Hyper Geometric Distribution :-

When the population is finite and the sampling is done without replacement, so that the events are stochastically independent, then we obtain a Hypergeometric distribution.

Consider an urn with N balls, M of which are white and $N-M$ are red balls. Suppose n balls are drawn at random from the box. (without replacement).

X = no of white balls drawn in the sample.

Then the pmf of X is

$$P[X=x] = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, & x \in [\max(0, n+M-N), \min(M, n)] \\ 0, & \text{Otherwise.} \end{cases}$$

Remark

(1) $0 \leq x \leq M$

(2) $0 \leq n-x \leq N-M$

$\max(0, n-N+M) \leq x \leq \min(M, n)$

Mean and variance of the Hypergeometric Distribution:

$$E(x) = \sum_{k=0}^{n} k \cdot p(x)$$

$$= \sum_{k=0}^{n} k \cdot \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$$= \sum_{k=0}^{n} \cancel{k} \cdot \frac{\frac{M}{\cancel{k}}\binom{M-1}{k-1}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$$= \frac{M}{\binom{N}{n}} \sum_{k=0}^{n} \binom{M-1}{k-1}\binom{N-M}{n-k},$$

Let $x = k-1, \; m = n-1, \; M-1 = A$

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

$N-M$

$N-1-A$

$N-(A+1)$

$n = m+1$

$k = x+1$

$m-x$

$$= \frac{M}{\binom{N}{n}} \cdot \binom{N-1}{m}$$

$$= \frac{M}{\binom{N}{n}} \binom{N-1}{n-1} = \frac{nM}{N}$$

$$E\{X(X-1)\} = \sum_{k=0}^{n} k(k-1)\left\{\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}\right\}$$

$$= \frac{M(M-1)}{\binom{N}{n}} \sum_{k=2}^{n} \left\{\binom{M-2}{k-2}\binom{N-M}{n-k}\right\}$$

$$E\{x(x-1)\} = \sum_{k=0}^{n} k(k-1)\left\{ \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \right\}$$

$$= \frac{M(M-1)}{\binom{N}{n}} \sum_{k=2}^{n} \binom{M-2}{k-2}\binom{N-M}{n-k}$$

$$= \frac{M(M-1)}{\binom{N}{n}} \binom{N-2}{n-2}$$

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

$$\therefore E(x^2) = E[x(x-1)] + E(x)$$

$$= \frac{M(M-1)\cdot n(n-1)}{N(N-1)} + \frac{n\cdot M}{N}$$

$$= \frac{NM\cdot(M-1)n(n-1) + nMN(N-1)}{N^2(N-1)}$$

$$V(x) = \frac{M(M-1)\,n(n-1)}{N(N-1)} + \frac{nM}{N} - \frac{n^2M^2}{N^2}$$

$$\frac{nMN - n^2M^2}{N^2}$$

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

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

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

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

$$= \frac{nM}{N}\left[\frac{\cancel{nMN}-MN-nN+\cancel{N}+N^2-\cancel{nMN}-\cancel{N}+nM\}}{N(N-1)}\right]$$

$$= \frac{nM}{N}\left(\frac{nM+N^2-N-nN}{N(N-1)}\right)$$

$$= \frac{nM}{N^2(N-1)}\left\{\frac{n(M-N)+N(N-1)}{\cancel{N(N-1)}}\right\}$$

$$= \frac{nM}{N^2(N-1)}\left\{\cancel{nMN}-MN-nN+\cancel{N}+N^2-\cancel{nMN}-\cancel{N}+nM\right\}$$

$$= \frac{nM}{N^2(N-1)}\left\{-MN-nN+N^2+nM\right\}$$

Approximation to Binomial Distribution:

Hypergeometric distribution tends to binomial distribution as $N \to \infty$, $\frac{M}{N} \to p$, we get

$$P(k: n, M, N) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$$= \frac{(M)_k}{k!} \cdot \frac{(N-M)_{n-k}}{(n-k)!} \Big/ \frac{(N)_n}{n!} \qquad \left[\because \binom{n}{k} = \frac{(n)_k}{k!}\right]$$

$$= \frac{(M)_k}{k!} \cdot \frac{(N-M)_{n-k}}{(n-k)!} \cdot \frac{n!}{(N)_n}$$

$$= \frac{n!}{k!} \cdot \frac{M(M-1)\cdots(M-k+1)\cdot(N-M)(N-M-1)\cdots(N-M-n+k-1)}{N(N-1)\cdots(N-n+1)}$$

$$= \binom{n}{k} \frac{\left(\frac{M}{N}-\frac{1}{N}\right)\cdots\left(\frac{M}{N}-\frac{k-1}{N}\right)\left(1-\frac{M}{N}\right)\left(1-\frac{M}{N}-\frac{1}{N}\right)\left(1-\frac{M}{n}-\frac{n}{N}+\frac{k}{N}-\frac{1}{N}\right)}{1\cdot\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{n-k}{N}\right)}$$

Considering the limits as $N \to \infty$ and $\frac{M}{N} \to p$, we get

$$\lim_{\substack{N\to\infty \\ \frac{M}{N}\to p}} P(k; n, M, N) = \binom{n}{k}(\underbrace{p \cdot p \cdots p}_{k \text{ times}} \cdot \underbrace{(1-p)(1-p)\cdots(1-p)}_{n-k \text{ times}})\cdots$$

$$= \binom{n}{k} p^k (1-p)^{n-k}$$ which is the pmf of Binomial distribution.

* **Approximation to Poisson distribution:-**

Considering the limit as $N \to \infty$, $n \to \infty$ $\frac{nM}{N} = \lambda$ as $\frac{M}{N} \to 0$. then the limiting distribution is known as Poisson distribution.

Recurrence Relation of Hypergeometric distribution:

The Recurrence Relation of the pmf of hyper geometric distribution is

$$\frac{P(k+1, n, M, N)}{P(k, n, M, N)} = \frac{(n-k)(M-k)}{(k+1)(N-M-n+k+1)}$$

From the above relation we get $P(k+1; n, M, N)$ is greater or less than $P(k, n, M, N)$ according as

$$\frac{(n-k)\cdot(M-k)}{(k+1)(N-M-n+k+1)} \gtrless 1$$

Application of hyper geometric distribution:-

1) Industrial quality control; where the number of defectives per sample is a hyper-geometric random variable.

2) The estimation of the size of animal population from "capture - re-capture" data.

3) The estimation of target population in epidemiological studies.

Scanned with CamScanner