Chapter

9

BINOMIAL AND HYPERGEOMETRIC DISTRIBUTIONS

9.1 INTRODUCTION

In random experiments, we can divide a sample space into two parts. Fruit in a basket can be divided into two parts sweet and sour, in a throw of die the six faces can be divided into even and odd faces or prime and non-prime faces. When two dice are thrown, the total on the dice may be divided into even or odd. These experiments have only two possible outcomes. A very simple example of this type of experiment is throw of a coin which shows a head or a tail. The probability of each outcome is not disturbed by repeating such experiment. In tossing a coin large number of times, the probability of head remains constant and every time the result of coin is independent of previous throws. We apply the binomial distribution in all these situations in which sample space is divided into two parts according to some definition of partition. Independent trials with only two possible outcomes and constant probabilities are called binomial or, sometimes, Bernoulli Trials, after J. Bernoulli (1654 – 1705), who was a pioneer in the field of probability.

9.2 BERNOULLI TRIALS

A student selected from a class may be intelligent or non-intelligent, a bulb selected from a big lot may be good or defective, a person may be smoker or non-smoker. These trials have only two possible outcomes. One of these outcomes is called success and the other is called failure. The word success is not always used for outcomes like "intelligent", "good" or "useful". This word is used for the outcome in which we are interested. It may be used for a non-intelligent student or for a defective bulb. The probability of success is usually denoted by p and the probability of failure is denoted by q ($q = 1 - p$). The term Bernoulli trials is used for these trials provided the probability of success p does not change from trial to trial.

9.3 BINOMIAL EXPERIMENT

An experiment with n independent trials in which the outcomes can always be classified as either a success or a failure and the probability of success remains constant from trial to trial is called a binomial experiment. A binomial experiment has the following properties.

(i) Each trial results in two outcomes which can be classified into success and failure.

(ii) The probability of success remains constant from trial to trial.

(iii) The successive trials are independent.

(iv) The experiment is repeated a fixed number of times say, n.

9.4 BINOMIAL RANDOM VARIABLE

Let X be the total number of successes in n independent Bernoulli trials, with p as the probability of success in a single trial, then X is called the binomial random variable with parameters n and p.

9.5 THE BINOMIAL DISTRIBUTION

The binomial distribution was discovered early in the eighteenth century by the Swiss mathematician, J. Bernoulli (1654 – 1705). It was published after his death in 1713. If p is the probability that an event will happen in any single trial and $q = 1 - p$ is the probability that it will fail to happen in any single trial, then the probability that the event will happen exactly X times in n trials (i.e. x successes and $n - x$ failures will occur) is given by

$$P(X = x) = f(x) = b(x; n, p) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

where $x = 0, 1, 2, 3, \ldots, n$ and $n! = n(n-1)(n-2) \ldots 3.2.1.$ $0! = 1$

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, 3, ..., x, ..., n successes are the respective terms in the binomial expansion $(q + p)^n$. The binomial distribution contains two independent constants, n and p. If n and p are known then we can determine all measures like mean, variance, coefficient of skewness etc. of the distribution. They are called parameters of the binomial distribution. If $p = q = 1/2$, the binomial distribution is a symmetrical distribution and when $p \neq q$, it is a skewed distribution.

9.6 GRAPH OF THE BINOMIAL DISTRIBUTION

The probability function will be symmetrical when $p = 1/2$ (Figure 9.1.). If $p > 1/2$ (Figure 9.2), the probability function will be negatively skewed. If $p < 1/2$ (Figure 9.3), it will be positively skewed. The greater the difference between p and $1 - p$, the greater the skewness of the probability function. However, as n increases, the probability function approaches symmetry regardless of the difference between p and $1 - p$. The degree of skewness can be measured with the help of a formula called

coefficient of skewness $= \dfrac{q - p}{\sqrt{npq}}$.

Figure 9.1. Binomial distribution for $n = 4$, $p = 1/2$.

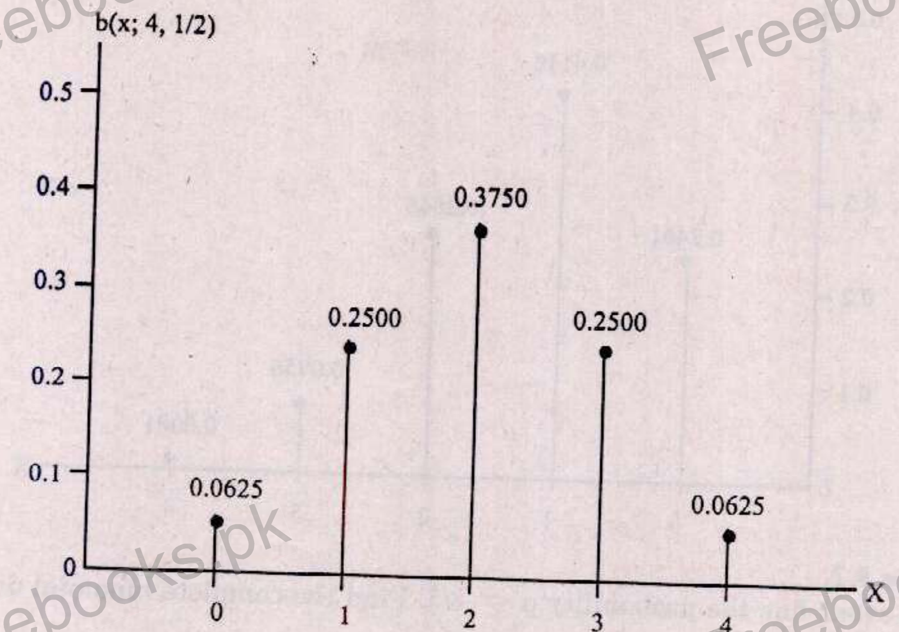


Figure 9.2. Binomial distribution for $n = 4$, $p = 7/10$.

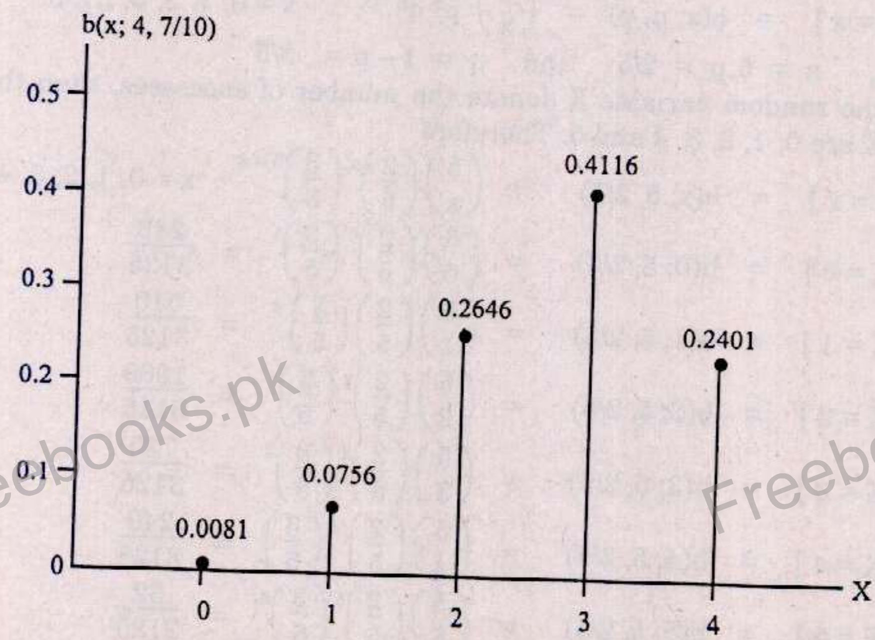
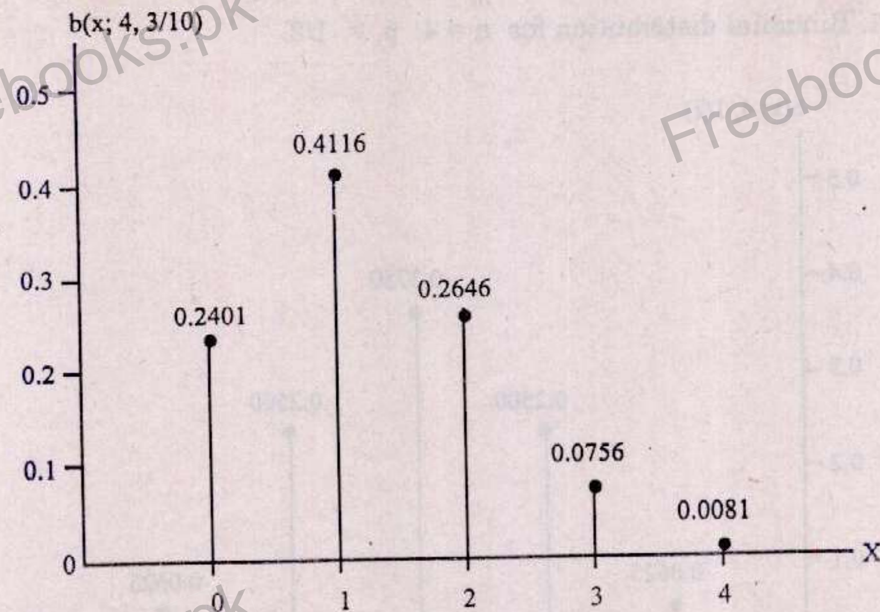


Figure 9.3. Binomial distribution for $n = 4$, $p = 3/10$.



Example 9.1.

An event has the probability $p = 2/5$. Find the complete binomial distribution for $n = 5$.

Solution:

The probability of x successes in a series of n trials is given by

$$P[X=x] = b(x; n, p) = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, 3, \ldots, n.$$

Here, $n = 5, p = 2/5$ and $q = 1 - p = 3/5$

Let the random variable X denote the number of successes, then the possible values of X are 0, 1, 2, 3, 4 and 5. Therefore

$$P[X=x] = b(x; 5, 2/5) = \binom{5}{x}\left(\frac{2}{5}\right)^x\left(\frac{3}{5}\right)^{5-x} \qquad x = 0, 1, 2, 3, 4, 5.$$

$$P[X=0] = b(0; 5, 2/5) = \binom{5}{0}\left(\frac{2}{5}\right)^0\left(\frac{3}{5}\right)^5 = \frac{243}{3125}$$

$$P[X=1] = b(1; 5, 2/5) = \binom{5}{1}\left(\frac{2}{5}\right)\left(\frac{3}{5}\right)^4 = \frac{810}{3125}$$

$$P[X=2] = b(2; 5, 2/5) = \binom{5}{2}\left(\frac{2}{5}\right)^2\left(\frac{3}{5}\right)^3 = \frac{1080}{3125}$$

$$P[X=3] = b(3; 5, 2/5) = \binom{5}{3}\left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^2 = \frac{720}{3125}$$

$$P[X=4] = b(4; 5, 2/5) = \binom{5}{4}\left(\frac{2}{5}\right)^4\left(\frac{3}{5}\right) = \frac{240}{3125}$$

$$P[X=5] = b(5; 5, 2/5) = \binom{5}{5}\left(\frac{2}{5}\right)^5\left(\frac{3}{5}\right)^0 = \frac{32}{3125}$$

Thus the complete binomial distribution with p = 2/5 and n = 5 in tabular form is given as follows:

x	0	1	2	3	4	5	Total
P[X = x]	$\frac{243}{3125}$	$\frac{810}{3125}$	$\frac{1080}{3125}$	$\frac{720}{3125}$	$\frac{240}{3125}$	$\frac{32}{3125}$	$\Sigma P[X = x] = 1$

Example 9.2.

The experience of a house-agent indicates that he can provide suitable accommodation for 75 percent of the clients who come to him. If on a particular occasion, 6 clients approach him independently, calculate the probability that:

(i) less than 4 clients will get satisfactory accommodation
(ii) exactly 4 clients will get satisfactory accommodation
(iii) at least 4 clients will get satisfactory accommodation
(iv) at most 4 clients will get satisfactory accommodation.

Solution:

The probability of x successes in a series of n trials is given by

$$P[X=x] = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, 3, \ldots, n.$$

Here, n = 6, p = 75/100 = 3/4 and q = 1 − p = 1/4

Let the random variable X denote the number of clients who will get satisfactory accommodation. Then the possible values of X are 0, 1, 2, 3, 4, 5, 6. Therefore

$$P[X=x] = \binom{6}{x}\left(\frac{3}{4}\right)^x\left(\frac{1}{4}\right)^{6-x} \qquad x = 0, 1, 2, 3, 4, 5, 6.$$

(i) $$P[X<4] = \binom{6}{0}\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^6 + \binom{6}{1}\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^5 + \binom{6}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^4 + \binom{6}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^3$$

$$= \frac{1}{4096} + \frac{18}{4096} + \frac{135}{4096} + \frac{540}{4096} = \frac{694}{4096} = 0.1694$$

(ii) $$P[X=4] = \binom{6}{4}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^2 = \frac{1215}{4096} = 0.2966$$

(iii) $$P[X\geq 4] = \binom{6}{4}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^2 + \binom{6}{5}\left(\frac{3}{4}\right)^5\left(\frac{1}{4}\right) + \binom{6}{6}\left(\frac{3}{4}\right)^6\left(\frac{1}{4}\right)^0$$

$$= \frac{1215}{4096} + \frac{1458}{4096} + \frac{729}{4096} = \frac{3402}{4096} = 0.8306$$

(iv) $$P[X\leq 4] = \binom{6}{0}\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^6 + \binom{6}{1}\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^5 + \binom{6}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^4 + \binom{6}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^3 + \binom{6}{4}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^2$$

$$= \frac{1}{4096} + \frac{18}{4096} + \frac{135}{4096} + \frac{540}{4096} + \frac{1215}{4096} = \frac{1909}{4096} = 0.4661$$

Example 9.3.

Large lots of incoming products at a manufacturing plant are inspected for defectives by means of a sampling scheme. Only 8 items are to be examined and the lot is rejected if 2 or more defectives are observed. If a lot contains 10 % defectives, what is probability that the lot will be:

(i) accepted? (ii) rejected?

Solution:

The probability of x successes in a series of n trials is given by

$$P[X=x] = b(x; n, p) = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, 3, \ldots, n.$$

Here, $n = 8, p = 10/100 = 1/10$ and $q = 1 - p = 9/10$

Let the random variable X denote the number of defective items, then the possible values of X are 0, 1, 2, 3, ..., 8. Therefore

$$P[X=x] = \binom{8}{x}\left(\frac{1}{10}\right)^x\left(\frac{9}{10}\right)^{8-x} \qquad x = 0, 1, 2, 3, \ldots, 8.$$

(i) $P[X<2] = \binom{8}{0}\left(\frac{1}{10}\right)^0\left(\frac{9}{10}\right)^8 + \binom{8}{1}\left(\frac{1}{10}\right)\left(\frac{9}{10}\right)^7$

$= 0.4305 + 0.3826 = 0.8131$

(ii) $P[X \geq 2] = 1 - P[X<2] = 1 - 0.8131 = 0.1869.$

Example 9.4

Assuming that each baby has probability 0.4 of being male, find the probability that a family of 4 children will have

(i) exactly one boy (ii) exactly one girl (iii) at least one boy
(iv) at least one girl (v) at most one boy (vi) at most one girl

Solution:

The probability of x successes in a series of n trials is given by

$$P(X=x) = b(x; n, p) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, 3, \cdots, n.$$

Here, $n = 4, p = 0.4, q = 1 - p = 1 - 0.4 = 0.6$

Let the random variable X denote the number of boys, then the possible values of X are 0, 1, 2, 3, 4. Therefore

$$P(X=x) = \binom{4}{x}(0.4)^x (0.6)^{4-x} \quad x = 0, 1, 2, 3, 4.$$

(i) $P(X=1) = \binom{4}{1}(0.4)^1(0.6)^3 = 0.3456$

(ii) $P(X=3) = \binom{4}{3}(0.4)^3(0.6)^1 = 0.1536$

(iii) $P(X \geq 1) = 1 - P(X=0) = 1 - \binom{4}{0}(0.4)^0(0.6)^4 = 1 - 0.1296 = 0.8704$

(iv) $P(X \le 3) = 1 - P(X = 4) = 1 - \binom{4}{4}(0.4)^4 (0.6)^0 = 1 - 0.0256 = 0.9744$

(v) $P(X \le 1) = \binom{4}{0}(0.4)^0 (0.6)^4 + \binom{4}{1}(0.4)^1 (0.6)^3$

$= 0.1296 + 0.3456 = 0.4752$

(vi) $P(X \ge 3) = \binom{4}{3}(0.4)^3 (0.6)^1 + \binom{4}{4}(0.4)^4 (0.6)^0$

$= 0.1536 + 0.0256 = 0.1792$

Example 9.5.

Given n = 6, p = 1/3. Find

(a) $P[X = -1]$ (b) $P[X = 2.5]$ (c) $P[X = 2]$ (d) $P[X = 10]$.

Solution:

The probability of x successes in a series of n trials is given by

$$P[X = x] = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, 3, \ldots, n.$$

Here, $n = 6, p = 1/3, \quad q = 1 - p = 2/3$ and $x = 0, 1, 2, 3, 4, 5, 6$.

$$P[X = x] = \binom{6}{x}\left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{6-x} \qquad x = 0, 1, 2, 3, \ldots, 6.$$

(a) $P[X = -1] = 0$, because a random variable X in a binomial distribution takes only positive integral values.

(b) $P[X = 2.5] = 0$, because X can take only integer values 0, 1, 2, 3, 4, 5, 6.

(c) $P[X = 2] = \binom{6}{2}\left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^4 = \frac{240}{729}$

(d) $P[X = 10] = 0$, because X can take only values 0, 1, 2, 3, 4, 5, 6.

9.7 MEAN, VARIANCE AND STANDARD DEVIATION OF THE BINOMIAL DISTRIBUTION

The mean, variance and standard deviation of the binomial distribution, means the mean, variance and standard deviation of the values taken by the variable in repeated binomial experiments. Instead of carrying out these experiments in order to calculate the mean, variance and standard deviation, it can be shown mathematically that the following formulae may be applied to any binomial distribution.

$$\text{Mean} = \mu = \Sigma x\, p(x) = np$$

$$\text{Variance} = \sigma^2 = \Sigma x^2 p(x) - [\Sigma x\, p(x)]^2 = npq$$

$$\text{S.D.} = \sigma = \sqrt{\Sigma x^2 p(x) - [\Sigma x\, p(x)]^2} = \sqrt{npq}$$

Example 9.6.

Find the mean, variance and standard deviation of the binomial distribution $(q+p)^2$.

Solution:

Here, $n = 2, \quad x = 0, 1, 2$. Therefore

$$P[X=x] = \binom{2}{x} p^x q^{2-x} \qquad x = 0, 1, 2.$$

$$P[X=0] = \binom{2}{0} p^0 q^2 = q^2 \qquad P[X=1] = \binom{2}{1} pq = 2pq$$

$$P[X=2] = \binom{2}{2} p^2 q^0 = p^2$$

Thus, the probability distribution in tabular form for the computation of mean, variance and standard deviation of X is given as follows:

x	$P[X=x] = p(x)$	$x\,p(x)$	x^2	$x^2\,p(x)$
0	q^2	0	0	0
1	$2pq$	$2pq$	1	$2pq$
2	p^2	$2p^2$	4	$4p^2$

$$E(X) = \mu = \Sigma x\, p(x) = 2pq + 2p^2$$

$$= 2p(q+p) = 2p(1) = 2p$$

$$E(X^2) = \Sigma x^2 p(x) = 2pq + 4p^2$$

$$Var(X) = \sigma^2 = E(X^2) - [E(X)]^2 = 2pq + 4p^2 - (2p)^2$$

$$= 2pq + 4p^2 - 4p^2 = 2pq$$

$$S.D(X) = \sigma = \sqrt{2pq}$$

Example 9.7.

Show that the mean = 3p, variance = 3pq and standard deviation = $\sqrt{3pq}$ for a binomial distribution in which n = 3.

Solution:

Here, $n = 3, \quad x = 0, 1, 2, 3$. Therefore

$$P[X=x] = \binom{3}{x} p^x q^{3-x} \qquad x = 0, 1, 2, 3.$$

$$P[X=0] = \binom{3}{0} p^0 q^3 = q^3 \qquad P[X=1] = \binom{3}{1} pq^2 = 3pq^2$$

$$P[X=2] = \binom{3}{2} p^2 q = 3p^2 q \qquad P[X=3] = \binom{3}{3} p^3 q^0 = p^3$$

Thus, the probability distribution in tabular form for the computation of mean. variance and standard deviation of X is given as follows:

x	$P[X=x]=p(x)$	$x\,p(x)$	x^2	$x^2\,p(x)$
0	q^3	0	0	0
1	$3pq^2$	$3pq^2$	1	$3pq^2$
2	$3p^2q$	$6p^2q$	4	$12p^2q$
3	p^3	$3p^3$	9	$9p^3$

$$E(X) = \mu = \Sigma x\,p(x) = 3pq^2 + 6p^2q + 3p^3 = 3p(q^2 + 2pq + p^2)$$
$$= 3p(q+p)^2 = 3p(1)^2 = 3p$$
$$E(X^2) = \Sigma x^2\,p(x) = 3pq^2 + 12p^2q + 9p^3 = 3p(q^2 + 4pq + 3p^2)$$
$$= 3p[\,q^2 + 2pq + p^2 + 2pq + 2p^2\,] = 3p[\,(q+p)^2 + 2p(q+p)\,]$$
$$= 3p[\,(1)^2 + 2p(1)\,] = 3p[\,1 + 2p\,] = 3p + 6p^2$$
$$Var(X) = E(X^2) - [\,E(X)\,]^2 = 3p + 6p^2 - (3p)^2 = 3p + 6p^2 - 9p^2$$
$$= 3p - 3p^2 = 3p(1-p) = 3pq$$
$$S.D(X) = \sigma = \sqrt{3pq}$$

Example 9.8.

Find the mean, variance and standard deviation of the binomial distribution $(q+p)^4$.

Solution:

Here, $n = 4$, $x = 0, 1, 2, 3, 4$. Therefore

$$P[X=x] = \binom{4}{x} p^x q^{4-x} \qquad x = 0, 1, 2, 3, 4.$$

$P[X=0] = q^4$ $\quad$ $P[X=1] = 4pq^3$ $\quad$ $P[X=2] = 6p^2q^2$

$P[X=3] = 4p^3q$ $\quad$ $P[X=4] = p^4$

Thus, the probability distribution in tabular form for the computation of mean, variance and standard deviation of X is given as follows:

x	$P[X=x]=p(x)$	$x\,p(x)$	x^2	$x^2\,p(x)$
0	q^4	0	0	0
1	$4pq^3$	$4pq^3$	1	$4pq^3$
2	$6p^2q^2$	$12p^2q^2$	4	$24p^2q^2$
3	$4p^3q$	$12p^3q$	9	$36p^3q$
4	p^4	$4p^4$	16	$16p^4$

$$E(X) = \mu = \Sigma x\, p(x) = 4pq^3 + 12p^2q^2 + 12p^3q + 4p^4$$
$$= 4p(q^3 + 3pq^2 + 3p^2q + p^3) = 4p(q + p)^3 = 4p(1)^3 = 4p$$
$$E(X^2) = \Sigma x^2\, p(x) = 4pq^3 + 24p^2q^2 + 36p^3q + 16p^4$$
$$= 4p[\, q^3 + 6pq^2 + 9p^2q + 4p^3]$$
$$= 4p[\, q^3 + 3pq^2 + 3p^2q + p^3 + 3pq^2 + 6p^2q + 3p^3]$$
$$= 4p[\, (q + p)^3 + 3p(q^2 + 2pq + p^2)]$$
$$= 4p[\, (q + p)^3 + 3p(q + p)^2] = 4p[\, (1)^3 + 3p(1)^2]$$
$$= 4p[\, 1 + 3p\,] = 4p + 12p^2$$
$$Var(X) = \sigma^2 = E(X)^2 - [E(X)]^2 = 4p + 12p^2 - (4p)^2$$
$$= 4p + 12p^2 - 16p^2 = 4p - 4p^2 = 4p(1 - p) = 4pq$$
$$S.D.(X) = \sigma = \sqrt{4pq}$$

Example 9.9.

Find the mean, variance and standard deviation of the binomial distribution $(q + p)^n$.

Solution:

The number of successes x will be 0, 1, 2, 3, 4, ..., n. We know that

$$(q + p)^n = \binom{n}{0} p^0q^n + \binom{n}{1} pq^{n-1} + \binom{n}{2} p^2q^{n-2} + \binom{n}{3} p^3q^{n-3} + \binom{n}{4} p^4q^{n-4} + \dots + \binom{n}{n} p^nq^0$$

Therefore the probabilities of 0, 1, 2, 3, 4, ..., n successes are respectively,

$$q^n, \binom{n}{1} pq^{n-1}, \binom{n}{2} p^2q^{n-2}, \binom{n}{3} p^3q^{n-3}, \binom{n}{4} p^4q^{n-4}, \dots, p^n$$

Thus, the probability distribution in tabular form for the computation of mean, variance and standard deviation of X is given as follows:

x	$P[X = x] = p(x)$	$x\,p(x)$	x^2	$x^2\,p(x)$
0	q^n	0	0	0
1	$\binom{n}{1} pq^{n-1}$	$\binom{n}{1} pq^{n-1}$	1	$\binom{n}{1} pq^{n-1}$
2	$\binom{n}{2} p^2q^{n-2}$	$2\binom{n}{2} p^2q^{n-2}$	4	$4\binom{n}{2} p^2q^{n-2}$
3	$\binom{n}{3} p^3q^{n-3}$	$3\binom{n}{3} p^3q^{n-3}$	9	$9\binom{n}{3} p^3q^{n-3}$
4	$\binom{n}{4} p^4q^{n-4}$	$4\binom{n}{4} p^4q^{n-4}$	16	$16\binom{n}{4} p^4q^{n-4}$
⋮	⋮	⋮	⋮	⋮
n	p^n	np^n	n^2	n^2p^n

$$E(X) = \mu = \Sigma x\, p(x) = \binom{n}{1} pq^{n-1} + 2\binom{n}{2} p^2q^{n-2} + 3\binom{n}{3} p^3q^{n-3} + 4\binom{n}{4} p^4q^{n-4} + \dots + np^n$$

$$= npq^{n-1} + 2\,\frac{n(n-1)}{2!}\,p^2q^{n-2} + 3\,\frac{n(n-1)(n-2)}{3!}\,p^3q^{n-3} + 4\,\frac{n(n-1)(n-2)(n-3)}{4!}\,p^4q^{n-4} + \dots + np^n$$

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

$$= np\left[q^{n-1} + \binom{n-1}{1} pq^{n-2} + \binom{n-1}{2} p^2q^{n-3} + \binom{n-1}{3} p^3q^{n-4} + \dots + p^{n-1}\right]$$

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

$$E(X^2) = \Sigma x^2\, p(x) = \binom{n}{1} pq^{n-1} + 4\binom{n}{2} p^2q^{n-2} + 9\binom{n}{3} p^3q^{n-3} + 16\binom{n}{4} p^4q^{n-4} + \dots + n^2p^n$$

$$= npq^{n-1} + 4\,\frac{n(n-1)}{2!}\,p^2q^{n-2} + 9\,\frac{n(n-1)(n-2)}{3!}\,p^3q^{n-3} + 16\,\frac{n(n-1)(n-2)(n-3)}{4!}\,p^4q^{n-4} + \dots + n^2p^n$$

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

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

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

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

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

$$= np\left[1 + (n-1)p\right] = np[\,1 + np - p\,] = np + n^2p^2 - np^2$$

$$\text{Var}(X) = \sigma^2 = E(X^2) - [E(X)]^2 = np + n^2p^2 - np^2 - (np)^2$$

$$= np + n^2p^2 - np^2 - n^2p^2 = np - np^2 = np(1-p) = npq$$

$$\text{S.D}(X) = \sigma = \sqrt{npq}$$

Alternative method

The binomial random variable X with parameters n and p has the probability distribution

$$P[X = x] = b(x;n,p) = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, 3, \ldots, n.$$

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

$$= \sum_{x=0}^{n} x\frac{n!}{x!(n-x)!}\,p^xq^{n-x} = \sum_{x=0}^{n} x\frac{n(n-1)!}{x(x-1)!\,(n-x)!}\,pp^{x-1}q^{n-x}$$

$$= np\sum_{x=1}^{n}\frac{(n-1)!}{(x-1)!\,(n-x)!}\,p^{x-1}q^{n-x} = np\sum_{x=1}^{n}\binom{n-1}{x-1} p^{x-1}q^{n-x}$$

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

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

$$E(X^2) = E[X(X-1)] + E(X) = E[X(X-1)] + np$$

Where
$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1)\,p(x) = \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)\,\frac{n!}{x!(n-x)!}\,p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)\,\frac{n(n-1)\,(n-2)!}{x(x-1)\,(x-2)!\,(n-x)!}\,p^2p^{x-2}q^{n-x}$$

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

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

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

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

So, $E(X^2) = n(n-1)p^2 + np = n^2p^2 - np^2 + np$

$Var(X) = \sigma^2 = E(X^2) - [E(X)]^2 = n^2p^2 - np^2 + np - (np)^2$

$= n^2p^2 - np^2 + np - n^2p^2 = np - np^2 = np(1-p) = npq$

$S.D(X) = \sigma = \sqrt{npq}$

Example 9.10.

If X is a binomial random variable with $E(X) = 1.44$ and $S.D(X) = 0.96$. Find the parameters of the binomial distribution. Also find $P[X = 2]$.

Solution:

Here, $E(X) = np = 1.44$, $S.D(X) = \sqrt{npq} = 0.96$

$Var(X) = npq = (0.96)^2 = 0.9216$

$1.44\,q = 0.9216$ (since $np = 1.44$)

or $q = \dfrac{0.9216}{1.44} = 0.64$ and $p = 1 - q = 0.36$

$np = 1.44$ or $n(0.36) = 1.44$ (since $p = 0.36$)

$n = \dfrac{1.44}{0.36} = 4$

$P[X = 2] = \binom{4}{2}(0.36)^2(0.64)^2 = 0.3185$

Hence, $n = 4$, $p = 0.36$ and $P(X = 2) = 0.3185$.

Example 9.11.

Is it possible to have a binomial distribution with mean = 10 and standard deviation = 6?

Solution:

Here, Mean $= \mu = np = 10$

$S.D(X) = \sigma = \sqrt{npq} = 6$

$Var(X) = \sigma^2 = npq = (6)^2 = 36$

$10\,q = 36$ (since $np = 10$) or $q = \dfrac{36}{10} = 3.6$

This is impossible because q is a probability which can never exceed one. Hence the given values of mean and standard deviation are wrong.

Example 9.12.

Given $n = 5$, $P(X = 1) = 5/32$ and $P(X = 2) = 10/32$. Find $P(X = 0)$ and $P(X = 3)$. Also find coefficient of skewness.

Solution:

$P[X = 1] = \binom{5}{1} pq^4 = 5pq^4 = \dfrac{5}{32}$ ...... (1)

$P[X = 2] = \binom{5}{2} p^2q^3 = 10p^2q^3 = \dfrac{10}{32}$ ...... (2)

Dividing equation (2) by equation (1)

$$\frac{10\,p^2q^3}{5\,pq^4} = \frac{10/32}{5/32} \text{ or } \frac{2p}{q} = 2 \text{ or } 2p = 2q \text{ or } p = q = \frac{1}{2}$$

$$P[X=0] = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

$$P[X=3] = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{10}{32}$$

Coefficient of skewness $= \frac{q-p}{\sqrt{npq}} = 0$, there is no skewness, thus the distribution is symmetrical.

9.8 PROPERTIES OF THE BINOMIAL DISTRIBUTION

Properties of the binomial distribution are listed in the following table.

Mean	$\mu = np$
Variance	$\sigma^2 = npq$
Standard Deviation	$\sigma = \sqrt{npq}$
Moment Coefficient of Skewness	$\gamma_1 = \sqrt{\beta_1} = \frac{q-p}{\sqrt{npq}}$
Moment Coefficient of Kurtosis	$\beta_2 = 3 + \frac{1-6pq}{npq}$

Note: Skewness is negative, zero or positive according as $p > 1/2$ or $p = 1/2$ or $p < 1/2$.

9.9 BINOMIAL FREQUENCY DISTRIBUTION

Let the n independent trials constitute one experiment and let this experiment be repeated N times. Then we expect x successes to occur $N\binom{n}{x}p^xq^{n-x}$ times. This will be called the expected frequency of x successes in N experiments and the possible number of successes together with the expected frequencies will be said to constitute the binomial (expected) frequency distribution. In practice, the observed frequencies will differ from the expected frequencies due to chance causes. For N sets, each of n trials the expected frequencies of 0, 1, 2, ..., n successes are given by the successive terms in the binomial expansion of $N(q+p)^n$, $q+p=1$.

9.10 FITTING OF THE BINOMIAL DISTRIBUTION

Suppose there are some intelligent and some non-intelligent students in a big college but we do not know the percentage of intelligent students. A random sample of 6 students is selected and the number of intelligent students in the sample is counted. This number will take any value between 0 to 6. We repeat the sample a large number of times say 100 and each time note the number of intelligent students. The frequencies of 0, 1, 2, ..., 6 intelligent students as observed in this experiment are, say, 14, 15, 25, 25, 10, 6, 5. The sample data can be written as below:

No. of intelligent students (X)	0	1	2	3	4	5	6
Frequency (f)	14	15	25	25	10	6	5

The sample mean $(\bar{X}) = \frac{\Sigma f X}{\Sigma f} = \frac{240}{100} = 2.4$. This observed data is also called experimental or actual data. From this data we want to estimate the proportion of intelligent students in the college. According to the theory of estimation, we find the sample mean = 2.4 and substitute it equal to np when n = 6. Thus 6p = 2.4 and p = 0.4. This value of p is called estimate of proportion of intelligent students in the college. When p = 0.4 then q = 1 – p = 1 – 0.4 = 0.6. Using n = 6, p = 0.4 and q = 0.6 we can find the probabilities of 0, 1, 2, ..., 6 intelligent students as given below. Each probability multiplied with 100 will give us expected frequencies of 0, 1, 2, ..., 6 intelligent students. These frequencies are called frequencies of the binomial distribution.

No. of intelligent students (X)	Probability $p(x) = \binom{6}{x}(0.4)^x(0.6)^{6-x}$	Expected Frequencies N.p(x)
0	$\binom{6}{0}(0.4)^0(0.6)^6 = 0.046656$	4.67
1	$\binom{6}{1}(0.4)(0.6)^5 = 0.186624$	18.66
2	$\binom{6}{2}(0.4)^2(0.6)^4 = 0.311040$	31.10
3	$\binom{6}{3}(0.4)^3(0.6)^3 = 0.276480$	27.65
4	$\binom{6}{4}(0.4)^4(0.6)^2 = 0.138240$	13.82
5	$\binom{6}{5}(0.4)^5(0.6) = 0.036864$	3.69
6	$\binom{6}{6}(0.4)^6(0.6)^0 = 0.004096$	0.41
Total	1	100

This procedure of finding expected frequencies of the binomial distribution is called fitting of the binomial distribution. It is basically the method of estimating probability of success p(parameter of population) from the sample data and then use it to find the expected frequencies (most likely) of the given experiment.

Example 9.13.

The probability of male birth is equal to the probability of female birth. Out of 400 families with 4 children each, find the expected number of families with 0, 1, 2, 3 and 4 males.

Solution:

The binomial frequency distribution is

$$N.P[X = x] = N.b(x; n, p) = N\binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, 3, \ldots, n.$$

Here, N = 400, n = 4, p = 1/2, q = 1 − p = 1/2 and x = 0, 1, 2, 3, 4.

$$400\, P[X = 0] = 400\binom{4}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 = 25 \text{ Families}$$

$$400\, P[X = 1] = 400\binom{4}{1}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 = 100 \text{ Families}$$

$$400\, P[X = 2] = 400\binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = 150 \text{ Families}$$

$$400\, P[X = 3] = 400\binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) = 100 \text{ Families}$$

$$400\, P[X = 4] = 400\binom{4}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = 25 \text{ Families}$$

Example 9.14.

Six dice are thrown 729 times. How many times do you expect at least three dice to show a five or a six?

Solution:

The binomial frequency distribution is

$$N.P[X = x] = N.b(x; n, p) = N\binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, 3, \ldots, n.$$

Here, p = 1/3, q = 1 − p = 2/3, n = 6 and N = 729.

Let the random variable X denote the number of dice to show a five or a six, then the possible values of X are 0, 1, 2, 3, 4, 5, 6. Therefore

$$729\, P[X = x] = 729\binom{6}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{6-x} \quad x = 0, 1, 2, \ldots, 6.$$

The expected number of times a least three dice will show a 5 or a 6 is

$$= 729\binom{6}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^3 + 729\binom{6}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2 + 729\binom{6}{5}\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)$$

$$+ 729\binom{6}{6}\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^0$$

$$= 729\left(\frac{160}{729}\right) + 729\left(\frac{60}{729}\right) + 729\left(\frac{12}{729}\right) + 729\left(\frac{1}{729}\right) = 233$$

Example 9.15.

Fit a binomial distribution to the following data, obtained by tossing a biased coin 5 times.

No. of heads	0	1	2	3	4	5	Total
Frequency	12	56	74	39	18	1	200

Solution: The value of p is not given. Let us first calculate p.

No. of heads (X)	0	1	2	3	4	5	Total
Frequency (f)	12	56	74	39	18	1	200
Product (f X)	0	56	148	117	72	5	398

Here, $\Sigma f = 200, \quad \Sigma fX = 398, \quad n = 5$

$$\bar{X} = np = \frac{\Sigma fX}{\Sigma f} = \frac{398}{200} = 1.99$$

or $5p = 1.99$ (since $n = 5$)

or $p = \frac{1.99}{5} = 0.398$ and $q = 1 - p = 1 - 0.398 = 0.602$

Hence the binomial distribution to be fitted to the data is

$$200P[X = x] = 200\binom{5}{x}(0.398)^x\,(0.602)^{5-x}$$

The expected or theoretical frequencies of 0, 1, 2, 3, 4, 5 successes are calculated as below:

No. of heads (X)	Probability $P[X = x]$	Expected Frequencies $N.P[X = x]$
0	$\binom{5}{0}(0.398)^0\,(0.602)^5 = 0.079065$	15.81 or 16
1	$\binom{5}{1}(0.398)\,(0.602)^4 = 0.261360$	52.27 or 52
2	$\binom{5}{2}(0.398)^2\,(0.602)^3 = 0.345586$	69.12 or 69
3	$\binom{5}{3}(0.398)^3\,(0.602)^2 = 0.228477$	45.70 or 46
4	$\binom{5}{4}(0.398)^4\,(0.602) = 0.075526$	15.10 or 15
5	$\binom{5}{5}(0.398)^5\,(0.602)^0 = 0.009987$	2.00 or 2
Total	1	200