STANDARD DISCRETE PROBABILITY DISTRIUTIONS

Introduction:

Probability Distributions are Theoretical Distributions

Consider a random variable X that measure the "number of heads" in a three-trial coin tossing experiment. The probability distribution of X will be

X	0	1	2	3
P(X)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Now imagine this experiment is repeated 200 times, we may expect 'no head' and 'three heads' will each occur 25 times; 'one head' and 'two heads' each will occur 75 times. Since these results are what we expect on the basis of theory, the resultant distribution is called a **theoretical or expected distribution.**

However, when the experiment is actually performed 200 times, the results, which we may actually obtain, will normally differ from the theoretically expected results. It is quite possible that in actual experiment 'no head' and 'three heads' may occur 20 and 28 times respectively and 'one head' and 'two heads' may occur 66 and 86 times respectively. The distribution so obtained through actual experiment is called the **empirical or observed distribution.**

In practice, however, assessing the probability of every possible value of a random variable through actual experiment can be difficult, even impossible, especially when the probabilities are very small. But we may be able to find out what type of random variable the one at hand is by examining the causes that make it random. Knowing the type, we can often approximate the random variable to a standard one for which convenient formulae are available.

What is Discrete Distribution?

A discrete distribution is a statistical distribution that shows the probabilities of discrete (countable) outcomes, such as 1, 2, 3... Statistical distributions can be either discrete or continuous.

Some of these distributions have mass (i.e. positive probability) at only a finite number of values, such as $\{1, 2, 3\}$ or $\{-2, -1, 0, 1, 2\}$.

Some of these discrete r.v. distributions have mass at a countably infinite number of values, like $\{0, 1, 2, 3, . . .\}$

Degenerate Distributions:

[i] A weighted die (or one that has a number 6 on all faces) always lands on the number six, so the probability of a six P(6) is 1.

[ii] A coin is double-sided with two heads (thousands of these "magician's coins" exist, but there are also real ones.

[iii] (Calculus): A random variable X that is distributed as the derivative of k when k=1. As k=1, the distribution can only have a value of 0 (because the derivative of any constant is 0).

Def: A random variable X is said to be follow a degenerate distribution if its probability mass function is given by

$$P(X = k) = 1 \qquad k \in R$$
$$P(X = x) = 0 \qquad x \neq k$$

- **Mean of the degenerate distribution**

  X follow degenerate distribution at X = k, then

  Mean of $X = E(k) = k$

- **Variance of the degenerate distribution**

  Variance of $X = V(k) = 0$

Use of Degenerate Distribution:

[i] Degenerate distributions are usually taught in advanced statistics courses like mathematical statistics.

[ii] They can be defined as special cases of the binomial distribution, normal and geometric distributions among others and are often used in queuing theory where service times or systems inter-arrival times are constant.

Uniform Distribution

Introduction:

In probability theory and **statistics**, the **discrete uniform distribution** is a symmetric probability **distribution** wherein a finite number of values are equally likely to be observed; every one of n values has equal probability 1/n. ...

A simple example of the **discrete uniform distribution** is throwing a fair die.

Definition: A random variable X is said to be follow a uniform distribution if its probability mass function is given by

$$P(X=x) = \frac{1}{n}, \quad x = 1, 2, \ldots\ldots, n$$

Where, n is called parameter of the distribution.

Mean of the uniform distribution

For a uniform distribution the probability function is given by

$$P(X=x) = \frac{1}{n}, \qquad x = 1, 2, \ldots\ldots, n$$

Now, the mean of the Uniform distribution is

$$E(X) = \sum_{x=1}^{n} x\, P(X = x)$$

$$= \sum_{x=1}^{n} x \frac{1}{n}$$

$$= \frac{1}{n}(1+2+\ldots+ n)$$

$$= \frac{1}{n}[\text{Sum of } n \text{ natural numbers}]$$

$$= \frac{1}{n}\frac{n(n+1)}{2}$$

$$= \frac{(n+1)}{2}$$

$\therefore$ The mean of the uniform distribution is $\dfrac{(n+1)}{2}$

Variance of the Uniform distribution:

The variance of the uniform distribution is

$$V(X) = E(X^2) - [E(X)]^2$$

$$V(X) = E(X^2) - \left[\frac{(n+1)}{2}\right]^2 \text{ ……………… (1)} \quad \because$$

$$\left[E(X) = \frac{(n+1)}{2}\right]$$

Now,

$$E(X^2) = \sum_{x=1}^{n} x^2 \frac{1}{n}$$

$$E(X^2) = \frac{1}{n}\left(1^2 + 2^2 + ... + n^2\right)$$

$$E(X^2) = \frac{1}{n}\left(\text{Sum of squares of } n \text{ natural numbers}\right)$$

$$E(X^2) = \frac{1}{n}\frac{n(n+1)(2n+1)}{6}$$

$$E(X^2) = \frac{(n+1)(2n+1)}{6}$$

$$E(X^2) = \frac{(n+1)(2n+1)}{6} \text{ ………… (2)}$$

Putting (2) in (1) we get

$$V(X) = \frac{(n+1)(2n+1)}{6} - \left[\frac{(n+1)}{2}\right]^2$$

$$V(X) = \left[\frac{(n+1)}{2}\right]\frac{(2n+1)}{3} - \frac{(n+1)}{2}$$

$$V(X) = \left[\frac{(n+1)}{2}\right]\frac{4n+2-3n-3}{6}$$

$$V(X) = \left[\frac{(n+1)}{2}\right]\frac{n-1}{6}$$

$$V(X) = \left[\frac{(n^2-1)}{12}\right]$$

$\therefore$ The variance of the Uniform distribution is $\frac{(n^2-1)}{12}$

BERNOULLI DISTRIBUTION

Introduction:

Suppose an operator uses a lathe to produce pins, and the lathe is not perfect in the sense that it does not always produce a good pin. Rather, it has a probability p of producing a good pin and (1 - p) of producing a defective one. Let us denote a good pin as “success” and a defective pin as “failure”.

Just after the operator produces one pin, it is inspected; let X denote the "number of good pins produced” i. e. “the number of successes”.

Now analyzing the trial- **“inspecting a pin”** and our random variable **X-“number of successes”**, we note two important points:

[i] The trial-“inspecting a pin” has only two possible outcomes, which are mutually exclusive. Such a trial, whose outcome can only be either a success or a failure, is a **Bernoulli trial.** In other words, the sample space of a Bernoulli trial is S = {success, failure}

[ii] The random variable, X, that measures number of successes in one Bernoulli trial, is a **Bernoulli random variable.** Clearly, X is 1 if the pin is good and 0 if it is defective.

It is easy to derive the probability distribution of Bernoulli random variable

X	0	1
P(X)	p	1-p

If X is a Bernoulli random variable, we may write X ~ BER (p) Where ~ is read as “is distributed as” and BER stands for Bernoulli.

Definition: A random variable X is said to be follow a Bernoulli distribution if its probability function is given by

$P(X=x) = p^x\ q^{1-x}$; $x = 0,1$ and $p + q = 1$; Where, p is called parameters of the Bernoulli distribution.

Mean of the Bernoulli distribution

For a Bernoulli distribution the probability function is given by

$$P(X=x) = p^x\ q^{1-x};\ x = 0,1$$

Now, the mean of the Bernoulli distribution is

$$E(X) = \sum_{x=0}^{1} x\, P(X = x)$$
$$E(X) = \sum_{x=0}^{1} x\, p^{x} q^{1-x}$$
$$E(X) = \sum_{x=0}^{1} x\, p^{x} q^{1-x}$$
$$E(X) = 0\, p^{0} q^{1-0} + 1 p^{1} q^{1-1}$$
$$E(X) = 0 + p^{1} q^{0}$$
$$E(X) = p$$

$\therefore$ The mean of the Bernoulli distribution is p

Variance of the Bernoulli distribution:

The variance of the Bernoulli distribution is

$$V(X)=E(X^{2})-[E(X)]^{2}$$
$$V(X)=E(X^{2})-(p)^{2} \quad \cdots(i)\left[\because E(X) = p\right]$$

Now,

$$E\left(X^{2}\right)=\sum_{x=0}^{1} x^{2}\, p^{x} q^{1-x}$$
$$E\left(X^{2}\right)= 0\, p^{0} q^{1-0} + 1^{2}\, p^{1}\, q^{1-1}$$
$$E\left(X^{2}\right)= 0 + p^{1} q^{0}$$
$$E\left(X^{2}\right)= p \qquad \ldots(ii)$$

Putting (ii) in (i) we get

$$V(X) = p - p^{2}$$
$$V(X) = p(1-p)$$
$$V(X) = pq$$

$\therefore$ The variance of the Bernoulli distribution is pq

Use of Bernoulli Distribution:

[i] In experiments and clinical trials, the Bernoulli distribution is sometimes used to model a single individual experiencing an event like death, a disease, or disease

exposure. The model is an excellent indicator of the probability a person has the event in question.

1 = “event” (P = p)

0 = “non- event” (P = 1 – p)

[ii] Bernoulli distributions are used in logistic regression to model disease occurrence.

BINOMIAL DISTRIBUTION

Introduction:

In the real world we often make several trials, not just one, to achieve one or more successes. Let us consider such cases of several trials.

Consider n number of identically and independently distributed Bernoulli random variables X_1, X_2 ………, X_n. Here, identically means that they all have the same p, and independently means that the value of one X does not in any way affect the value of another. For example, the value of X_2 does not affect the value of X_3 or X_8 and so on. Such a sequence of identically and independently distributed Bernoulli variables is called a **Bernoulli Process.**

Suppose an operator produces n pins, one by one, on a lathe that has probability p of making a good pin at each trial, the sequence of numbers (1 or 0) denoting the good and defective pins produced in each of the n trials is a Bernoulli process.

For example, in the sequence of nine trials denoted by 001011001 the third, fifth, sixth and ninth are good pins, or successes. The rest are failures.

In practice, we are usually interested in the total number of good pins rather than the sequence of 1's and 0's. In the example above, four out of nine are good. In the general case, let X denote the total number of good pins produced in n trials. We then have

$X = X_1 + X_2 + \ldots\ldots\ldots + X_n$ where all $X_i \sim BER(p)$ and are independent.

The random variable that counts the number of successes in many independent, identical Bernoulli trials is called a Binomial Random Variable.

Conditions for a Binomial Random Variable:

We may appreciate that the condition to be satisfied for a binomial random variable is that **the experiment should be a Bernoulli Process.**

Any uncertain situation or experiment that is marked by the following three properties is known as a Bernoulli Process:

[i] There are only two mutually exclusive and collectively exhaustive outcomes in the experiment i.e. S = {success, failure}

[ii] In repeated trials of the experiment, the probabilities of occurrence of these events remain constant

[iii] The outcomes of the trials are independent of one another

The probability distribution of Binomial Random Variable is called the Binomial Distribution.

Sir Arthur Bowley
(1869-1957)

Probability Mass Function of Binomial Distribution.

Suppose an experiment is repeated 'n' times and each trail is independent.

Let us assume that each trail results in two possible mutually exclusive and exhaustive outcomes i.e. success and failure.

Let X is random variable represents total no. of successes in 'n' trails. Let the probability of success in each trail is p and the probability of failure is q=1-p and p remains constant from trail to trail.

Now, we have to find out the probability of x successes in n trails.

Let us suppose that a particular order of outcomes of x successes in n repetitions be as follows

SSSSSFFFSSFS………FS(x number of successes and n-x failures)

Since, the trails are all independent the probability for the joint occurrence of the event is pppppqqqppqp……..qp

= (pppppp…..x times)(qqqqqq…… (n-x) times)

$= p^x q^{n-x}$

Further in a series of n trails x successes and n-x failures can occur in ${}^{n}c_{x}$ ways. So, the required probability is

Probability of x successes in n trails is

$$P(X=x) = {}^{n}c_{x} p^{x} q^{n-x}, \qquad x = 0,1,2,........,n$$

This is called probability distribution of Binomial random variable X or simply Binomial distribution. Symbolically this can be written as

B(X; n, p)

Definition: A random variable X is said to be follow a binomial distribution if its probability function is given by

$$P(X=x) = {}^{n}c_{x}p^{x}q^{n-x}, \qquad x = 0,1,2,\ldots\ldots,n. \quad \text{and } p + q = 1$$

Where, n and p are called parameters of the binomial distribution.

The sum of the probabilities of the binomial distribution is unity.

Proof:

For a binomial distribution the probability function is given by

$$P(X=x) = {}^{n}c_{x}p^{x}q^{n-x}, \qquad x = 0,1,2,\ldots\ldots,n$$

Now,

$$\sum_{x=0}^{n} P(X=x) = \sum_{x=0}^{n} {}^{n}c_{x}p^{x}q^{n-x}$$

$$\sum_{x=0}^{n} P(X=x) = {}^{n}c_{0}p^{0}q^{n-0} + {}^{n}c_{1}p^{1}q^{n-1} + {}^{n}c_{2}p^{2}q^{n-2} + \ldots + {}^{n}c_{n}p^{n}q^{n-n}$$

$$\sum_{x=0}^{n} P(X=x) = (q+p)^{n} = 1 \ (\because q+p=1)$$

- **Uses of Binomial Distribution**:

[i] It has major application in the field of industrial quality control when items are classified as defective and non- defective

[ii] This distribution is used when we like to know the opinion of the public when the voters may be in favour of or against a candidate.

[iii] This distribution is also used in market researches where a consumer may prefer the product of brand A or brand B

[iv] This distribution is used in medical research where a particular drug might cure a person or not

[v] This distribution also used in economic survey where respondents are in for or against a certain economic policy of the govt.

- **Mean of the binomial distribution**

For a binomial distribution the probability function is given by

$$P(X = x) = {}^{n}c_{x} p^{x} q^{n-x}, \qquad x = 0,1,2,........,n$$

Now, the mean of the Binomial distribution is

$$E(X) = \sum_{x=0}^{n} x\, P(X = x)$$

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

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

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

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

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

$$E(X) = np \sum_{x=1}^{n} {}^{n-1}c_{x-1}\, p^{x-1} q^{n-x}$$

$$E(X) = np(q+p)^{n-1}$$

$$E(X) = np(1)^{n-1} \qquad [\because q+p=1]$$

$$E(X) = np$$

$\therefore$ The mean of the binomial distribution is np

Variance of the Binomial distribution:

The variance of the Binomial distribution is

$$V(X) = E(X^2) - [E(X)]^2$$
$$V(X) = E(X^2) - (np)^2 \qquad \ldots(i)\left[\because E(X) = np\right]$$

Now,

$$E(X^2) = \sum_{x=0}^{n} x^2\, {}^n c_x p^x q^{n-x}$$
$$E(X^2) = \sum_{x=0}^{n} [x(x-1) + x]\, {}^n c_x p^x q^{n-x}$$
$$E(X^2) = \sum_{x=0}^{n} x(x-1)\frac{n!}{x!(n-x)!} p^x q^{n-x} + \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x}$$
$$E(X^2) = \sum_{x=0}^{n} x(x-1)\frac{n(n-1)(n-2)!}{x(x-1)(x-2)!(n-x)!} p^2 p^{x-2} q^{n-x} + E(X)$$
$$E(X^2) = n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x} + np$$
$$E(X^2) = n(n-1)p^2 \sum_{x=2}^{n} {}^{n-2}c_{x-2}\, p^{x-2} q^{n-x} + np$$
$$E(X^2) = n(n-1)p^2 (q+p)^{n-2} + np$$
$$E(X^2) = n(n-1)p^2 (1)^{n-2} + np \qquad \ldots(ii)\left[\because q+p = 1\right]$$
$$E(X^2) = n(n-1)p^2 + np$$

Putting (ii) in (i) we get

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

$\therefore$ The variance of the Binomial distribution is npq

Note: In B.D. mean > variance, since mean = np and variance = npq and p + q = 1, therefore

Recurrence relation:

In order to calculate the probabilities one most evaluate ${}^n c_x$ which is tedious to be especially when n and x are large. There is chain relation between the successive probabilities using which the calculation become easily. This relation is called recurrence relation.

The binomial p. m. f. is given by

$$P(x) = {}^{n}c_{x}p^{x}q^{n-x} \qquad x = 0,1,2,...,n$$

$$P(x+1) = {}^{n}c_{x+1}\,p^{x+1}q^{n-(x+1)}$$

$$P(x+1) = {}^{n}c_{x+1}\,p^{x+1}q^{n-x-1} \qquad x = 0,1,2,...,n-1$$

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

$$\frac{P(x+1)}{P(x)} = \frac{\dfrac{n!}{(x+1)1[n-(x+1)]!}pp^{x}q^{n-x}q^{-1}}{\dfrac{n!}{x!(n-x)!}p^{x}q^{n-x}}$$

$$= \frac{n!pp^{x}q^{n-x}q^{-1}}{(x+1)\,x![n-x-1]!} \times \frac{x!(n-x)(n-x-1)!}{n!p^{x}q^{n-x}}$$

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

$$P(x+1) = \left(\frac{n-x}{x+1}\right)\frac{p}{q}P(x)$$

Which is the required recurrence formula for the probabilities of binomial distribution Put x = 0, we get

$$P(0+1) = \left(\frac{n-0}{0+1}\right)\frac{p}{q}P(0)$$

$$P(1) = \frac{np}{q}P(0) \qquad \ldots(i)$$

$$P(x) = {}^{n}c_{x}p^{x}q^{n-x}$$

$$\therefore P(0) = {}^{n}c_{0}p^{0}q^{n-0}$$

$$P(x) = {}^{n}c_x p^x q^{n-x}$$
$$P(0) = {}^{n}c_0 p^0 q^{n-0}$$
$$P(0) = 1 \times 1 \times q^n$$
$$P(0) = q^n$$
$$P(0) = q^n$$

Using equation (1) and expected frequency f(x) = NP(x), we can find remaining probabilities

Remark-

$N(q + p)^n$ this frequencies are called expected or theoretical frequencies. On the other hand the frequencies obtained by making experiments are called actual or observed frequencies. As N increases expected and observed difference is smaller.

Mode of the binomial distribution

Let

$P[x = x]$ is the maximum

$\therefore P[x = (x-1] \le P(x = x) \ge P[x = (x+1)]$

If x is maximum , $x-1$ & $x+1$ is minimum

$$\frac{n!}{(x-1)![n-(x-1)]!} p^{x-1} q^{n-x+1} \le \frac{n!}{x!(n-x)!} p^x q^{n-x} \ge \frac{n!}{(x+1)![n-(x+1)]!} p^{x+1} q^{n-x-1}$$

Multiplying by $\dfrac{x!(n-x)!}{n! p^x q^{n-x}}$

$$\therefore \frac{n!}{(x-1)![n-x+1]!} p^x p^{-1} q^{n-x} q \times \frac{x!(n-x)!}{n! p^x q^{n-x}} \le \frac{n!}{x!(n-x)!} p^x q^{n-x} \times \frac{x!(n-x)!}{n! p^x q^{n-x}} \ge$$

$$\frac{n!}{(x+1)![n-x-1]!} p^x p q^{n-x} q^{-1} \times \frac{x!(n-x)!}{n! p^x q^{n-x}}$$

$$\frac{n!}{(x+1)![n-x-1]!} p^x p q^{n-x} q^{-1} \times \frac{x!(n-x)!}{n! p^x q^{n-x}} \frac{p^{-1} q x}{(n-x+1)} \le 1 \ge \frac{p q^{-1}(n-x)}{x+1}$$

$$\frac{x}{(n-x+1)}\frac{q}{p} \leq 1 \geq \frac{(n-x)}{(x+1)}\frac{p}{q}$$

when,

$$\frac{x}{(n-x+1)}\frac{q}{p} \leq 1$$

$$xq \leq (n-x+1)p$$

$$xq \leq np-xp+p$$

$$xq+xp \leq np+p$$

$$x(q+p) \leq p(n+1)$$

$$x \leq p(n+1) \qquad ...(i)(\because q+p=1)$$

when

$$1 \geq \frac{(n-x)}{(x+1)}\frac{p}{q}$$

$$(x+1)q \geq (n-x)p$$

$$xq+q \geq np-xp$$

$$xq+xp \geq np-q$$

$$x(q+p) \geq np-(1-p) .....(\because q+p=1, \therefore q=1-p)$$

$$x \geq np-1+p$$

$$x \geq p(n+1)-1 \qquad ....(ii)$$

From (i) and (ii), we get

$P(n+1)-1 \leq x \leq (n+1)P$ ---(iii)

Case –(i).

When $(n+1)P-1 = r$ say an integer.

$\therefore \quad (n+1)P = r + 1$

Since, x is always integral in binomial distribution.

Equation (iii) becomes

$r \leq x \leq r+1$

Which shows that distribution is bimodal.

Case-(ii).

When (n+1)P-1 is not an integer (n+1)P is not also an integer and

(n+1) P-1 $\le x \le$ (n+1)P shows f(x) is maximum for greatest integer within (n+1)P:

i.e. $4.3 \le x \le 5.2$

Mode = 5

Moment generating function of binomial distribution:

$\Rightarrow$ Let $x \sim B(n,p)$ then

$$M_X(t) = E(e^{tX})$$
$$M_X(t) = \sum_{x=0}^{n} e^{tx}\, P(X)$$
$$M_X(t) = \sum_{x=0}^{n} e^{tx}\, {}^{n}c_x p^x q^{n-x}$$
$$M_X(t) = \sum_{x=0}^{n} {}^{n}c_x\, (pe^t)^x\, q^{n-x}$$
$$M_X(t) = {}^{n}c_o\, (pe^t)^o\, q^{n-o} + {}^{n}c_1 (pe^t)^1 q^{n-1} + \dots + {}^{n}c_n (pe^t)^n\, q^{n-n}$$
$$M_X(t) = (q+pe^t)^n$$
$$\therefore M_X(t) = (q+pe^t)^n$$

Median:

If M be the median of the distribution, then for x = m, the distribution function is

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

Examples:-

[1] Ten coins are thrown simultaneously find the probability of getting at least seven heads.

Answer:-

Let p be the probability getting a head and therefore p=1/2

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

The probability of getting x heads in q random throw of 10 coins is

$$P(x) = {}^{10}c_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{10-x}$$
$$P(x) = {}^{10}c_x\left(\frac{1}{2}\right)^{10}; \qquad x = 0\ ,\ 1,\ 2\ ,\ldots,10$$

$\therefore$ Probability of getting at least seven head is

$$P(x \geq 7) = P(x = 7) + P(x = 8) + P(x = 9) + P(x = 10)$$

$$P(x) = {}^{10}c_7\left(\frac{1}{2}\right)^{10} + {}^{10}c_8\left(\frac{1}{2}\right)^{10} + {}^{10}c_9\left(\frac{1}{2}\right)^{10} + {}^{10}c_{10}\left(\frac{1}{2}\right)^{10}$$
$$P(x) = \left(\frac{1}{2}\right)^{10}\left[{}^{10}c_7 + {}^{10}c_8 + {}^{10}c_9 + {}^{10}c_{10}\right]$$
$$P(x) = \frac{1}{1024}\left[\frac{10\times9\times8\times7!}{7!3!} + \frac{10\times9\times8!}{8!2!} + \frac{10\times9!}{9!1!} + \frac{10!}{10!0!}\right]$$

$$P(x) = \frac{1}{1024}\left[10\times3\times4 + 5\times9 + 10 + 1\right]$$

$$P(x) = \frac{120 + 45 + 10 + 1}{1024}$$
$$P(x) = \frac{176}{1024}$$
$$\boxed{P(x) = \frac{176}{1024}}$$

[2] A coin is tossed 10 times, what is the probability of getting (i) exactly 3 heads (ii) at least 1 head

Solution-Here, $n = 10, x = 3, p = q = \frac{1}{2}$

Probability of getting exactly 3 heads

$$P(x = 3) = {}^{10}c_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{10-3}$$

$$P(x) = \frac{10 \times 9 \times 8 \times 7!}{3! \times 7!} \left(\frac{1}{2}\right)^{10}$$

$$P(x) = \frac{10 \times 3 \times 4}{1024} = \frac{120}{1024}$$

$$\boxed{P(x) = \frac{15}{128}}$$

(ii) The event at least one is always complementary to the event ‘n one’.So initially we will find the probability for no head i.e. x=0,which is

$$P(x = 0) = {}^{n}c_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{10-0}$$

$$P(x = 0) = \left(\frac{1}{2}\right)^{10}$$

So the probability for at least one head

$$P(x = 0) = 1 - \frac{1}{2^{10}} = \frac{1023}{1024} = 0.99$$

[3] For a binomial distribution with mean 5 and S.D. 2, find the mode

Solution:-

Mean= E(x) = np = 5 and S.D.= $\sqrt{npq}$ = 2

$\sqrt{npq} = 2$

variance = npq = 4

$\frac{npq}{np} = \frac{4}{5}$

$\boxed{q = \frac{4}{5}}$

so

$p = 1 - q = 1 - \frac{4}{5} = \frac{1}{5}$

np = 5

$n \times \frac{1}{5} = 5$

n = 25

Now

$(n+1)p = (25+1)\frac{1}{5} = \frac{26}{5} = 5.2$

Which is not an integer, then

$(n+1)p\text{-}1 \leq x \leq (n+1)p$

$4.2 \leq x \leq 5.2$

Mode = 5

[4] Determine the binomial distribution for which the mean is 4 &variance 3,

find its mode

Answer:- Let $x \sim B(n,p)$,we are given that

Mean = E(x) = np = 4   …(i)

&v(x) = npq = 3   …(ii)

Dividing (ii) by (i),we get

$$\frac{v(x)}{E(x)} = \frac{npq}{np} = \frac{3}{4}$$

$$q = \frac{3}{4}$$

$$p = 1 - q = 1 - \frac{3}{4} = \frac{1}{4}$$

$$np = 4$$

$$n \times \frac{1}{4} = 4$$

$$n = 16$$

Thus the given binomial distribution has parameters n=16 & p=$\frac{1}{4}$

Now

$$\left(n+1\right)p = \left(16+1\right)\frac{1}{4} = \frac{17}{4} = 4.25$$

which is not at integer then

$$\left(n+1\right)p\text{-}1 \le x \le \left(n+1\right)p$$

$$3.25 \le x \le 4.25$$

Mode = 4

[5] If $x \sim B\left(7, \frac{1}{2}\right)$ find mode

Answer:- Given that $n = 7$, $p = \frac{1}{2}$

Now

$$\left(n+1\right)p = \left(7+1\right)\frac{1}{2} = \frac{8}{2} = 4$$

Which is an integer

$$\left(n+1\right)p\text{-}1 \le x \le \left(n+1\right)p$$

$$3 \le x \le 4$$

Modes are 3&4

[6] Comment on the following mean of binomial distribution is 3 & variance 4

Answer: If $x \sim B(n,p)$

Mean = E(x) = np = 3 …(i)

&v(x) = npq = 4 …(ii)

Dividing (ii) by (i),we get

$$\frac{v(x)}{E(x)} = \frac{npq}{np} = \frac{4}{3}$$

$$q = \frac{4}{3}$$

This is impossible, since probability cannot exceed unity.

Hence the given statement is wrong

[7] The mean and variance of a binomial distribution are 4 & 3 res. Find the parameters n & p

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

We are given that

Mean = E(x) = np = 4 …(i)

&v(x) = npq = 3 …(ii)

Dividing (ii) by (i),we get

$$\frac{v(x)}{E(x)} = \frac{npq}{np} = \frac{3}{4}$$

$$q = \frac{3}{4}$$

$$p = 1 - q = 1 - \frac{3}{4} = \frac{1}{4}$$

$$np = 4$$

$$n\times\frac{1}{4} = 4$$

$$n = 16$$

Thus the given binomial distribution has parameters n=16 & $p=\frac{1}{4}$

[8] If the independent random variables X, Y are binomially distributed res. With $n = 3,\ p=\frac{1}{3}$ and $n=5,\ p=\frac{1}{3}$, write down the probability that $X+Y \geq 1$

Solution:- We are given that

$$X \sim B\left(3, \frac{1}{3}\right) \quad \text{and} \quad Y \sim B\left(5, \frac{1}{3}\right)$$

Since X and Y are independent binomial random variable, with $p_1 = p_2 = \frac{1}{3}$

By the additive property of binomial distribution, we get $X+Y \sim B\left(8, \frac{1}{3}\right)$

$$P(X+Y=r) = {}^{n}c_r p^r q^{n-r}$$

$$P(X+Y=r) = {}^{8}c_r \left(\frac{1}{3}\right)^r \left(\frac{2}{3}\right)^{8-r}$$

$$P(X+Y\geq 1) = 1 - P(X+Y<1)$$

$$P(X+Y\geq 1) = 1 - P(X+Y=0)$$

$$P(X+Y\geq 1) = 1 - {}^{8}c_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^{8-0}$$

$$P(X+Y\geq 1) = 1 - 1\times 1\left(\frac{2}{3}\right)^8$$

$$P(X+Y\geq 1) = 1 - \left(\frac{256}{6561}\right)$$

$$P(X+Y\geq 1) = 1 - 0.039$$

$$P(X+Y\geq 1) = 0.961$$

[9] X is binomially distributed with parameters n & P. What is distribution of Y= n-x?

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

It represents the number of success of n independent trials with constant

probability 1 of success for each trial

$\therefore$ Y = n-x represent the number of failure of n independent trials

With constant probability 'q' of failure of each trial. Hence,

$$Y = (n - x) \sim B(n,q)$$

Additive property

[10] Let $X \sim B(n_1, p_1)$ & $Y \sim B(n_2, p_2)$ be independent random variables then

$$M_X(t) = (q + p_1 e^t)^{n_1} \ \& \ M_Y(t) = (q + p_2 e^t)^{n_2}$$

Then

$$M_{X+Y}(t) = M_X(t) M_Y(t) \ (\because X \ \& \ Y \text{ are independent})$$

$$M_{X+Y}(t) = (q + p_1 e^t)^{n_1} (q + p_2 e^t)^{n_2} \qquad \ldots(i)$$

This is not expressed in the form of $(q + pe^t)^n$. From uniqueness theorem of m.g.f. It follows that X + Y is not a binomial variats

i. e. Sum of two independent binomial variates is not a binomial variate.

i. e. Binomial Distribution does not possess the additive or reproductive properties.

However, If we take $p_1 = p_2 = p$ then equation (i)

$$M_{X+Y}(t) = (q + pe^t)^{n_1 + n_2}$$

Which is the m.g.f. of a binomial variate, with parameters. Hence,

$X + Y \sim B(n_1 + n_2, p)$ Thus the binomial distribution possesses the additive reproductive property if $p_1 = p_2$

[11] The mean and variance of binomial distribution are 4 and 4/3 respectively find P(X>1).

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

We are given that

Mean = E(x) = np = 4 …(i)

&V(x) = npq = 4/3 …(ii)

Dividing (ii) by (i),we get

$$\frac{V(x)}{E(x)} = \frac{npq}{np} = \frac{\frac{4}{3}}{4}$$
$$q = \frac{1}{3}$$
$$p = 1 - q = 1 - \frac{1}{3} = \frac{2}{3}$$
$$np = 4$$
$$n \times \frac{2}{3} = 4$$
$$n = 6$$

$$P(x > 1) = 1 - P(x < 1$$
$$P(x > 1) = 1 - P(X = 0)$$
$$P(x > 1) = 1 - {}^{6}c_{0}\left(\frac{2}{3}\right)^{0}\left(\frac{1}{3}\right)^{6-0}$$
$$P(x > 1) = 1 - \left(\frac{1}{3}\right)^{6}$$
$$P(x > 1) = 1 - \frac{1}{729}$$
$$P(x > 1) = 1 - 0.0014$$
$$P(x > 1) = 0.9986$$

[12] If $X \sim B(n,p)$ then show that

$$[i]\ E\left(\frac{x}{n}-p\right)^2=\frac{pq}{n}$$

$$[ii]\ Cov\left(\frac{x}{n},\frac{n-x}{n}\right)=\frac{-pq}{n}$$

Solution:- Since, $X \sim B(n,p)$

$E(x)=np \quad \& \qquad V(x)=npq$

$$\therefore E\left(\frac{x}{n}\right)=\frac{1}{n}E(x)=\frac{1}{n}\times np=p \qquad \ldots(i)$$

$$\& V\left(\frac{x}{n}\right)=\frac{1}{n^2}v(x)=\frac{1}{n^2}\times npq=\frac{pq}{n} \qquad \ldots(ii)$$

[i] Now, $E\left(\frac{x}{n}-p\right)^2=E\left[\frac{x}{n}-E\left(\frac{x}{n}\right)\right]^2$

$$E\left(\frac{x}{n}-p\right)^2=Var\left(\frac{x}{n}\right) \qquad (\because \text{by definition of variance})$$

$$E\left(\frac{x}{n}-p\right)^2=\frac{pq}{n} \qquad (\because (ii))$$

Hence,

$$E\left(\frac{x}{n}-p\right)^2=\frac{pq}{n}$$

Now,

(ii)

$$Cov\left(\frac{x}{n},\frac{n-x}{n}\right)=E\left[\left\{\frac{x}{n}-E\left(\frac{x}{n}\right)\right\}\left\{\frac{n-x}{n}-E\left(\frac{n-x}{n}\right)\right\}\right]$$

$$=E\left(\frac{x}{n}-p\right)\left\{\left(1-\frac{x}{n}\right)-E\left(1-\frac{x}{n}\right)\right\} \qquad (\because (i))$$

$$= E\left[\left(\frac{x}{n}-p\right)\left\{\left(1-\frac{x}{n}\right)-\left\{1-E\left(\frac{x}{n}\right)\right\}\right\}\right]$$

$$= E\left[\left(\frac{x}{n}-p\right)\left\{\left(1-\frac{x}{n}\right)-(1-p)\right\}\right] \qquad (\because (i))$$

$$= E\left[\left(\frac{x}{n}-p\right)\left(1-\frac{x}{n}-1+p\right)\right]$$

$$= E\left[\left(\frac{x}{n}-p\right)\left(-\frac{x}{n}+p\right)\right]$$

$$= -E\left[\left(\frac{x}{n}-p\right)\left(\frac{x}{n}-p\right)\right]$$

$$= -E\left(\frac{x}{n}-p\right)^2$$

$$= -E\left[\frac{x}{n}-E\left(\frac{x}{n}\right)\right]^2$$

$$= -Var\left(\frac{x}{n}\right)$$

$$= -\frac{pq}{n}$$

$$\therefore Cov\left(\frac{x}{n},\frac{n-x}{n}\right)=\frac{-pq}{n}$$

[13] Assuming the probability of male birth as $\frac{1}{2}$, find the probability distribution of number of boys out of 5 births.

(a) Find the probability that a family of 5 children have

(i) at least one boy

(ii) at most 3 boys

(b) Out of 960 families with 5 children each find the expected number of families of (i) at least one boy (ii) at most 3 boys

Solution: Let the random variable *X* measures the number of boys out of 5 births. Clearly X is a binomial random variable. So we apply the Binomial probability function to calculate the required probabilities.

$$X \sim B\left(5, \frac{1}{2}\right)$$

$$P(X) = {}^{n}C_{x} p^{x} q^{n-x} \quad ; x = 0,1,2,3,4,5$$

The probability distribution of X given below:

X	0	1	2	3	4	5
P(X)	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

(a) The required probabilities are

(i) $P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{1}{32} = \frac{31}{32}$

(ii) $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

$$P(X \leq 3) = \frac{1}{32} + \frac{5}{32} + \frac{10}{32} + \frac{10}{32} = \frac{26}{32}$$

$$P(X \leq 3) = \frac{26}{32}$$

(b) Out of 960 families with 5 children, the expected number of families with

(i) at least one boy = $960 \times P(X \geq 1)$

$= 960 \times \frac{31}{32}$

$= 930$

(ii) at most 3 boys $= 960 \times P(X \leq 3)$

$= 960 \times \frac{31}{32}$

$= 780$

Fitting Of Binomial Distribution:

When we fit a theoretical or expected distribution to a given observed or actual distribution our purpose is to examine whether the observe distribution may be regarded as the distribution in a random sample from the population characterized by the theoretical distribution.

Procedure

When binomial distribution is to be fitted to observe data the following steps are

(i) Determine the values of p and q

(ii) Find P(0) by using relation $P(0) = q^n$

(iii) Find expected frequency using $f(x) = N \times q^n$ or $f(x) = N \times P(x)$

Write down the variable values and expected frequencies i.e. Table

[13] Seven coins are tossed and number of heads noted. The experiment is repeated 128 times and the following distribution is obtained.

No. of heads	0	1	2	3	4	5	6	7	total
frequencies	7	6	19	35	30	23	7	1	128

Fit a binomial distribution assuming the coin is unbalanced.

Solution:-Here,

$$p = q = \frac{1}{2} \quad \therefore \frac{p}{q} = 1, \quad n = 7, \quad N=128$$

Now,

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

$$P(x) = {}^n c_x p^x q^{n-x}$$

$$P(x) = {}^n c_0 p^0 q^{n-0} \qquad \left(\because {}^n c_0 = 1, \quad p^0 = 1\right)$$

$$P(x) = q^n$$

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

$$Nq^n = 128\left(\frac{1}{2}\right)^7 = 128 \times \frac{1}{128} = 1$$

Put x =0 in recurrence relation

$$P(x+1) = \left(\frac{n-x}{x+1}\right)\frac{p}{q}P(x)......1$$

$$P(0+1) = \left(\frac{n-0}{0+1}\right)(1)\,P(0)$$

$$P(1) = \left(\frac{7}{128}\right)$$

Put x =1 in recurrence relation

$$P(x+1) = \left(\frac{n-x}{x+1}\right)\frac{p}{q}P(x)$$

$$P(1+1) = \left(\frac{7-1}{1+1}\right)(1)\,P(1)$$

$$P(2) = \left(\frac{6}{2}\right)\times\left(\frac{7}{128}\right)$$

$$P(2) = \frac{21}{128}$$

Put x =2 in recurrence relation

$$P(x+1) = \left(\frac{n-x}{x+1}\right)\frac{p}{q}P(x)$$

$$P(2+1) = \left(\frac{7-2}{2+1}\right)(1)\,P(2)$$

$$P(3) = \left(\frac{5}{3}\right)\times\left(\frac{21}{128}\right)$$

$$P(3) = \frac{35}{128}$$

Put x =3 in recurrence relation

$$P(x+1) = \left(\frac{n-x}{x+1}\right)\frac{p}{q}P(x)$$

$$P(3+1) = \left(\frac{7-3}{3+1}\right)(1)\,P(3)$$

$$P(4) = \left(\frac{4}{4}\right)\times\left(\frac{35}{128}\right)$$

$$P(4) = \frac{35}{128}$$

HYPERGEOMETRIC DISTRIBUTION:

Difference between Binomial and Hypergeometric Distribution:

The **hypergeometric distribution**, intuitively, is the probability distribution of the number of red marbles drawn from a set of red and blue marbles, without replacement of the marbles. In contrast, the binomial distribution measures the probability distribution of the number of red marbles drawn with replacement of the marbles.

The following conditions should be satisfied for the application of hypergeometric distribution:

[i] The population is divided into two mutually exclusive categories.

[ii] The successive outcomes are dependent.

[iii] The probability of 'success' changes from trial to trial.

[iv] The number of draws are fixed.

Introduction:

Suppose a bag contains balls of which M are red and N-M are black. A sample of n balls is drawn without replacement from the N boys.

Let X denote the number of red balls in the sample hence possible values of X are 0,1,2-------,n (assuming n < m)

The A class contain 10 male and 17 female student suppose 6 student are drawn at random from this class without replacement and we are interested in the numbers of male student drawn. Clearly at the first drawn, probability of getting a male student is $\frac{10}{27}$

Suppose male student is selected as the first draw, because it would we keep aside, the probability of getting a student of the second draw would be $\frac{9}{26}$ Thus, p does not remain constant also the successive draws are not independent (dependent) probability of getting male students in the second draw is dependent on which student you have drawn at the first draw

Sampling without replacement, the hyper geometric distribution used.

Probability Mass Function of Hypergeometric distribution:

$$P\left[X=x\right]=P\left(x\right)=\frac{{}^{M}C_{x}\;{}^{N-M}C_{n-x}}{{}^{N}C_{n}}, \qquad x=0,1,2,...,n$$

$$= \quad 0 \qquad \text{otherwise}$$

X is hyper geometrically distributed with parameters N, M and n.

$\therefore X \sim H\left(N, M, n\right)$

Note:

M-male

x-no of male students in the sample

N-M-female

N-total students

n-sample students

(n-x)- no of female students in the sample

Mean and variance

Let

$$X \sim H(N, M, n)$$

$$\therefore P(x) = \frac{{}^{M}c_x \; {}^{N-M}C_{n-x}}{{}^{N}C_n}; \qquad x = 1, 2, \ldots\ldots. n$$

$$\text{mean} = \mu_1' = E(x) = \sum_{x=0}^{n} X\,P(X)$$

$$= \sum_{x=0}^{n} X \frac{{}^{M}c_x \; {}^{N-M}c_{n-x}}{{}^{N}c_n}$$

$$= \frac{1}{{}^{N}c_n} \sum_{x=0}^{n} X \frac{M!}{x!(M-x)!} \; {}^{N-M}c_{n-x}$$

$$= \frac{M}{{}^{N}c_n} \sum_{x=1}^{n} \frac{(M-1)!}{(x-1)!(M-x)!} \; {}^{N-M}c_{n-x}$$

$$= \frac{M}{{}^{N}c_n} \sum_{x=1}^{n} \frac{(M-1)!}{(x-1)![M-1+1-x]!} \; {}^{N-M+1-1}c_{n-x+1-1}$$

$$= \frac{M}{{}^{N}c_n} \sum_{x=1}^{n} \frac{(M-1)!}{(x-1)![(M-1)-(x-1)]!} \; {}^{N-1-M+1}c_{n-1-x+1}$$

$$\text{Mean} = \frac{M}{{}^{N}c_n} \sum_{x=1}^{n} {}^{M-1}c_{x-1} \; {}^{[(N-1)-(M-1)]}c_{[(n-1)-(x-1)]}$$

$$\text{Mean} = \frac{M}{{}^{N}c_n} \; {}^{N-1}c_{n-1} \left(\because \sum_{x=1}^{n} {}^{N-1}c_{M-1} \; {}^{n+1}c_{n+1} = {}^{N+n}c_{M+n} \right)$$

$$\text{Mean} = \frac{M}{\frac{N!}{n!(N-n)!}} \frac{(N-1)!}{(n-1)![N-1-(n-1)]!}$$

$$\text{Mean} = \frac{M \times n \times (n-1)!(N-n)!}{N(N-1)!} \times \frac{(N-1)!}{(n-1)!(N-n)!}$$

$$\text{Mean} = \frac{Mn}{N}$$

$$\text{Mean} = \frac{nM}{N}$$

$$\text{Variance} = \sigma^2 = \mu_2' - (\mu_1')^2$$

$$\text{Variance} = E(X^2) - [E(X)]^2$$

$$\text{Variance} = E[X^2 - X + X] - \left[\frac{nM}{N}\right]^2 \ldots\ldots\ldots.(i)$$

Now,

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

$$E[X(X-1) + X] = \sum_{x=0}^{n} X(X-1)P(X) + \sum_{x=0}^{n} XP(X)$$

$$= \frac{1}{{}^N c_n} \sum_{x=0}^{n} X(X-1)\, {}^M c_x \; {}^{N-M} c_{n-x} + \frac{nM}{N}$$

$$= \frac{1}{{}^N c_n} \sum_{x=0}^{n} X(X-1) \frac{M!}{x!(M-x)!} \; {}^{N-M} c_{n-x} + \frac{nM}{N}$$

$$= \frac{1}{{}^N c_n} \sum_{x=0}^{n} X(X-1) \frac{M(M-1)(M-2)!}{x(x-1)(x-2)![M-2+2-x]!} \; {}^{N-M} c_{n-x} + \frac{nM}{N}$$

$$= \frac{M(M-1)}{{}^N c_n} \sum_{x=2}^{n} \frac{(M-2)!}{(x-2)![(M-2)-(x-2)]!} \; {}^{[(N-2)-(M-2)]} c_{[(n-2)-(x-2)]} + \frac{nM}{N}$$

$$= \frac{M(M-1)}{{}^N c_n} \; {}^{N-2} c_{n-2} + \frac{nM}{N}$$

$$= \frac{M(M-1)}{\frac{N!}{n!(N-n)!}} \frac{(N-2)!}{(n-2)![(N-2)-(n-2)]!} + \frac{nM}{N}$$

$$= M(M-1) \times \frac{n(n-1)(n-2)!(N-n)!}{N(N-1)(N-2)!} \times \frac{(N-2)!}{(n-2)!(N-n)!} + \frac{nM}{N}$$

$$= \frac{nM}{N}\left[\frac{(M-1)(n-1)}{(N-1)} + 1\right]$$

Put this vaiue in (i) we get

$$\text{variance} = \frac{nM}{N}\left[\frac{(M-1)(n-1)}{(N-1)} + 1\right] - \frac{n^2M^2}{N^2}$$

$$\text{variance} = \frac{nM}{N}\left[\frac{(M-1)(n-1)}{(N-1)} + 1 - \frac{nM}{N}\right]$$

Examples:

[1] Among the 200 employees of a company, 160 are union members and the others are non-union. If four employees are to be chosen to serve on the staff welfare committee, find the probability that two of them will be union members and the others non-union, using (i) hypergeometric distribution, (ii) binomial distribution as an approximation.

Solution: Let X denote number of union members selected in the sample.

$\therefore\ X \sim H\left(N = 200, M = 160, n = 4\right).$

(i) The required probability is

$$P\left[X = x\right] = P\left(x\right) = \frac{{}^{M}c_{x}\ {}^{N\text{-}M}c_{n\text{-}x}}{{}^{N}c_{n}}, \qquad x = 0,1,2,...,n$$

$$= \quad 0 \quad \text{otherwise}$$

$$P\left[X = 2\right] = \frac{{}^{160}c_{2}\ {}^{200\text{-}160}c_{4\text{-}2}}{{}^{200}c_{4}}$$

$$P\left[X = 2\right] = \frac{{}^{160}c_{2}\ {}^{40}c_{2}}{{}^{200}c_{4}}$$

$$P\left[X = 2\right] = \frac{\frac{160!}{(160\text{-}2)!2!} \times \frac{40!}{(40\text{-}2)!2!}}{\frac{200!}{(200\text{-}4)!196!}} \qquad \left[\because\ {}^{n}c_{r} = \frac{n!}{(n\text{-}r)!r!}\right]$$

$$P\left[X = 2\right] = \frac{\frac{160\times159\times158!}{158!\times2!} \times \frac{40\times39\times38!}{38!\times2!}}{\frac{200\times199\times198\times197\times196!}{196!\times4!}}$$

$$P\left[X = 2\right] = \frac{\frac{160\times159}{2!} \times \frac{40\times39}{2!}}{\frac{200\times199\times198\times197}{4!}}$$

$$P\left[X = 2\right] = \frac{160\times159}{2\times1} \times \frac{40\times39}{2\times1} \times \frac{4\times3\times2\times1}{200\times199\times198\times197}$$

$$P\left[X = 2\right] = \frac{952473600}{6209755200}$$

$$P\left[X = 2\right] = 0.1534$$

(ii) binomial distribution as an approximation

$$\therefore\ X \sim B\left(n = 4,\ p = \frac{M}{N} = \frac{160}{200} = 0.8\right)$$

$$P(X = x) = {}^{n}c_{x}\left(p\right)^{x}\left(q\right)^{n\text{-}x}$$

So the probability for union members

$$P(X = 2) = {}^{4}c_{2}(0.8)^{2}(0.2)^{2}$$
$$P(X = 2) = 0.1536$$

[2] A room has 4 sockets. From a collection of 12 bulbs, of which only 5 are good, a person selects 4 bulbs at random (without replacement) and puts them in the sockets. Find the probability that

(i) the room is lighted, (ii) exactly one bulb in the selected bulbs is good. Solution: Notice that N = 12, M = 5, n = 4, X = number of good bulbs in the sample.

$\therefore\ X \sim H\,(N = 12, M = 5, n = 4)$.

(i) The required probability is

$$P[X = x] = P(x) = \frac{{}^{M}c_{x}\ {}^{N\text{-}M}c_{n\text{-}x}}{{}^{N}c_{n}}, \qquad x = 0,1,2,\ldots,n$$
$$= \quad 0 \quad \text{otherwise}$$

$$P[X = x] = \frac{{}^{5}c_{x}\ {}^{7}c_{4\text{-}x}}{{}^{12}c_{4}}; \qquad x = 0,1,2,3,4$$

(i) The room is lighted even if a single bulb is good. Therefore, the required probability is

$$P[X \geq 1] = 1 - P[X = 0]$$
$$P[X \geq 1] = 1 - \frac{{}^{5}c_{0}\ {}^{7}c_{4}}{{}^{12}c_{4}}$$

$$P[X \geq 1] = 1 - \frac{1 \times \dfrac{7!}{(7\text{-}4)! \times 4!}}{\dfrac{12!}{(12\text{-}4)! \times 4!}} \qquad \left[\because\ {}^{n}c_{0} = 1 \text{ and } {}^{n}c_{r} = \frac{n!}{(n\text{-}r)!r!}\right]$$

$$P[X \geq 1] = 1 - 1 \times \frac{7!}{3! \times 4!} \times \frac{8! \times 4!}{12!}$$

$$P[X \geq 1] = 1 - 1 \times \frac{7 \times 6 \times 5}{3 \times 2 \times 1} \times \frac{4 \times 3 \times 2 \times 1}{12 \times 11 \times 10 \times 9}$$

$$P[X \geq 1] = 1 - \frac{5040}{71280}$$

$$P[X \geq 1] = 1 - 0.07$$

$$P[X \geq 1] = 1 - 0.93$$

(ii) Exactly one bulb in the selected bulbs is good

$$P[X = 1] = \frac{{}^{5}c_{1}\,{}^{7}c_{3}}{{}^{12}c_{4}}$$

$$P[X = 1] = \frac{\frac{5!}{(5\text{-}1)!\times 1!} \times \frac{7!}{(7\text{-}3)!\times 3!}}{\frac{12!}{(12\text{-}4)!\times 4!}} \left[\because {}^{n}c_{r} = \frac{n!}{(n\text{-}r)!r!}\right]$$

$$P[X = 1] = \frac{5!}{4!\times 1!} \times \frac{7!}{4!\times 3!} \times \frac{8!\times 4!}{12!}$$

$$P[X = 1] = \frac{5\times 7\times 6\times 5}{3\times 2} \times \frac{4\times 3\times 2\times 1}{12\times 11\times 10\times 9}$$

$$P[X = 1] = \frac{25200}{71280}$$

$$P[X = 1] = 0.3535$$

Use of Hypergeometric Distribution:

[i] It is useful for situations in which observed information cannot re-occur, such as poker (and other card games) in which the observance of a card implies it will not be drawn again in the hand.

[ii] It is also applicable to many of the same situations that the binomial distribution is useful for, including risk management and statistical significance.

[iii] The **hypergeometric test** is used to determine the statistical significance of having drawn k objects with a desired property from a population of size N with K total objects that have the desired property. In other words, it tests to see whether a sample is truly random or whether it over-represents (or under-represents) a particular demographic.

[iv] In quality control department a random sample of items is inspected from a consignment containing defective and non-defective items.

[v] In opinion survey, where the persons have to give answers of YES, NO type questions.

Multiple Choice Questions (MCQ)

Choose the correct alternative from the following:

[1] A theoretical probability distribution.

(a) does not exist. (b) exists only in theory.

(c) exists in real life. (d) both (b) and (c).

Answer: (d) both (b) and (c).

[2] Probability distribution may be

(a) discrete. (b) continuous.

(c) infinite. (d) (a) or (b).

Answer: (d) (a) or (b).

[3] An important discrete probability distribution is

(a) Poisson distribution. (b) Normal distribution.

(c) Cauchy distribution. (d) Log normal distribution.

Answer: (a) Poisson distribution.

[4] An important continuous probability distribution

(a) Binomial distribution. (b) Poisson distribution.

(c) Geometric distribution (d) Normal distribution.

Answer: (d) Normal distribution.

[5] Parameter is a characteristic of

(a) Population. (b) Sample.

(c) Probability distribution. (d) both (a) and (b).

Answer: (a) Population.

[6] An example of a parameter is

(a) Sample mean. (b) Population mean.

(c) Binomial distribution. (d) Sample size.

Answer: (b) Population mean.

[7] A trial is an attempt to

(a) Make something possible

(b) Make something impossible

(c) Prosecute an offender in a court of law

(d)Produce an outcome which is neither certain nor impossible

Answer: (d)Produce an outcome which is neither certain nor impossible

[8] The important characteristic(s) of Bernoulli trials

(a) Each trial is associated with just two possible outcomes.

(b) Trials are independent.

(c) Trials are infinite.

(d) Both (a) and (b).

Answer: (d) Both (a) and (b).

[9] The probability mass function of binomial distribution is given by

(a) $P(X) = p^x q^{n-x}$ (b) $P(X) = {}^n c_x p^x q^{n-x}$

(c) $P(X) = {}^n c_x p^{n-x} q^x$ (d) NOTA

Answer: (a) $P(X) = p^x q^{n-x}$

[10] If x is a binomial variable with parameters n and p, then x can assume

(a) Any value between 0 and n.

(b) Any value between 0 and n, both inclusive.

(c) Any whole number between 0 and n, both inclusive.

(d) Any number between 0 and infinity.

Answer: (c) Any whole number between 0 and n, both inclusive.

[11] A binomial distribution is

(a) Never symmetrical. (b) Never positively skewed.

(c) Never negatively skewed. (d) Symmetrical-when p = 0.5.

Answer: (d) Symmetrical-when p = 0.5.

[12] The mean of a binomial distribution with parameter n and p is

(a) $n(1\text{-}p)$ (b) $np(1\text{-}p)$

(c) np (d) $\sqrt{np(1\text{-}p)}$

Answer: (c) np

[13] The variance of a binomial distribution with parameters n and p is

(a) $np^2(1\text{-}p)$ (b) $\sqrt{np(1\text{-}p)}$

(c) $nq(1\text{-}q)$ (d) $n^2p^2(1\text{-}p)^2$

Answer: (c) $nq(1\text{-}q)$

[14] An example of a bi-parametric discrete probability distribution is

(a) Binomial distribution (b) Poisson distribution

(c) Normal distribution (d) Both (a) and (b)

Answer: (a) Binomial distribution

[15] For a binomial distribution, mean and mode

(a) Are never equal. (b) Are always equal.

(c) Are equal when (d) Do not always exist.

Answer: (c) Are equal when

[16] The mean of binomial distribution is

(a) Always more than its variance

(b) Always equal to its variance

(c) Always less than its variance

(d) Always equal to its standard deviation

Answer: (a) Always more than its variance

[17] For a binomial distribution, there may be

(a) One mode (b) Two modes

(c) (a) or (b) (d) NOTA

Answer: (c) (a) or (b)

[18] The maximum value of the variance of a binomial distribution with parameters n and p is

(a) $\frac{n}{2}$ (b) $\frac{n}{4}$

(c) $np(1-p)$ (d) $2n$

Answer: (b) $\frac{n}{4}$

[19] The method usually applied for fitting a binomial distribution is known as

(a) Method of least squares

(b) Method of moments

(c) Method of probability distribution

(d) Method of deviation

Answer: (b) Method of moments

<u>Self-Assessment Questions:</u>

[1] "Binomial random variable measures the number of successes in a Bernoulli Process". Explain this statement. Also develop and generalize Binomial probability rule with the help of an example.

[2] State the important properties of a Binomial distribution. Give examples of

some of the important area where Binomial distribution is used.

[3] A gardener knows from his personal experiences that 2% of seedlings fail to service on transplantation. Find the mean, standard deviation and moment coefficient of skewness of the distribution of rate of failure to service in a sample of 400 seedlings.

[4] If the sum of mean and variance of a binomial distribution of 5 trials is 9/5, find the binomial distribution.

[5] The mean and variance of a binomial distribution are 2 and 1.5 respectively. Find the probability of

(a) 2 successes (b) at least 2 successes (c) at most 2 successes.

[6] 150 random samples of 4 units each are inspected for number of defective item. The results are:

Number of defective items	0	1	2	3	4
Number of Samples	28	62	46	10	4

Fit a binomial distribution to the observed data.

[7] The probability that a particular injection will have reaction to an individual is 0.002. Find the probability that out of 1000 individuals (a) no, (b) 1, (c) at least 1, and (d) almost 2; individuals will have reaction from the injection.

[8] In a razor blades manufacturing factory, there is small chance of 1/500 for any blade to be defective. The blades are supplied in packets of 10. Find the approximate number of packets containing (a) no, (b) 1, and (c) 2 defective blades in a consignment of 10,000 packets.

[B] Previous Year University Questions

[1] Give an expression for m.g.f. of a binomial distribution with mean 2.4 and variance 1.48. **(April-2015)**

[2] If X follows discrete uniform distribution on 1, 2, ... n. Mean of X . **(April-2015)**

(a) $\frac{n}{2}$ (b) $\frac{n+1}{2}$

(C) $n+\frac{1}{2}$ (D) $1+\frac{n}{2}$

[3] For a degenerate distribution at X = c, mean and variance are. **(April-2015)**

(a) c, c (b) c, 0

(C) 0, 0 (D) 0, c

[4] Find mean of a Bernoulli distribution with parameter 'p' **(April-2015)**

[5] Suppose X follows binomial distribution with parameters n and p. Obtain the probability distribution of Y = n- x using its m. g. f. **(April-2015)**

[6] Define discrete uniform distribution on integers 1 to n. Obtain its variance and also comment on its median. **(April-2015)**

[7] Let $X \rightarrow B\left(n_1, p\right),\ Y \rightarrow B\left(n_2, p\right)$ and are independent r.v.'s. Obtain the conditional distribution of X given X + Y = n. Identify it. **(April-2015)**

[8] Define a binomial distribution. Obtain its mode, it is unique. **(April-2015)**

[9] Define a degenerate distribution. **(October-2015) (October-2016)**

[10] Give one real-life situation where binomial distribution can be applied. **(October-2015)**

[11] If X follows discrete uniform distribution on 0, 1, 2, ...,n and the mean of the distribution is 6. Hence the value of n is. **(October-2015)**

(a) 11 (b) 6

(C) 36 (D) 12

[12] Let $X \sim B(10, 0.6)$, find mode of X. **(October-2015)**

[13] For a Bernoulli r.v. X, $\mu_3' = 0.6$, find Var (X). **(October-2015)**

[14] Let $X \sim B(n, p)$, find the m. g. f. of X. **(October-2015)**

[15] Define hypergeometric distribution and find its mean. **(October-2015)**

[16] Define a discrete uniform distribution with parameter 'n', also find its variance. **(October-2015)**

[17] Let X and Y be two independent binomial variables with parameters (n_1 = 6, p = 0.6) and (n_2 = 7, P = 0.6) respectively. Find **(October-2015)**

(A) P[X + Y = 6] (B) P[X = 2 / X + Y = 9]

[18] State moment generating function (m.g.f.) of a binomial distribution with parameters n and p. **(April-2016)**

[19] Let $X \sim B\left(n=10, p=\frac{1}{3}\right)$, then state the mode of X. **(April-2016)**

[20] Define a discrete uniform distribution with parameter 'n'. Find its mean and variance. **(April-2016)**

[21] Let $X \sim B(n, p)$,. State the c. g. f. of X, hence find mean of X. **(April-2016)**

[22] State and prove binomial approximation to hypergeometric distribution. **(April-2016) (October-2016)**

[23] State the p. m. f. of hypergeometric distribution. Find mean and variance of the distribution. **(April-2016)**

[24] Let $X \sim B\left(n, p=\frac{1}{3}\right)$, and mean of X is 5 then the value of n is ….

(October-2016)

(a) 15	(b) 5
(C) 5/3	(D) 10/3

[25] Explain with an illustration, what is meant by a Bernoulli trial.

[26] If $X \sim B(n, p)$, with n = 20, E(X) = 8, find parameter p and var (X). **(April-2017)**

[27] Give one real-life situation where hypergeometric distribution can be applied. **(April-2017)**

[28] Define Bernoulli distribution with parameter p. Obtain μ_3', for a Bernoulli r.v. **(April-2017)**

[29] Find recurrence relation between the successive probabilities of binomial distribution with parameters n and p. **(April-2017)**

[30] Define a binomial distribution and find its mean. **(April-2017)**

[31] A lot contains 10 items of which 3 are defective. Two items are drawn at random from the lot one after other without replacement. Find the probability that both items are non-defective. **(April-2017)**

[32] State the p. m. f. of H(N, M, n) variable. Obtain its mean. **(April-2017)**