STA2001 Probability and Statistics I

Computer-based Exercise 4

Clement Leung

February 2021

The goal of this exercise consists of (i) to simulate and observe the outcome of the hypergeometric distribution using an urn model, and (ii) to see if the hypergeometric distribution can be approximated by the binomial distribution when the number of balls in the urn is large.

Problem

Consider an urn consists of 100 balls. Suppose half of these balls are red, and half are black. You are to randomly pick 5 balls from the urn. Let X be the number of red balls in such a sample. Then X has a hypergeometric distribution with pmf

$$f(x) = \frac{\binom{50}{x}\binom{50}{5-x}}{\binom{100}{5}} \tag{1}$$

1. Simulate this sampling experiment 1,000 times.
2. Produce a relative frequency histogram of X
3. Compare the actual observed values of X with the theoretical frequency obtained from equation (1) above by plotting the theoretical frequency in the same figure as the simulation results

You may use either Python or Matlab. In addition to the mechanisms for random number generation indicated in previous exercises, the plotting functions (such as plot.hist in Python, or plot in Matlab) may also be used.