Hypergeometric Distribution

Esra Eryılmaz
171044046

26 June 2021

1 INTRODUCTION

1.1 DEFINITION

In probability theory and statistics, the **hypergeometric distribution** is a discrete probability distribution that describes the probability of $\boldsymbol{k}$ 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 **K** objects with that feature, wherein each draw is either a success or a failure.

1.2 PROPERTIES

- The hypergeometric distribution is discrete. It is similar to the binomial distribution. Both describe the number of times a particular event occurs in a fixed number of trials. However, binomial distribution trials are independent, while hypergeometric distribution trials change the success rate for each subsequent trial and are called "trials without replacement".

- Applications for the hypergeometric distribution are found in many areas, with heavy use in acceptance sampling, electronic testing, and quality assurance.

- Why is it called hypergeometric distribution?
  The hypergeometric distribution is so named because its probability generating function (PGF), i.e. the function whose coefficients are the probabilities, is a hypergeometric function.

- What is the hypergeometric distribution used for?
  The hypergeometric distribution can be used for sampling problems such as the chance of picking a defective part from a box (without returning parts to the box for the next trial).

1.3 Conditions

The hypergeometric distribution is used under these conditions :

- Total number of items (population) is fixed.
- Sample size (number of trials) is a portion of the population.
- Probability of success changes after each trial.

1.4 Formula

A random variable X follows the hypergeometric distribution if its probability mass function (pmf) is given by

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \tag{1.1}$$

where

- **N** is the population size,
- **n** is the number of draws
- **K** is the number of success states in the population,
- **k** is the number of observed successes.

(**A probability mass function** is a function that indicates that the probability of a discrete random variable is exactly equal to some value.)

2 PROBLEMS

2.1 PROBLEM 1

- **Assume that a bag contains 8 blue, 12 red balls,**

  a) What is the probability of getting red ball among 5 of the 8 selected balls, without replacement?

  $$P(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

  - N = 20 (Total 20 balls exist)
  - n = 8 (8 balls are drawn from the bag)
  - K = 12 (Since the red ball will be drawn, 12 red balls exist in bag)
  - k = 5 (Question asks the probability of 5 selected balls being red)

  $$P(X=5)=\frac{\binom{12}{5}\binom{20-12}{8-5}}{\binom{20}{8}}=\frac{\binom{12}{5}\binom{8}{3}}{\binom{20}{8}}=\frac{\frac{12!}{5!7!}\cdot\frac{8!}{3!5!}}{\frac{20!}{8!12!}}=\frac{792\cdot 56}{125970}\approx 0,352$$

  b) What is the probability of getting at most 1 red ball among selected 4 balls, without replacement?

  $$P(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

  - N = 20 (Total 20 balls exist)
  - n = 4 (Choose 4 balls from the bag)
  - K = 12 (Since the red ball will be drawn, 12 red balls exist in bag)
  - k = 0 and 1 (Question asks the probability of at most 1 selected balls being red)

  $$P(X\le 1)=[P(X=0)+P(X=1)]=\left[\frac{\binom{12}{0}\binom{20-12}{4-0}}{\binom{20}{4}}+\frac{\binom{12}{1}\binom{20-12}{4-1}}{\binom{20}{4}}\right]$$
  $$=\frac{1\cdot 70+12\cdot 56}{4845}\approx 0,153$$

References

[1] Walpole, Myers, Myers, Ye, Probability & Statistics for Engineers & Scientists - *9th Edition*

[2] Wikipedia: Hypergeometric distribution

[3] BUders : http://www.buders.com/

[4] Oracle : Hypergeometric Distribution