The hypergeometric distribution is used for sampling without replacement.

p(x) = choose(m, x) choose(n, k-x) / choose(m+n, k)

for x = 0, …, k.

Note that p(x) is non-zero only for max(0, k-n) <= x <= min(k, m).

E[X] = μ = k p

and variance

Var(X) = k p (1 - p) * (m+n-k)/(m+n-1)

$N \rightarrow$ sample size of distribution

$m \rightarrow$ subclass-1 variable

$n \rightarrow$ subclass-2 variable

$k \rightarrow$ sample of element drawn

$p \rightarrow$ probability.

Proof:

$$p(x=i) = \frac{\binom{m}{i}\binom{n}{k-i}}{\binom{m+n}{k}}$$

The i^{th} selection has an equal likelihood of being in any trial, so the fraction of acceptable selection p is

$$p = \frac{m}{m+n} \quad \text{i.e.,} \; P(x_i = 1) = \frac{m}{n+m}$$

$$E(x) = \mu = \left\langle \sum_{i=1}^{k} x_i \right\rangle = \sum_{i=1}^{k} \langle x_i \rangle$$

$$= \sum_{i=1}^{k} \frac{m}{m+n} = \frac{km}{m+n} = k*p$$

$$var(x) = \sum_{i=1}^{k} var(x_i) + \sum_{i=1}^{k} \sum_{j=1}^{k} cov(x_i, x_j).$$

$\because$ x_i is a Bernoulli variable,

$$var(x_i) = p(1-p)$$

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

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

so, $\sum_{i=1}^{k} var(x_i) = \frac{knm}{(m+n)^2}$

for $i<j$, the covariance is

$$\mathrm{cov}(x_i, x_j) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle.$$

The probability that both i, j are succesful for $i \neq j$ is

$$P(x_i = 1, x_j = 1) = P(x_i = 1) * P(x_j = 1 \mid x_i = 1)$$

$$= \frac{m}{m+n} * \frac{m-1}{m+n-1}$$

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

$\because$ x_i, x_j are Bernoulli variables, their product is also a Bernoulli variable.

$$\langle x_i, x_j \rangle = P(x_i = 1, x_j = 1) = \frac{m(m-1)}{(m+n)(m+n-1)}$$

$$\langle x_i \rangle \langle x_j \rangle = \frac{m}{m+n} * \frac{m}{m+n} = \frac{m^2}{(m+n)^2}$$

$$\therefore \operatorname{cov}(x_i, x_j) = \frac{m(m-1)}{(m+n)(m+n-1)} - \frac{m^2}{(m+n)^2}$$

$$= \frac{(m+n)(m^2-m) - m^2(m+n-1)}{(m+n)^2(m+n-1)}$$

$$= \frac{\cancel{m^2(m+n)} \cancel{- m^2} - mn - \cancel{m^2(m+n)} \cancel{+ m^2}}{(m+n)^2(m+n-1)}$$

$$= \frac{-mn}{(m+n)^2(m+n-1)}$$

$$\sum_{i=1}^{K} \sum_{j=1}^{K} \operatorname{cov}(x_i, x_j) = \frac{-K(K-1)mn}{(m+n)^2(n+m-1)}$$

$$\text{var}(x) = \frac{kmn}{(m+n)^2} - \frac{k(k-1)mn}{(n+m)^2(n+m-1)}$$

$$= \frac{kmn(m+n-k)}{(n+m)^2(n+m-1)}$$

$$= kp(1-p)\left(\frac{m+n-k}{n+m-1}\right)$$

$\because$ $p = \frac{m}{n+m}$ $\quad 1-p = \frac{n}{n+m}$

In the given problem, we need to find N, m(or)n given a distribution & k value.

N $\rightarrow$ length of distribution.

$P = \mu/k$ so we can find p here.

substituting p in variance we get m+n

$$m+n = \frac{k - \left(\frac{\text{var}(x)}{k*p(1-p)}\right)}{1 - \left(\frac{\text{var}(x)}{k*p(1-p)}\right)}$$

we know,

$P = \frac{m}{n+n} \Rightarrow m = p*(m+n)$