Monte Carlo Simulations for Finite Populations

Dick Startz

November 2020

Doug Steigerwald and I are writing a paper on statistical inference when draws are taken from a finite population. (Most statistics assumes an infinite population, so that observations are not repeated across different samples.) We have extended what is known about this situation by applying theoretical analysis. To check our theory, we need a set of “Monte Carlo simulations.” The way such a simulation works is

1. Choose a set of parameters to control the simulation. (Often there may be a vector with a variety of values for one or more parameters.)
2. Generate a finite population according to a given statistical specification.
3. Draw a sample from the population. Generate the statistic of interest and record it. This is called a Monte Carlo draw.
4. Summarize the Monte Carlo draws and compare them to a pre-specified theoretical value.

In all this, documentation is very important because we may need to share the programs with other people several years down the road. Do the material should be completely self-explanatory without the need for discussion.

In general, a simulation will require five pieces.

1. A program in which one sets parameters.
2. A program to generate the finite population.
3. A program to draw a sample, compute and record the statistic of interest
4. A program to summarize the results.
5. An R Markdown file to present the results of the simulation. This program can either call the other programs, including (3) repeatedly, or there can be another program which does this and which is called from the R Markdown file.