GEOMETRIC DISTRIBUTION :

Consider the geometric series,

$$\sum_{x=0}^{\infty} q^x = \frac{1}{1-q}, \quad \text{if } |q|<1$$

$$\Rightarrow \sum_{x=0}^{\infty} pq^x = 1, \quad \text{where } p+q=1.$$

For $0<q<1$, $p+q=1$,

the terms of the geometric series forms a pm,. Hence, the distribution given by the pmf is known as Geometric Distribution.

Definition: A non-negative integer valued random variable X is said to follow Geometric Distribution if it's pmf is given by,

$$f(x) = \begin{cases} pq^x, & x = 0,1,2,\ldots \\ 0, & \text{otherwise.} \end{cases} \qquad \begin{matrix} 0<p<1 \\ p+q=1 \end{matrix}$$

Remark:
Consider a sequence of independent Bernoulli trail with probability of success 'p'. Let X be the number of failures before the 1st success,

$$P(X=r) = P(\underbrace{F\,F \ldots F}_{(x-1)} \underset{x\text{th}}{S}) = q^x p \quad [\text{Due to independence}]$$

A random variable X that has a Geometric Distribution is often referred to a discrete waiting time random variable as it represents how long (in terms of failure) one has to wait to get a success.

Mean and Variance :

$$\sum_{x=0}^{\infty} q^x = \frac{1}{1-q}, \quad 0<q<1$$

Differentiating with respect to q, $\sum_{x=1}^{\infty} xq^{x-1} = (-1)(1-q)^{-2}(-1) = p^{-2}$

Again, differentiating wrt q, $\sum_{x=2}^{\infty} x(x-1)q^{x-2} = (-2)(1-q)^{-3}(-1) = 2p^{-3}$

$$E(X) = \sum_{x=0}^{\infty} x q^x p.$$

$$= pq \sum_{x=1}^{\infty} xq^{x-1} = pq\,p^{-2} = q/p.$$

$$V(X) = E(X^2) - E(X)$$
$$= E(X(X-1)) + E(X) - E^2(X)$$
$$= \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2}$$
$$= \frac{q^2}{p^2} + \frac{q}{p} = \frac{q}{p}\left(\frac{q}{p}+1\right)$$
$$= \frac{q}{p}\,\frac{q+p}{p} = \frac{q}{p^2}$$

$$E(X(X-1)) = \sum x(x-1)q^x p$$
$$= pq^2 \sum x(x-1)q^{x-2}$$
$$= pq^2\, 2p^{-3}$$
$$= 2p^{-2}q^2$$

Remark :

Let, Y denote the no of trail required to get the 1st success, In a sequence of independent Bernoulli trails,

$$Y = X+1.$$
$$P(Y=y) = P(X+1=y) = P(X=y-1) = pq^{y-1}. \quad y = 1, 2, \ldots \infty.$$
$$E(Y) = E(X)+1 = \frac{q}{p}+1 = \frac{1}{p}$$
$$V(Y) = V(X) = \frac{q}{p^2}.$$

In a random sampling WR from a population with N distinct members. Let, X denotes no of drawings needed to get n distinct elements in the sample. Find E(X) and var(X).

Let Z_i $\{i = 1(1)n\}$ denotes the no of drawing needed to get new distinct elements when $(i-1)$ distinct elements have already obtained.

It is to be noted that $Z_1 = 1$

$$X = \sum_{i=1}^{n} Z_i = 1 + Z_1 + Z_2 + \ldots + Z_i + \ldots + Z_n$$

Now, Z_i's are independently distributed with pmf.

$P(Z_i = z) = P$[exactly z drawings are necessary to get new distinct element when $(i-1)$ distinct has already been obtained].

$$= \underbrace{\left(\frac{i-1}{N}\right)\left(\frac{i-1}{N}\right)\cdots\left(\frac{i-1}{N}\right)}_{(z-1)} \frac{N-(i-1)}{N}$$

$$= \left(\frac{i-1}{N}\right)^{z-1}\left(1 - \frac{i-1}{N}\right) = q_i^{z-1} p_i$$

$$E(z_i) = \frac{q_i}{p_i} = \frac{N}{(N-i+1)}$$

$$V(z_i) = \frac{q_i}{p_i^2} = \frac{N(N-i)}{(N-i+1)^2}$$

$$E(X) = \sum_{i=1}^{n} E(z_i) = \sum_{i=1}^{n} \frac{N}{(N-i+1)}$$

$$V(X) = \sum_{i=1}^{n} V(z_i) = \sum_{i=1}^{n} \frac{N(N-i)}{(N-i+1)^2}$$

CDF : $F_X(x) = P(X \leq x) = \sum_{t=0}^{[x]} pq^t$

$$= p\left(1+q+q^2+\dots+q^{[x]}\right)$$

$$= p\,\frac{1-q^{[x]+1}}{1-q}$$

$$= 1-q^{[x]+1}$$

Tail function $-\ P(X>x)$

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

$$= q^{[x]+1}$$

NOTE : ① If $X_i \overset{iid}{\sim} Geo(p)$, $\sum_{i=1}^{r} X_i \sim NB(r,p)$

② If $X_i \overset{iid}{\sim} Geo(p)$, $i=1,2$, Find the conditional distribution of $(X_1 | X_1+X_2)$

$$\frac{P(X_1=x,\ X_2=y-x)}{P(X_1+X_2=y)} = \frac{P(X_1=x)\ P(X_2=y-x)}{P(X_1+X_2=y)} \sim NB(2,p)$$

$$= \frac{\cancel{p}\cancel{q^x}\ \cancel{p}q^{\cancel{y}-x}}{\binom{y+2-1}{y}\cancel{p^2}\cancel{q^y}}$$

$$= \frac{1}{\binom{y+1}{y}} = \frac{1}{y+1} \sim DU(0,1,2,\dots y)$$

③ $X_1, X_2, \dots X_n \overset{iid}{\sim} Geo(p)$, find the distribution of $Y = \min(X_i)$ $i=1(1)n$.

$$F_Y(y) = P(Y\leq y) = 1-P(Y>y)$$

$$= 1 - P(\min\{X_i\} > y)$$

$$= 1 - P(\min\{X_i\} > y)$$
$$= 1 - P(X_1 > y, X_2 > y, \dots X_n > y)$$
$$= 1 - (P(X_i > y))^n \quad [\because \text{ iid}]$$
$$= 1 - (q^{[y]+1})^n$$

$$P(Y = y) = P(Y \leq y) - P(Y \leq y-1)$$
$$= 1 - (q^{[y]+1})^n - 1 + (q^{[y]})^n$$
$$= (q^y)^n - (q^{y+1})^n$$
$$= (q^n)^y (1 - q^n)$$
$$= Q^y P. \qquad Q = q^n, \quad P = 1 - q^n.$$

$Y = \min\{X_i\} \sim Geo(P = 1 - q^n)$.

Lack of Memory Property :-

If X has a Geometric Distribution, then $P(X > m+n \mid X > m) = P(X \geq n)$. $\forall$ natural no.s m, n.

$$P(X > m+n \mid X > m) = \frac{P(X > m+n, X > m)}{P(X > m)}$$
$$= \frac{P(X > m+n)}{P(X > m)}$$
$$= \frac{q^{m+n+1}}{q^{m+1}} = q^n = q^{n-1+1} = P(X > n-1)$$
$$= P(X \geq n)$$

Interpretation:
This theorem states that the probability that more than (m+n+1) trails required for the 1st success given that there have been more than ‘m’ failures is equal to the unconditional probability of at least n trails are needed before the first success ie, P(X>m+nlX>m) is independent of m ie, the information of no of success in (m+1) trails has been forgotten in the subsequent calculation.

Theorem : Let X be a non-negative integer valued random variable satisfying $P(X > m+n \mid X > m) = P(X \geq n)$ $\forall m, n$. Then, X must have a Geometric Distribution.

Let, $p_k = P(X = k)$, $k = 0, 1, 2, \dots$ be the pmf of X,

Define, $q_m = P(X > m) = \sum_{k=m+1}^{\infty} p_k$

$$P(X > m+n \mid X > m) = P(X \geq n)$$

$$\Rightarrow \quad P(X > m+n) = P(X > n-1)\,P(X > m)$$

$$\Rightarrow \quad q_{m+n} = q_{n-1}\,q_m.$$

Set $n=1$, $\quad q_{m+1} = q_m q_0$

$$\Rightarrow \frac{q_{m+1}}{q_m} = q_0 \quad \text{—} \ (i)$$

Now, $$q_m = \underbrace{\frac{q_m}{q_{m-1}}\;\frac{q_{m-1}}{q_{m-2}}\;\cdots\;\frac{q_1}{q_0}}_{m}\;q_0$$

$$= \underbrace{q_0\;q_0\;\cdots\;q_0}_{m}\;q_0 \qquad (\text{by } (i))$$

$$= q_0^{m+1}$$

Again, $$q_{m-1} = p_m + p_{m+1} + p_{m+2} + \cdots$$
$$-\;q_m = \quad -\,p_{m+1} - p_{m+2} - \cdots$$

$$q_{m-1} - q_m = p_m$$

$$\therefore \; p_m = q_0^m - q_0^{m+1} = q_0^m (1-q_0)$$

Now, $q_0 = P(X>0) = 1 - P(X \leq 0) = 1 - p_0$

So, $$p_m = \begin{cases} p_0 (1-p_0)^m & m = 0,1,2,\ldots \\ 0 & , \text{ow}, \end{cases}$$

$\therefore\; X \sim \text{Geo}(p_0).$

Remark:
Among all non-negative integer valued random variables X has the lack of memory property if X has a Geometric Distribution. Hence, the lack of memory property is a characteristic of geometric random variable in the class of all no -negative integer valued random variables.

Show that $E\left(\frac{1}{1+X}\right) = -\frac{p}{q}\log p$

$$E\left(\frac{1}{1+X}\right) = \sum_{x=0}^{\infty} \frac{1}{(1+x)}\,p q^x$$

$$= \frac{p}{q} \sum_{x=0}^{\infty} \frac{1}{1+x}\,q^{x+1}$$

$$= \frac{p}{q} \left\{ \frac{q^1}{1} + \frac{q^2}{2} + \cdots \cdots \right\}$$

$$= -\frac{p}{q} \left[-q - \frac{q^2}{2} - \frac{q^3}{3} - \cdots \right]$$

$$= -\frac{p}{q} \log (1-q)$$

$$= -\frac{p}{q} \log p.$$