# Hypergeometric Distribution :

A discrete random variable X is said to have a Hypergeometric Distribution if it's pmf is given by,

$$f(x) = \begin{cases} \dfrac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} , & x = 0, \dots n \;.\; \left. \begin{matrix} 0<p<1 \\ p+q=1 \end{matrix} \right\} \\ 0 & \text{otherwise} \end{cases}$$

Note that, $0 \le x \le Np, \quad 0 \le n-x \le Nq$

$$\Rightarrow n - Nq \le x \le n$$

$$\Rightarrow \max\{0, n-Nq\} \le x \le \min\{Np, n\}$$

We know, $\sum_x \binom{M}{x}\binom{N-M}{n-x} = \binom{N}{n}$

$$\Rightarrow \frac{\sum_x \binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = 1 \quad \text{(this term will form a pmf)}$$

$$\therefore \sum_x \frac{\binom{Np}{x}\binom{N-Np}{n-x}}{\binom{N}{n}} = \sum_x \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} = 1$$

## Moments :-

(factorial moments)

$$E(X_r) = \sum_{x=0}^{n} (x)_r \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}$$

$x! = (x-r)!\,(x)_r$

$(Np)! = (Np-r)!\,(Np)_r$

$$= \sum_{x=0}^{n} (x)_r \frac{\frac{(Np)!}{x!\,(Np-x)!}\binom{Nq}{n-x}}{\binom{N}{n}}$$

$$= \frac{(Np)_r}{(x)_r} \sum_{x=r}^{n} \frac{\frac{(Np-r)!}{(x-r)!\,(Np-x)!}\binom{Nq}{n-x}}{\binom{N}{n}}$$

$$= \frac{(Np)_r}{\binom{N}{n}} \sum_{x=r}^{n} \binom{Np-r}{x-r}\binom{Nq}{n-x}$$

$$= \frac{(Np)_r}{\binom{N}{n}} \sum_{y=x-r=0}^{n-r} \binom{Np-r}{y}\binom{Nq}{n-r-y}$$

$$= \frac{(Np)_r}{\binom{N}{n}} \binom{N-r}{n-r} \underbrace{\sum_{y=0}^{n-r} \frac{\binom{Np-r}{y}\binom{Nq}{n-r-y}}{\binom{N-r}{n-r}}}$$

pmf of Hypergeometric $\therefore = 1.$

$$= \frac{(Np)_r}{\binom{N}{n}} \binom{N-r}{n-r}$$

$$E(X_r) = \frac{(Np)_r}{\binom{N}{n}} \binom{N-r}{n-r}$$

$$r=1, \quad E(X) = \frac{Np}{\binom{N}{n}} \binom{N-1}{n-1} = \frac{\cancel{N}p}{\cancel{N!}_{\cancel{N}}} \frac{\cancel{n!}^{\,n} (N-\cancel{n})!}{(N-\cancel{n})!} \frac{(\cancel{N-1})!}{(\cancel{n-1})!} = np$$

$$r=2, \quad E(X_2) = E(x(x-1)) = \frac{\cancel{N}p(Np-1)}{\cancel{N!}_{\cancel{N}(N-1)}} \frac{\cancel{n!}^{\,n(n-1)} (N-\cancel{n})!}{(\cancel{n-2})!} \frac{(\cancel{N-2})!}{(N-\cancel{n})!}$$

$$= \frac{p(Np-1)\,n(n-1)}{(N-1)}$$

$$\text{var}(x) = E(x^2) - E^2(x)$$

$$= E(x(x-1)) + E(x) - E^2(x)$$

$$= \frac{p(Np-1)n(n-1)}{(N-1)} + np - n^2p^2$$

$$= \frac{p(Np-1)n(n-1) + np(N-1) - n^2p^2(N-1)}{(N-1)}$$

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

$$= \frac{np}{(N-1)} \left[ Np\cancel{n} - n - Np + \cancel{1} + N - \cancel{1} - np\cancel{N} + np \right]$$

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

$$= \frac{np}{(N-1)} (N-n)(1-p)$$

$$= npq \frac{(N-n)}{(N-1)}$$

Suppose there are N individuals in a population of which Np individuals possess a character A and No do not possess the character A. Let X be the number of individuals with characteristics A in a random sample of size n. Find P(X=x) under the following scheme WR and WOR.

WR (with replacement) :-

X denote the no of individual in the sample who possess the character A.

$P(X=x) = P($ x no of individuals with character A and $(n-x)$ with character $A^c$ in a random sample of size n $)$

$$= \binom{n}{x}\left(\frac{Np}{N}\right)^x\left(\frac{Nq}{N}\right)^{n-x}$$

$Np \to A$, $Nq \to A^c$ WR

$$= \frac{\binom{n}{x}(Np)^x(Nq)^{n-x}}{N^n} = \binom{n}{x}p^x q^{n-x}.$$

$X \underset{WR}{\sim} Bin(n, p)$

WOR (without replacement):

$$P(X=x) = \frac{\binom{n}{x}(Np)_x (Nq)_{n-x}}{\binom{N}{n}}$$ (considering ordered sample)

$$= \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}$$

$X \underset{WOR}{\sim} Hyp(n, N, p)$

Show that, $\dfrac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} \xrightarrow{N\to\infty} \binom{n}{x}p^x q^{n-x}$

$N \to \infty$. $\frac{n}{N} \to 0$.

$$\frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} = \binom{n}{x}\frac{(Np)_x\,(Nq)_{n-x}}{(N)_n}$$

$$= \binom{n}{x}\frac{Np(Np-1)\cdots(Np-x+1)\,Nq(Nq-1)\cdots(Nq-n+x)}{N(N-1)(N-2)\cdots(N-n+1)}$$

$$= \binom{n}{x}\frac{p\left(p-\frac{1}{N}\right)\left(p-\frac{2}{N}\right)\cdots\left(p-\frac{x-1}{N}\right)q\left(q-\frac{1}{N}\right)\cdots\left(q-\frac{n-x-1}{N}\right)}{1\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)}$$

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

as $N \to \infty$, but $\frac{n}{N} \to 0$.

Interpretation:
As N tends to infinity, finite population become infinite population. From a p infinite population drawing sample under WOR is practically equivalent to draw sample under WR. As a sample under WR can also be obtained under WOR.
Hence, as N tends to infinity, the distribution of WOR becomes the distribution of WR,
=> Hypergeometric distribution becomes Binomial distribution.

Remark :

$$E(X_{WR}) = E(X_{WOR}) = np.$$

$$Var(X_{WR}) = npq > npq\underbrace{\left(\frac{N-n}{N-1}\right)}_{<1} = Var(X_{WOR})$$

**Examples:**
The factor in the variance $\left(\frac{N-n}{N-1}\right)$ as a multiplier using due to the use of WOR schemes in a finite population instead of an infinite population. It is known as FPC. (Finite population correction)

1. An unknown number of animals inhabit in a certain region. First, catch a number of 'm' of these animals, mark them in some manner and release them. After allowing the marked animals time to dispersed through at the region. A new catch of size 'n' is made. Let X denotes the number of marked animals in the second capture. Find the probability distribution of X. Find the probability distribution of X. Find E(X) and var (X).

Let there are total N number of unknown animals.
Out of these m are marked, so, (N-m) are unmarked.

$$P(X=x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}, \quad x = 0(1)n.$$

$\therefore X \sim \text{Hypergeometric}(m, N, p)$

$$E(X) = \frac{nm}{N}, \quad \frac{m}{N} = p$$

$$V(X) = npq\left(\frac{N-n}{N-1}\right) = n\frac{m}{N}\left(1-\frac{m}{N}\right)\left(\frac{N-n}{N-1}\right)$$

2) If $X \sim Hyp(n, N, p)$. show that $E(n-X)_r = (n)_r \frac{(Nq)_r}{(N)_r}$

$$X \sim Hyp(n, N, p)$$

$$Y = n - X \sim Hyp(n, N, 1-p)$$

$$P(Y=y) = P(n-X=y) = P(X=n-y)$$

$$= \frac{\binom{Np}{n-y}\binom{Nq}{y}}{\binom{N}{n}}$$

$$= \frac{\binom{Nq}{y}\binom{Np}{n-y}}{\binom{N}{n}} \sim Hyp(n, N, q)$$

We got, $E(X_r) = \frac{(Np)_r}{\binom{N}{n}} \binom{N-r}{n-r}$

$$E(Y_r) = E[(n-X)_r] = \frac{(Nq)_r}{\binom{N}{n}} \binom{N-r}{n-r}$$

$$= (Nq)_r \frac{(N-r)!}{(n-r)!\,(N-\cancel{n})!} \frac{n!\,(N-\cancel{n})!}{N!}$$

$$= (Nq)_r \frac{(n)_r}{(N)_r}$$

3) $X \sim Hyp(n, N, m)$, show that $E(X^k) = \frac{nm}{N} E(Y+1)^{k-1}$, where, $Y \sim Hyp(n-1, N-1, m-1)$. Find $E(X)$, $var(X)$

$$X \sim Hyp(n, N, m),$$

$$E(X^k) = \sum_{x=0}^{n} x^k \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

$$= \sum_{x=0}^{n} x^{k-1} \frac{\binom{m-1}{x-1} m \binom{N-m}{n-x}}{\binom{N}{n}}$$

$$= \sum_{x=1}^{n} x^{k-1} \frac{m \binom{m-1}{x-1}\binom{(N-1)-(m-1)}{(n-1)-(x-1)}}{\frac{N}{n}\binom{N-1}{n-1}}$$

$$= \sum_{x-1=y=0}^{n} (y+1)^{k-1} \frac{\binom{m-1}{y}\binom{N-1-y}{n-1-y}}{\binom{N-1}{n-1}} \frac{nm}{N}$$

$$= \frac{nm}{N} \quad E(Y+1)$$

$Y \sim Hyp(n-1, N-1, m-1)$

$$K=1, \quad E(X) = \frac{nm}{N} \quad E(Y+1)^0 = \frac{nm}{N}$$

$$K=2, \quad E(X^2) = \frac{nm}{N} \quad E(Y+1) = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{(N-1)} + 1\right]$$

$$= \frac{nm}{N}\left(\frac{nm-m-n}{N-1}\right)$$

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

$$= \frac{nm}{N}\left(\frac{nm-m-n}{N-1}\right) - \frac{n^2m^2}{N^2}$$

$$= \frac{nm}{N} \frac{(N-m)(N-n)}{N(N-1)}$$

4) Let $X_1, \ldots X_n$ be $n$ independent Bernoulli($p$) RV. Let $S_j = \sum_{i=1}^{j} X_i$ where $j = 1(1)n$, Show that, $P(S_m = r \mid S_n = s)$ does not depend on $p$, where $m<n$. Identify the distribution and obtain first two moments of it.

$X_1, X_2 \ldots X_n \overset{iid}{\sim} Bern(p)$

$$S_j = \sum_{i=1}^{j} X_i \sim Bin(j, p), \quad j = 1(1)n.$$

$$S_m = \sum_{i=1}^{m} X_i \sim Bin(m, p).$$

$$S_n = \sum_{i=1}^{n} X_i \sim Bin(n, p)$$

$$\begin{cases} S_n = X_1 + \cdots + X_m + X_{m+1} + \cdots + X_n \\ S_m = X_1 + \cdots + X_m \end{cases}$$
$n > m$

$$S_n - S_m = \sum_{i=m+1}^{n} X_i \sim Bin(n-m, p)$$

$$P(S_m = r \mid S_n = s) = \frac{P(S_m = r, S_n = s)}{P(S_n = s)}$$

$$= \frac{P(S_m = r, S_n - S_m = s - r)}{P(S_n = s)}$$

$$= \frac{P(S_m = r)\, P(S_n - S_m = s - r)}{P(S_n = s)}$$

$$= \frac{\binom{m}{r} p^r q^{m-r} \binom{n-m}{s-r} p^{s-r} q^{n-m-s+r}}{}$$

$$\binom{n}{s} p^s q^{n-s}$$

$$= \frac{\binom{m}{r}\binom{n-m}{s-r}}{\binom{n}{s}}, \quad r = 0(1)s.$$

$$E(X) = \frac{sm}{n}$$

$$var(X) = s\,\frac{m}{n}\left(1-\frac{m}{n}\right)\left(\frac{n-s}{n-1}\right)$$

## 5) Mode of Hypergeometric Distribution :-

pmf, $f(x) = \dfrac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, \ldots n$

Note that, $$\frac{f(x)}{f(x-1)} = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\cancel{\binom{N}{n}}} \cdot \frac{\cancel{\binom{N}{n}}}{\binom{Np}{x-1}\binom{Nq}{n-x+1}}$$

$$= \frac{\cancel{(Np)!}\;\cancel{(Nq)!}\;\cancel{(Nq-n+x-1)!}\;\overset{(n-x+1)}{\cancel{(n-x+1)!}}\;\cancel{(x-1)!}\;\overset{(Np-x+1)}{\cancel{(Np-x+1)!}}}{\underset{x}{\cancel{x!}}\;\cancel{(Np-x)!}\;\cancel{(n-x)!}\;\underset{(Nq-n+x)}{\cancel{(Nq-n+x)!}}\;\cancel{(Np)!}\;\cancel{(Nq)!}}$$

$$= \frac{(Np-x+1)(n-x+1)}{(Nq-n+x)\,x} \gtrless 1 \text{ if}$$

$$x \lessgtr \frac{(n+1)(Np+1)}{(N+2)}$$

Case I –

$\dfrac{(n+1)(Np+1)}{(N+2)}$ is not an integer,

let $\left[\dfrac{(n+1)(Np+1)}{(N+2)}\right] = K$

$$\frac{f(1)}{f(0)} > 1, \quad \frac{f(2)}{f(1)} > 1, \quad \ldots, \quad \frac{f(K)}{f(K-1)} > 1, \quad \ldots \quad \frac{f(K+1)}{f(K)} < 1, \ldots$$

$\Rightarrow f(0) < f(1) < f(2) < \cdots < f(K-1) < f(K) > f(K+1) > \cdots$

Clearly, $f(K)$ is the maximum,

$\therefore K = \left[\dfrac{(n+1)(Np+1)}{(N+2)}\right]$ is the mode.

Case II :-

$K = \dfrac{(n+1)(Np+1)}{(N+2)}$ is an integer.

$$\frac{f(1)}{f(0)} > 1, \quad \frac{f(2)}{f(1)} > 1, \quad \ldots, \quad \frac{f(K)}{f(K-1)} = 1, \quad \frac{f(K+1)}{f(K)} < 1, \ldots$$

$f(0) \quad f(1) \quad f(k-1) \quad f(k)$

$f(0) < f(1) < f(2) \ldots < f(k-1) = f(k) > f(k+1) > f(k+2) > \ldots$

Clearly, $f(k-1)$ and $f(k)$ is the maximum,

Hence, $k, \& (k-1) = \dfrac{(n+1)(Np+1)}{(N+2)}$ are the mode.