3.5 Hypergeometric and Negative Binomial Distributions

The hypergeometric and negative binomial distributions are both related to the binomial distribution.

Hypergeometric distributions:

The hypergeometric distribution is a discrete probability distribution. The hypergeometric distribution describes the probability of x successes (random draws for which the object drawn has a specified feature) in n draws, without replacement, from a finite population of size N that contains exactly M objects with that feature, wherein each draw is either success or failure.

In contrast, the binomial distribution describes the probability of x successes in draws with replacement.

. Sampling with replacement: In sampling with replacement, the object that was drawn at random is placed back to the given set and the set is mixed thoroghly. Then we draw the next object at random.

Sampling without replacement: In sampling without replacement the object that was drawn is put aside.

Example: A box contains 10 screws, 3 of which are defective. Two screws are drawn at random. Let A = 1st drawn screw non defective and
B = 2nd " " "

Then

with replacement

$P(A) = \frac{7}{10}$

$P(B) = \frac{7}{10}$

without replacement

$P(A) = 7/10$

$P(B) = 6/9$

Assumptions :

The assumptions leading to the hypergeometric distribution are as follows:

1) The population consists of N individuals, objects, or elements (a finite population)
2) Each individual can be characterized as a success(S) or 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.

The random variable of interest is X = the number of success in the sample. The probability distribution of X depends on the parameters n, M and N, so we wish to obtain
$P(X=x) = h(x; n, M, N)$

Probability mass function of Hypergeometric distribution:

If X is the number of success(S) in a completely random sample of size n drawn from a population consisting of M successes and (N−M) failures, then the probability distribution of X, called the hypergeometric distribution, is given by,

$$P(X=x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

for x, an integer, satisfying

$$\max(0, n-N+M) \le x \le \min(n, M).$$

Example of Hypergeometric distribution

Suppose a large urn contains 400 red marbles and 600 blue marbles. A random sample of 10 marbles is drawn without replacement. What is the probability exactly 3 are red?

Ans: Here N = population size $= 400 + 600 = 1000$

n = no. of draws = sample size $= 10$

M = Number of observed successes in the population $= 400$.

The random variable is X = no. of successes in the sample

Here $x = 3$. The Hypergeometric distribution is

$$P(X=x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$\Rightarrow P(X=3) = h(3; 10, 400, 1000) = \frac{\binom{400}{3}\binom{1000-400}{10-3}}{\binom{1000}{10}}$$

$$= \frac{\binom{400}{3}\binom{600}{7}}{\binom{1000}{10}} = 0.2155$$

Example: 1

Five individuals from an animal population thought to be near extinction in a certain region have been caught, tagged, and released to mix into the population. After they have had an opportunity to mix, ~~an~~ a random ~~variable~~ sample of 10 of these animals is selected. Let $X =$ the number of tagged animals in the second sample. If there are actually 25 animals of this type in the region, what is the probability that a) $X = 2$?
b) $X \leq 2$?

Ans: The parameters values are $n = 10$, $M = 5$, (5 tagged animal in the population)
$N = 25$. $\Rightarrow N - M = 25 - 5 = 20$

So, the p.m.f is

$$h(x; 10, 5, 25) = \frac{\binom{5}{x}\binom{20}{10-x}}{\binom{25}{10}}, \quad x = 0, 1, 2, 3, 4, 5$$

(a) $\therefore P(X = 2) = h(2; 10, 5, 25) = \dfrac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}} = 0.385$

(b) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$= h(0; 10, 5, 25) + h(1; 10, 5, 25) + h(2; 10, 5, 25)$

$= 0.057 + 0.257 + 0.385$

$= 0.699$

Mean and variance of hypergeometric distribution.

The mean and variance of the hypergeometric r.v. X having p.m.f $h(x; n, M, N)$ are

$$E(X) = n \cdot \frac{M}{N}$$

$$V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N}\left(1-\frac{M}{N}\right)$$

Note:

The ratio M/N is the proportion of successes in the population. If we replace M/N by p in $E(X)$ and $V(X)$, we get

$$E(X) = np$$

$$V(X) = \left(\frac{N-n}{N-1}\right) \cdot np(1-p)$$

The above expressions shows that the means of the binomial and hypergeometric r.v.' are same, whereas the variances of the two random variables differ by the factor $\frac{N-n}{N-1}$, often called the finite population correction factor. This factor is less than 1, so the hypergeometric variable has smaller variance then does the binomial r.v.

The correction factor can be written as $\dfrac{1-\frac{n}{N}}{1-\frac{1}{N}}$, which is approximately 1 when n is smaller relative to N.

Example:

In the animal-tagging example, $n=10$, $M=5$, and $N=25$ so, $p = \frac{5}{25} = 0.2$

$$\therefore E(X) = np = 10 \times 0.2 = 2$$

$$V(X) = \frac{25-10}{25-1} \times 10 \times 0.2 \times 0.8 = 1$$

Q. (69) [3.5]

Each of 12 refrigerators of a certain type has been returned to a distributor because of an audible, high-pitched, oscilating noise when the refrigerators are running. Suppose that 7 of these refrigerators have a defective compressor and the other 5 have less serious problems. If the refrigerators are examined in random order, let the X be the number among the first 6 examined that have a defective compressor. Compute the following:

a) $P(X=5)$, b) $P(X \leq 4)$.

c) The probability that X exceeds its mean value by more than 1 standard deviation.

Ans: Given $N =$ population size $= 12$

$n =$ Number of draws $=$ Sample size $= 6$

$M =$ Number of observed successes $= 7$

The p.m.f of hyper-geometric distribution is

$$P(X=x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Since a finite population has been given, we need to use the hypergeometric distribution.

(a) $$P(X=5) = h(5; 6, 7, 12) = \frac{\binom{7}{5}\binom{5}{1}}{\binom{12}{6}}$$

$$= \frac{5}{44} = 0.1136 = 11.36\%.$$

(b) $P(X \le 4)$

$= P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$\max(0, n-N+M) \le x \le \min(n, M)$

$\Rightarrow \max(0, 6-12+7) \le x \le \min(6,7)$

$\Rightarrow 1 \le x \le 6$

$\Rightarrow x = 1,2,3,4,5,6$

$$= \frac{\binom{7}{1}\binom{5}{5}}{\binom{12}{6}} + \frac{\binom{7}{2}\binom{5}{4}}{\binom{12}{6}} + \frac{\binom{7}{3}\binom{5}{3}}{\binom{12}{6}} + \frac{\binom{7}{4}\binom{5}{2}}{\binom{12}{6}}$$

$$= \frac{1}{132} + \frac{5}{44} + \frac{25}{66} + \frac{25}{66} = \frac{29}{33} = 0.8788 = 87.88\%$$

(c) $E(X) = np = n.\frac{M}{N} = 6 \times \frac{7}{12} = 3.5 \quad [\because p = \frac{M}{N}]$

$$\sigma^2 = V(X) = \frac{N-n}{N-1}\, np(1-p)$$

$$\Rightarrow \sigma = \sqrt{\frac{N-n}{N-1}\, np(1-p)}$$

$$= \sqrt{\frac{12-6}{12-1} \times 6 \times \frac{7}{12}\left(1-\frac{7}{12}\right)}$$

$$= 0.8919$$

$\therefore$ P(X is more than one standard deviation above the mean)

$= P(X > \mu + \sigma)$

$= P(X > 3.5 + 0.8919) = P(X > 4.3919)$

$= P(X > 4)$

$= 1 - P(X \le 4)$

$= 1 - 29/33 = \frac{4}{33} \approx 0.1212 = 12.12\%$

Q. 70 [3.5]

An instructor who taught two sections of engineering statistics last term, the first with 20 students and the second with 30, decided to assign a term project. After all projects had been turned in, the instructor randomly ordered them before grading. Consider the first 15 graded projects.

a) what is the probability that exactly ~~not least~~ 10 of these are from second section?

b) what is the probability that at least 10 of these are from the second section?

c) what is the probability that at least 10 of these are from the same section?

d) what are the mean value and standard deviation of the number of projects ~~not~~ among these first 15 that are from ~~the~~ the second section?

e) what are the mean value and standard deviation of the number of projects not among these first 15 that are from the second section?

Ans:, Here

N = population size = 50

n = Number of draws = 15

M = Number of observed successes = 30

vow Hypergeometric distribution :

$$P(X=x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

where $\max(0, n-N+M) \le x \le \min(n, M)$

$\Rightarrow 0 \le x \le 15$

(a) $P(X=10)$

$= h(10; 15, 30, 50) = \dfrac{\binom{30}{10}\binom{50-30}{15-10}}{\binom{50}{15}} = \dfrac{\binom{30}{10}\binom{20}{5}}{\binom{50}{15}} = 0.2070$

(b) $P(X \geq 10)$

$= P(X=10) + P(X=11) + P(X=12) + P(13) + P(X=14) + P(X=15)$

$= \sum_{x=10}^{15} \dfrac{\binom{30}{x}\binom{50-30}{15-x}}{\binom{50}{15}} = \sum_{x=10}^{15} \dfrac{{}^{30}C_x \times {}^{20}C_{15-x}}{{}^{50}C_{15}} = 0.3799$

(c) P(at least 10 of these are from the same section)

= P(at least 10 of these are from 1st section or at least 10 of these are from 2nd section)

= P (at least 10 of these are from 1st section) + P (at least 10 of these from 2nd section)

Now P(at least 10 of these are from 1st section)

$= P(X \geq 10)$, but here $M=20$ $N=50$ $n=15$

$= \sum_{x=10}^{15} \dfrac{\binom{20}{x}\binom{50-20}{15-x}}{\binom{50}{15}} = \sum_{x=10}^{15} \dfrac{\binom{20}{x}\binom{30}{15-x}}{\binom{50}{15}} = \cancel{0.0140} = 0.0140$

and P(at least 10 of these from 2nd section)

$= 0.3799$ [from part (b)]

$\therefore$ The required probability $= 0.0140 + 0.3799$

$= 0.3939$

(d) mean $= \mu = E(x) = np = n \cdot \frac{M}{N} = 15 \times \frac{30}{50} = 9$

S.D, $\sigma = \sqrt{V(x)} = \sqrt{\frac{N-n}{N-1} np(1-p)}$, $p = \frac{M}{N} = \frac{30}{50} = 3/5$

$= \sqrt{\frac{50-15}{50-1} \times 15 \times \frac{3}{5}(1-3/5)} = 1.6036$

(e) Given that the distribution of the number of projects Not among from the second section. That means they are from the 1st section.

$\therefore$ In this case, $N = 50$, $n = 15$, $M = 20$

$\therefore$ mean, $\mu = np = 15 \times \frac{2}{5} = 6$ | Here $p = \frac{M}{N} = \frac{20}{50} = 2/5$

S.D, $\sigma = \sqrt{V(x)} = \sqrt{\frac{N-n}{N-1} np(1-p)}$

$= \sqrt{\frac{50-15}{50-1} \times 15 \times \frac{2}{5}(1-2/5)} = 1.6036$

Q.(74) [3.5]

A second-Stage smog alert has been called in a certain area of Los Angeles County in which there are 50 industrial firms. An inspector will visit 10 randomly selected firms to check for violations of regulations.

a) If 15 of the firms are actually violating at least one regulation, what is the pmf of the number of firms visited by the inspector that are in violation of at least one regulation?

b) If there are 500 firms in the area, of which 150 are in violation, approximate the pmf of part (a) by a simpler pmf.

c) For X = the number among the 10 visited that are in violation, compute E(X) and V(X) both for the exact pmf and the approximating pmf in part (b).

We will use the hypergeometric distribution as the sample of 10 is more than10 % of the population so, we cannot use the Binomial distribution.

Ans (a) Here $N =$ population size $= 50$

$n =$ Number of draws $= 10$

$M =$ Number of observed Success $= 15$

$p =$ probability of success $= \frac{\text{No. of favorable outcomes}}{\text{No. of possible outcomes}} = \frac{15}{50} = 0.3$

∴ The p.m.f of Hyper geometric distribution is given by

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

$$= h(x; 10, 15, 50) = \frac{\binom{15}{x}\binom{35}{10-x}}{\binom{50}{10}}$$

(b) $N = 500$, $n = 10$, $M = 150$ $\therefore p = \frac{150}{500} = \frac{3}{10} = 0.3$

~~Here, N = 500 is large and p = 0.3 is small~~

We can use the binomial distribution, because the sample of 10 is less than 10% of the population.

The p.m.f of binomial distribution is

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\Rightarrow b(x; 10, 0.3) = \binom{10}{x} (0.3)^x (1-0.3)^{10-x}$$

$$\Rightarrow b(x; 10, 0.3) = \binom{10}{x} (0.3)^x (0.7)^{10-x}$$

(c) For Hypergeometric distribution,

mean, $\mu = E(X) = np = n\frac{M}{N}$, $\left[\because p = \frac{M}{N}\right]$

$$V(X) = \frac{N-n}{N-1}\, np(1-p)$$

$$\therefore E(X) = 10 \times \frac{15}{50} = 3$$

$$v(x) = \frac{50-10}{50-1}\, 3(1-0.3)$$

$$= \frac{40}{49} \times 3 \times \frac{7}{10} = \frac{12}{7} \simeq 1.7143$$

For approximate

$$\mu = E(X) = np = 10 \times 0.3 = 3$$

$$V(X) = \sigma^2 = np(1-p) = 10 \times 0.3\,(1-0.3)$$

$$= 3 \times 0.7 = 2.1$$