Hypergeomtric Distribution

Young W. Lim

2018-02-23 Fri

Outline

1. Hypergeometric Distribution
   - Based on
   - Examples
   - Assumptions

Based on

"Probability with R: An Introduction with Computer Science Applications"
Jane Horgan
https://en.wikipedia.org/wiki/Hypergeometric_distribution

Example (1)

The classical application of the hypergeometric distribution is sampling without replacement.
Think of an urn with two types of marbles, red ones and green ones.
Define drawing a green marble as a success and drawing a red marble as a failure (analogous to the binomial distribution).

- the variable N describes the number of all marbles in the urn (see contingency table below)
- K describes the number of green marbles
- then, N K corresponds to the number of red marbles.
- In this example, X is the random variable whose outcome is k, the number of green marbles actually drawn in the experiment

Example (2)

- the variable N describes the number of all marbles in the urn (see contingency table below)
- K describes the number of green marbles
- then, N K corresponds to the number of red marbles.
- In this example, X is the random variable whose outcome is k, the number of green marbles actually drawn in the experiment

	drawn	not drawn	total
green marbles	k	K - k	K
red marbles	n - k	N + k - n - K	N - K
total	n	N - n	N

Example (3)

- assume (for example) that there are 5 green and 45 red marbles in the urn.
- Standing next to the urn, you close your eyes and draw 10 marbles without replacement.
- What is the probability that exactly 4 of the 10 are green?
- Note that although we are looking at success/failure, the data are not accurately modeled by the binomial distribution,
- because the probability of success on each trial is not the same, as the size of the remaining population changes as we remove each marble.

	drawn	not drawn	total
green marbles	k = 4	K-k = 1	K = 5
red marbles	n-k = 6	N+k-n-K = 39	N-K = 45
total	n = 10	N-n = 40	N = 50

Assumptions

- The result of each draw (the elements of the population being sampled) can be classified into one of two mutually exclusive categories (e.g. Pass/Fail or Employed/Unemployed).
- The probability of a success changes on each draw, as each draw decreases the population (sampling without replacement from a finite population).