AI1110
Assignment 13

Ankit Saha
AI21BTECH11004

26 May 2022

Outline

1. Question
2. First Part
3. Second Part
4. Third Part

Papoulis Exercise 5-51

A box contains N identical items of which $M < N$ are defective ones. A sample of size n is taken from the box, and let X represent the number of defective items in the sample.

1. Find the distribution function of X if the n samples are drawn with replacement.
2. If the n samples are drawn without replacement, then show that

$$\Pr(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \quad \max(0, n + M - N) \leq k \leq \min(M, n) \tag{1}$$

   Find the mean and variance of X. This distribution is known as the hypergeometric distribution.

3. In the previous equation, let $N \to \infty, M \to \infty$ such that $\frac{M}{N} \to p, 0 < p < 1$. Then show that the hypergeometric random variable can be approximated by a binomial random variable with parameters n and p provided $n \ll N$

First Part

Let there be k defective items among the n chosen samples. Since we are drawing with replacement, the distribution will be a binomial distribution with parameters n and $p = \frac{M}{N}$ where p is the probability that a randomly chosen sample is defective.

$$
\begin{aligned}
f_X(k) &= \binom{n}{k} p^k (1-p)^{n-k}, \ k \in \{0, 1, \dots, n\} && (2) \\
&= \binom{n}{k} \left(\frac{M}{N}\right)^k \left(\frac{N-M}{N}\right)^{n-k}, \ k \in \{0, 1, \dots, n\} && (3) \\
&= \binom{n}{k} \frac{M^k (N-M)^{n-k}}{N^n}, \ k \in \{0, 1, \dots, n\} && (4)
\end{aligned}
$$

Second Part

Now we are drawing without replacement. n samples can be chosen from N items in $\binom{N}{n}$ ways. Out of these, k defective samples can be chosen from a total of M defective items in $\binom{M}{k}$ ways and the remaining $n-k$ samples can be chosen from the $N-M$ non-defective items in $\binom{N-M}{n-k}$ ways. Thus, the total number of ways of choosing k defective and $n-k$ non-defective items is

$$\binom{M}{k}\binom{N-M}{n-k} \tag{5}$$

Therefore, the probability is given by

$$\Pr(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \tag{6}$$

The constraints on k can be obtained as follows:

$$k \geq 0, n - k \leq N - M \tag{7}$$

$$\implies k \geq \max(0, n + M - N) \tag{8}$$

$$k \leq M, n - k \geq 0 \tag{9}$$

$$\implies k \leq \min(M, n) \tag{10}$$

$$\therefore \max(0, n + M - N) \leq k \leq \min(M, n) \tag{11}$$

Mean

The mean of X is given by

$$E[X] = \sum_k k \Pr(X = k) \tag{12}$$

$$= \sum_k \frac{k\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \tag{13}$$

$$k\binom{M}{k} = k\frac{M!}{k!(M-k)!} = M\frac{(M-1)!}{(k-1)!(M-k)!} = M\binom{M-1}{k-1} \tag{14}$$

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N}{n}\frac{(N-1)!}{(n-1)!(N-n)!} = \frac{N}{n}\binom{N-1}{n-1} \tag{15}$$

$$\implies E[X] = \sum_k \frac{M\binom{M-1}{k-1}\binom{N-M}{n-k}}{\frac{N}{n}\binom{N-1}{n-1}} \tag{16}$$

$$E[X] = \frac{Mn}{N} \frac{\sum_k \binom{M-1}{k-1}\binom{N-M}{n-k}}{\binom{N-1}{n-1}} \tag{17}$$

By Vandermonde's identity,

$$\sum_i \binom{a}{i}\binom{b}{r-i} = \binom{a+b}{r} \tag{18}$$

Combinatorially, this can be shown by counting the number of ways to select r fruits from a basket consisting of a apples and b oranges. This is equal to the sum over all possible values of i, of the number of ways to select i apples and $r-i$ oranges.

$$\implies \sum_k \binom{M-1}{k-1}\binom{N-M}{n-k} = \sum_k \binom{M-1}{k-1}\binom{N-M}{(n-1)-(k-1)} \tag{19}$$

$$= \binom{N-1}{n-1} \tag{20}$$

$$\implies \frac{\sum_k \binom{M-1}{k-1}\binom{N-M}{n-k}}{\binom{N-1}{n-1}} = 1 \tag{21}$$

Therefore,

$$E[X] = \frac{Mn}{N} \tag{22}$$

Variance

The variance of X is given by

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 \tag{23}$$

$$E[X^2] = \sum_k k^2 \Pr(X = k) \tag{24}$$

$$= \sum_k \frac{k \cdot k\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \tag{25}$$

$$= \sum_k \frac{k \cdot M\binom{M-1}{k-1}\binom{N-M}{n-k}}{\binom{N}{n}} \tag{26}$$

$$= M\sum_k \frac{(k - 1 + 1)\binom{M-1}{k-1}\binom{N-M}{n-k}}{\binom{N}{n}} \tag{27}$$

$$\tag{28}$$

$$E[X^2] = M\sum_k \frac{(k-1)\binom{M-1}{k-1}\binom{N-M}{n-k}}{\frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}} + M\sum_k \frac{\binom{M-1}{k-1}\binom{N-M}{n-k}}{\frac{N}{n}\binom{N-1}{n-1}} \tag{29}$$

$$= \frac{Mn(n-1)}{N(N-1)}\sum_k \frac{(M-1)\binom{M-2}{k-2}\binom{N-M}{n-k}}{\binom{N-2}{n-2}} + \frac{Mn}{N} \tag{30}$$

$$= \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{Mn}{N} \tag{31}$$

following a similar computation as that done for the mean

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 \tag{32}$$

$$= \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{Mn}{N} - \left(\frac{Mn}{N}\right)^2 \tag{33}$$

$$\mathrm{Var}(X) = \frac{Mn}{N}\left(\frac{(M-1)(n-1)}{N-1} + 1 - \frac{Mn}{N}\right) \tag{34}$$

$$= \frac{Mn}{N}\frac{N(M-1)(n-1) + N(N-1) - Mn(N-1)}{N(N-1)} \tag{35}$$

$$= \frac{Mn}{N^2(N-1)}\left(NMn + N - NM - Nn + N^2 - N - NMn + Mn\right) \tag{36}$$

$$= \frac{Mn}{N^2(N-1)}(N(N-M) - n(N-M)) \tag{37}$$

Therefore,

$$\mathrm{Var}(X) = \frac{Mn(N-n)(N-M)}{N^2(N-1)} \tag{38}$$

$$\Pr\left(X = k\right) \tag{39}$$

$$= \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \tag{40}$$

$$= \frac{M!}{k!(M-k)!} \frac{(N-M)!}{(n-k)!(N-M-n+k)!} \frac{n!(N-n)!}{N!} \tag{41}$$

$$= \frac{n!}{k!(n-k)!} \frac{M\cdots(M-k+1)}{N\cdots(N-k+1)} \frac{(N-M)\cdots(N-M-n+k+1)}{(N-k)\cdots(N-n+1)} \tag{42}$$

$$= \binom{n}{k} \frac{M^k\left(1-\frac{1}{M}\right)\cdots\left(1-\frac{k-1}{M}\right)}{N^k\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{k-1}{N}\right)} \frac{(N-M)^{n-k}\left(1-\frac{1}{N-M}\right)\cdots\left(1-\frac{n-k-1}{N-M}\right)}{(N-k)^{n-k}\left(1-\frac{1}{N-k}\right)\cdots\left(1-\frac{n-k-1}{N-k}\right)} \tag{43}$$

Now, $N \to \infty, M \to \infty$ such that $\frac{M}{N} \to p, 0 < p < 1$

$$\implies N - M \simeq N(1 - p) \to \infty \tag{44}$$

$$\text{and } k \leq n \ll N \implies N - k \simeq N \to \infty \tag{45}$$

Thus,

$$\Pr(X = k) \simeq \binom{n}{k} \frac{M^k}{N^k} \frac{(N - M)^{n-k}}{N^{n-k}} \tag{46}$$

$$= \binom{n}{k} p^k (1 - p)^{n-k} \tag{47}$$

where $p = \frac{M}{N}$
Therefore, the hypergeometric random variable can be approximated by a binomial random variable with parameters n and k. $\square$