Illustration 4. The incidence of occupational disease in an industry is such that the workmen have a 25% chance of suffering from it. What is the probability that out of six workmen 4 or more will contact the disease ?

Solution. Let q denote chance of suffering and p chance of not suffering.

$\therefore\ q = 25\% = \frac{1}{4}$ and $p = \frac{3}{4}$.

The binomial expression is

$$(q+p)^6 = [(1/4) + (3/4)]^6$$
$$= q^6 + 6q^5p + 15q^4p^2 + 20q^3p^3 + 15q^2p^4 + 6qp^5 + p^6.$$

The probability of 4 or more (that is 4, 5 and 6 successes) is

$$= 15q^4p^2 + 6q^5p + q^6$$

Substituting the values we get

$$= 15\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^2 + 6\left(\frac{1}{4}\right)^5\left(\frac{3}{4}\right) + \left(\frac{1}{4}\right)^6$$

$$= \frac{15 \times 9}{4096} + \frac{6 \times 3}{4096} + \frac{1}{4096} = \frac{135 + 18 + 1}{4096} = \frac{154}{4096} = 0{\cdot}0376$$

Fitting a Binomial Distribution

When a binomial distribution is to be fitted to the observed data the following procedure is adopted :

1. Determine the values of p and q. If one of these values is known the other can be found out by the simple relationship $p = (1 - q)$, and $q = (1 - p)$. when p and q are equal the distribution is symmetric for p and q may be interchanged without altering the value of any term, and consequently terms equidistant from the two ends of the series are equal. If p and q are unequal, the distribution is skew. If p is less than $\frac{1}{2}$ the distribution is positively skewed and when p is more than $\frac{1}{2}$ the distribution is negatively skewed.

2. Expand the binomial $(q + p)^n$. The power n is equal to one less than the number of terms in the expanded binomial. Thus when two coins are tossed ($n = 2$) there will be three terms in the binomial. Similarly, when four coins are tossed ($n = 4$) there will be five terms, and so on.

3. Multiply each term of the expanded binomial by N (the total frequency) in order to obtain the expected frequency in each category.

The following example shall illustrate the procedure :

Illustration 5. Eight coins are tossed at a time, 256 times. Number of heads observed at each throw is recorded and the results are given below. Find the expected frequencies. What are the theoretical values of mean and standard deviation ? Calculate also the mean and S.D. of the observed frequencies.

No. of heads at a throw	*Frequency*	*No. of heads at a throw*	*Frequency*
0	2	5	56
1	6	6	32
2	30	7	10
3	52	8	1
4	67		

Solution. The chance of getting a head in a single throw of one coin is $\frac{1}{2}$.

Hence $p = \frac{1}{2}, q = \frac{1}{2}, n = 8, N = 256.$

$N = 256,\ n = 8$

$N(q+p)^n$

By expanding $256\left(\frac{1}{2}+\frac{1}{2}\right)^8$ we shall get the expected frequencies of 0, 1, 2, ——, 8 heads (successes).

No of heads (X)	Frequency $= N \times {}^nC_r q^{n-r} p^r$ $N \times {}^nC_r p q$
0	$256\left(\frac{1}{2}\right)^8 = 1$
1	$256 \times {}^8C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^7 = 8$
2	$256 \times {}^8C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^6 = 28$
3	$256 \times {}^8C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^5 = 56$
4	$256 \times {}^8C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^4 = 70$
5	$256 \times {}^8C_5\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^3 = 56$
6	$256 \times {}^8C_6\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^2 = 28$
7	$256 \times {}^8C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^1 = 8$
8	$256 \times \left(\frac{1}{2}\right)^8 = 1$
	Total = 256

The mean of the above distribution is $np = 8 \times \frac{1}{2} = 4$.

The standard deviaton is $\sqrt{npq} = \sqrt{\frac{1}{2} \times \frac{1}{2} \times 8} = \sqrt{2} = 1.414$.

These are the mean and standard deviation of the expected frequency distribution. The mean and standard deviation of the observed frequency distribution shall be :

X	f	d	fd	fd^2
0	2	–4	–8	32
1	6	–3	–18	54
2	30	–2	–60	120
3	52	–1	–52	52
4	67	0	0	0
5	56	+1	+56	56
6	32	+2	+64	128
7	10	+3	+30	90
8	1	+4	+4	16
	$N = 256$	$\Sigma d = 0$	$\Sigma fd = 16$	$\Sigma fd^2 = 548$

$$\bar{X} = A + \frac{\Sigma fd}{N} = 4 + \frac{16}{256} = 4.0625$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = \sqrt{\frac{548}{256} - \left(\frac{16}{256}\right)^2}$$

$$= \sqrt{2.141 - 0.004} = \sqrt{2.137} = 1.462.$$

Illustration 6. The following data show the number of seeds germinating out of 10 on a damp filter for 80 set of seeds. Fit a binomial distribution to this 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. FITTING BINOMIAL DISTRIBUTION

X	f	fX
0	6	0
1	20	20
2	28	56
3	12	36
4	8	32
5	6	30
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
	$N = 80$	$\Sigma fX = 174$

$$\bar{X} = \frac{174}{80} = 2.175$$

But mean $= np = \frac{174}{80}$ $\quad (\because n = 10)$

$$p = \frac{174}{800} = 0.2175$$

$$\therefore \quad q = 1 - p = 0.7825.$$

Hence the bionomial distribution to be fitted to the data is :

$$80(0.7825 + 0.2175)^{10}$$

The theoretical frequencies are the successive terms in the expansion of $80\,(.7825 + 0.2175)^{10}$ and are tabulated below :

X	Theoretical frequencies $N \times {}^nC_r q^{n-r} p^r$	fe
0	$80 \times (.7825)^{10}$	= 6.9
1	$80 \times 10\,(.7825)^9\,(.2175)^1$	= 19.1
2	$80 \times 45\,(.7825)^8\,(.2175)^2$	= 24.0
3	$80 \times 120\,(.7825)^7\,(.2175)^3$	= 17.8
4	$80 \times 210\,(.7825)^6\,(.2175)^4$	= 8.6
5	$80 \times 252\,(.7825)^5\,(.2175)^5$	= 2.9
6	$80 \times 210\,(.7825)^4\,(.2175)^6$	= 0.7
7	$80 \times 120\,(.7825)^3\,(.2175)^7$	= 0.1
8	$80 \times 45\,(.7825)^2\,(.2175)^8$	= 0.0
9	$80 \times 10\,(.7825)^1\,(.2175)^9$	= 0.0
10	$80 \times (.2175)^{10}$	= 0.0
	Total	= 80.1

12·3. POISSON DISTRIBUTION

Poisson distribution is a discrete probability distribution and is very widely used in statistical work. It was originated by a French mathematician, Simeon Denis Poisson (1781-1840) in 1837. Strictly speaking, the Poisson distribution is the limiting form of the binomial distribution as n becomes infinitely large