BITS Pilani
K K Birla Goa Campus

Probability & Statistics,

Dr. Jajati Keshari Sahoo
Department of Mathematics

Geometric Distribution

Assumptions

1. The experiment consists of a series of Bernoulli trials.
2. The probability of a success is the same for each trial.
3. The trials are independent.
4. X denote the no. of trial needed to obtain the first success.

Geometric Distribution

Definition

A random variable X is said to have geometric distribution with parameter p if its density is given

$$f(x) = (1-p)^{x-1} p, \quad x = 1, 2, 3, \cdots$$

where p is the probability of getting success.

Mean, Variance and Moment generating function of Geometric distribution

Mean

$$\mu = \frac{1}{p}$$

$p \rightarrow$ probability of success

Variance

$$\sigma^2 = \frac{(1-p)}{p^2} = \frac{q}{p^2}$$

Moment generating function

$$M_X(t) = \frac{pe^t}{1-qe^t}.$$

Example-1

Suppose a light switch is turned on and off until it fails. If the probability that the switch will fail any time it is turned on or off is 0.001, what is the probability that the switch will fail *after* it has turned on or off 1,200 times? Assume that the conditions underlying the geometric distribution are met.

Example-2

Suppose independent identical laboratory experiments are to be undertaken. Each Experiment is externally sensitive to environmental conditions and there is only a probability '*r*' that it will be completed successfully. Find the value of 'r' that maximize the probability of 5^{th} trial being the first unsuccessful trial.

Memory Loss Property

Example-3

If X is geometric random variable with parameter p. Show that

$$P(X > m+n \mid X > m) = P(X > n).$$

Note: The converse also true. Verify ?

Memory Loss Property

Remark:

- If X represents the lifetime of a device, then memory loss property states that if the device has been working for time m, then the probability that it will survive an additional time n depends only on n (not on m) and is identical to the probability of survival for time n of a new device.

- The equipment does not remember that it has been in use for time m.

Hypergeometric distribution

Assumptions

1. The experiment consists of drawing n objects without replacement from a collection of N objects. (without regard to the order)
2. Of the N objects, r have a trait of interest to us (success) and remaining N-r do not have the trait (failure).
3. X denote the no. of successes in the sample n.

Hypergeometric Distribution

Sampling without replacement

Number of ways in which x successes can be chosen is $\binom{r}{x}$

Number of ways in which $n - x$ failures be chosen is $\binom{N-r}{n-x}$

Hence number of ways x successes and $n - x$ failures can be chosen is

$$\binom{r}{x}\binom{N-r}{n-x}$$

Hypergeometric Distribution

Number of ways n objects can be chosen from N objects is $\binom{N}{n}$

If all the possibilities are equally likely then for sampling without replacement the probability of getting "x successes in n trials" is given by

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}.$$

Q. Range of X ?

Hypergeometric Distribution

Definition

A random variable X is said to have hypergeometric distribution with parameters N, n, and r if its density is given by

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}},$$

where $\max\{0, n-(N-r)\} \le x \le \min(n, r).$

Example-4

A job plant can pass inspection if there are no defects in 3 randomly chosen sample from a lot of size 20(without replacement).

(a) Find the probability of the plant passing the inspection if there are 2 defects in the lot.

Example-5

20 microprocessor chips are in stock. Out of 20, 3 have etching errors that can not detected by naked eye. 5 chips are selected at random without replacement. Find the density for X, the number of chips selected that have etching errors.

Example-6

A numerical comparison between hypergeometric and binomial distributions

A shipment of 100 tape recorders contains 25 that are defective. If 10 of them are randomly chosen for inspection, what is the probability that 2 of the 10 will be defective find it by using:

(a) hypergeometric distribution;

(b) binomial distribution as an approximation ?

Ans: $h(2;10,25,100) = 0.292$. and

$b(2;10,0.25) = 0.282$.

Hypergeometric Distribution an approximation to Binomial distribution

- In general it can be shown that

$$h(x; n, r, N) \to b(x; n, p)$$

with $p = (r/N)$ when $N \to \infty$.

- A good rule of thumb is to use the binomial distribution as an approximation to the hyper-geometric distribution if $n/N \leq 0.05$

Example-7

Among the 300 employees of a company, 240 are union members, while the others are not, If 8 of the employees are chosen to serve on the administrative committee, find the probability that 5 of them will be union member while the others are not.

Solution:

$$N = 300, \quad r = 60$$

$$n = 8, \quad x = 3$$

$$h(3;8,60,300) = \frac{\binom{60}{3}\binom{240}{5}}{\binom{300}{8}} = 0.1470$$

$$b(3,8,0.2) = 0.1466$$

Hypergeometric Distribution

Mean of hypergeometric distribution

$$\mu = n \cdot \frac{r}{N}$$

$n \rightarrow$ sample size
N $\rightarrow$ population size
$r \rightarrow$ number of success

Proof:

$$\mu = \sum_{x=0}^{n} x.f(x) = \sum_{x=1}^{n} x.\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Hypergeometric Distribution

$$\binom{r}{x} = \frac{r!}{x!(r-x)!} = \frac{r}{x}\frac{(r-1)!}{(x-1)!(r-x)!} = \frac{r}{x}\binom{r-1}{x-1}$$

$$\mu = \sum_{x=1}^{n} r.\frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{r}{\binom{N}{n}}\sum_{x=1}^{n}\binom{r-1}{x-1}\binom{N-r}{n-x}$$

Hypergeometric Distribution

Put $x-1=y$

$$\mu=\frac{r}{\binom{N}{n}}\sum_{y=0}^{n-1}\binom{r-1}{y}\binom{N-r}{n-1-y}$$

$$k=n-1$$
$$m=r-1$$
$$r=y$$
$$s=N-r$$

Use the identity $\sum_{r=0}^{k}\binom{m}{r}\binom{s}{k-r}=\binom{m+s}{k}$

Hypergeometric Distribution

We get

$$\mu = \frac{r}{\binom{N}{n}} \binom{N-1}{n-1}$$

$$= n \cdot \frac{r}{N}$$

Hypergeometric Distribution

Variance of hypergeometric distribution

$$\sigma^2 = \frac{n \cdot r \cdot (N-r) \cdot (N-n)}{N^2 \cdot (N-1)}$$

Proof:

$$\mu_2' = \sum_{x=0}^{n} x^2 f(x) = \sum_{x=1}^{n} x^2 . \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

$$\mu_2' = r\sum_{x=1}^{n} x.\frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\binom{N}{n}}$$

$$= \frac{r}{\binom{N}{n}}\sum_{x=1}^{n}(x-1+1).\binom{r-1}{x-1}\binom{N-r}{n-x}$$

Hypergeometric Distribution

$$\mu_2' = \frac{r(r-1)}{\binom{N}{n}} \sum_{x=2}^{n} \cdot \binom{r-2}{x-2}\binom{N-r}{n-x}$$

$$+ \frac{r}{\binom{N}{n}} \sum_{x=1}^{n} \binom{r-1}{x-1}\binom{N-r}{n-x}$$

Put $x - 2 = y$ in 1st summation

Hypergeometric Distribution

$$\mu_2' = \frac{r(r-1)}{\binom{N}{n}} \sum_{y=0}^{n-2} \cdot \binom{r-2}{y} \binom{N-r}{n-2-y} + r\frac{n}{N}$$

Use the identity $\sum_{r=0}^{k} \binom{m}{r} \binom{s}{k-r} = \binom{m+s}{k}$

$$k = n-2,\ m = r-2$$

$$r = y,\ s = N-r$$

Hypergeometric Distribution

$$\mu_2' = \frac{r(r-1)}{\binom{N}{n}}\binom{N-2}{n-2} + r\frac{n}{N}$$

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

$$\boxed{\sigma^2 = \mu_2' - \mu^2 = \frac{n \cdot r \cdot (N-r) \cdot (N-n)}{N^2 \cdot (N-1)}}$$