Hypergeometric Distribution :- We know that binomial distribution is applied whenever we draw a random sample with replacement. This is because, in sampling with replacement, the probability of getting 'success' p, remains same at every draw, Also the successive draws remain independent. Thus the assumptions of binomial experiment are satisfied. Now consider the following situation.

A bag contains 4 red and 5 black balls. Suppose 3 balls are drawn at random from this bag without replacement and we are interested in the number of red balls drawn. Clearly at the first draw probability of getting a red ball is $\frac{4}{9}$. Now, suppose a red ball is selected at the first draw. Because, it would be kept aside, the probability of getting a red ball at the second draw would be $\frac{3}{8}$. Thus 'p' does not remain constant. Also, the successive draws are not independent. Probability of getting red balls in the second draw is dependent on which ball you have drawn at the first draw. Thus, in case of sampling without replacement, the binomial distribution cannot be applied.

In such situations, the hypergeometric distribution is used. Consider the following situation.

Suppose a bag contains N balls of which M are red and N−M are black. A sample of 'n' balls is drawn without replacement from the N balls. Let X denote the number of red balls in the sample. Hence the possible values of X are 0, 1, 2 - - - n (assuming $n \leq M$)

The p.m.f is obtained in the following manner.

We want to get $P[X=x]$ p.m.f (Probability Mass function)

$$\begin{tikzcd} & N \arrow[dl] \arrow[dr] & \\ M \arrow[d] & & N-M \arrow[d] \\ x & & n-x \end{tikzcd}$$

If the sample of 'n' balls contains 'x' red balls, then it will contain 'n-x' black balls. Hence, number of ways in which x red balls can be selected from M red balls is $\binom{M}{x}$ and number of ways in which n-x black balls can be selected from N-M black balls is $\binom{N-M}{n-x}$. The sample contains both red and black balls. Therefore, the total number of ways in which the above event can occur is $\binom{M}{x}\binom{N-M}{n-x}$

In all 'n' balls are selected from N balls. Therefore the total number of possible selections is $\binom{N}{n}$

Using the definition of probability of an event, we get

$$P(x) = P[X=x] = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2 \ldots n$$

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

The above $P(x)$ is called as the p.m.f of hypergeometric distribution with parameters N, M and n.

Notations:- $X \rightarrow H(N, M, n)$

If we don't assume $n \leq M$, then the range X is 0, 1, 2 ... min(n, M). This is because at the most M red balls can be there in the sample.

Ex. 1 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.

Sol^n. Notice that $N=12$, $M=5$, $n=4$, X = Number of good bulbs in the sample.

$$\therefore \quad X \rightarrow H(N=12, M=5, n=4)$$

$$\therefore \quad P(x) = \frac{\binom{5}{x}\binom{7}{4-x}}{\binom{12}{4}}, \quad x = 0, 1, \ldots 4.$$

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

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

$$= 1 - \frac{\binom{5}{0}\binom{7}{4}}{\binom{12}{4}} = 0.9292$$

(ii) $$P[X=1] = \frac{\binom{5}{1}\binom{7}{3}}{\binom{12}{4}} = 0.357$$

Q. 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 hypergeometric distribution.

Sol^n Let X denote number of union members selected in the sample

$\therefore X \rightarrow H(N = 200, M = 160, n = 4)$

(i) The required Probability is

$$P[X=2] = \frac{\binom{160}{2}\binom{40}{2}}{\binom{200}{4}}$$

$$= \frac{\frac{160 \times 159}{2} \times \frac{40 \times 39}{2}}{\frac{200 \times 199 \times 198 \times 197}{4 \times 3 \times 2}}$$

$$= 0.1534$$

Geometric Distribution:- If repeated independent trials can result in a success with probability P and a failure with probability $q = 1-P$, then the probability distribution of the random variable X, the number of the trial on which the first success occurs, is

$$g(x;P) = Pq^{x-1}, \quad x = 1,2,3---$$

Q.1 For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

Solⁿ Using the geometric distribution with

$x = 5$, and $P = 0.01$

$$g(5;0.01) = (0.01)(0.99)^4$$
$$= 0.0096$$

Q.2 At a 'busy time' a telephone exchange is very near capacity, so callers have difficulty placing their calls. It may be of interest to know the number of attempts necessary in order to make a connection. Suppose that we let $p = 0.05$ be the probability of a connection during a busy time. We are interested in knowing the probability that 5 attempts are necessary for a successful call.

Solⁿ Using the geometric distribution with $x = 5$ and $P = 0.05$

$$P(X=x) = g(5;0.05) = (0.05)(0.95)^4 = 0.0411$$

Quite often, in applications dealing with the geometric distribution, the mean and variance are important. For example, in ~~Ex~~ this example, the expected number of calls necessary to make a connection is quite important. The following theorem states without proof the mean and variance of the geometric distribution.

<u>Theorem:-</u> The Mean and Variance of a random variable following the geometric distribution are

$$\mu = \frac{1}{p} \quad \text{and} \quad \sigma^2 = \frac{1-p}{p^2}$$

Negative Binomial Distribution :- A negative binomial experiment is a statistical experiment that has the following properties:

* The experiment consists of 'x' repeated trials
* Each trial can result in just two possible outcomes, we call one of these outcome a success and the other, a failure.
* The probability of success, denoted by P, is the same on every trial.
* The trials are independent, that is the outcome on one trial does not affect the outcome on other trials.
* The experiment continues until 'r' successes are observed where r is specified in advance.

Consider the following statistical experiment you flip a coin repeatedly and count the number of times the coin lands on heads. You continue flipping the coin until it has landed 5 times on Heads. This is negative binomial experiment because

* The experiment consists of repeated trials. We flip a coin repeatedly until it has landed 5 times on heads.
* Each trial can result in just two possible outcomes heads or tails.
* The probability of success is constant - 0.5 on every trial.
* The trials are independent that is getting heads on one trial does not affect whether we get heads on other trials.
* The experiment continues until a fixed number of successes have occured in this case, 5 heads

Notation:- The following notation is helpful, when we talk about negative binomial probability.

* x: The number of trials required to produce r successes in a negative binomial experiment.

* r: The number of successes in the negative binomial experiment

* P:

Negative Binomial Distribution:- The negative binomial Experiment is almost the same as a Binomial Experiment with one difference. a binomial Experiment has a fixed number of trials.

If the following five Conditions are true the Experiment is binomial:

(1) fixed Number of n trials.

(2) Each trial is independent.

(3) Only two outcomes are possible (success and failure)

(4) Probability of success (p) for each trial is Constant.

(5) A random variable Y = the number of successes.

Example:- Take a standard deck of cards, shuffle them, and choose a card. Replace the card and repeat twenty times. Y is the number of aces you draw.

The Negative binomial is similar to the binomial with two differences (specifically to numbers 1 and 5 in the list above)

- The number of trials, n is not fixed.
- A random variable Y = the number of trials needed to make r successes.

Example:- Take a Standard deck of card, shuffle them, and Choose a card. Replace the card and repeat until you have drawn two aces. Y is the number of draws needed to draw two aces. As the number of trials is not fixed (ie, you stop when you draw the second ace) this makes it a negative binomial distribution.

What is a Negative Binomial distribution?

A negative binomial distribution (also called the Pascal Distribution) is a discrete probability distribution for random variables in a negative binomial experiment.

The random variable is the number of repeated trials, X that produce a certain number of successes r, In other words, its the number of failure before a success. This is the main difference from the binomial distribution: with a negative binomial distribution, its the number of failures that counts.

Why is it called Negative Binomial? When you hear the term negative you might think that a positive distribution is flipped over the x-axis, making it negative binomial actually stems from the fact that one facet of the binomial distribution is reversed: in a binomial experiment, you count the number of successes in a fixed number of trials; in the above Example, you' are counting how many aces you draw. In a negative binomial Experiment, you are counting the failures, or how many cards it takes you to pick two aces.

Negative Binomial Distribution: – If repeated independent trials can result in a success with probability p and a failure with probability $q = 1-p$, then the probability distribution of the random variable X, the number of the trials on which the k^{th} success occurs, is

$$b^*(x; k, p) = \binom{x-1}{k-1} p^k q^{x-k}, \quad x = k, k+1, k+2 \ldots$$

Q.1 In an NBA (National Basketball Association) Championship series, the team that wins four games out of seven is the winner. Suppose that teams A and B face each other in the championship games and that team A has probability 0.55 of winning a game over team B.

(a) What is the probability that team A will win the series in 6 games?

(b) What is the probability that team A will win the series?

(c) If teams A and B were facing each other in a regional playoff series, which is decided by winning three out of five games, what is the probability that team A would win the series?

Sol^n (a) $b^*(6; 4, 0.55) = \binom{5}{3} 0.55^4 (1-0.55)^{6-4}$

$= 0.1853$

(b) P(team A wins the championship series) is

$b^*(4; 4, 0.55) + b^*(5; 4, 0.55) + b^*(6; 4, 0.55)$
$+ b^*(7; 4, 0.55)$

$= 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083$

(c) P(team A wins the playoff) is

$$b^*(3;3,0.55) + b^*(4;3,0.55) + b^*(5;3,0.55)$$

$$= 0.1664 + 0.2246 + 0.2021 = 0.5931$$

The Negative binomial distribution derives its name from the fact that each term in the expansion of $p^k(1-q)^{-k}$ corresponds to the values of $b^*(x;k,p)$ for $x = k, k+1, k+2 \ldots$ if we consider the special case of the negative binomial distribution where $k=1$, we have a probability distribution for the number of trials required for a single success. An example would be the tossing of a coin until a head occurs. We might be interested in the probability that the first head occurs on the fourth toss. The negative binomial distribution reduces to the form

$$b^*(x;1,p) = pq^{x-1}, \quad x = 1,2,3 \ldots$$

Since the successive terms constitute, a geometric progression, it is customary to refer to this special case as the geometric distribution and denote its values by $g(x;p)$