Statistical Inference Course Project: Distribution Simulation and Basic Inferential Data Analysis - Part 1

Alejandro Montoya

May 8, 2017

Part 1: Simulation Exercise

Simulation

We need to simulate the event of generating multiple exponential distributions (in this case they have to be 1,000) with 40 exponentials each, in order to be able to compare population estimated statistics (mean and variance) from samples, vs. the theoretical values

Let's start by creating the simulation and storing it into a matrix:

Sample Mean versus Theoretical Mean:

Show the sample Mean and compare it to the theoretical mean of the distribution.

We have to calculate the mean of the exponentials for each one of the 1,000 simulations, and then we can estimate the mean of the distribution by calculating the mean of the simulations' means:

Table 1: Mean Comparison

Statistic	Simulation.Value	Theoretical.Value	Difference
Mean	5.001	5	-0.001

The difference between the simulated (5.001) and the theoretical (1 / 0.2 = 5) means of the exponential distribution with lambda = 0.2 is -0.001… Since they are that close, we can safely say that the simulation of multiple exponential distributions can accurately estimate the population.

Sample Variance versus Theoretical Variance:

Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

Since we already have the vector with the means of the 1,000 simulations, we can now use the R function var to find the variance between them and then compare it with the theoretical one, which is given by the formulae Var(X) = s^2 / n, with s = 1 / lambda:

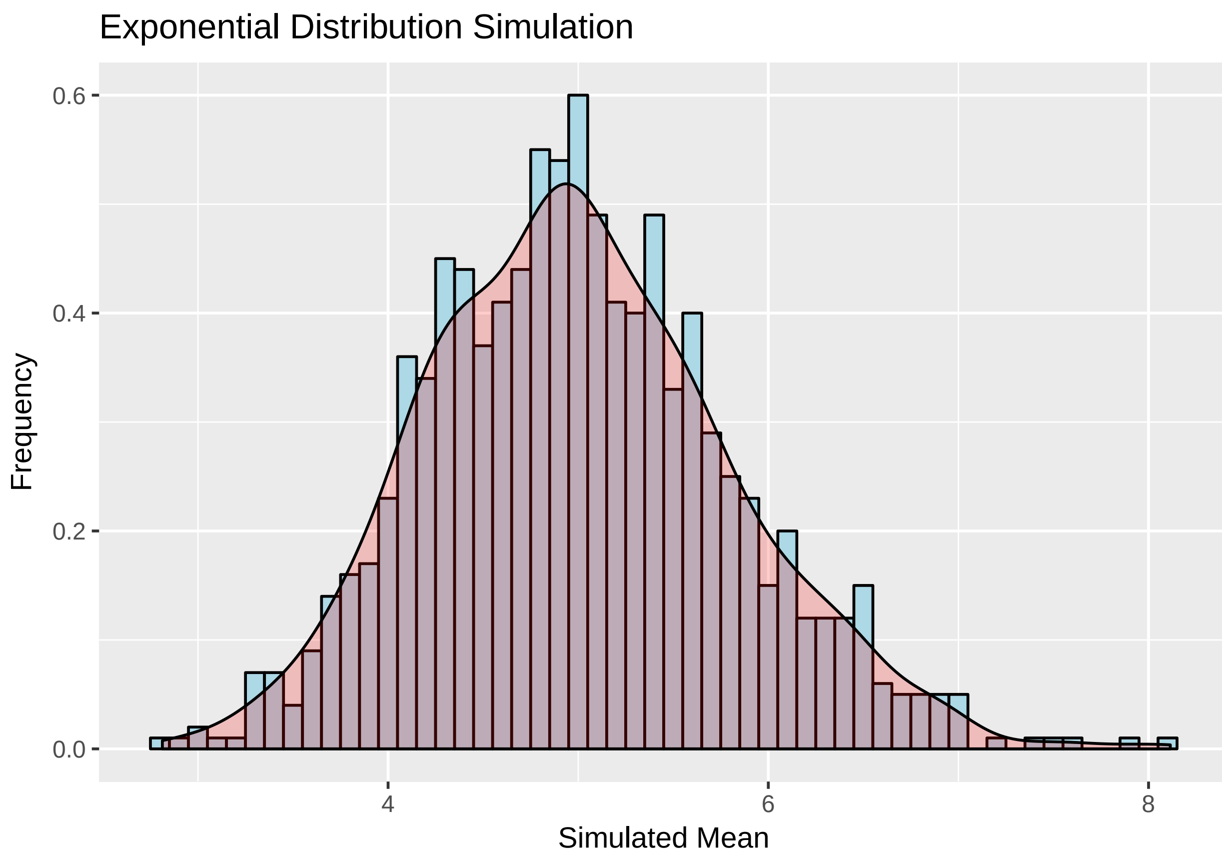
Table 2: Variance Comparison

Statistic	Simulation.Value	Theoretical.Value	Difference
Variance	0.637	0.625	0.012

Distribution

Show that the distribution is approximately normal.

Now, let's plot a histogram of the simulation with a line on the theoretical mean:



As we can see with this histogram and the density plot, the simulation distribution is aproximately normal with the following statistics:

Table 3: Distribution Statistics

Statistic	Simulation.Value	Theoretical.Value
Mean	5.001	5
Variance	0.637	0.625
Standard Deviation	0.798	0.791
95th Quantile	6.314	6.3
95th Quantile in Std Devs	1.645	1.645
97.5th Quantile	6.565	6.549
97.5th Quantile in Std Devs	1.96	1.96
25th Quantile	4.463	4.467
25th Quantile in Std Devs	-0.674	-0.674

Appendix

The R Code used to generate all these results is located at Exponential Distribution Simulation