Frequently Asked Questions

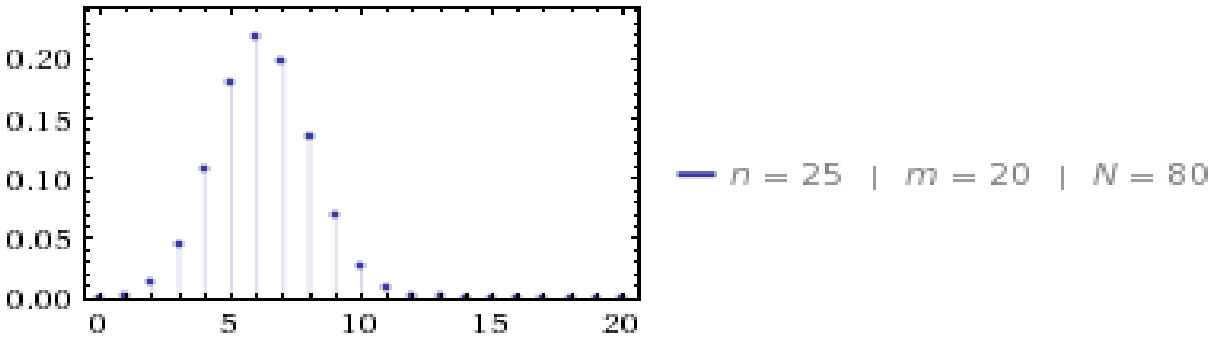
1. Define hypergeometric distribution?

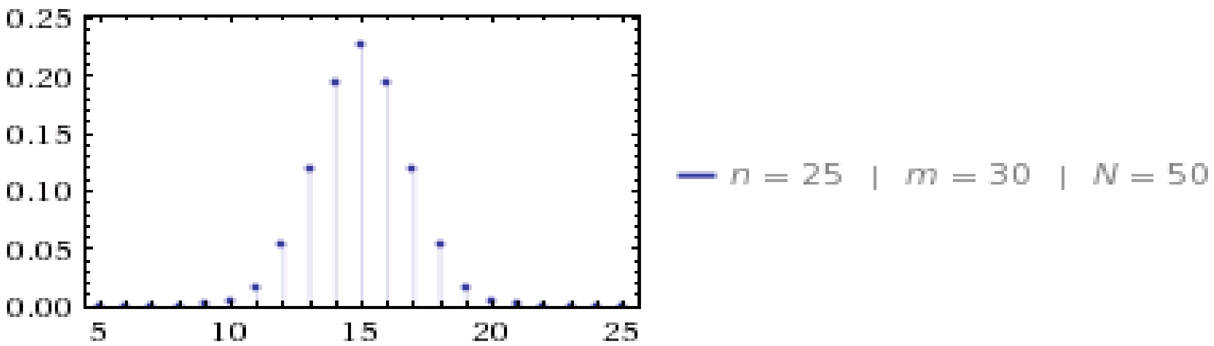
Answer: A random variable X is said to follow Binomial distribution with parameters n and p if its probability mass function is given by

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, k = 0,1,2,\ldots \min(n, M)$$

2. Draw the graph of probability mass function of Hypergeometric distribution.

Answer:





3. Show how the Hypergeometric distribution arises, by giving an example.

Answer:

When the population is finite and the sampling is done without replacement, so that the events are stochastically dependent, although random, we obtain Hypergeometric distribution.

Consider an urn with N balls, M of which are white and N minus M are red. Suppose that we draw a sample of n balls at random (without replacement) from the urn, then the probability of getting k white balls out of n (k<n) is

$$\binom{M}{k}\binom{N-M}{n-k} \div \binom{N}{n}$$

4. Write the basic characteristics of Hypergeometric distribution

Answer:

1. It models that the total number of successes in a size sample drawn without replacement from a finite population.
2. It differs from the binomial only in that the population is finite and the sampling from the population is without replacement.
3. Trials are dependent

5. Give examples of Hypergeometric distribution.

Answer:

1. Choose a team of 8 from a group of 10 boys and 7 girls.
2. Choose a committee of five from the legislature consisting of 52 Democrats and 48 Republicans.
3. Choosing exactly two marbles of each colour from 5 black, 10 white and 15 red marbles.
4. Choosing at least one defective chip when a batch of 100 computer chips containing 10 defective chips and 5 chips are chosen at random.
5. If a pond contains 1000 fish and 200 are tagged and a sample of size 20 is taken then getting 5 tagged fishes.

6. Find mean of Hypergeometric distribution.

Answer:

$$E(X) = \sum_{k=0}^{n} k.P(X = k)$$

$\mu_1' =$

$$= \sum_{k=0}^{n} k \left\{ \binom{M}{k} \binom{N-M}{n-k} \div \binom{N}{n} \right\}$$

$$= \frac{M}{\binom{N}{n}} \sum_{k=1}^{n} \binom{M-1}{k-1} \binom{N-M}{n-k}$$

$$= \frac{M}{\binom{N}{n}} \sum_{x=0}^{m} \binom{A}{x} \binom{N-A-1}{m-x}$$

, Where x=k-1, m=n-1 and M-1=A

$$= \frac{M}{\binom{N}{n}} \binom{N-1}{m}$$

$$= \frac{M}{\binom{N}{n}} \binom{N-1}{n-1} = \frac{nM}{N}$$

7. Obtain variance of the distribution

Answer:

$$E(X) = \sum_{k=0}^{n} k.P(X=k)$$

μ_1' =

$$= \sum_{k=0}^{n} k \left\{ \binom{M}{k} \binom{N-M}{n-k} \div \binom{N}{n} \right\}$$

$$= \frac{M}{\binom{N}{n}} \sum_{k=1}^{n} \binom{M-1}{k-1} \binom{N-M}{n-k}$$

$$= \frac{M}{\binom{N}{n}} \sum_{x=0}^{m} \binom{A}{x} \binom{N-A-1}{m-x}$$

, Where x=k-1, m=n-1 and M-1=A

$$= \frac{M}{\binom{N}{n}} \binom{N-1}{m}$$

$$= \frac{M}{\binom{N}{n}} \binom{N-1}{n-1} = \frac{nM}{N}$$

μ_2' = $E(X^2)$

Consider E[X(X-1)]

$$= \sum_{k=0}^{n} k(k-1).P(X=k)$$

$$= \sum_{k=0}^{n} k(k-1) \left\{ \binom{M}{k} \binom{N-M}{n-k} \div \binom{N}{n} \right\}$$

$$= \frac{M(M-1)}{\binom{N}{n}} \sum_{k=2}^{n} \binom{M-2}{k-2}\binom{N-M}{n-k}$$

$$= \frac{M(M-1)}{\binom{N}{n}} \binom{N-2}{n-2}$$

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

$E(X^2) = E[X(X-1)] + E(X)$

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

Variance is given by,

$$V(X) = \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{nM}{N} + \left(\frac{nM}{N}\right)^2$$

$\mu_2 = \mu'_2 - \mu'^2_1 =$

$$= \frac{nM}{N}\left[\frac{(M-1)(n-1)}{(N-1)} + 1 + \frac{nM}{N}\right] = \frac{nM(N-M)(N-n)}{N^2(N-1)}$$

8. Find the limiting distribution of Hypergeometric distribution

Answer: Hypergeometric distribution tends to binomial distribution as N—›∞ and M/N —›p

$$P(X = k) = \binom{M}{k}\binom{N-M}{n-k} \div \binom{N}{n}$$

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

$$= \frac{M(M-1)(M-2)...(M-k+1)}{k!} \times \frac{(N-M)(N-M-1)...(N-M-k+1)}{(n-k)!}$$

$$\times \frac{n!}{N(N-1)(N-2)...(N-n+1)}$$

$$= \frac{n!}{k!(n-k)!}\frac{M}{N}\left(\frac{M}{N}-\frac{1}{N}\right)\left(\frac{M}{N}-\frac{2}{N}\right)\dots\left(\frac{M}{N}-\frac{k-1}{N}\right)$$

$$\times\frac{\left(1-\frac{M}{N}\right)\left(1-\frac{M}{N}-\frac{1}{N}\right)\dots\left(1-\frac{M}{N}-\frac{n-k-1}{N}\right)}{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\dots\left(1-\frac{n-1}{N}\right)}$$

$$p(x)=\binom{n}{k}\frac{M}{N}\left(\frac{M}{N}-\frac{1}{N}\right)\left(\frac{M}{N}-\frac{2}{N}\right)\left(\frac{M}{N}-\frac{k-1}{N}\right)\times\frac{\left(1-\frac{M}{N}\right)\left(1-\frac{M}{N}-\frac{1}{N}\right)\dots\left(1-\frac{M}{N}-\frac{n-k-1}{N}\right)}{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\dots\left(1-\frac{n-1}{N}\right)}$$

Proceeding to the limit as N—›∞ and M/N —›p, we get,

$$\lim_{N\to\infty} p(x)=\binom{n}{k}\lim_{N\to\infty}\{\frac{M}{N}\left(\frac{M}{N}-\frac{1}{N}\right)\left(\frac{M}{N}-\frac{2}{N}\right)\left(\frac{M}{N}-\frac{k-1}{N}\right)$$

$$\times\frac{\left(1-\frac{M}{N}\right)\left(1-\frac{M}{N}-\frac{1}{N}\right)\dots\left(1-\frac{M}{N}-\frac{n-k-1}{N}\right)}{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\dots\left(1-\frac{n-1}{N}\right)}\}$$

$$=\binom{n}{k}\frac{p.p\dots p}{k-times}\times\frac{(1-p)(1-p)\dots(1-p)}{(n-k)-times}=\binom{n}{k}p^k(1-p)^{n-k}$$

, which is the probability mass function of binomial distribution with parameters n and k

9. Obtain the recurrence relation for the probabilities of Hypergeometric distribution.

Answer:

Recurrence relation for the probabilities of Hypergeometric distribution can be found as follows.

W.K.T

$$P(X=k)=\binom{M}{k}\binom{N-M}{n-k}\div\binom{N}{n}$$

Consider

$$P(X=k+1)=\binom{M}{k+1}\binom{N-M}{n-k-1}\div\binom{N}{n}$$

Consider the ratio,

$$\frac{P(X=k+1)}{P(X=k)}=\frac{\binom{M}{k+1}\binom{N-M}{n-k-1}\div\binom{N}{n}}{\binom{M}{k}\binom{N-M}{n-k}\div\binom{N}{n}}=\frac{(n-k)(M-k)}{(k+1)(N-M-n+k+1)}$$

Hence the recurrence relation is given by

$$P(X=k+1)=\frac{(n-k)(M-k)}{(k+1)(N-M-n+k+1)}.P(X=k)$$

, for k=1,2,3 …

10. Find an expression for finding factorial moments of Hypergeometric distribution

Answer

$$E[X^{(r)}]=\sum_{k=r}^{n}k^{(r)}P(X=k)$$

$$=\sum_{k=r}^{n}k^{(r)}\left\{\binom{M}{k}\binom{N-M}{n-k}\div\binom{N}{n}\right\}$$

$$=\sum_{k=r}^{n}M^{(r)}\left\{\binom{M-r}{k-r}\binom{N-M}{n-k}\div\binom{N}{n}\right\}$$

$$=M^{(r)}\sum_{j=0}^{n-r}\left\{\binom{M-r}{j}\binom{(N-r)-(M-r)}{(n-r)-j}\div\binom{N}{n}\right\}$$

$$=\frac{M^{(r)}n^{(r)}}{N^{(r)}}\sum_{j=0}^{n-r}\left\{\binom{M-r}{j}\binom{(N-r)-(M-r)}{(n-r)-j}\div\binom{N-r}{n-r}\right\}$$

$$=\frac{M^{(r)}n^{(r)}}{N^{(r)}}\times 1=\frac{M^{(r)}n^{(r)}}{N^{(r)}}$$

11. Explain how you will use Hypergeometric model to estimate the number of fish in a lake.

Answer

Let us suppose that in a lake there are N fish, N unknown. The problem is to estimate N. A catch of r fish, all at the same time, is made and these fish are returned alive into the lake after marking each with a red spot. After a reasonable period of time, during which are ‘marked’ fish are assumed to have distributed themselves ‘at random’ in the lake, another catch of ‘s’ fish, again all at once, is made. Here r and s are regarded as fixed predetermined constants. Among these ‘s’ fish caught, there will be, say, x marked fish, where X is a random variable following discrete probability function given by Hypergeometric model:

$$p(N) = \binom{r}{x}\binom{N-r}{s-x} \div \binom{N}{s}$$

, where x is an integer such that max(0, s-N+r)≤x≤min(r,s)

= 0 otherwise.

The value of N is estimated by the principle of Maximum Likelihood, that is we find the value $\hat{N}$ of N which maximizes p(N).
Since N is a discrete random variable, the principle of maxima and minima in calculus cannot be used here. Here we proceed as follows

$$\lambda(N) = \frac{p(N)}{p(N-1)} = \frac{(N-r)(N-s)}{N(N-r-s+x)}$$

$$N > \frac{rs}{x}$$

Suppose λ(N)>1, then

$$N > \frac{rs}{x}$$

Therefore p(N)>p(N-1) iff

$$N < \frac{rs}{x}$$

Suppose λ(N)<1, then

$$N < \frac{rs}{x}$$

Therefore p(N)<p(N-1) iff

From above inequalities we see that p(N) reaches the maximum value, when N is approximately equal to rs/x. Hence maximum likelihood estimate of N is given by,

$$\hat{N} = \frac{rs}{x}$$

12. If X and Y are independent binomial variates with parameters n_1, p and n_2, p respectively, then show that P(X=r/X+Y=n) is Hypergeometric.

Answer

W.K.T, X~B(n_1, p) and Y ~B(n_2, p).

Since X and Y are independent,

X+Y ~ B(n_1+n_2, p).

P[X=r/X+Y = n] = $\frac{P(X=r \cap X+Y=n)}{P(X+Y=n)}$

$$= \frac{P(X=r \cap Y=n-r)}{P(X+Y=n)}$$

$$= \frac{P(X=r)P(Y=n-r)}{P(X+Y=n)}$$

$$= \frac{\binom{n_1}{r} p^r (1-p)^{n_1-r} . \binom{n_2}{n-r} p^{n-r} (1-p)^{n_1-n+r}}{\binom{n_1+n_2}{n} p^n (1-p)^{n_1+n_2-n}}$$

$$= \frac{\binom{n_1}{r} . \binom{n_2}{n-r}}{\binom{n_1+n_2}{n}}$$

, which is probability mass function of Hypergeometric distribution.

13. Suppose that from a population of N elements of which M are defective and N-M are non-defective, a sample of size n is drawn without replacement. What is the probability that the sample contains exactly x defectives? Name this probability distribution.

Answer:

The probability that the sample contains exactly x defectives is

$$\binom{M}{x}\binom{N-M}{n-x} \div \binom{N}{n}$$

This is Hypergeometric distribution.

14. Show that for Hypergeometric distribution, $E(X) = \frac{nM}{N}$ and $V(X) = \frac{nM}{N}\left(1-\frac{M}{N}\right)\left(1-\frac{n-1}{N-1)}\right)$

Answer:

$$E(X) = \sum_{k=0}^{n} k.P(X=k)$$

$$= \sum_{k=0}^{n} k \left\{ \binom{M}{k}\binom{N-M}{n-k} \div \binom{N}{n} \right\}$$

$$= \frac{M}{\binom{N}{n}} \sum_{k=1}^{n} \binom{M-1}{k-1}\binom{N-M}{n-k}$$

$$= \frac{M}{\binom{N}{n}} \sum_{x=0}^{m} \binom{A}{x} \binom{N-A-1}{m-x}$$

, Where x=k-1, m=n-1 and M-1=A

$$= \frac{M}{\binom{N}{n}} \binom{N-1}{m}$$

$$= \frac{M}{\binom{N}{n}} \binom{N-1}{n-1} = \frac{nM}{N}$$

$\mu_2' = E(X^2)$

Consider E[X(X-1)]

$$= \sum_{k=0}^{n} k(k-1).P(X=k)$$

$$= \sum_{k=0}^{n} k(k-1) \left\{ \binom{M}{k} \binom{N-M}{n-k} \div \binom{N}{n} \right\}$$

$$= \frac{M(M-1)}{\binom{N}{n}} \sum_{k=2}^{n} \binom{M-2}{k-2} \binom{N-M}{n-k}$$

$$= \frac{M(M-1)}{\binom{N}{n}} \binom{N-2}{n-2}$$

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

$E(X^2) = E[X(X-1)] + E(X)$

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

Variance is given by,

$$V(X) = \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{nM}{N} + \left(\frac{nM}{N}\right)^2$$

$\mu_2 = \mu'_2 - \mu'^2_1 =$

$$= \frac{nM}{N}\left[\frac{(M-1)(n-1)}{(N-1)} + 1 + \frac{nM}{N}\right]$$

$$= \frac{nM(N-M)(N-n)}{N^2(N-1)}$$

$$= \frac{nM}{N}\left(1-\frac{M}{N}\right)\left(1-\frac{n-1}{N-1)}\right)$$

15. What is the mode of Hypergeometric distribution.

Answer:

Mode of Hypergeometric distribution is the value for which p(x) is maximum and is given by,

$$\frac{(n+1)(M+1)}{N+2}$$