UNIT 9 BINOMIAL DISTRIBUTION

Structure

9.1 INTRODUCTION

In Unit 5 of the Course, you have studied random variables, their probability functions and distribution functions. In Unit 8 of the Course, you have come to know as to how the expectations and moments of random variables are obtained. In those units, the definitions and properties of general discrete and probability distributions have been discussed.

The present block is devoted to the study of some special discrete distributions and in this list, Bernoulli and Binomial distributions are also included which are being discussed in the present unit of the course.

Sec. 9.2 of this unit defines Bernoulli distribution and its properties. Binomial distribution and its applications are covered in Secs. 9.3 and 9.4 of the unit.

Objectives

Study of the present unit will enable you to:

- define the Bernoulli distribution and to establish its properties;
- define the binomial distribution and establish its properties;
- identify the situations where these distributions are applied;
- know as to how binomial distribution is fitted to the given data; and
- solve various practical problems related to these distributions.

9.2 BERNOULLI DISTRIBUTION AND ITS PROPERTIES

There are experiments where the outcomes can be divided into two categories with reference to presence or absence of a particular attribute or characteristic. A convenient method of representing the two is to designate either of them as success and the other as failure. For example, head coming up in the toss of a fair coin may be treated as a success and tail as failure, or vice-versa. Accordingly, probabilities can be assigned to the success and failure.

Suppose a piece of a product is tested which may be defective (failure) or non-defective (a success). Let p the probability that it found non-defective and q = 1 – p be the probability that it is defective. Let X be a random variable such that it takes value 1 when success occurs and 0 if failure occurs. Therefore,

$P[X=1]=p,$ and

$P[X=0]=q=1-p$.

The above experiment is a Bernoulli trial, the r.v. X defined in the above experiment is a Bernoulli variate and the probability distribution of X as specified above is called the Bernoulli distribution in honour of J. Bernoulli (1654-1705).

Definition

A discrete random variable X is said to follow Bernoulli distribution with parameter p if its probability mass function is given by

$$P[X=x]=\begin{cases} p^x(1-p)^{1-x} & ; x=0,1 \\ 0 & ; \text{elsewhere} \end{cases}$$

i.e. $P[X=1]=p^1(1-p)^{1-1}=p$ [putting x = 1]

and $P[X=0]=p^0(1-p)^{1-0}=1-p$ [putting x = 0]

The Bernoulli probability distribution, in tabular form, is given as

X	0	1
$p(x)$	$1-p$	p

Remark 1: The Bernoulli distribution is useful whenever a random experiment has only two possible outcomes, which may be labelled as success and failure.

Moments of Bernoulli Distribution

The r^{th} moment about origin of a Bernoulli variate X is given as

$$\mu_r' = E(X^r)$$

$$= \sum_{x=0}^{1} x^r p(x) \qquad \text{[See Unit 8 of this course]}$$

$$= (0)^r p(0) + (1)^r p(1)$$

$$= (0)(1-p) + (1)p$$

$$= p$$

$\Rightarrow \mu_1' = p,\ \mu_2' = p,\ \mu_3' = p,\ \mu_4' = p$.

Hence,

Mean = $\mu_1' = p$,

Variance $(\mu_2) = \mu_2' - (\mu_1')^2 = p - p^2 = p(1-p)$,

Third order central moment $(\mu_3) = \mu_3' - 3\mu_2'(\mu_1') + 2(\mu_1')^3$

$$= p - 3pp + 2(p)^3$$

$$= p - 3p^2 + 2p^3$$
$$= p\left(2p^2 - 3p + 1\right) = p\left(2p - 1\right)\left(p - 1\right)$$
$$= p\left(1 - p\right)\left(1 - 2p\right)$$

Fourth order central moment $(\mu_4) = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\left(\mu_1'\right)^2 - 3\left(\mu_1'\right)^4$

$$= p - 4p.p + 6p\left(p\right)^2 - 3\left(p\right)^4$$
$$= p - 4p^2 + 6p^3 - 3p^4$$
$$= p\left[1 - 4p + 6p^2 - 3p^3\right]$$
$$= p\left(1 - p\right)\left(1 - 3p + 3p^2\right)$$

[**Note:** For relations of central moments in terms of moments about origin, see Unit 3 of MST-002.]

Example 1: Let X be a random variable having Bernoulli distribution with paramete p = 0.4. Find its mean and variance.

Solution:

Mean = p = 0.4,
Variance = p(1 – p) = (0.4) (1 –0.4) = (0.4) (0.6) = 0.24

Single trial is taken into consideration in Bernoulli distribution. But, if trials are performed repeatedly a finite number of times and we are interested in the distribution of the sum of independent Bernoulli trials with the same probability of success in each trial, then we need to study binomial distribution which has been discussed in the next section.

9.3 BINOMIAL PROBABILITY FUNCTION

Here, is this section, we will discuss binomial distribution which was discovered by J. Bernoulli (1654-1705) and was first published eight years after his death i.e. in 1713 and is also known as “Bernoulli distribution for n trials”. Binomial distribution is applicable for a random experiment comprising a finite number (n) of independent Bernoulli trials having the constant probability of success for each trial.

Before defining binomial distribution, let us consider the following example: Suppose a man fires 3 times independently to hit a target. Let p be the probability of hitting the target (success) for each trial and $q\left(=1-p\right)$ be the probability of his failure.

Let S denote the success and F the failure. Let X be the number of successes in 3 trials,

P[X = 0] = Probability that target is not hit at all in any trial
= P [Failure in each of the three trials]

$$= P\left(F \cap F \cap F\right)$$
$$= P\left(F\right).P\left(F\right).P\left(F\right) \quad [\because \text{ trials are independent}]$$
$$= q.q.q$$
$$= q^3$$

This can be written as

$$\boxed{P[X=0] = {}^3C_0\, p^0 q^{3-0}}$$

$\left[\because\ {}^3C_0 = 1,\ p^0 = 1,\ q^{3-0} = q^3\right.$. Recall ${}^nC_x = \dfrac{\lfloor n}{\lfloor x \lfloor n-x}$ (see Unit 4 of MST-001)]

P[X = 1] = Probability of hitting the target once

= [(Success in the first trial and failure in the second and third trial) or (success in the second trial and failure in the first and third trials) or (success in the third trial and failure in the first two trials)]

$$= P\left[(S\cap F\cap F) \text{ or } (F\cap S\cap F) \text{ or } (F\cap F\cap S)\right]$$

$$= P(S\cap F\cap F) + P(F\cap S\cap F) + P(F\cap F\cap S)$$

$$= P(S).P(F).P(F) + P(F).P(S).P(F) + P(F).P(F).P(S)$$

[$\because$ trials are independent]

$$= p.q.q + q.p.q + q.q.p$$

$$= pq^2 + pq^2 + pq^2$$

$$= 3pq^2$$

This can also be written as

$$\boxed{P[X=1] = {}^3C_1 p^1 q^{3-1}} \qquad [\because\ {}^3C_1 = 3,\ p^1 = p,\ q^{3-1} = q^2]$$

P[X = 2] = Probability of hitting the target twice

= P[(Success in each of the first two trials and failure in the third trial) or (Success in first and third trial and failure in the second trial) or (Success in the last two trials and failure in the first trial)]

$$= P\left[(S\cap S\cap F)\cup(S\cap F\cap S)\cup(F\cap S\cap S)\right]$$

$$= P[S\cap S\cap F] + P[S\cap F\cap S] + P[F\cap S\cap S]$$

$$= P(S).P(S).P(F) + P(S).P(F).P(S) + P(F).P(S).P(S)$$

$$= p.p.q + p.q.p + q.p.p$$

$$= 3\,p^2 q$$

This can also be written as

$$\boxed{P[X=2] = {}^3C_2 p^2 q^{3-2}} \qquad [\because\ {}^3C_2 = 3, q^{3-2} = q]$$

P[X = 3] = Probability of hitting the target thrice

= [Success in each of the three trials]

$$= P[S\cap S\cap S]$$

$$= P(S).P(S).P(S)$$

$$= p.p.p$$

$$= p^3$$

This can also be written as

$\boxed{P[X=3] = {}^3C_3\, p^3 q^{3-3}}$ $\quad [\because\ {}^3C_3 = 1, q^{3-3} = 1]$

From the above four enrectangled results, we can write

$P[X=r] = {}^3C_r p^r q^{3-r}$; r = 0, 1, 2, 3.

which is the probability of r successes in 3 trials. ${}^3C_r$, here, is the number of ways in which r successes can happen in 3 trials.

The result can be generalized for n trials in the similar fashion and is given as
$P[X=r] = {}^nC_r p^r q^{n-r}$; r = 0, 1, 2,..., n.

This distribution is called the binomial probability distribution. The reason behind giving the name binomial probability distribution for this probability distribution is that the probabilities for x = 0, 1, 2, …, n are the respective probabilities ${}^nC_0 p^0 q^{n-0}$, ${}^nC_1 p^1 q^{n-1}$, ..., ${}^nC_n p^n q^{n-n}$ which are the successive terms of the binomial expansion $(q + p)^n$.

$[\because\ (q+p)^n = {}^nC_0 q^n p^0 + {}^nC_1 q^{n-1} p^1 + \ldots + {}^nC_n q^0 p^n\,]$

Binomial Expansion:
'Bi' means 'Two'. 'Binomial expansion' means 'Expansion of expression having two terms, e.g.

$$(X+Y)^2 = X^2 + 2XY + Y^2 = {}^2C_0 X^2 Y^0 + {}^2C_1 X^{2-1} Y^1 + {}^2C_2 X^{2-2} Y^2,$$

$$(X+Y)^3 = X^3 + 3X^2Y + 3XY^2 + Y^3$$

$$= {}^3C_0 X^3 Y^0 + {}^3C_1 X^{3-1} Y^1 + {}^3C_2 X^{3-2} Y^2 + {}^3C_3 X^{3-3} Y^3$$

So, in general,

$$(X+Y)^n = {}^nC_0\, X^n Y^0 + {}^nC_1\, X^{n-1} Y^1 + {}^nC_2 X^{n-2}\, Y^2 + \ldots + {}^nC_n X^{n-n} Y^n$$

The above discussion leads to the following definition.

Definition:

A discrete random variable X is said to follow binomial distribution with parameters n and p if it assumes only a finite number of non-negative integer values and its probability mass function is given by

$$P[X=x] = \begin{cases} {}^nC_x p^x q^{n-x}; & x = 0, 1, 2, \ldots, n \\ 0; & \text{elsewhere} \end{cases}$$

where, n is the number of independent trials,

x is the number of successes in n trials,

p is the probability of success in each trial, and

q = 1 – p is the probability of failure in each trial.

Remark 2:

i) The binomial distribution is the probability distribution of sum of n independent Bernoulli variates.

ii) If X is binomially distributed r.v. with parameters n and p, then we may write it as X ~ B(n, p).

iii) If X and Y are two binomially distributed independent random variables with parameters (n_1, p) and (n_2, p) respectively then their sum also follows a binomial distribution with parameters $n_1 + n_2$ and p. But, if the probability of success is not same for the two random variables then this property does not hold.

Example 2: An unbiased coin is tossed six times. Find the probability of obtaining

(i) exactly 3 heads

(ii) less than 3 heads

(iii) more than 3 heads

(iv) at most 3 heads

(v) at least 3 heads

(vi) more than 6 heads

Solution: Let p be the probability of getting head (success) in a toss of the coin and n be the number of trials.

$\therefore$ n = 6, p = $\frac{1}{2}$ and hence q = 1 – p = $1 - \frac{1}{2} = \frac{1}{2}$.

Let X be the number of successes in n trials,

$\therefore$ by binomial distribution, we have

$$P\left[X = x\right] = {}^nC_x p^x q^{n-x}; x = 0, 1, 2, ..., n$$

$$= {}^6C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{6-x} \;;\; x = 0, 1, 2, ..., 6$$

$$= {}^6C_x \left(\frac{1}{2}\right)^6 ; x = 0, 1, 2, ..., 6 .$$

$$= \frac{1}{64}.\,{}^6C_x \,;\, x = 0, 1, 2, ..., 6.$$

Therefore,

(i) P[exactly 3 heads] = P [X = 3]

$$= \frac{1}{64}\left({}^6C_3\right) = \frac{1}{64}\left[\frac{6\times5\times4}{3\times2}\right] = \frac{5}{16}$$

$\left[\because \text{ Recall } {}^nC_x = \frac{\lfloor n}{\lfloor x \lfloor n-x}\right.$ (see Unit 4 of MST- 001)]

(ii) P[less than 3 heads] = P[X < 3]

$$= P\left[X = 2 \text{ or } X = 1 \text{ or } X = 0\right]$$

$$= P\left[X = 2\right] + P\left[X = 1\right] + P\left[X = 0\right]$$

$$= \frac{1}{64}.\,{}^6C_2 + \frac{1}{64}.\,{}^6C_1 + \frac{1}{64}.\,{}^6C_0$$

$$= \frac{1}{64}\left[{}^6C_2 + {}^6C_1 + {}^6C_0\right] = \frac{1}{64}\left[\frac{6\times5}{2} + 6 + 1\right]$$

$$= \frac{22}{64} = \frac{11}{32}.$$

(iii) P[more than 3 heads] = P[X > 3]

$$= P[X = 4 \text{ or } X = 5 \text{ or } X = 6] \quad \left[\begin{array}{l}\because \text{ in 6 trials one can} \\ \text{have at most 6 heads}\end{array}\right]$$

$$= P[X = 4] + P[X = 5] + P[X = 6]$$

$$= \frac{1}{64}.{}^6C_4 + \frac{1}{64}.{}^6C_5 + \frac{1}{64}.{}^6C_6$$

$$= \frac{1}{64}\left[{}^6C_4 + {}^6C_5 + {}^6C_6\right]$$

$$= \frac{1}{64}\left[\frac{6\times5}{2} + 6 + 1\right] = \frac{22}{64} = \frac{11}{32}.$$

(iv) P[at most 3 heads] = P [3 or less than 3 heads]

$$= P[X = 3] + P[X = 2] + P[X = 1] + P[X = 0]$$

$$= \frac{1}{64}.{}^6C_3 + \frac{1}{64}.{}^6C_2 + \frac{1}{64}.{}^6C_1 + \frac{1}{64}.{}^6C_0$$

$$= \frac{1}{64}\left[{}^6C_3 + {}^6C_2 + {}^6C_1 + {}^6C_0\right]$$

$$= \frac{1}{64}\left[20 + 15 + 6 + 1\right] = \frac{42}{64} = \frac{21}{32}.$$

(v) P[at least 3 heads] = P[3 or more heads]

$$= P[X = 3] + P[X = 4] + P[X = 5] + P[X = 6]$$

or

$$= 1 - \left(P[X = 0] + P[X = 1] + P[X = 2]\right)$$

$$\left[\begin{array}{l}\because \text{ sum of probabilities of all possible} \\ \text{values of a random variable is 1}\end{array}\right]$$

$$= 1 - \left(\frac{11}{32}\right) \quad \left[\begin{array}{l}\text{Already obtained in} \\ \text{part (ii) of this example}\end{array}\right]$$

$$= \frac{21}{32}.$$

(vi) P [more than 6 heads] = P [7 or more heads]

$$= P\text{ [an impossible event]} \quad \left[\begin{array}{l}\because \text{ in six tosses, it} \\ \text{is impossible to get} \\ \text{more than six heads}\end{array}\right]$$

$$= 0$$

Example 3: The chances of catching cold by workers working in an ice factory during winter are 25%. What is the probability that out of 5 workers 4 or more will catch cold?

Solution: Let catching cold be the success and p be the probability of success for each worker.

$\therefore$ Here, n = 5, p = 0.25, q = 0.75 and by binomial distribution

$$P\left[X = x\right] = {}^{n}C_{x}\, p^{x} q^{n-x} \; ; \; x = 0, 1, 2, ..., n$$

$$= {}^{5}C_{x}\left(0.25\right)^{x}\left(0.75\right)^{5-x} ; \; 0, 1, 2, ..., 5$$

Therefore, the required probability = P[X ≥ 4]

$$= p\left[X = 4 \text{ or } X = 5\right]$$

$$= P\left[X = 4\right] + P\left[X = 5\right]$$

$$= {}^{5}C_{4}\left(0.25\right)^{4}\left(0.75\right)^{1} + {}^{5}C_{5}\left(0.25\right)^{5}\left(0.75\right)^{0}$$

$$= \left(5\right)\left(0.002930\right) + \left(1\right)\left(0.000977\right)$$

$$= 0.014650 + 0.000977$$

$$= 0.015627$$

Example 4: Let X and Y be two independent random variables such that X ~ B(4, 0.7) and Y ~ B(3, 0.7). Find $P[X + Y \le 1]$.

Solution: We know that if X and Y are independent random variables each following binomial distribution with parameters (n_1, p) and (n_2, p), then X + Y ~ B($n_1 + n_2$, p).

Therefore, here X + Y follows binomial distribution with parameters 4 + 3 and 0.7, i.e. 7 and 0.7. So, here, n = 7 and p = 0.7.

Thus, the required probability = $[X + Y \le 1]$

$$= P[X + Y=1] + P[X + Y= 0]$$

$$= {}^{7}C_{1}\left(0.7\right)^{1}\left(0.3\right)^{6} + {}^{7}C_{0}\left(0.7\right)^{0}\left(0.3\right)^{7}$$

$$= 7(0.7)(0.000729) + 1(1)(0.0002187)$$

$$= 0.0035721 + 0.0002187$$

$$= 0.0037908$$

Now, we are sure that you can try the following exercises:

E1) The probability of a man hitting a target is $\frac{1}{4}$. He fires 5 times. What is the probability of his hitting the target at least twice?

E2) A policeman fires 6 bullets on a dacoit. The probability that the dacoit will be killed by a bullet is 0.6. What is the probability that the dacoit is still alive?

9.4 MOMENTS OF BINOMIAL DISTRIBUTION

The r^{th} order moment about origin of a binomial variate X is given as

$$\mu_r' = E\left(X^r\right) = \sum_{x=0}^{n} x^r.P\left[X = x\right]$$

$$\therefore \quad \mu_1' = E\left(X\right) = \sum_{x=0}^{n} x\ .P\left[X = x\right]$$

$$= \sum_{x=0}^{n} x.\,{}^{n}C_x\, p^x q^{n-x} \qquad \left[\because P\left[X = x\right] = {}^{n}C_x p^x q^{n-x};\ x = 0, 1, 2, ..., n\right]$$

$$= \sum_{x=1}^{n} x.\,{}^{n}C_x\, p^x q^{n-x} \qquad \left[\begin{array}{l}\because \text{first term with } x = 0 \text{ will be zero} \\ \text{and hence we may start from } x = 1\end{array}\right]$$

$$= \sum_{x=1}^{n} x.\frac{n}{x}.\ {}^{n-1}C_{x-1}\, p^x q^{n-x}$$

$$\left[\begin{array}{l}\because\ {}^{n}C_x = \dfrac{\underline{|n}}{\underline{|x}\,\underline{|n-x}} = \dfrac{n\underline{|n-1}}{x\underline{|x-1}\,\underline{|(n-1)-(x-1)}} = \dfrac{n}{x}\ {}^{n-1}C_{x-1}, \\ \left(\text{see Unit 4 of MST} - 001\right)\end{array}\right]$$

$$= \sum_{x=1}^{n} n.\ {}^{n-1}C_{x-1}\, p^{x-1}.p^1.q^{(n-1)-(x-1)} \qquad [n - x = (n - 1) - (x - 1)]$$

$$= np\sum_{x=1}^{n}\ {}^{n-1}C_{x-1}\, p^{x-1}.q^{(n-1)-(x-1)}$$

$$= np\left[\begin{array}{r}{}^{n-1}C_0\, p^0\, q^{(n-1)-0} + {}^{n-1}C_1\, p^1\, q^{(n-1)-1} + {}^{n-1}C_2\, p^2 q^{(n-1)-2} + ... \\ + {}^{n-1}C_{n-1}.\,p^{n-1}\, q^{(n-1)-(n-1)}\end{array}\right]$$

$$= np \times \left[\begin{array}{c}\text{Sum of probabilities of all possible values of a} \\ \text{binomial variate with parameters } n-1 \text{ and } p\end{array}\right]$$

$$= np \times 1 \qquad \left[\begin{array}{l}\because \text{sum of probabilities of all possible} \\ \text{values of a random variable is 1}\end{array}\right]$$

$$= np.$$

$\therefore$ Mean = First order moment about origin

$= \mu_1'$

$= np.$

$$\boxed{\text{Mean} = np}$$

$$\mu_2' = E\left(X^2\right) = \sum_{x=0}^{n} x^2.P\left[X = x\right] = \sum_{x=0}^{n} x^2.\ {}^{n}C_x p^x q^{n-x}$$

Here, we will write x^2 as $x\left(x-1\right)+x$ $\quad [\because x\left(x-1\right)+x = x^2 - x + x = x^2]$

This is done because in the following expression, we get $x\left(x-1\right)$ in the denominator:

$$\left[\because \ {}^nC_x = \frac{\lfloor n}{\lfloor x \times \lfloor n-x} = \frac{n(n-1)\lfloor n-2}{x(x-1)\lfloor x-2 \times \lfloor (n-2)-(x-2)} = \frac{n(n-1)}{x(x-1)}.{}^{n-2}C_{x-2}\right]$$

$$\therefore \ \mu_2' = \sum_{x=0}^{n}\left[x(x-1)+x\right]{}^nC_x p^x q^{n-x}$$

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

$$= \left[\sum_{x=2}^{n} x(x-1)\,{}^nC_x p^x q^{(n-2)-(x-2)}\right] + \left(\mu_1'\right)$$

$$= \left[\sum_{x=2}^{n} x(x-1).\frac{n(n-1)}{x(x-1)}\ {}^{n-2}C_{x-2} p^x q^{n-x}\right] + \mu_1'$$

$$= \left[\sum_{x=2}^{n} n(n-1)\,{}^{n-2}C_{x-2} p^{x-2}.p^2 q^{(n-2)-(x-2)}\right] + \mu_1'$$

$$= \left[n(n-1)p^2 \sum_{x=2}^{n} {}^{n-2}C_{x-2} p^2 q^{(n-2)-(x-2)}\right] + \mu_1'$$

$$= n(n-1)p^2 \times \left[\begin{array}{c}\text{Sum of probabilities of all possible values of a} \\ \text{binomial variate with parameters } n-2 \text{ and } p\end{array}\right] + \mu_1'$$

$$= n(n-1)\,p^2\,(1) + np \qquad \left[\because \mu_1' = np\right]$$

$$= n^2 p^2 - np^2 + np$$

$\therefore$ Variance $(\mu_2) = \mu_2' - \left(\mu_1'\right)^2$ [See Unit 3 of MST-002]

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

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$= np - np^2$$

$$= np(1-p)$$

$$= npq$$

$\therefore$ $\boxed{\text{Variance} = npq}$

$$\mu_3' = \sum_{x=0}^{n} x^3.P\left[X = x\right]$$

Here, we will write x^3 as $x(x-1)(x-2)+3x(x-1)+x$

> Let $x^3 = x(x-1)(x-2)+Bx(x-1)+Cx$
> Comparing coefficients of x^2, we have
> $0 = -3 + B \Rightarrow B = 3$
> Comparing coeffs of x, we have
> $0 = 2 - B + C \Rightarrow C = B - 2 = 3 - 2 \Rightarrow C = 1$

$$\therefore \quad \mu_3' = \sum_{x=0}^{n}\left[x(x-1)(x-2)+3x(x-1)+x\right].{}^nC_x p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)(x-2)\,{}^nC_x p^x q^{n-x} + 3\sum_{x=0}^{n} x(x-1)\,{}^nC_x p^x q^{n-x} + \sum_{x=0}^{n} x.\,{}^nC_x p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)(x-2)\frac{n}{x}.\frac{n-1}{x-1}.\frac{n-2}{x-2}\,{}^{n-3}C_{x-3}p^x q^{n-x} + 3\left[n(n-1)p^2\right] + [np]$$

[The expression within brackets in the second term is the first term of R.H.S. in the derivation of μ_2' and the expression in the third term is μ_1' as already obtained.]

$$\left[\because {}^nC_x = \frac{\lfloor n}{\lfloor x \lfloor n-x} = \frac{n(n-1)(n-2)\lfloor n-3}{x(x-1)(x-2)\lfloor x-3 \lfloor (n-3)-(x-3)} = \frac{n(n-1)(n-2)}{x(x-1)(x-2)}.{}^{n-3}C_{x-3}\right]$$

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

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

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

$\therefore$ Third order central moment is given by

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\left(\mu_1'\right)^3$ [See Unit 4 of MST-002]

$= npq(q-p)$ [On simplification]

$$\boxed{\mu_3 = npq(q-p)}$$

$$\mu_4' = \sum_{x=0}^{n} x^4 P(X=x)$$

Writing

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

and proceeding in the similar fashion as for $\mu_1', \mu_2', \mu_3',$ we have

$$\mu_4' = n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np$$

and hence

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\left(\mu_1'\right)^2 - \left(\mu_1'\right)^4$$

$\mu_4 = npq\left[1 + 3(n-2)pq\right]$ [On simplification]

Now, recall the measures of skewness and kurtosis which you have studied in Unit 4 of MST-002

These measures are given as follows:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{[npq(q-p)]^2}{[npq]^3} = \frac{(q-p)^2}{npq},$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{npq[1+3(n-2)pq]}{[npq]^2} = 3 + \frac{1-6pq}{npq},$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{q-p}{\sqrt{npq}} = \frac{1-2p}{\sqrt{npq}}, \text{ and}$$

$$\gamma_2 = \beta_2 - 3 = \frac{1-6pq}{npq}$$

Remark 3:

(i) As $0 < q < 1$

$\Rightarrow q < 1$

$\Rightarrow npq < np$ [Multiplying both sides by $np > 0$]

$\Rightarrow$ Variance < Mean

Hence, for binomial distribution

Mean > Variance

(ii) As variance of $X \sim B(n, p)$ is npq,

$\therefore$ its standard deviation is $\sqrt{npq}$.

Example 4: For a binomial distribution with $p = \frac{1}{4}$ and $n = 10$, find mean and variance.

Solution: As $p = \frac{1}{4}$, $\therefore$ $q = 1 - \frac{1}{4} = \frac{3}{4}$.

Mean = $np = 10 \times \frac{1}{4} = \frac{5}{2}$,

Variance = $npq = 10 \times \frac{1}{4} \times \frac{3}{4} = \frac{15}{8}$.

Example 5: The mean and standard deviation of binomial distribution are 4 and $\frac{2}{\sqrt{3}}$ respectively. Find $P[X \geq 1]$.

Solution: Let $X \sim B(n, p)$, then

Mean = $np = 4$

and variance = $npq = \left(\frac{2}{\sqrt{3}}\right)^2$ [$\because$ S.D. = $\frac{2}{\sqrt{3}}$ and variance is square of S.D.]

Dividing second equation by the first equation, we have

$$\frac{npq}{np} = \frac{\frac{4}{3}}{4}$$

$$\Rightarrow q = \frac{1}{3}$$

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

Putting p = $\frac{2}{3}$ in the equation of mean, we have

$$n\left(\frac{2}{3}\right) = 4 \Rightarrow n = 6$$

$\therefore$ by binomial distribution,

$$P[X = x] = {}^{n}C_{x}p^{x}q^{n-x}$$

$$= {}^{6}C_{x}\left(\frac{2}{3}\right)^{x}\left(\frac{1}{3}\right)^{6-x} \; ; \; x = 0, 1, 2, \ldots, 6.$$

Thus, the required probability

$$P[X \geq 1] = P[X = 1] + P[X = 2] + P[X = 3] + \ldots + P[X = 6]$$

$$= 1 - P[X = 0]$$

$$= 1 - {}^{6}C_{0}\left(\frac{2}{3}\right)^{0}\left(\frac{1}{3}\right)^{6-0} = 1 - (1)(1)\frac{1}{729} = \frac{728}{729}.$$

Example 6: If X ~ B (n, p). Find p if n = 6 and 9P[X = 4] = P[X = 2].

Solution: As X ~ B(n, p) and n = 6,

$$\therefore \; P[X = x] = {}^{6}C_{x}p^{x}(1-p)^{6-x} \; ; x = 0, 1, 2, \ldots, 6.$$

Now, $9P[X = 4] = P[X = 2]$

$$\Rightarrow 9 \times {}^{6}C_{4} \times p^{4}(1-p)^{6-4} = {}^{6}C_{2} \times p^{2}(1-p)^{4}$$

$$\Rightarrow 9 \times \frac{6 \times 5}{2} \times p^{4} \times (1-p)^{2} = \frac{6 \times 5}{2}p^{2}(1-p)^{4}$$

$$\Rightarrow 9p^{2} = (1-p)^{2}$$

$$\Rightarrow 9p^{2} = 1 + p^{2} - 2p$$

$$\Rightarrow 8p^{2} + 2p - 1 = 0$$

$$\Rightarrow 8p^{2} + 4p - 2p - 1 = 0$$

$$\Rightarrow 4p(2p+1) - 1(2p+1) = 0$$

$$\Rightarrow (2p+1)(4p-1) = 0$$

$$\Rightarrow (2p+1) = 0 \text{ or } (4p-1) = 0$$

$$\Rightarrow p = -\frac{1}{2} \text{ or } \frac{1}{4}$$

But $p = -\frac{1}{2}$ rejected [$\because$ probability can never be negative]

Hence, $p = \frac{1}{4}$

Now, you can try the following exercises:

E3) Comment on the following:

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

E4) Find the binomial distribution when sum of mean and variance of 5 trails is 4.8.

E5) The mean of a binomial distribution is 30 and standard deviation is 5. Find the values of

i) n, p and q,

ii) Moment coefficient of skewness, and

iii) Kurtosis.

9.5 FITTING OF BINOMIAL DISTRIBUTION

To fit a binomial distribution, we need the observed data which is obtained from repeated trials of a given experiment. On the basis of the observed data, we find the theoretical (or expected) frequencies corresponding to each value of the binomial variable. Process of finding the probabilities corresponding to each value of the binomial variable becomes easy if we use the recurrence relation for the probabilities of Binomial distribution. So, in this section, we will first establish the recurrence relation for probabilities and then define the binomial frequency distribution followed by process of fitting a binomial distribution.

Recurrence Relation for the Probabilities of Binomial Distribution

You have studied that binomial probability function is

$$p(x) = P[X = x] = {}^{n}C_{x} p^{x} q^{n-x} \qquad \ldots (1)$$

If we replace x by x + 1, we have

$$p(x+1) = {}^{n}C_{x+1} p^{x+1} q^{n-(x+1)} \qquad \ldots (2)$$

Dividing (2) by (1), we have

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

$$= \frac{\lfloor n \, \lfloor x \, \lfloor n-x}{\lfloor x+1 \, \lfloor n-x-1 \, \lfloor n} \times \frac{p}{q} \qquad \left[\because \ {}^{n}C_{x+1} = \frac{\lfloor n}{\lfloor x+1 \, \lfloor n-x-1} \text{ and } {}^{n}C_{x} = \frac{\lfloor n}{\lfloor x \, \lfloor n-x} \right]$$

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

$$\Rightarrow p(x+1) = \frac{n-x}{x+1} \frac{p}{q} p(x) \qquad \ldots (3)$$

Putting $x = 0$, 1, 2, 3,…in this equation, we get p(1) in terms of p(0), p(2) in terms of p(1), p(3) in terms of p(2), and so on. Thus, if p(0) is known, we can find p(1) then p(2), p(3) and so on.

So, eqn. (3) is the recurrence relation for finding the probabilities of binomial distribution. The initial probability i.e. p(0) is obtained from the following formula:

$$p(0) = q^n$$

[$\because$ $p(x) = {}^nC_x p^x q^{n-x}$ putting x = 0, we have p(0) = ${}^nC_0 p^0 q^n = q^n$]

Binomial Frequency Distribution

We have studied that in a random experiment with n trials and having p as the probability of success in each trial,

$$P[X = x] = {}^nC_x p^x q^{n-x}; x = 0, 1, 2, ..., n$$

where x is the number of successes. Now, if such a random experiment of n trials is repeated say N times, then the expected (or theoretical) frequency of getting x successes is given by

f(x) = N.P$[X = x] = N.\,{}^nC_x p^x q^{n-x}; x = 0, 1, 2, ..., n$

i.e. probability is multiplied by N to get the corresponding expected frequency.

Process of Fitting a Binomial Distribution

Suppose we are given the observed frequency distribution. We first find the mean from the given frequency distribution and equate it to np. From this, we can find the value of p. After having obtained the value of p, we obtain $p(0) = q^n$, where q = 1 – p.

Then the recurrence relation i.e. $p(x+1) = \frac{n-x}{x+1} p(x)$ is applied to find the values of p(1), p(2),…. After that, the expected (theoretical) frequencies f(0), (1), f(2), … are obtained on multiplying each of the corresponding probabilities i.e. p(0), p(1), p(2), … by N.

In this way, the binomial distribution is fitted to the given data. Thus, fitting of a binomial distribution involves comparing the observed frequencies with the expected frequencies to see how best the observed results fit with the theoretical (expected) results.

Example 7: Four coins were tossed and number of heads noted. The experiment is repeated 200 times.

The number of tosses showing 0, 1, 2, 3 and 4 heads were found distributed as under. Fit a binomial distribution to these observed results assuming that the nature of the coins is not known.

Number of heads:	0	1	2	3	4
Number of tosses	15	35	90	40	20

Solution: Here n = 4, N = 200.

First, we obtain the mean of the given frequency distribution as follows:

Number of head X	Number of tosses f	fX
0	15	0
1	35	35
2	90	180
3	40	120
4	20	80
Total	200	415

$\therefore$ Mean = $\frac{\sum f(x)}{\sum f}$ [See Unit 1of MST-002]

$= \frac{415}{200}$

$= 2.075$

As mean for binomial distribution is np,

$\therefore$ np = 2.075

$\Rightarrow p = \frac{2.075}{4}$

$= 0.5188$

$\Rightarrow q = 1 - p$

$= 1 - 0.5188$

$= 0.4812$

$\therefore$ p(0) = q^n

$= (0.4812)^4$

$= 0.0536$

Now, using the recurrence relation

$$p(x+1) = \frac{n-x}{x+1}.\frac{p}{q}p(x); x = 0, 1, 2, 3, 4;$$

we obtain the probabilities for different values of the random variable X i.e. p(1) is obtained on multiplying p(0) with $\frac{4-0}{0+1}$, p(2) is obtained on multiplying p(1) with $\frac{4-1}{1+1}$, and so on; i.e. the values as shown in col. 3 of the following table are obtained on multiplying the preceding values of col. 2 and col 3, except the first value which has been obtained using p(0) = q^n as above.

Number of Heads (X) (1)	$\frac{n-x}{x+1}\cdot\frac{p}{q} = \frac{4-x}{x+1}\left(\frac{0.5188}{0.4812}\right) = \frac{4-x}{x+1}(1.07814)$ (2)	$p(x)$ (3)	Expected or theoretical frequency $f(x)$ (4)
0	$\frac{4-0}{0+1}(1.07814) = 4.31256$	p(0) = 0.0536	10.72□ 11
1	$\frac{4-1}{1+1}(1.07814) = 1.61721$	p(1) = 4.31256×0.0536 = 0.23115	46.23□ 46
2	$\frac{4-2}{2+1}(1.07814) = .071876$	p(2) =1.61721×0.23115 = 0.37382	74.76□ 75
3	$\frac{4-3}{3+1}(1.07814) = 0.26954$	p(3) = 0.71876 × 0.37382 = 0.26869	53.73 □ 54
4	$\frac{4-4}{4+1}(1.07814) = 0$	p(4) = 0.26954 × .26869 = 0.0724	14.48 □ 14

Remark 3: In the above example, if the nature of the coins had been known e.g. if it had been given that “the coins are unbiased” then we would have taken

$p = \frac{1}{2}$ and then the observed data would not have been used to find p. Such a situation can be seen in the problem **E6).**

Here are two exercises for you:

E6) Seven coins are tossed and number of heads noted. The experiment is repeated 128 times and the following distribution is obtained:

Number of heads	0	1	2	3	4	5	6	7
Frequencies	7	6	19	35	30	23	7	1

Fit a binomial distribution assuming the coin is unbiased.

E7) Out of 800 families with 4 children each, how many families would you expect to have 3 boys and 1 girl, assuming equal probability of boys and girls?

Now before ending this unit, let's summarize what we have covered in it.

9.6 SUMMARY

The following main points have been covered in this unit:

1) A discrete random variable X is said to follow **Bernoulli distribution** with parameter p if its probability mass function is given by

$$P[X = x] = \begin{cases} p^x (1-p)^{1-x} & ; \; x = 0, 1 \\ 0; & \text{elsewhere} \end{cases}$$

Its **mean** and **variance** are **p** and **p(1 − p)**, respectively. **Third** and **fourth central moments** of this distribution are $p(1-p)(1-2p)$ and $p(1-p)(1-3p+3p^2)$ respectively.

2) A discrete random variable X is said to follow **binomial distribution** if it assumes only a finite number of non-negative integer values and its probability mass function is given by

$$P[X = x] = \begin{cases} {}^nC_x p^x q^{n-x} & ; \; x = 0, 1, 2, ..., n \\ 0; & \text{elsewhere} \end{cases}$$

where, n is the number of independent trials,

x is the number of successes in n trial,

p is the probability of success in each trial, and

q = 1 – p is the probability of failure in each trial.

3) The **constants of Binomial distribution** are:

Mean= np, Variance= npq,

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

$\beta_1 = \dfrac{(q-p)^2}{npq}$, $\beta_2 = 3 + \dfrac{1-6pq}{npq}$,

$\gamma_1 = \dfrac{1-2p}{\sqrt{npq}}$, and $\gamma_2 = \dfrac{1-6pq}{npq}$

4) For a binomial distribution, **Mean** > **Variance.**

5) **Recurrence relation for the probabilities of binomial distribution** is

$$p(x+1) = \frac{n-x}{x+1} \cdot \frac{p}{q} \cdot p(x), \; x = 0, 1, 2, \ldots, n-1$$

6) The **expected frequencies of the binomial distribution** are given by

$$f(x) = N.P[X = x] = N. {}^nC_x p^x q^{n-x} ; \; x = 0, 1, 2, ..., n$$

9.7 SOLUTIONS/ANSWERS

E1) Let p be the probability of hitting the target (success) in a trial.

$\therefore$ n = 5, $p = \frac{1}{4}$, $q = 1 - \frac{1}{4} = \frac{3}{4}$,

and hence by binomial distribution, we have

$$P[X = x] = {}^nC_x p^x q^{n-x} = {}^5C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{5-x} ; x = 0, 1, 2, 3, 4, 5.$$

$\therefore$ Required probability = $P[X \geq 2]$

$$= P[X = 2] + P[X = 3] + P[X = 4] + P[X = 5]$$

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

$$= 1 - \left[{}^5C_0 \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^{5-0} + {}^5C_1 \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^{5-1}\right]$$

$$= 1 - \left[\frac{243}{1024} + \frac{405}{1024}\right] = \frac{376}{1024} = \frac{47}{128}$$

E2) Let p be the probability that the dacoit will be killed (success) by a bullet.

$\therefore$ n = 6, p = 0.6, q = 1 – p = 1 – 0.6 = 0.4, and hence by binomial distribution, we have

$$P[X = x] = {}^nC_x\, p^x\, q^{n-x} ; x = 0, 1, 2, ..., n$$

$$= {}^6C_x (0.6)^x (0.4)^{6-x} ; x = 0, 1, 2, ..., 6 .$$

$\therefore$ The required probability = P[The dacoit is still alive]

= P[No bullet kills the dacoit]

= P[Number of successes is zero]

= P[X = 0] = ${}^6C_0 (0.6)^0 (0.4)^6$

= 0.0041

E3) Mean = np = 3 ... (1)

Variance = npq = 4 ... (2)

$\therefore$ Dividing (2) by (1), we have

$q = \frac{4}{3} > 1$ and hence not possible

[$\because$ q, being probability, cannot be greater than 1]

E4) Let $X \sim B(n, p)$, then

n = 5 and

$np + npq = 4.8$ [$\because$ given that Mean + Variance = 4.8]

$\Rightarrow 5p + 5pq = 4.8$

$\Rightarrow 5[p + p(1-p)] = 4.8$

$\Rightarrow 5[p + p - p^2] = 4.8$

$\Rightarrow 5p^2 - 10p + 4.8 = 0$

$\Rightarrow 25p^2 - 50p + 24 = 0$ [Multiplying by 5]

$\Rightarrow 25p^2 - 30p - 20p + 24 = 0$

$\Rightarrow 5p(5p - 6) - 4(5p - 6) = 0$

$\Rightarrow (5p - 6)(5p - 4) = 0$

$\Rightarrow p = \frac{6}{5}, \frac{4}{5}$

The first value $p = \frac{6}{5}$ is rejected [$\because$ probability can never exceed 1]

$\therefore\ p = \frac{4}{5}$ and hence $q = 1 - p = \frac{1}{5}$.

Thus, the binomial distribution is

$$P[X = x] = {}^nC_x p^x q^{n-x}$$

$$= {}^5C_x \left(\frac{4}{5}\right)^x \left(\frac{1}{5}\right)^{5-x}; \ x = 0, 1, 2, 3, 4, 5.$$

The binomial distribution in tabular form is given as

X	p(x)
0	${}^5C_0 \left(\frac{4}{5}\right)^0 \left(\frac{1}{5}\right)^5 = \frac{1}{3125}$
1	${}^5C_1 \left(\frac{4}{5}\right)^1 \left(\frac{1}{5}\right)^4 = \frac{20}{3125}$
2	${}^5C_2 \left(\frac{4}{5}\right)^2 \left(\frac{1}{5}\right)^3 = \frac{160}{3125}$
3	${}^5C_3 \left(\frac{4}{5}\right)^3 \left(\frac{1}{5}\right)^2 = \frac{640}{3125}$
4	${}^5C_4 \left(\frac{4}{5}\right)^4 \left(\frac{1}{5}\right)^1 = \frac{1280}{3125}$
5	${}^5C_5 \left(\frac{4}{5}\right)^5 \left(\frac{1}{5}\right)^0 = \frac{1024}{3125}$

E5) Given that Mean = 30 and S.D. = 5

Thus, $np = 30$, $\sqrt{npq} = 5$

$\Rightarrow np = 30,\ npq = 25$

i) $\frac{npq}{np} = \frac{25}{30} = \frac{5}{6} \Rightarrow q = \frac{5}{6}, p = 1 - q = 1 - \frac{5}{6} = \frac{1}{6},\ n\left(\frac{1}{6}\right) = 30 \Rightarrow n = 180$

ii) $\mu_2 = npq = 180 \times \frac{1}{6} \times \frac{5}{6} = 25$

$$\mu_3 = npq(q - p) = 25\left(\frac{5}{6} - \frac{1}{6}\right) = \frac{50}{3}$$

$$\Rightarrow \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{4}{225}$$

$\therefore$ Moment coefficient of skewness is given by

$$\gamma_1 = \sqrt{\beta_1} = \frac{2}{15}$$

iii) $\beta_2 = 3 + \frac{1 - 6pq}{npq} = 3 + \frac{1 - 6 \times \frac{1}{6} \times \frac{5}{6}}{25} = 3 + \frac{1}{150}$

$$\Rightarrow \gamma_2 = \beta_2 - 3 = \frac{1}{150} > 0$$

So, the curve of the binomial distribution is leptokurtic.

E6) As the coin is unbiased, $\therefore\ p = \frac{1}{2}$.

Here, n = 7, N = 128, $p = \frac{1}{2}, q = 1 - p = \frac{1}{2}$.

$\Rightarrow p(0) = q^n = \left(\frac{1}{2}\right)^7 = \frac{1}{128}$.

Expected frequencies are, therefore, obtained as follows:

Number of heads (X)	$\frac{n-x}{x+1}.\frac{p}{q} = \frac{7-x}{x+1}.\frac{\frac{1}{2}}{\frac{1}{2}}$ $= \frac{7-x}{x+1}$	$p(x)$	Expected or theoretical Frequency $f(x) = N.p(x)$ $= 128.p(x)$
0	$\frac{7-0}{0+1} = 7$	$\frac{1}{128}$	1
1	$\frac{7-1}{1+1} = 3$	$7 \times \frac{1}{128} = \frac{7}{128}$	7
2	$\frac{7-2}{2+1} = \frac{5}{3}$	$3 \times \frac{7}{128} = \frac{21}{128}$	21
3	$\frac{7-3}{3+1} = 1$	$\frac{5}{3} \times \frac{21}{128} = \frac{35}{128}$	35

4	$\frac{7-4}{4+1}=\frac{3}{5}$	$1\times\frac{35}{128}=\frac{35}{128}$	35
5	$\frac{7-5}{5+1}=\frac{1}{3}$	$\frac{3}{5}\times\frac{35}{128}=\frac{21}{128}$	21
6	$\frac{7-6}{6+1}=\frac{1}{7}$	$\frac{1}{3}\times\frac{21}{128}=\frac{7}{128}$	7
7	$\frac{7-7}{7+1}=0$	$\frac{1}{7}\times\frac{7}{128}=\frac{1}{128}$	1

E7) Here, probability (p) to have a boy is $\frac{1}{2}$ and the probability (q) to have a girl is $\frac{1}{2}$, n = 4, N = 800.

Let X be the number of boys in a family.

$\therefore$ by binomial distribution, the probability of having 3 boys in a family of 4 children

$= P[X = 3] \qquad [\because\ P[X = x] = {}^{n}C_{x}p^{x}q^{n-x}]$

$$= {}^{4}C_{3}\left(\frac{1}{2}\right)^{3}\left(\frac{1}{2}\right)^{4-3} = 4\left(\frac{1}{2}\right)^{4}$$

Hence, the expected number of families having 3 boys and 1 girl

$$= N.p(3) = 128\left(\frac{1}{4}\right) = 32$$