Hypergeometric distribution

The sampling with replacement is binomial distribution. and sampling without replacement is hypergeometric distribution. Hypergeometric distribution is based on conditional probability theorems.

Ex-1 Three screws are drawn at random from a lot of 100 screws. Ten of which are defective ⓐ Find the probability that all the screws drawn will be defective in drawing without replacement. and P{at most two defectives}

Sol^n Let D = drawing a defective screw

N = drawing a nondefective screw.

The event is $E = \{DDD\}$, $|D| = 10$

$|N| = 90$

ⓐ Prob. that all 3 screws are defective

$= P(E) = \frac{10}{100} \times \frac{9}{99} \times \frac{8}{98} = 0.0007$

Alternative

If A_1 = 1st drawn is defective

A_2 = 2nd " " "

A_3 = 3rd " " "

then $P(E) = P(DDD) = P(A_1 A_2 A_3) = P(A_1)\, P(A_2|A_1)\, P(A_3|A_1 A_2)$

$= \frac{10}{100} \times \frac{9}{99} \times \frac{8}{98} = 0.0007$

(b) Probability of getting at most two defective without replacement

$= P\{DDN, DND, NDD\}$

$= P(DDN) + P(DND) + P(NDD)$

$= \frac{10}{100} \times \frac{9}{99} \times \frac{90}{98} + \frac{10}{100} \times \frac{90}{99} \times \frac{9}{98} + \frac{90}{100} \times \frac{10}{99} \times \frac{9}{98}$

$= 3 \times \frac{10 \times 9 \times 90}{100 \times 99 \times 98} = 0.025$

Assumption of Hypergeometric distribution.

1. The population or set to be sampled consists of N individuals, or objects or elements is finite population

2. Each individual can be characterised as a success (S) or a failure (F) and there are M successes in the population.

3. A sample of n individuals is selected without replacement in such a way that each subset of size n is equally likely to be chosen.

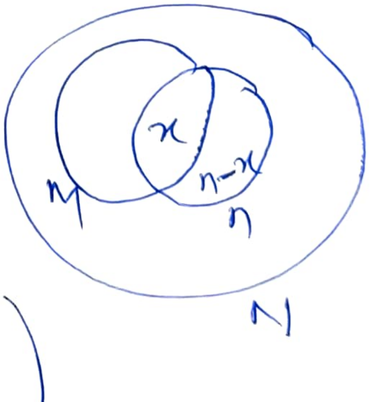
pmf of hypergeometric distribution

Let there are N number of things, out of which M things are defective, i.e. no. of nondefective things are $N-M$.

x number of defective things can be chosen from M things in $\binom{M}{x}$ ways.

If there are n trials to choose x defectives then $n-x$ nondefective can be chosen from $N-M$ nondefectives in $\binom{N-M}{n-x}$ ways.

Now n things can be chosen from N things in $\binom{N}{n}$ ways.



The probability of getting x defectives from trials is

$$p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2 \ldots n \text{ and } \max\{0, n+M-N\} \le x \le \min(n, M)$$

Ex-2 from ex-1, we have

$N = 150$, $M = 10$, $n = 3$, $x = 3$

To find X = counts no. of defectives

$$P(3) = \frac{\binom{10}{3}\binom{90}{0}}{\binom{150}{3}} = 0.0007$$

(b) $P(2) = \dfrac{\binom{10}{2}\binom{90}{1}}{\binom{100}{3}} = 0.025$

Def^n If X is the number of successes (S) in a completely random sample of size n drawn from a population consisting of M no. of successes (S) and $N-M$ no. of failures (F). Then the probability distribution of X is called hypergeometric distribution and is defined by

$$p(x) = P[X=x] = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \quad x = \ldots$$

$$\boxed{X \sim Hyp(n, M, N)}$$

$0 \le x \le M,\ n+M-N \le x \le n$

$= h(x; n, M, N) \sim b(x; n, \frac{M}{N})$ — M, N are large, $p = \frac{M}{N} < 0.5$

where $x \in X$ satisfies the cond^n: $\max\{0, n-N+M\} \le x \le \min\{n, M\}$

For ex-2 X = No of defected screws selected without replacements

(a) $P(3) = h(3; 3, 10, 100) = 0.0007$

(b) $P(2) = h(2; 3, 10, 100) = 0.025$

Mean and Variance

If the r.v. X is distributed by hypergeometric distribution, then

$\text{mean}(X) = \mu = E(X) = n \cdot \frac{M}{N} = np$

$\text{variance} = \sigma^2 = V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) = \left(\frac{N-n}{N-1}\right) npq, \quad p+q=1$

Q. Prove that hypergeometric distribution approaches to binomial distribution if no. of population (N) and no. of success (M) out of N are large enough such that $p = \frac{M}{N} < 0.5$, ...

Proof To prove

$$\lim_{N\to\infty} h(x;n,M,N) = b(x;n,p) \text{ if}$$

M, N are large and $P = \frac{M}{N} < 0.5$

We have

$$h(x;n,M,N) = \frac{\binom{M}{C_x}\binom{N-M}{C_{n-x}}}{\binom{N}{C_n}}$$

$$= \frac{M!}{x!(M-x)!} \times \frac{(N-M)!}{(n-x)!\,(N-M-n+x)!} \Big/ \frac{N!}{n!(N-n)!}$$

$$= \frac{M!}{x!(M-x)!} \times \frac{(N-M)!}{(n-x)!(N-M-n+x)!} \times \frac{n!(N-n)!}{N!}$$

$$= \frac{n!}{x!(n-x)!} \frac{M!\,(N-M)!\,(N-n)!}{(M-x)!\,(N-M-n+x)!\,N!}$$

$$= \frac{n!}{x!(n-x)!} \frac{[M(M-1) \overset{x\text{-terms}}{\cdots} (M-x+1)]\,[(N-M)(N-M-1) \overset{(n-x)\text{ terms}}{\cdots} (N-M-n+x+1)]}{N(N-1)(N-2) \cdots (N-n+1)}$$

$$= \frac{n!}{x!(n-x)!} \frac{\frac{M}{N}\left(\frac{M}{N}-\frac{1}{N}\right)\left(\frac{M}{N}-\frac{2}{N}\right)\cdots\left(\frac{M}{N}-\frac{x-1}{N}\right)\left(1-\frac{M}{N}\right)\left(1-\frac{M}{N}-\frac{1}{N}\right)\cdots\left(1-\frac{M}{N}-\frac{n-x-1}{N}\right)}{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)}$$

* Devide x N's then $(n-x)$ N's numerator

Now $\lim_{N \to \infty} h(x; n, M, N)$

$= \binom{n}{x} p^x (1-p)^{n-x}$ as $p = \frac{M}{N}$

$= b(x; n, p)$ where x is relatively small to M

Ex-
$$h(3; 10, 150, 500) = \frac{\binom{150}{3}\binom{500-150}{10-3}}{\binom{500}{10}}$$

$= 0.2695$

$$b\left(3; 10, \frac{150}{500}\right) = b(3, 10, 0.3)$$

$= \binom{10}{3}(0.3)^3(0.7)^7$

$= 0.2668$

Ex If $X \sim Hyp(n, M, N)$, then evaluate

(a) $F(4)$ if $n = 6$, $M = 10$, $N = 15$

(b) $F(4)$ if $n = 5$, $M = 10$, $N = 16$

Soln (a) $F(4) = \sum_{x=x_1}^{4} h(x; 6, 10, 15)$ where $x_1 = \max\{0, n+M-N\}$

$= \max\{0, 1\} = 1$

$= \sum_{x=1}^{4} h(x; 6, 10, 15) = 0.7063$

(b) $F(4) = \sum_{x=x_1}^{4} h(x; 5, 10, 16)$, $x_1 = \max\{0, n+M-N\}$

$= \max\{0, -1\} = 0$

$= \sum_{x=0}^{4} h(x; 5, 10, 16) = 0.9423$