Introduction to Statistics and Data Science using *eStat*

Chapter 5 Probability Distribution

5.3 Discrete Random Variable - Hypergeometric Distribution -

Jung Jin Lee
Professor of Soongsil University, Korea
Visiting Professor of ADA University, Azerbaijan

5.3 Discrete Random Variable

5.3.4 Hypergeometric Distribution

- Consider a box consisting of 20 products and 15 of them are normal products and 5 are defective products.
- When three of the 20 products are sampled, the probability of having one normal products and two defective product is $\frac{{}_{15}C_1 \times {}_5C_2}{{}_{20}C_3}$
- The random variable that counts the number of 'success' in the finite population consisting of only 'success' and 'failure' is called

  ⇨ **hypergeometric random variable**

  ⇨ **hypergeometric distribution**.

5.3 Discrete Random Variable

5.3.4 Hypergeometric Distribution

- Consider a population of size N which consists of D 'success' and $N - D$ 'failure'.
- If we collect a sample of size n without replacement and X is the number of 'success' in the sample, then the distribution of X is called hypergeometric distribution

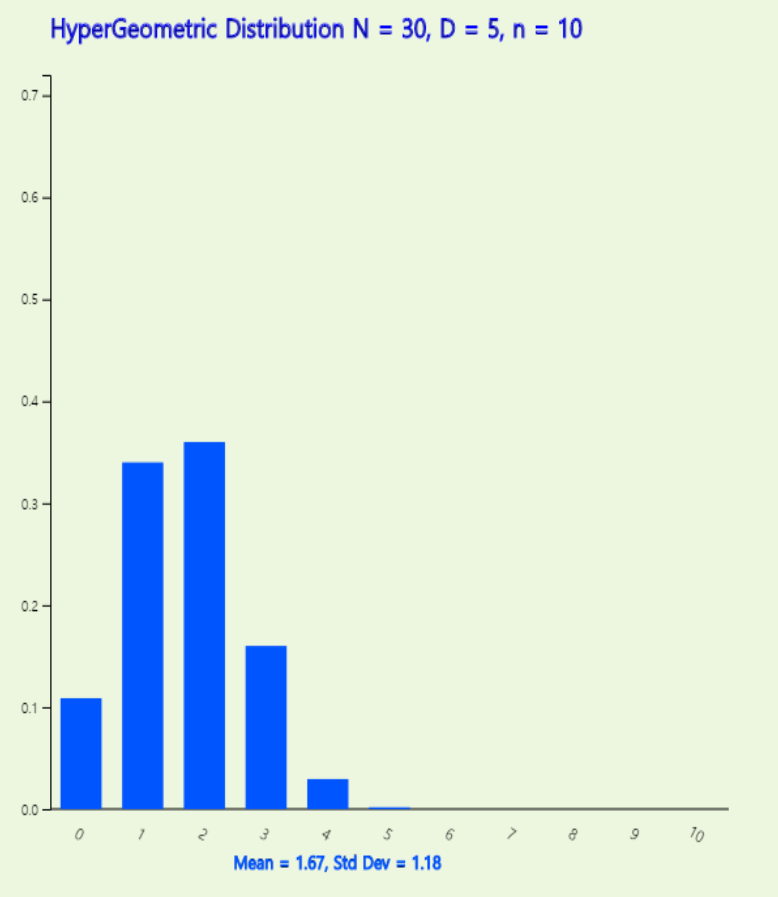
$$f(x) = \frac{{}_{D}C_{x} \times {}_{N-D}C_{n-x}}{{}_{N}C_{n}}, \; x = 0, 1, \ldots, n$$

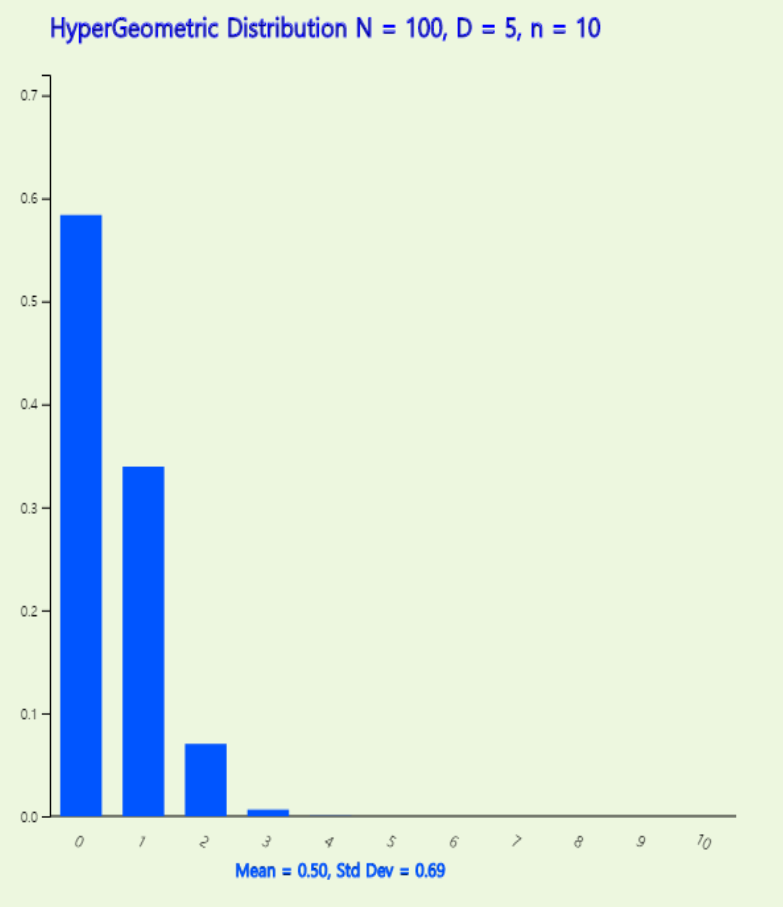
- If we let $p = \frac{D}{N}$,

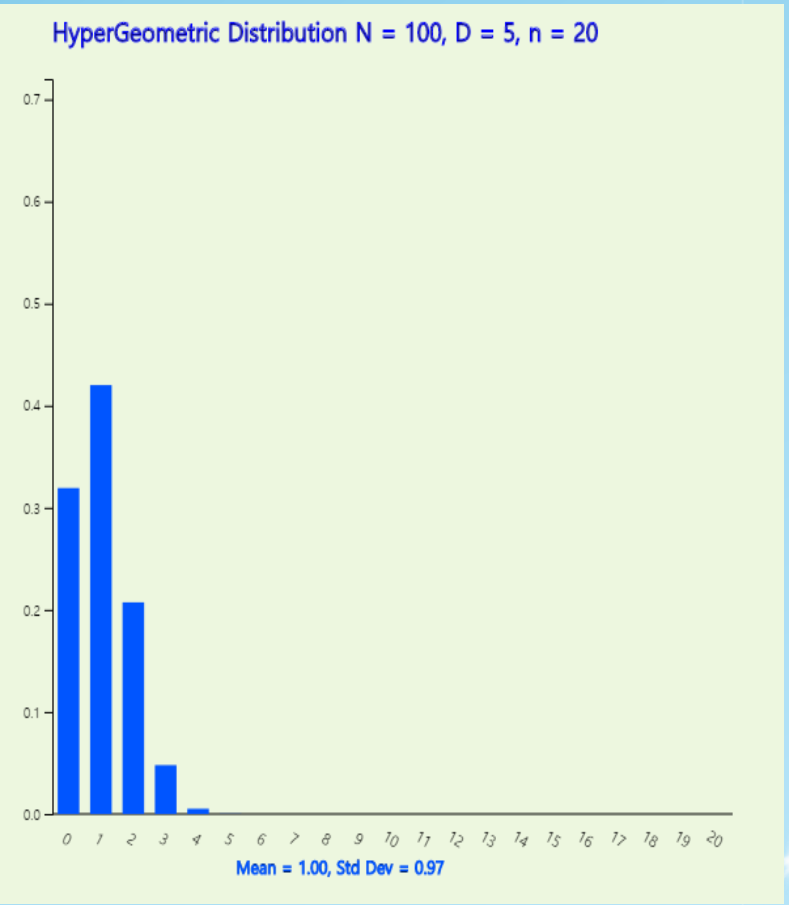
$$E(X) = np, \quad V(X) = np(1-p)\frac{N-n}{N-1}$$

5.3 Discrete Random Variable

5.3.4 HyperGeometric Distribution







5.3 Discrete Random Variable

5.3.4 HyperGeometric Distribution

[Ex 5.3.12] Sample of size 3 is selected from a box containing 20 tobacco products of which there are 15 normal products and 5 defective products. What is the probability of having one, two, or three defectives in the sample?

<Answer>

- These probability calculations have already been learned using combinations in section 5.1. This is the hypergeometric distribution with N = 20, D = 15, n = 3 , so it is as follows.

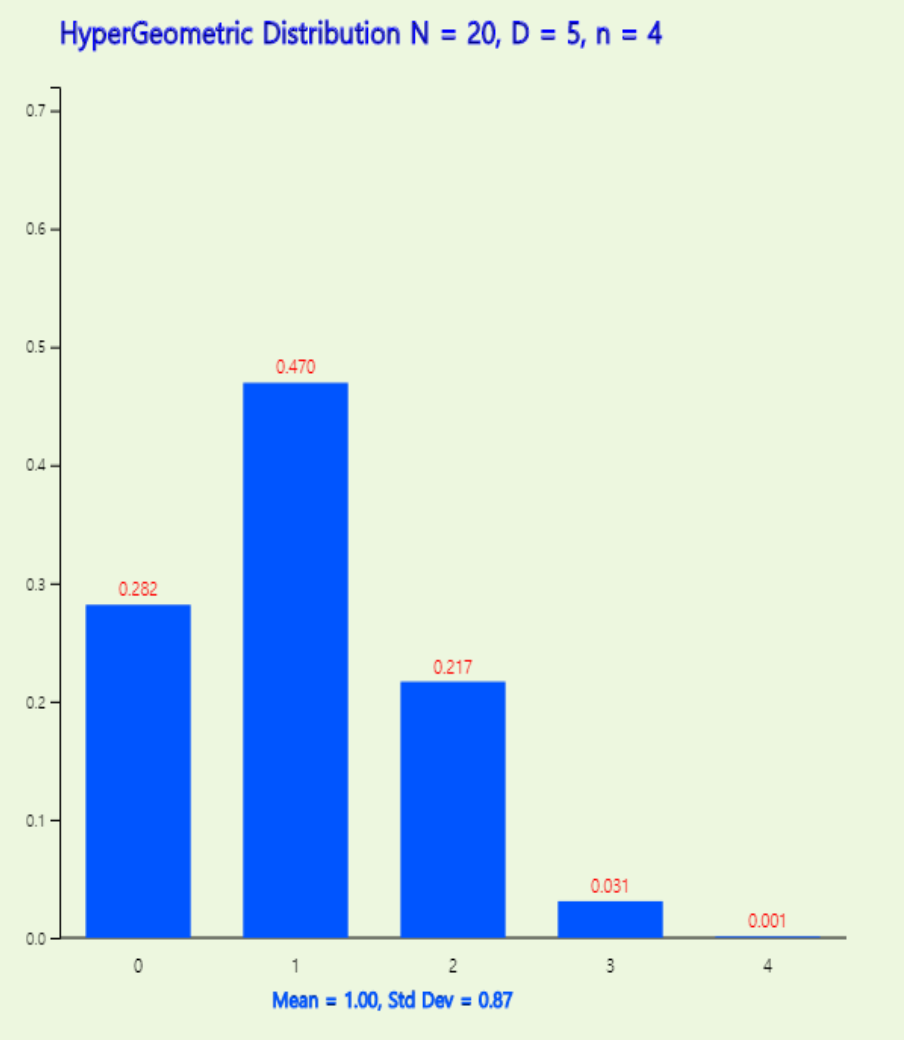
$$P(X=1) = \frac{{}_{15}C_2 \times {}_5C_1}{{}_{20}C_3} = \frac{15 \times 10}{1140} = 0.460$$

$$P(X=2) = \frac{{}_{15}C_1 \times {}_5C_2}{{}_{20}C_3} = \frac{105 \times 5}{1140} = 0.132$$

$$P(X=3) = \frac{{}_{15}C_0 \times {}_5C_3}{{}_{20}C_3} = \frac{455 \times 1}{1140} = 0.099$$

5.3 Discrete Random Variable

5.3.3 HyperGeometric Distribution



N = 20	D = 5	n = 3	
x	P(X = x)	P(X x)	P(X x)
0	0.3991	0.3991	1.0000
1	0.4605	0.8596	0.6009
2	0.1316	0.9912	0.1404
3	0.0088	1.0000	0.0088

Thank you