item which lies in 90–100 group.

$$\therefore \quad Q_1 = L + \frac{N/4 - C}{f} \times h = 90 + \frac{57.5 - 30}{35} \times 10 = 97.85$$

Similarly, $3N/4 = 172.5$ *i.e.* Q_3 is 173rd item which lies in 120–130 group.

$$\therefore \quad Q_3 = L + \frac{3N/4 - C}{f} \times h = 120 + \frac{172.5 - 157}{45} \times 10 = 123.44$$

Hence quartile coefficient of skewness $= \dfrac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1}$

$$= \frac{97.85 + 123.44 - 2 \times 111.6}{123.44 - 97.85} = -0.07 \text{ (approx.)}.$$

PROBLEMS 25.3

1. Calculate the first four moments of the following distribution about the mean :

x :	0	1	2	3	4	5	6	7	8
f :	1	8	28	56	70	56	28	8	1

Also evaluate β_1 and β_2. (*V.T.U., 2004 ; Madras, 2003*)

2. The following table gives the monthly wages of 72 workers in a factory. Compute the standard deviation, quartile deviation, coefficients of variation and skewness. (*V.T.U., 2001*)

Monthly wages (in ₹)	*No. of workers*	*Monthly wages (in ₹)*	*No. of workers*
12.5—17.5	2	37.5—42.5	4
17.5—22.5	22	42.5—47.5	6
22.5—27.5	19	47.5—52.5	1
27.5—32.5	14	52.5—57.5	1
32.5—37.5	3		

3. Find Pearson's coefficient of skewness for the following data :

Class :	10–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89
Frequency :	5	9	14	20	25	15	8	4

4. Compute the quartile coefficient of skewness for the following distribution :

x :	3–7	8–12	13–17	18–22	23–27	28–32	33–37	38–42
f :	2	108	580	175	80	32	18	5

(*Madras, 2002*)

Also compute the measure of skewness based on the third moment.

5. The first three moments of a distribution about the value 2 of the variable are 1, 16 and – 40. Show that the mean = 3, the variance = 15 and $\mu_3 = -86$. (*V.T.U., 2003 S*)

6. Compute skewness and kurtosis, if the first four moments of a frequency distribution $f(x)$ about the value $x = 4$ are respectively 1, 4, 10 and 45.

7. The first four moments of a distribution about the value '4' of the variable are – 1.5, 17, – 30 and 108. State whether the distribution is leptokurtic or platykurtic. (*U.P.T.U., 2007*)

8. The first four moments of a distribution about the value '0' are – 0.20, 1.76, – 2.36 and 10.88. Find the moments about the mean and the kurtosis. (*U.P.T.U., 2009*)

3. The frequency function of a continuous random variable is given by
$$f(x) = y_0 x(2 - x),\ 0 \le x \le 2.$$
Find the value of y_0, mean and variance of x. (Kerala, 2005 ; J.N.T.U., 2003)

4. The probability density $p(x)$ of a continuous random variable is given by
$$p(x) = y_0 e^{-|x|},\ -\infty < x < \infty.$$
Prove that $y_0 = 1/2$. Find the mean and variance of the distribution.

(Calicut, 2012 ; C.S.V.T.U., 2008 ; Kurukshetra, 2007)

5. $f(x) = \begin{cases} \frac{1}{2}(x+1), & -1 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$

represents the density of a random variable X, find $E(X)$ and Var (X).

6. A function is defined as under :
$$f(x) = 1/k,\ x_1 \le x \le x_2$$
$$= 0, \text{ elsewhere.}$$
Find the cumulative distribution of the variate x when k satisfies the requirements for $f(x)$ to be a density function.

26.13 REPEATED TRIALS

We know that the probability of getting a head or a tail on tossing a coin is $\frac{1}{2}$. If the coin is tossed thrice, the probability of getting one head and two tails can be combined as $H—T—T$, $T—H—T$, $T—T—H$. The probability of each one of these being $\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}$, *i.e.*, $\left(\frac{1}{2}\right)^3$, their total probability shall be $3(1/2)^3$.

Similarly if a trial is repeated n times and if p is the probability of a success and q that of a failure, then the probability of r successes and $n - r$ failures is given by $p^r q^{n-r}$.

But these r successes and $n - r$ failures can occur in any of the ${}^nC_r$ ways in each of which the probability is same.

Thus the probability of r successes is ${}^nC_r\, p^r q^{n-r}$.

Cor. *The probabilities of at least r successes in n trials*

$= $ the sum of the probabilities of $r, r+1, \ldots, n$ successes

$= {}^nC_r\, p^r q^{n-r} + {}^nC_{r+1}\, p^{r+1} q^{n-r-1} + \ldots + {}^nC_n\, p^n$.

26.14 (1) BINOMIAL DISTRIBUTION*

It is concerned with trials of a repetitive nature in which only the occurrence or non-occurrence, success or failure, acceptance or rejection, yes or no of a particular event is of interest.

If we perform a series of independent trials such that for each trial p is the probability of a success and q that of a failure, then the probability of r successes in a series of n trials is given by ${}^nC_r\, p^r q^{n-r}$, where r takes any integral value from 0 to n. The probabilities of 0, 1, 2, ... r, ..., n successes are, therefore, given by
$$q^n,\ {}^nC_1 pq^{n-1},\ {}^nC_2 p^2 q^{n-2}, \ldots,\ {}^nC_r\, p^r q^{n-r}, \ldots,\ p^n.$$
The probability of the number of successes so obtained is called *the* **binomial distribution** for the simple reason that the probabilities are the successive terms in the expansion of the binomial $(q + p)^n$.

$\therefore$ the sum of the probabilities
$$= q^n + {}^nC_1 pq^{n-1} + {}^nC_2 p^2 q^{n-2} + \ldots + p^n = (q+p)^n = 1.$$

(2) **Constants of the binomial distribution.** The moment generating function about the origin is
$$M_0(t) = E(e^{tx}) = \Sigma\, {}^nC_x\, p^x q^{n-x} e^{tx}$$ [By (1) § 26.11]
$$= \Sigma\, {}^nC_x\, (pe^t)^x q^{n-x} = (q + pe^t)^n$$

*It was discovered by a Swiss mathematician *Jacob Bernoulli* and was published posthumously in 1713.

Differentiating with respect to t and putting $t = 0$ and using (3) § 26.11, we get the mean

$$\mu'_1 = np.$$

Since $M_a(t) = e^{-at} M_0(t)$, the m.g.f. of the binomial distribution about its mean $(m) = np$, is given by

$$M_m(t) = e^{-npt}(q + pe^t)^n = (qe^{-pt} + pe^{qt})^n$$

$$= \left(1 + pq\frac{t^2}{2!} + pq(q^2 - p^2)\frac{t^3}{3!} + pq(q^3 + p^3)\frac{t^4}{4!} + \cdots\right)^n$$ (Anna, 2012

or
$$1 + \mu_1 t + \mu_2\frac{t^2}{2!} + \mu_3\frac{t^3}{3!} + \mu_4\frac{t^4}{4!} + \cdots$$

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

Equating the coefficients of like powers of t on either side, we have

$$\mu_2 = npq,\ \mu_3 = npq(q - p),\ \mu_4 = npq\,[1 + 3(n-2)pq].$$

Also
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(q-p)^2}{npq} = \frac{(1-2p)^2}{npq} \quad \text{and} \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1-6pq}{npq}$$

Thus *mean* $= np$, *standard deviation* $= \sqrt{(npq)}$.

skewness $= (1 - 2p)/\sqrt{(npq)}$, *kurtosis* $= \beta_2$.

Obs. The skewness is positive for $p < \frac{1}{2}$ and negative for $p > \frac{1}{2}$. When $p = \frac{1}{2}$, the skewness is zero, *i.e.*, the probability curve of the binomial distribution will be symmetrical (bell-shaped).

As n the number of trials increase indefinitely, $\beta_1 \to 0$, and $\beta_2 \to 3$.

(3) **Binomial frequency distribution.** If n independent trials constitute one experiment and thi experiment be repeated N times, then the frequency of r successes is $N\,{}^nC_r\,p^r\,q^{n-r}$. The possible number of successes together with these expected frequencies constitute the *binomial frequency distribution.*

(4) **Applications of Binomial distribution.** This distribution is applied to problems concerning

(*i*) Number of defectives in a sample from production line,

(*ii*) Estimation of reliability of systems,

(*iii*) Number of rounds fired from a gun hitting a target,

(*iii*) Radar detection.

Example 26.38. *The probability that a pen manufactured by a company will be defective is 1/10. If 12 such pens are manufactured, find the probability that*

(*a*) *exactly two will be defective.* (*b*) *at least two will be defective.*

(*c*) *none will be defective.* (*V.T.U., 2012 S ; Burdwan, 2003*)

Solution. The probability of a defective pen is $1/10 = 0.1$

$\therefore$ The probability of a non-defective pen is $1 - 0.1 = 0.9$

(*a*) *The probability that exactly two will be defective*

$$= {}^{12}C_2\,(0.1)^2\,(0.9)^{10} = 0.2301$$

(*b*) *The probability that at least two will be defective*

$= 1 -$ (prob. that either none or one is non-defective)

$$= 1 - [{}^{12}C_0(0.9)^{12} + {}^{12}C_1(0.1)\,(0.9)^{11}] = 0.3412$$

(*c*) *The probability that none will be defective*

$$= {}^{12}C_{12}(0.9)^{12} = 0.2833.$$

Example 26.39. *In 256 sets of 12 tosses of a coin, in how many cases one can expect 8 heads and 4 tails.* (*J.N.T.U., 2003*)

Solution. $P(\text{head}) = \frac{1}{2}$ and $P(\text{tail}) = \frac{1}{2}$

By binomial distribution, probability of 8 heads and 4 tails in 12 trials is

$$P(X=8) = {}^{12}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^4 = \frac{12!}{8!\,4!} \cdot \frac{1}{2^{12}} = \frac{495}{4096}$$

$\therefore$ the expected number of such cases in 256 sets

$$= 256 \times P(X=8) = 256\,\frac{495}{4096} = 30.9 = 31 \text{ (say)}.$$

Example 26.40. *In sampling a large number of parts manufactured by a machine, the mean number of defectives in a sample of 20 is 2. Out of 1000 such samples, how many would be expected to contain at least 3 defective parts.* *(V.T.U., 2004)*

Solution. Mean number of defectives $= 2 = np = 20p$.

$\therefore$ The probability of a defective part is $p = 2/20 = 0.1$.

and the probability of a non-defective part $= 0.9$

$\therefore$ The probability of at least three defectives in a sample of 20.

$$= 1 - \text{(prob. that either none, or one, or two are non-defective parts)}$$
$$= 1 - [{}^{20}C_0(0.9)^{20} + {}^{20}C_1(0.1)(0.9)^{19} + {}^{20}C_2(0.1)^2(0.9)^{18}]$$
$$= 1 - (0.9)^{18} \times 4.51 = 0.323.$$

Thus *the number of samples having at least three defective parts* out of 1000 samples

$$= 1000 \times 0.323 = 323.$$

Example 26.41. *The following data are the number of seeds germinating out of 10 on damp filter paper for 80 sets of seeds. Fit a binomial distribution to these data :*

x:	0	1	2	3	4	5	6	7	8	9	10
f:	6	20	28	12	8	6	0	0	0	0	0

Solution. Here $n = 10$ and $N = \Sigma f_i = 80$

$$\therefore \quad \text{mean} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{20 + 56 + 36 + 32 + 30}{80} = \frac{174}{80} = 2.175$$

Now the mean of a binomial distribution $= np$

i.e., $\quad np = 10p = 2.175 \quad \therefore \quad p = 0.2175, q = 1 - p = 0.7825$

Hence the binomial distribution to be fitted is

$$N(q+p)^n = 80\,(0.7825 + 0.2175)^{10}$$
$$= 80.{}^{10}C_0\,(0.7825)^{10} + 80.{}^{10}C_1\,(0.7825)^9\,(0.2175)^1 + {}^{10}C_2\,(0.7825)^8\,(0.2175)^2 + \ldots + {}^{80}C_9\,(0.7825)^1\,(0.2175)^9 + {}^{80}C_{10}\,(0.2175)^{10}$$
$$= 6.885 + 19.13 + 23.94 + \ldots + 0.0007 + 0.00002$$

$\therefore$ the successive terms in the expansion give the expected or theoretical frequencies which are

x:	0	1	2	3	4	5	6	7	8	9	10
f:	6.9	19.1	24.0	17.8	8.6	2.9	0.7	0.1	0	0	0

PROBLEMS 26.5

1. Determine the binomial distribution for which mean = 2 (variance) and mean + variance = 3. Also find $P(X \le 3)$. *(Kerala, 2005)*
2. (*a*) An ordinary six-faced die is thrown four times. What are the probabilities of obtaining 4, 3, 2, 1 and 0 faces ?
   (*b*) A die is thrown five times. If getting an odd number is a success, find the probability of getting at least four successes. *(U.P.T.U., 2012)*
3. If the chance that one of the ten telephone lines is busy at an instant is 0.2.
   (*a*) What is the chance that 5 of the lines are busy ?
   (*b*) What is the most probable number of busy lines and what is the probability of this number ?
   (*c*) What is the probability that all the lines are busy ? *(V.T.U., 2002 S)*
4. If the probability that a new-born child is a male is 0.6, find the probability that in a family of 5 children there are exactly 3 boys. *(Kurukshetra, 2005)*

5. If on an average 1 vessel in every 10 is wrecked, find the probability that out of 5 vessels expected to arrive, at least 4 will arrive safely. (P.T.U., 2005)
6. The probability that a bomb dropped from a plane will strike the target is 1/5. If six bombs are dropped, find the probability that (*i*) exactly two will strike the target, (*ii*) at least two will strike the target.
7. A sortie of 20 aeroplanes is sent on an operational flight. The chances that an aeroplane fails to return is 5%. Find the probability that (*i*) one plane does not return (*ii*) at the most 5 planes do not return, and (*iii*) what is the most probable number of returns ?
8. The probability that an entering student will graduate is 0.4. Determine the probability that out of 5 students (*a*) none (*b*) one and (*c*) at least one will graduate.
9. Out of 800 families with 5 children each, how many would you expect to have (*a*) 3 boys, (*b*) 5 girls, (*c*) either 2 or 3 boys ? Assume equal probabilities for boys and girls. (V.T.U., 2004)
10. If 10 per cent of the rivets produced by a machine are defective, find the probability that out of 5 rivets chosen at random (*i*) none will be defective, (*ii*) one will be defective, and (*iii*) at least two will be defective. (*Kurukshetra*, 2013)
11. In a bombing action there is 50% chance that any bomb will strike the target. Two direct hits are needed to destroy the target completely. How many bombs are required to be dropped to give a 99% chance or better of completely destroying the target. (V.T.U., 2003 S)
12. A product is 0.5% defective and is packed in cartons of 100. What percentage contains not more than 3 defectives ?
13. If in a lot of 500 solenoids 25 are defective, find the probability of 0, 1, 2, 3 defective solenoids in a random sample of 20 solenoids.
14. 500 articles were selected at random out of a batch containing 10,000 articles, and 30 were found to be defective. How many defectives articles would you reasonably expect to have in the whole batch ? (J.N.T.U., 2003)
15. Fit a binomial distribution for the following data and compare the theoretical frequencies with the actual ones : (*Bhopal*, 2006)

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

16. Fit a binomial distribution to the following frequency distribution :

(*a*) (U.P.T.U., 2011)

x :	0	1	3	4
y :	28	62	10	4

(*b*) (*Kurukshetra*, 2009 ; C.S.V.T.U., 2007)

x :	0	1	2	3	4	5	6
f :	13	25	52	58	32	16	4

26.15 (1) POISSON DISTRIBUTION*

It is a distribution related to the probabilities of events which are extremely rare, but which have a large number of independent opportunities for occurrence. The number of printing mistakes per page, number of accidents on a highway, number of dishonoured cheques at a bank, number of persons born blind per year in a large city and the number of deaths by horse kick in an army corps are some of the phenomena, in which this law is followed.

This distribution can be derived as a limiting case of the binomial distribution by making n very large and p very small, keeping np fixed (= m, say).

The probability of r successes in a binomial-distribution is

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

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

As $n \to \infty$, $p \to 0$ $(np = m)$, we have

$$P(r) = \frac{m^r}{r!} \underset{n\to\infty}{\text{Lt}} \frac{(1-m/n)^n}{(1-m/n)^r} = \frac{m^r}{r!} e^{-m}$$

so that the probabilities of 0, 1, 2..., r,... successes in a Poisson distribution are given by

$$e^{-m},\ me^{-m},\ \frac{m^2e^{-m}}{2!}, \ldots, \frac{m^re^{-m}}{r!}, \ldots$$

The sum of these probabilities is unity as it should be.

---

* It was discovered in 1837 by a French mathematician Simon Denius Poisson (1781 – 1840).

(2) Constants of the Poisson distribution. These constants can easily be derived from the corresponding constants of the binomial distribution simply by making $n \to \infty, p \to 0, (q \to 1)$ and noting that $np = m$

$$Mean = \text{Lt}\,(np) = m$$
$$\mu_2 = \text{Lt}\,(npq) = m\ \text{Lt}\,(q) = m$$

$\therefore$ $$Standard\ deviation = \sqrt{m}$$ (V.T.U., 2014)

Also $\mu_3 = m, \mu_4 = m + 3m^2$

$\therefore$ $$Skewness\ (= \sqrt{\beta_1}\,) = 1/m,\ Kurtosis\ (= \beta_2) = 3 + 1/m.$$

Since μ_3 is positive, Poisson distribution is positively skewed and since $\beta_2 > 3$, it is *Leptokurtic.*

(3) Applications of Poisson distribution. This distribution is applied to problems concerning :

(i) Arrival pattern of 'defective vehicles in a workshop', 'patients in a hospital' or 'telephone calls'.

(ii) Demand pattern for certain spare parts.

(iii) Number of fragments from a shell hitting a target.

(iv) Spatial distribution of bomb hits.

Example 26.42. *If the probability of a bad reaction from a certain injection is 0.001, determine the chance that out of 2,000 individuals more than two will get a bad reaction.* (V.T.U., 2008 ; Kottayam, 2005)

Solution. It follows a Poisson distribution as the probability of occurrence is very small.

Mean $m = np = 2000(0.001) = 2$

Probability that more than 2 will get a bad reaction

= 1 – [prob. that no one gets a bad reaction + prob. that one gets a bad reaction + prob. that two get bad reaction]

$$= 1 - \left[e^{-m} + \frac{m^1 e^{-m}}{1!} + \frac{m^2 e^{-m}}{2!}\right] = 1 - \left[\frac{1}{e^2} + \frac{2}{e^2} + \frac{2}{e^2}\right] \qquad [\because\ m = 2]$$

$$= 1 - \frac{5}{e^2} = 0.32. \qquad [\because\ e = 2.718\,]$$

Example 26.43. *In a certain factory turning out razor blades, there is a small chance of 0.002 for any blade to be defective. The blades are supplied in packets of 10, use Poisson distribution to calculate the approximate number of packets containing no defective, one defective and two defective blades respectively in a consignment of 10,000 packets.* (Kurukshetra, 2009 S; Madras, 2006 ; V.T.U., 2004)

Solution. We know that $m = np = 10 \times 0.002 = 0.02$

$$e^{-0.02} = 1 - 0.02 + \frac{(0.02)^2}{2!} - \ldots = 0.9802 \text{ approximately}$$

Probability of no defective blade is $e^{-m} = e^{-0.02} = 0.9802$

$\therefore$ no. of packets containing no defective blade is

$$10{,}000 \times 0.9802 = 9802$$

Similarly the number of packets containing one defective blade $= 10{,}000 \times me^{-m}$

$$= 10{,}000 \times (0.02) \times 0.9802 = 196$$

Finally the number of packets containing two defective blades

$$= 10{,}000 \times \frac{m^2 e^{-m}}{2!} = 10{,}000 \times \frac{(0.02)^2}{2!} \times 0.9802 = 2 \text{ approximately.}$$

Example 26.44. *Fit a Poisson distribution to the set of observations :*

x :	0	1	2	3	4
f :	122	60	15	2	1

(U.T.U., 2010 ; Bhopal, 2007 S ; V.T.U., 2004)

Solution. Mean $= \dfrac{\Sigma f_i x_i}{\Sigma f_i} = \dfrac{60 + 36 + 6 + 4}{200} = 0.5.$

∴ mean of Poisson distribution *i.e.*, $m = 0.5$.

Hence the theoretical frequency for r successes is

$$\frac{Ne^{-m}(m)^r}{r!} = \frac{200e^{-0.5}(.5)^r}{r!} \quad \text{where } r = 0, 1, 2, 3, 4$$

∴ the theoretical frequencies are

x:	0	1	2	3	4
f:	121	61	15	2	0

PROBLEMS 26.6

1. If a random variable has a Poisson distribution such that $P(1) = P(2)$, find (*i*) mean of the distribution. (*Rohtak, 2011 S*) (*ii*) $P(4)$. (*V.T.U., 20..*)
2. X is a Poisson variable and it is found that the probability that $X = 2$ is two-thirds of the probability that $X = 1$. Find the probability that $X = 0$ and the probability that $X = 3$. What is the probability that X exceeds 3?
3. For Poisson distribution, prove that $m\, \mu_2\, \gamma_1\, \gamma_2 = 1$, where symbols have their usual meanings. (*C.S.V.T.U., 20..*)
4. A certain screw making machine produces on average of 2 defective screws out of 100, and packs them in boxes of 500. Find the probability that a box contains 15 defective screws. (*Kurukshetra, 20..*)
5. Using Poisson distribution, find the probability that all of the spades will be drawn from a pack of well-shuffled cards atleast once in 104 consecutive trials. (Given $e^{-2} = 0.136$). (*U.T.U., 201..*)
6. A manufacturer knows that the condensers he makes contain on the average 1 % defectives. He packs them in boxes of 100. What is the probability that a box picked at random will contain 3 or more faulty condensers?
7. A car-hire firm has two cars which it hires out day by day. The number of demands for a car on each day is distributed as a Poisson distribution with mean 1.5. Calculate the proportion of days (*i*) on which there is no demand, (*ii*) on which demand is refused. ($e^{-1.5} = 0.2231$). (*Bhopal, 2008 S ; J.N.T.U., 200..*)
8. The incidence of occupational disease in an industry is such that the workmen have a 10% chance of suffering from it. What is probability that in a group of 7, five or more will suffer from it?
9. The frequency of accidents per shift in a factory is as shown in the following table :

Accidents per shift :	0	1	2	3	4
Frequency :	180	92	24	3	1

Calculate the mean number of accidents per shift and the corresponding Poisson distribution and compare with actual observations.

10. A source of liquid is known to contain bacteria with the mean number of bacteria per cubic centimetre equal to 3. Ten 1 c.c., test-tubes are filled with the liquid. Assuming that Poisson distribution is applicable, calculate the probability that all the test-tubes will show growth *i.e.*, contain atleast 1 bacterium each.
11. Find the expectation of the function $\phi(x) = xe^{-x}$ in a Poisson distribution. (*V.T.U., 2003*)

[Hint : If m be the mean of the Poisson distribution, then expectation of

$$\phi(x) = \sum_{x=0}^{\infty} \frac{\phi(x) \cdot m^x e^{-m}}{x!} = m \exp. m\,(e^{-1} - m - 1)$$

12. Fit a Poisson distribution to the following :

x:	0	1	2	3	4
f:	46	38	22	9	1

(*Rohtak, 2011 ; Bhopal, 2008 ; V.T.U., 2003 S*)

13. Fit a Poisson distribution to the following data given the number of yeast cells per square for 400 squares :

No. of cells per sq. :	0	1	2	3	4	5	6	7	8	9	10
No. of squares :	103	143	98	42	8	4	2	0	0	0	0

(*C.S.V.T.U., 2007*)

26.16 (1) NORMAL DISTRIBUTION*

Now we consider a continuous distribution of fundamental importance, namely the normal distribution. *Any quantity whose variation depends on random causes is distributed according to the normal law.* Its importance lies in the fact that a large number of distributions approximate to the normal distribution and as such it is a corner stone of modern statistics.

* In 1924, *Karl Pearson* found this distribution which *Abraham De Moivre* had discovered as early as 1733. It is also known as *Gaussian* distribution due to *Karl Friedrich Gauss*. See footnote p. 843, 647 and 37.

Let us define a variate $z = \dfrac{x - np}{\sqrt{(npq)}}$ ...(1)

...x is a binomial variate with mean np and S.D. $\sqrt{(npq)}$ so that z is a variate with mean zero and variance unity. ...the limit as n tends to infinity, the distribution of z becomes a continuous distribution extending from $-\infty$ to ∞.

It can be shown that *the limiting form of the binomial distribution (1) for large values of n when neither p ...is very small, is the normal distribution.* The normal curve is of the form

$$y = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-\mu)^2/2\sigma^2} \quad ...(2)$$

...μ and σ are the mean and standard deviation respectively.

(2) Properties of the normal distribution

I. The normal curve (2) is bell-shaped and is symmetrical about its mean. It is unimodal with ordinates ...sing rapidly on both sides of the mean (Fig. 26.3). The maximum ordinate is $1/\sigma\sqrt{(2\pi)}$, found by putting ... in (2).

As it is symmetrical, its mean, median and mode are the same. Its points of inflexion (found by putting ...$=0$ and verifying that at these points $d^3y/dx^3 \neq 0$) are given by $x = \mu \pm \sigma$, *i.e.*, these points are equidistant ... the mean on either side.

II. Mean deviation from the mean μ

$$= \int_{-\infty}^{\infty} |x-\mu| \cdot \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-\mu)^2/2\sigma^2}\, dx \qquad [\text{Put } z = (x-\mu)/\sigma]$$

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

$$= \frac{\sigma}{\sqrt{(2\pi)}} \left[\int_{-\infty}^{0} -ze^{-z^2/2}dz + \int_{0}^{\infty} ze^{-z^2/2}dz\right] = \frac{2\sigma}{\sqrt{(2\pi)}} \int_0^{\infty} ze^{-z^2/2}\, dz$$

$$= \frac{2\sigma}{\sqrt{(2\pi)}} \left| -e^{-z^2/2} \right|_0^{\infty} = -\sqrt{\left(\frac{2}{\pi}\right)}\, \sigma(0-1) = 0.7979\,\sigma \approx (4/5)\,\sigma$$

(*Rohtak, 2011 S*)

III. Moments about the mean

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

$$= \frac{\sigma^{2n+1}}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} z^{2n+1} e^{-z^2/2}\, dz \text{ where } z = (x-\mu)/\sigma$$

= 0, since the integral is an odd function.

Thus all odd order moments about the mean vanish.

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

[Integrate by parts]

$$= \frac{\sigma^{2n}}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} z^{2n-1} e^{-z^2/2} \,.\, zdz$$

$$= \frac{\sigma^{2n}}{\sqrt{(2\pi)}} \left[\left| -z^{2n-1} e^{-z^2/2} \right|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} (2n-1) z^{2n-2} e^{-z^2/2}\, dz \right]$$

$$= \frac{\sigma^{2n}}{\sqrt{(2\pi)}} (0-0) + (2n-1)\,\sigma^2 \mu_{2n-2}$$

Repeated application of this reduction formula, gives

$$\mu_{2n} = (2n-1)(2n-3) \ldots 3 \,.\, 1\sigma^{2n}$$

In particular, $\mu_2 = \sigma^2$, $\mu_4 = 3\sigma^4$.

Hence $\quad \beta_1 = \dfrac{{\mu_3}^2}{{\mu_2}^3} = 0$ and $\beta_2 = \dfrac{\mu_4}{{\mu_2}^2} = 3$

i.e., the coefficient of skewness is zero (*i.e.* the curve is symmetrical) and the Kurtosis is 3. This is the basis for the choice of the value 3 in the definitions of platykurtic and leptokurtic (page 844).

IV. The probability of x lying between x_1 and x_2 is given by the area under the normal curve from x_1 to x_2, *i.e.*, $P(x_1 \le x \le x_2)$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{x_1}^{x_2} e^{-(x-\mu)^2/2\sigma^2}\, dx$$

$$= \frac{1}{\sqrt{(2\pi)}} \int_{z_1}^{z_2} e^{-z^2/2}\, dz \text{ where } z = (x-\mu)/\sigma,\ dz = dx/\sigma \text{ and } z_1 = (x_1-\mu)/\sigma,\ z_2 = (x_2-\mu)/\sigma.$$

$$= \frac{1}{\sqrt{(2\pi)}} \left\{ \int_0^{z_2} e^{-z^2/2}\, dz - \int_0^{z_1} e^{-z^2/2}\, dz \right\} = P_2(z) - P_1(z)$$

The values of each of the above integrals can be found from the table IV–Appendix 2, which gives the values of

$$P(z) = \frac{1}{\sqrt{(2\pi)}} \int_0^{z} e^{-z^2/2}\, dz$$

for various values of z. This integral is called the *probability integral* or the *error function* due to its use in the theory of sampling and the theory of errors.

Using this table, we see that the area under the normal curve from $z = 0$ to $z = 1$, *i.e.* from $x = \mu$ to $\mu + \sigma$ is 0.3413.

∴ (*i*) The area under the normal curve between the ordinates $x = \mu - \sigma$ and $x = \mu + \sigma$ is 0.6826, ~ *68% nearly. Thus approximately 2/3 of the values lie within these limits.*

(*ii*) The area under the normal curve between $x = \mu - 2\sigma$ and $x = \mu + 2\sigma$ is 0.9544 ~ 95.5%, which implies that *about $4\frac{1}{2}$ % of the values lie outside these limits.*

(*iii*) *99.73% of the values lie between $x = \mu - 3\sigma$ and $x = \mu + 3\sigma$ i.e., only a quarter % of the whole lies outside these limits.*

(*iv*) *95% of the values lie between $x = \mu - 1.96\sigma$ and $x = \mu + 1.96\,\sigma$ i.e., only 5% of the values lie outside these limits.*

(*v*) *99% of the values lie between $x = \mu - 2.58\sigma$ and $x = \mu + 2.58\sigma$ i.e., only 1% of the values lie outside these limits.*

(*vi*) 99.9% of the values lie between $x = \mu - 3.29\sigma$ and $x = \mu + 3.29\sigma$.

In other words, a value that deviates more than σ from μ occurs about once in 3 trials. A value that deviates more than 2σ or 3σ from μ occurs about once in 20 or 400 trials. Almost all values lie within 3σ of the mean.

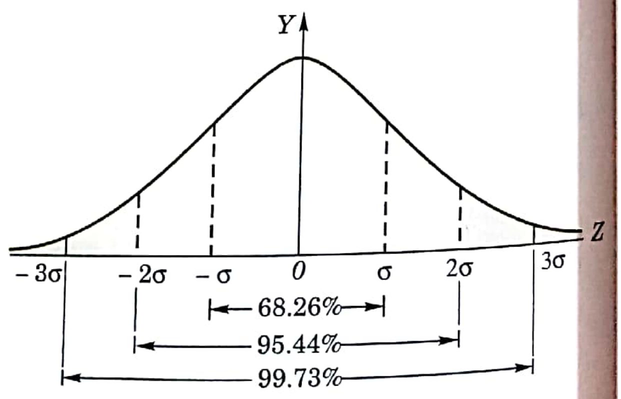


Fig. 26.4

The shape of the standardised normal curve is

$$y = \frac{1}{\sqrt{(2\pi)}} e^{-z^2/2} \text{ where } z = (x-\mu)/\sigma \qquad ...(3)$$

and the respective areas are shown in Fig. 26.4. 'z' is called a *normal variate*. Areas under the Normal curve are given in Table IV.

Obs. 50% area lies in the z-interval (– 0.745, 0.745) and 99% area lies in the z-interval (– 2.58, 2.58).

(3) **Normal frequency distribution.** We can fit a normal curve to any distribution. If N be the total *frequency*, μ the mean and σ the standard deviation of the given distribution then the curve

$$y = \frac{N}{\sigma\sqrt{(2\pi)}} e^{-(x-\mu)^2/2\sigma^2} \qquad ...(4)$$

will fit the given distribution as best as the data will permit. The frequency of the variate between x_1 and x_2 as *given by the fitted curve,* will be the area under (1) from x_1 to x_2.

(4) **Applications of normal distribution.** This distribution is applied to problems concerning :

(*i*) *Calculation of errors made by chance* in experimental measurements.

(*ii*) *Computation of hit probability of a shot.*

(*iii*) *Statistical inference* in almost every branch of science.