Discrete Probability Distribution

Binomial Probability Distribution:

Bernoulli trial:

A trail having only two possible results success or failure is called Bernoulli Trail. E.g tossing a coin.

Binomial Distribution:

Binomial Distribution is a discrete probability distribution and is used to find the probability of "X" number of success of an event in n trails of the same experiment when.

i. When there only two mutually exclusive and possible outcomes.
ii. The numbers of trails "n" are independent.
iii. The probability of success, "P" remains constant in each trail.

$F(x) = \binom{n}{x} p^x (1-p)^{n-x}$

$X = 0, 1, 2, \ldots\ldots., n$

Here n and p are known as parameters of the distributions.

A binomial experiment involves n independent and identical trails such that each trail can result into one of the two possible outcomes, namely, success or failure. If P is the probability of observing a success in each trail then the number of success X that can be observed out of this n trail is referred to as the binomial random variable with n trials and success probability P. the probability of observing k success out of these trails is given by the probability mass function.

$P(x = k/n, p) = \binom{n}{k} p^k (1-p)^{n-k}$

$k = 0, 1, 2, \ldots\ldots., n$

The cumulative distribution function of X is given by

$P(x < k/n , p) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$

$k = 0, 1, 2, \ldots\ldots., n$

The binomial distribution is not appropriate if the sample was drawn without replacement from a finite population. We denote a binomial distribution with n trails and success probability p by binomial (n;P). This distribution is right-skewed when p< 0.5 and left skewed when p> 0.5 and symmetric when p = 0.5. For large n, binomial distribution is approximately symmetric about its mean np.

Mean:

np

Variance:

np (1 – p)

Mode:

The largest integer ≤ np

Properties of the Binomial Distribution:

i. Mean = np and variance = npq and mean > variance.
ii. Area under the Binomial curve is unity.
iii. If p = q = ½ then this distribution is symmetrical about its origin and when p≠ q this distribution is skewed.
iv. As n →∞ then binomial distribution tends to normality.
v. Probability generating function.
   G (Ө) = (p Ө + q) n
vi. Moment generating function,
   M (t) = (p e^t + q) n
vii. Characteristics function,
   Ø (t) = (p e^{it} + q) n
viii. Mode is p. (n + 1) – 1 ≤ x ≤ p.(n + 1)

As discussed earlier, a discrete probability distribution gives the probability of every possible value of a discrete random variable. We shall introduce here some important discrete probability distribution which is often used in statistical theory and analysis.

Binomial Probability Distribution:

Many experiments consist of repeated independent trails, each trail having only two possible outcomes.

For example. Two possible outcomes of a trail maybe head and tail. Success and failure, right and wrong, alive and dead, good and defective, infected and not infected, and so forth. If the probability of each outcome remain the same throughout the trails and such trails are called Bernoulli Trails is called binomial experiment. In other words, an experiment is called a binomial probability experiment if it possesses the following four properties.

i. The outcomes of each trail maybe classified into one of two categories, conventionally called success and failure. It is to be noted that the outcomes of interest is called a success and the other a failure.

ii. The probability of success, denoted by “p” remains constant for all trails.
iii. The successive trails are all independent.
iv. The experiment is repeated a fixed no of times, say n.

When X denoted the no. of successes in n trials of a binomial probability experiment, it is called a binomial random variable and its probability distribution is called Binomial probability distribution. The r.v. X can obviously take on any of the (n + 1) integer value 0,1,2,3……, n

When the binomial r.v X assumes a value X, the binomial p.d is given by:

$$F(x) = P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{where } x = 0, 1, 2, \ldots, n$$

Where; $q = 1 - p$, the probability of failure in each trail. The binomial p.d has two parameters n and p and is generally denoted by b(x; n,p). the binomial probability distribution is appropriate when a random sample of size n is drawn with replacement from a finite population of size, N or sampling is done from an infinite population.

The binomial p.d , which is the most widely used distribution in two- outcomes situations, was discovered by Swiss mathematician Jakob Bernoulli. (1654 – 1704) whose main work on probability.

Probability of the Binomial Probability Distribution:

The properties of the binomial probability distribution include the mean number of successes, the variance of the number of success measures of skewness and kurtosis etc. and the shape of the distribution. Some of the properties are described below;

1. Let X be a random variable with the binomial distribution b(x; n,p) then its mean and variance are given by $\mu = np$ and $\sigma^2 = npq$ respectively.

Now Mean, $\mu = E(X)$

$$= \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}, \quad where\ x = 0, 1, 2, \ldots, n.$$

$$= 0.q^n + 1.\binom{n}{1} q^{n-1} p + 2\binom{n}{2} q^{n-2} p^2 + \ldots + np^n$$

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

Alternative Method:

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

$$= np, \text{ because } q + p = 1.$$

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

But $$x\binom{n}{x} = \frac{x(n)(n-1)!}{x(x-1)!(n-x)!} = n\binom{n-1}{x-1}$$

$\because$ $$E(X) = n\sum_{x=1}^{n}\binom{n-1}{x-1} p^x q^{n-x}, \; for\ x = 1, 2, \ldots, n$$ (since the first term in the summation being zero (x = 0) is omitted).

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

Substituting $y = x - 1$ and $m = n - 1$ in the summation, we get

$$E(X) = np\sum_{y=0}^{m}\binom{m}{y} p^y q^{m-y}$$ (as x ranges from 1 to n, so y (= x – 1) must range from 0 to n – 1 i.e. m)

$= np$ ($\because$ summation is the expansion of $(q+p)^m$)

Hence mean = np , in other words, the mean number of successes is np. Similarly the mean number of failure is nq.

By definition, the variance σ^2, is given by

$$\sigma^2 = E[X-\mu]^2 = E(X^2) - [E(X)]^2$$

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

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

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

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

(x starts at 2 since x = 0, 1 add nothing to the sum)

The term (n – x) may be written as [(n – 2) – (x – 2)]

Substituting y = x – 2 and m = n – 2 in the summation we obtain.

$$E[X(X-1)] = n(n-1)p^2 \sum_{y=0}^{m} \frac{m!}{y!(m-y)!} p^y q^{m-y}$$

$$= E[X(X-1)] + E(X) - [E(X)]^2$$
$$= n(n-1)p^2 + np - (np)^2$$
$$= n^2p^2 - np^2 + np - n^2p^2$$
$$= np - np^2 = np(1-p) = npq, \text{ and}$$
$$\sigma = \sqrt{npq}$$

Hence the variance of the number of successes is npq, and the standard deviation is $\sqrt{npq}$

Example:

A Fair coin is tossed 5 times. Find the probabilities of obtaining various numbers of head.

Let us regard the tossing of a coin as an experiment. Then we observe that.

1. Each toss of coin has two possible outcomes, head and tail.
2. The probability of a head (success) is p =1/2 and remain the same for successive tosses.
3. The successive tosses of the coin are independent
4. The coin is tossed 5 times.

Therefore the r.v X which denotes the numbers of heads (successes) has a binomial probability distribution with p = ½ and n = 5, the possible value of X are 0,1,2,3,4, and 5 hence.

$$P(\text{no head}) = P(X=0) = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 \quad = 1\times\left(\frac{1}{2}\right)^5 = \frac{1}{32}.$$

$$P(1\text{ head}) = P(X=1) = \binom{5}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{5-1} = 5\times\left(\frac{1}{2}\right)^5 = \frac{5}{32},$$

$$P(2\text{ heads}) = P(X=2) = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{5-2} = 10\times\left(\frac{1}{2}\right)^5 = \frac{10}{32},$$

$$P(3\text{ heads}) = P(X=3) = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{5-3} = 10\times\left(\frac{1}{2}\right)^5 = \frac{10}{32},$$

$$P(4\text{ heads}) = P(X=4) = \binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{5-4} = 5\times\left(\frac{1}{2}\right)^5 = \frac{5}{32}, \text{ and}$$

$$P(5\text{ heads}) = P(X=5) = \binom{5}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0 \quad = 1\times\left(\frac{1}{2}\right)^5 = \frac{1}{32}.$$

These probabilities can also be obtained by expanding the binomial (½ + ½)^5. The binomial p.d for the number of heads obtained in 5 tosses of fair coin is.

x	0	1	2	3	4	5
$f(x)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

Example: A event has the probability p = 3/8. Find the complete binomial distribution for n= 5 trails.

Here $p = \frac{3}{8}$ so that $q = 1 - p = \frac{5}{8}$; and $n = 5$.

Hence the desired probabilities are the successive terms in the binomial expansion of $\left(\frac{5}{8} + \frac{3}{8}\right)^5$, *i.e.*

$$\left[\left(\frac{5}{8}\right)^5 + \binom{5}{1}\left(\frac{5}{8}\right)^4\left(\frac{3}{8}\right) + \binom{5}{2}\left(\frac{5}{8}\right)^3\left(\frac{3}{8}\right)^2 + \binom{5}{3}\left(\frac{5}{8}\right)^2\left(\frac{3}{8}\right)^3 + \binom{5}{4}\left(\frac{5}{8}\right)\left(\frac{3}{8}\right)^4 + \left(\frac{3}{8}\right)^5\right]$$

i.e. $\frac{1}{(8)^5}\left[(5)^5 + 5.(5)^4(3) + 10.(5)^3(3)^2 + 10(5)^2(3)^3 + 5(5)(3)^4 + (3)^5\right]$

i.e. $\frac{1}{32768}\left[3125 + 9375 + 11250 + 6750 + 2025 + 243\right]$

i.e. $[0.0954 + 0.2861 + 0.3433 + 0.2060 + 0.0618 + 0.0074]$

We can now write these probabilities in the form of a probability table as below:

x	0	1	2	3	4	5
$P(X=x)$	0.0954	0.2861	0.3433	0.2060	0.0618	0.0074

Example:

Let X have a binomial distribution with n = 4 and p = 1/3. Find P(x=1) , P=(X=3/2) p(x=3), p(x=6) and P(x ≤ 2).

The binomial probability distribution for n = 4 and $p = \frac{1}{3}$, is

$$f(x) = \binom{4}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{4-x} \quad \text{for } x = 0, 1, 2, 3, 4.$$

Now $P(X = 1) = \binom{4}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^{4-1} = \frac{32}{81}$;

$P\left(X=\frac{3}{2}\right)=f\left(\frac{3}{2}\right)=0$; because a r.v. X with a binomial distribution takes only one of the integer values 0, 1, 2, …, n.

$$P(X=3)=f(3)=\binom{4}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^{4-3}=\frac{8}{81};$$

P(X = 6) = $f(6)$ = 0, because X can take only values 0, 1, 2, 3, 4.

and $$P(X\le 2)=\sum_{x=0}^{2}f(x)=f(0)+f(1)+f(2)$$

$$=\binom{4}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^4+\binom{4}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^3+\binom{4}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^2$$

$$=\frac{16}{81}+\frac{32}{81}+\frac{24}{81}=\frac{72}{81}=\frac{8}{9}.$$

Example:

A and B play a game in which A's probability of winning is 2/3. In a series of 8 games what is the probability of A will win (i) exactly 4 games, (ii) at least 4 games, (iii) 6 or more games, and (iv) from 3 to 6 games.

We observe that

a) there are two possible outcomes, i.e. A will win or will not win the game;

b) the probability of A's winning in each game is p = 2/3;

c) the successive games are independently won or lost; and

d) there are 8 games.

Therefore the Binomial probability distribution with n = 8 and p=2/3 is appropriate.

Let X denote the number of games won by A. Then

i) $$P(X=4)=\binom{8}{4}\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^4=\frac{1120}{6561}=0.1707$$

ii) $P(X\geq 4)=1-P(X<4)$; ($\because$ at least 4 means 4 or more)

$$=1-\sum_{x=0}^{3}\binom{8}{x}\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{8-x}$$

$$=1-\left[\left(\frac{1}{3}\right)^8+8\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^7+28\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^6+56\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^5\right]$$

$$=1-\frac{1}{6561}[1+16+112+448]$$

$$=1-\frac{577}{6561}=\frac{5984}{6561}=0.9121$$

iii) $P(X \geq 6) = \sum_{x=6}^{8} \binom{8}{x}\left(\frac{2}{3}\right)^{x}\left(\frac{1}{3}\right)^{8-x}$

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

$$= \frac{64}{6561}[28+16+4] = \frac{64 \times 48}{6561} = \frac{1024}{2187} = 0.4682$$

iv) $P(3 \leq X \leq 6) = \sum_{x=3}^{6} \binom{8}{x}\left(\frac{2}{3}\right)^{x}\left(\frac{1}{3}\right)^{8-x}$

$$= \binom{8}{3}\left(\frac{2}{3}\right)^{3}\left(\frac{1}{3}\right)^{5} + \binom{8}{4}\left(\frac{2}{3}\right)^{4}\left(\frac{1}{3}\right)^{4} + \binom{8}{5}\left(\frac{2}{3}\right)^{5}\left(\frac{1}{3}\right)^{3} + \binom{8}{6}\left(\frac{2}{3}\right)^{6}\left(\frac{1}{3}\right)^{2}$$

$$= \frac{2^3}{(3)^8}[56+140+224+224]$$

$$= \frac{8 \times 644}{6561} = \frac{5152}{6561} = 0.7852$$

Example:

The experience of a house agent indicated that he can provide suitable accommodation for 75 percentage 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, (ii) exactly 4 clients (iii) at least 5 clients, will get satisfactory accommodation.

We observe that

a) there are two possible outcomes, *i.e.* each client will get *or* will not get accommodation,

b) on each occasion, probability of getting accommodation is $p = 3/4$,

c) clients approach the house-agent independently, and

d) there are 6 clients.

Therefore the binomial probability distribution with n=6 and p=3/4 is appropriate to calculate the desired probabilities.

Let X denote the number of clients who get satisfactory accommodation. Then we need to calculate (i) $P(X<4)$, (ii) $P(X=4)$ and (iii) $P(X \geq 5)$. Hence,

i) $P(X<4) = \sum_{x=0}^{3} \binom{6}{x}\left(\frac{3}{4}\right)^{x}\left(\frac{1}{4}\right)^{6-x}$

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

$$= \left(\frac{1}{4}\right)^6 [1+(6)(3)+(15)(9)+(20)(27)]$$

$$= \frac{694}{4096} = 0.169$$

ii) $$P(X=4) = \binom{6}{4}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^2 = \frac{15\times 81}{(4)^6} = \frac{1215}{4096} = 0.297$$

iii) $$P(X\geq 5) = \sum_{x=5}^{6}\binom{6}{x}\left(\frac{3}{4}\right)^x\left(\frac{1}{4}\right)^{6-x}$$

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

$$= \left(\frac{1}{4}\right)^6 [(6)(3)^5 + (3)^6] = \frac{2187}{4696} = 0.534$$

Binomial Frequency Distribution:

If the binomial probability distribution is multiplied by N, the no of experiment or sets, the resulting distribution is known as the binomial frequency distribution. Thus the expected frequency of x success is N. $\binom{n}{k}$ p^k $(1-p)^{n-k}$. it should be noted that the n independent trails constitute one experiment or one set.

Example:

Six dice are thrown 729 times. How many time do you expect at least 3 dice to show a 5 or a 6.?

The probability of getting a 5 or a 6 with one dice is p = 2:6. Since 6 dices are thrown and there are 729 sets, binomial frequency distribution is given by.

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

Hence the expected number of times *at least* 3 dice showing 5 or 6

$$= 729\left[\sum_{x=3}^{6}\binom{6}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{6-x}\right]$$

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

$$= \frac{729}{(3)^6}[160+60+12+1] = 233.$$

Example:

A certain event is believed to follow the binomial distribution. In 1024 samples of 5, the result was observed once 405 times and twice 270 times. Find p and q.

The first three term in the expansion of the Binomial Frequency Distribution $N(q+p)^n$ corresponding to x = 0, 1, and 2 are Nq^n, $N\binom{n}{1}q^{n-1}p$ and $N\binom{n}{2}q^{n-2}p^2$.

We are given N = 1024, n = 5 and the following information:

$$1024\binom{5}{1}q^{5-1}p = 405,$$

$$1024\binom{5}{2}q^{5-2}p^2 = 270,$$

Dividing the second equation by the first, we get

$$\frac{10q^3p^2}{5q^4p} = \frac{270}{405} \text{ or } \frac{2p}{q} = \frac{2}{3}$$

or 3p = q or 3p = 1 – p or 4p = 1

Hence $p = \frac{1}{4}$ and $q = \frac{3}{4}$