Recall :- **Hypergeometric distribution** (N, r, m)

$N > \max(r, m)$

$S = \max\{0, m-(N-r)\}, \dots, \min\{m, r\}$.

$\mathbb{P}: \mathcal{F} \to [0,1] \qquad \mathbb{P}(\{k\}) = \dfrac{\binom{r}{k}\binom{N-r}{m-k}}{\binom{N}{m}}$

Choosing k people, with characteristic r, from a population of N people. [Sampling was done without replacement]

2.3.2 Hypergeometric distribution as a Series of Dependent Trials

Example 2.3.1 (Contd.) :- 5000 people, 1000 under age of 18.

– Sampling without replacement of 4 people was done by choosing every group of 4 in an equally likely manner.

$A_j = \{$ the j^{th} selection is a person younger than 18 $\}$, $j = 1, 2, 3, 4$

– Want to determine Probability: Sample size of 4 will produce only two people below age of 18

$\mathbb{P}(A_1) = \dfrac{1000}{5000} = \dfrac{1}{5}$

$$\mathbb{P}(A_2 \mid A_1) = \frac{999}{4999} \quad ; \quad \mathbb{P}(A_4^c \mid A_1 \cap A_2 \cap A_3^c) = \frac{3999}{4997}$$

$$\mathbb{P}(A_3^c \mid A_1 \cap A_2) = \frac{4000}{4998}$$

View sampling WITHOUT replacement as DEPENDENT Bernoulli trials

The above calculation: Success — Choose one person below age of 18

failure — otherwise

$$\cdot \ \underbrace{\mathbb{P}(\text{Success}, \text{Success}, \text{Failure}, \text{Failure})}_{\text{one way of achieving two people under age of 18 in the sample of size 4}} = \mathbb{P}(A_1 \cap A_2 \cap A_3^c \cap A_4^c)$$

$$= \mathbb{P}(A_1)\, \mathbb{P}(A_2 \mid A_1)\, \mathbb{P}(A_3^c \mid A_1 \cap A_2)\, \mathbb{P}(A_4^c \mid A_1 \cap A_2 \cap A_3)$$

$$= \frac{1000}{5000} \cdot \frac{999}{4999} \cdot \frac{4000}{4998} \cdot \frac{3999}{4997}.$$

$$\cdot \ \mathbb{P}(\text{Success}, \text{Failure}, \text{Success}, \text{Failure}) = \mathbb{P}(A_1 \cap A_2^c \cap A_3 \cap A_4^c)$$

$$= \mathbb{P}(A_1)\, \mathbb{P}(A_2^c \mid A_1)\, \mathbb{P}(A_3 \mid A_1 \cap A_2^c)\, \mathbb{P}(A_4^c \mid A_1 \cap A_2^c \cap A_3)$$

$$= \frac{1000}{5000} \cdot \frac{4000}{4999} \cdot \frac{999}{4998} \cdot \frac{3999}{4997}$$

If we want to achieve exactly 2 successes in sample of size 4 WITHOUT replacement - The ordering of success does not matter for the value of the probability.

$\therefore$ $\mathbb{P}$(Exactly Two under age of 18 in a sample of size 4 WITHOUT replacement)

$=$ # of ordering of Success, Success, Failure, Failure $\times$ $\mathbb{P}$(of any order)

$\boxed{\text{understanding - Dependent Bernoulli trials}}$

$$= \binom{4}{2} \times \frac{1000}{5000} \cdot \frac{4000}{4999} \cdot \frac{999}{4998} \cdot \frac{3999}{4997}$$

$$= \binom{4}{2} \times \frac{1000 \times 999 \times 4000 \times 3999}{(5000)(4999)(4998)(4997)}$$

(Ex.) $= \dfrac{\binom{1000}{2}\binom{4000}{2}}{\binom{5000}{4}}$

N , $r \equiv$ specific characteristic , $m \equiv$ sample size without replacement

$S = \{\max\{0, m-(N-r)\}, \dots, \min\{m, r\}\}$

$$\boxed{\mathbb{P}(\{k\}) = \frac{\binom{r}{k}\binom{N-r}{m-k}}{\binom{N}{m}}}$$

$k \in S$

$$\mathbb{P}(\{k\}) = \frac{\dfrac{r!}{k!\,(r-k)!} \cdot \dfrac{(N-r)!}{(m-k)!\,(N-r-(m-k))!}}{\dfrac{N!}{m!\,(N-m)!}}$$

$$= \binom{m}{k} \frac{\prod_{i=0}^{k-1}(r-i) \prod_{i=0}^{m-1-k}(N-r-i)}{\prod_{i=0}^{m-1}(N-i)}$$

$$= \binom{m}{k} \prod_{i=0}^{k-1}\frac{(r-i)}{(N-i)} \prod_{i=0}^{m-1-k}\frac{(N-r-i)}{(N-k-i)}$$

$$\boxed{\mathbb{P}(\{k\}) = \binom{m}{k} \prod_{i=0}^{k-1}\frac{r-i}{N-i} \prod_{i=0}^{m-1-k}\frac{N-r-i}{N-k-i}}$$

– Dependent Bernoulli trial

$\tilde{Q}$:- $\dfrac{\binom{1000}{2}\binom{4000}{2}}{\binom{5000}{4}} \cong \cdots$ Very close $0.153\ldots$ $\cong \binom{4}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^2$

WITHOUT replacement

WITH replacement

Ex:-

Hypergeometric (N, ℓ, m)

$$P(\{k\}) = \binom{m}{k} \prod_{i=0}^{k-1} \frac{r-i}{N-i} \prod_{i=0}^{m-1-k} \frac{N-\ell-i}{N-k-i}$$

$p = \frac{r}{N}$, $p_1 = \frac{r-k}{N-k}$, $p_2 = \frac{r-k}{N-m}$

$$\binom{m}{k} p_1^k (1-p_2)^{m-k} \le P(\{k\}) \le \binom{m}{k} p^k (1-p_1)^{m-k}$$

Probability of $\{k\}$
Sampling without replacement $\Rightarrow$ $\mathbb{P}(\{k\})$

close to $\Longleftrightarrow$ $\frac{r}{N} \approx \frac{r-k}{N-k} \approx \frac{r-k}{N-m}$

$p_1 \approx p \approx p_2$

Probability of $\{k\}$
Sampling with replacement