Bernoulli distribution :-

A random variable with two possible values 0 and 1, is called a Bernoulli variable, its distribution is Bernoulli distribution, and any eperiment with a binary outcome is called a Bernoulli trial.

examples) heads and tails, good or defective components, pass or fail tests, transmitted or lost signals .etc

Let p be the probability of success (ie, $x=1$) and $q=1-p$ be the probability of failure (ie, $x=0$),

then $E(x) = \Sigma x\, p(x)$

$= 1 \times p + 0 \times q$

$E(x) = p$

$Var(x) = \Sigma (x-p)^2 p(x)$

$= (1-p)^2 \times p + (0-p)^2 \times q$

$= p(1-p)^2 + p^2(1-p)$

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

$= p(1-p)$

$Var(x) = pq$

Geometric distribution

Consider a sequence of independent Bernoulli trials. Each trial results in a success or a failure.

The number of Bernoulli trials needed to get the first success has Geometric distribution.

example: A search engine goes through a list of sites looking for a given key phrase. Suppose the search terminates as soon as the key phrase is found. The number of sites visited is Geometric.

A hiring manager interviews candidates, one by one, to fill a vacancy. The number of candidates interviewed until one candidate receives an offer has Geometric distribution.

Geometric random variables can take any integer value from 1 to infinity, because one needs at least 1 trial to have the first success, and the number of trials needed is not limited by any specific number. The only parameter is p, the probability of a success.

Geometric probability mass function has the form,

$$P(x) = P\{\text{the } 1^{st} \text{ success occurs on the } x^{th} \text{ trial}\}$$

$$= (1-p)^{x-1} p, \quad x = 1, 2, \cdots,$$

which is the probability of $(x-1)$ failures followed by 1 success.

The mean of the Geometric distribution is given by:

$$\mu = E(X) = \frac{1}{p}$$

The variance of the Geometric distribution is given by:

$$\sigma^2 = Var(X) = \frac{1-p}{p^2}$$

The CDF of the Geometric distribution is given by:

$$P(X \le x) = 1 - (1-p)^x$$

We have the geometric series, $s(q) = \sum_0^\infty q^x = (1-q)^{-1}$

Differentiating w.r.t $q \Rightarrow s'(q) = \frac{1}{(1-q)^2}$

or $\frac{1}{(1-q)^2} = s'(q)$

$= \left(\sum_0^\infty q^x\right)'$

$= \sum_0^\infty x q^{x-1}$

$\frac{1}{(1-q)^2} = \sum_1^\infty x q^{x-1}$

$s''(q) = \frac{2}{(1-q)^3}$

or $\frac{2}{(1-q)^3} = \left(\sum_0^\infty q^x\right)''$

$\frac{2}{(1-q)^3} = \sum_0^\infty x(x-1) q^{x-2}$

$\frac{2}{(1-q)^3} = \frac{1}{q}\sum_1^\infty x^2 q^{x-1} - \frac{1}{q}\sum_1^\infty x q^{x-1}$

$\Rightarrow \sum_1^\infty x^2 q^{x-1} = \frac{2q}{(1-q)^3} + \frac{1}{(1-q)^2}$

$= \frac{1+q}{(1-q)^3}$

$\sum_1^\infty x^2 q^{x-1} = \frac{1+q}{p^3}$

$\therefore E(X) = \sum_{x=1}^\infty x(1-p)^{x-1} p$

$= p\sum_{x=1}^\infty x q^{x-1}$

$= p \times \frac{1}{(1-q)^2}$

$= \frac{p}{p^2}$

$E[X] = \frac{1}{p}$

$\therefore Var[X] = E[X^2] - (E[X])^2$

$= \sum_{x=1}^\infty x^2 (1-p)^{x-1} p - \frac{1}{p^2}$

$= p\,\frac{(1+q)}{p^3} - \frac{1}{p^2}$

$= \frac{q}{p^2}$

$Var[X] = \frac{1-p}{p^2}$

$P(X \geq x) = \sum_x^\infty (1-p)^{x-1} p$

$= (1-p)^{x-1}p + (1-p)^x p + (1-p)^{x+1} p + \cdots$

$P(X \geq x) = \frac{(1-p)^{x-1} p}{1-(1-p)} = (1-p)^{x-1}$

$P(X > x) = \sum_{x+1}^\infty (1-p)^{x-1} p$

$= (1-p)^x p + (1-p)^{x+1} p + \cdots$

$P(X > x) = \frac{(1-p)^x p}{1-(1-p)} = (1-p)^x$

$P(X \leq x) = 1 - (1-p)^x$

* An internet search engine looks for a certain keyword in a sequence of independent web sites. It is believed that 20% of the sites contain this keyword.

(i) Compute the probability that atleast 5 of the first 10 sites contain the given keyword.

(ii) Compute the probability that the search engine had to visit at least 5 sites in order to find the first occurrance of the keyword.

Sol^n (i) $p = 0.2$, $n = 10$, $q = 0.8$

$$P(X \geq 5) = 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)]$$
$$= 1 - [(0.8)^{10} + 5 \times 0.2 \times (0.8)^9 + 10 \times (0.2)^2 \times (0.8)^8 + 10 \times (0.2)^3 (0.8)^7 + 5 \times (0.2)^4 (0.8)^6]$$
$$= 1 - 0.9672$$
$$= 0.0328$$

(ii) $p = 0.2$
$$P(X \geq 5) = (1-p)^{x-1}$$
$$= (1-0.2)^{5-1}$$
$$= 0.4096$$

* About ten percent of users donot close windows properly. Suppose that Windows is intstalled in a public library that is used by ~~certai~~ random people in a random order.

(i) On the average, how many users of this computer do not close Windows properly before someone does close it properly?

(ii) What is the probability that exactly 8 of the next 10 users will close Windows properly?

Sol^n (i) $\mu = 0.1$, $n = 1 \Rightarrow p = 0.1$ $\therefore E[X] = \frac{1}{p} = \frac{1}{0.1} = 10$

(ii) $p = 0.1$, $q = 0.9$ $\therefore P(X=8) = {}^{10}C_8 (0.9)^8 (0.1)^2 = 0.1937.$