LECTURE NOTE

ON

PROBABILITY AND STATISTICS 2

BY

ASSIST. PRF. DR. MUSTAFA I. NAIF

DEPARTMENT OF MATHEMATICS
COLLEGE OF EDUCATION FOR PURE SCIENCE
UNIVERISTY OF ANBAR

LECTURE 8#

➢ Outline :-

✓ Discrete distributions

5- Hypergeometric distribution

Definition

Expected value and Variance

Moment generating function

Characteristic function

Distribution function

Solved exercises

Exercises

5-Hypergeometric Distribution

Consider a collection of n objects which can be classified into two classes, say class 1 and class 2. Suppose that there are n_1 objects in class 1 and n_2 objects in class 2. A collection of r objects is selected from these n objects at random and without replacement. We are interested in finding out the probability that exactly x of these r objects are from class 1. If x of these r objects are from class 1, then the remaining r − x objects must be from class 2. We can select x objects from class 1 in any one of $\binom{n_1}{x}$ways. Similarly, the remaining r − x objects can be selected in $\binom{n_2}{r-x}$ ways. Thus, the number of ways one can select a subset of r objects from a set of n objects, such that x number of objects will be from class 1 and r − x number of objects will be from class 2, is given by $\binom{n_1}{x}\binom{n_2}{r-x}$ Hence,

$$P(X = x) = \frac{\binom{n_1}{x}\binom{n_2}{r-x}}{\binom{n}{r}},$$

where $x \leq r$, $x \leq n_1$ and $r - x \leq n_2$.

Hypergeometric Distribution

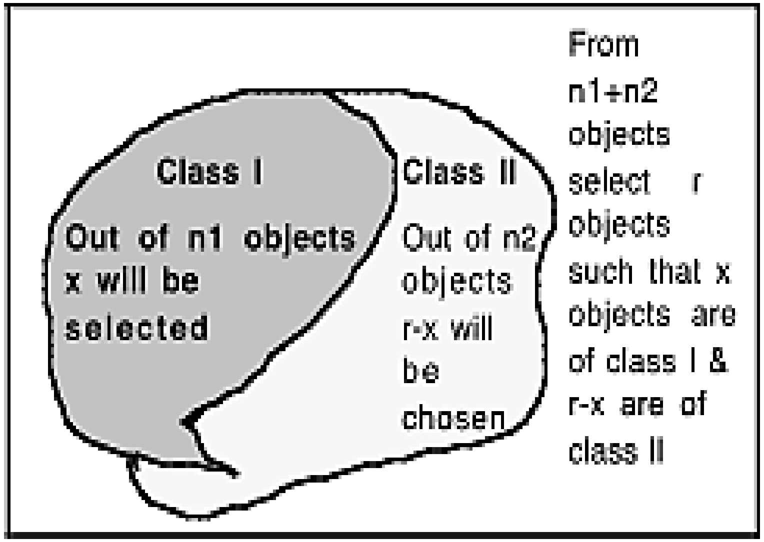
Definition : A random variable X is said to have a hypergeometric distribution if its probability mass function is of the form:

$$f(x) = \frac{\binom{n_1}{x}\binom{n_2}{r-x}}{\binom{n_1+n_2}{r}}, \qquad x = 0, 1, 2, ..., r$$

where $x \leq n_1$ and $r - x \leq n_2$ with n_1 and n_2 being two positive integers.

We shall denote such a random variable by writing $X \sim HYP(n_1, n_2, r)$.

Example :Suppose there are 3 defective items in a lot of 50 items. A sample of size 10 is taken at random and without replacement. Let X denote the number of defective items in the sample. What is the probability that the sample contains at most one defective item?



Hypergeometric Distribution

Answer: Clearly, $X \sim HYP(3, 47, 10)$. Hence the probability that the sample contains at most one defective item is

$$
\begin{aligned}
P(X \leq 1) &= P(X = 0) + P(X = 1) \\
&= \frac{\binom{3}{0}\binom{47}{10}}{\binom{50}{10}} + \frac{\binom{3}{1}\binom{47}{9}}{\binom{50}{10}} \\
&= 0.504 + 0.4 \\
&= 0.904.
\end{aligned}
$$

Theorem If $X \sim HYP(n_1, n_2, r)$, then

$$
\begin{aligned}
E(X) &= r\,\frac{n_1}{n_1 + n_2} \\
Var(X) &= r\left(\frac{n_1}{n_1 + n_2}\right)\left(\frac{n_2}{n_1 + n_2}\right)\left(\frac{n_1 + n_2 - r}{n_1 + n_2 - 1}\right).
\end{aligned}
$$

Hypergeometric Distribution

Proof: Let $X \sim HYP(n_1, n_2, r)$. We compute the mean and variance of X by computing the first and the second factorial moments of the random variable X. First, we compute the first factorial moment (which is same as the expected value) of X. The expected value of X is given by

$$\begin{aligned}
E(X) &= \sum_{x=0}^{r} x\, f(x) \\
&= \sum_{x=0}^{r} x \frac{\binom{n_1}{x}\binom{n_2}{r-x}}{\binom{n_1+n_2}{r}} \\
&= n_1 \sum_{x=1}^{r} \frac{(n_1-1)!}{(x-1)!\,(n_1-x)!} \frac{\binom{n_2}{r-x}}{\binom{n_1+n_2}{r}} \\
&= n_1 \sum_{x=1}^{r} \frac{\binom{n_1-1}{x-1}\binom{n_2}{r-x}}{\frac{n_1+n_2}{r}\binom{n_1+n_2-1}{r-1}} \\
&= r\,\frac{n_1}{n_1+n_2} \sum_{y=0}^{r-1} \frac{\binom{n_1-1}{y}\binom{n_2}{r-1-y}}{\binom{n_1+n_2-1}{r-1}}, \qquad \text{where } y = x-1 \\
&= r\,\frac{n_1}{n_1+n_2}.
\end{aligned}$$

The last equality is obtained since $\sum_{y=0}^{r-1} \frac{\binom{n_1-1}{y}\binom{n_2}{r-1-y}}{\binom{n_1+n_2-1}{r-1}} = 1.$ where $\sum_{i=0}^{n}\binom{a}{i}\binom{b}{n-i} = \binom{a+b}{n}$

Similarly, we find the second factorial moment of X to be

$E(X(X-1)) = \frac{r(r-1)\,n_1\,(n_1-1)}{(n_1+n_2)\,(n_1+n_2-1)}$. Therefore, the variance of X is

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= E(X(X-1)) + E(X) - E(X)^2 \\ &= \frac{r(r-1)\,n_1\,(n_1-1)}{(n_1+n_2)\,(n_1+n_2-1)} + r\,\frac{n_1}{n_1+n_2} - \left(r\,\frac{n_1}{n_1+n_2}\right)^2 \\ &= r\left(\frac{n_1}{n_1+n_2}\right)\left(\frac{n_2}{n_1+n_2}\right)\left(\frac{n_1+n_2-r}{n_1+n_2-1}\right). \end{aligned}$$

Distribution Function:The distribution function of a discrete hypergeometric random variable X is:

$F(X) = P(X \leq x) = \sum_{k=c}^{x} \frac{\binom{n_1}{x}\binom{n_2}{r-x}}{\binom{n_1+n_2}{r}}$, where c=max(0,r-$n_1 + n_2$)

Moment generating function :

The m g. f. of a discrete hypergeometric random variable X is:

$$M_X(t) = \frac{(n_1 - r)!\,(n_1 - n_2)!}{n_1} . H(-r; -n_2; n_1 - n_2 + 1; e^t)$$

where $H(-r; -n_2; n_1 - n_2 + 1; e^t)$=$\sum_{j=0}^{\infty} \frac{(-r)^{[j]}(-n_2)^{[j]}(e^t)^j}{(n_1-n_2-r+1)^{[j]}j!}$ and in general ,

for any number a , then :

$$a^{[j]} = a(a+1)(a+2)\ldots(a+j-1).$$

Note: Let X1, X2 are r.v's distributed as Ber(p). If X2 is not independent of X1, and we should not expect X to have a binomial distribution. (why?)

Hypergeometric Distribution

Example : A random sample of 5 students is drawn without replacement from among 300 seniors, and each of these 5 seniors is asked if she/he has tried a certain drug. Suppose 50% of the seniors actually have tried the drug. What is the probability that two of the students interviewed have tried the drug?

Answer: Let X denote the number of students interviewed who have tried the drug. Hence the probability that two of the students interviewed have tried the drug is

$$P(X=2) = \frac{\binom{150}{2}\binom{150}{3}}{\binom{300}{5}}$$
$$= 0.3146.$$

Hypergeometric Distribution

Example: A box contains 20 balls , 12 is red and others are black , if we select 8 ball a r.s. form this box, what is the probability of:

1- to get 3 red balls from this sample

2- At least two red balls have been got.

Sol: let X be the number of red balls selected from the sample.

So, X~HYP(20,12,8). And that means,

$$p(x) = \frac{\binom{12}{x}\binom{8}{8-x}}{\binom{20}{8}}, \qquad 0 \le x \le 8$$

So,

$$1- \ p(3) = \frac{\binom{12}{3}\binom{8}{5}}{\binom{20}{8}} = 0.098801$$

2- $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) = 1 - [P(X = 0) + P(X = 1)]$

$$=1- \left[\frac{\binom{12}{0}\binom{8}{8}}{\binom{20}{8}} + \frac{\binom{12}{1}\binom{8}{7}}{\binom{20}{8}} \right]=1-0.0008=0.9992$$

See you next Lecture