

Simulation of the Exponential Distribution

Yolande Korsten

Friday, August 22, 2014

Simulation of the exponential distribution to illustrate the tenet of the **Central Limit Theorem** that states *that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases*.

Note that the `echo = FALSE` parameter has been added to the code chunks to prevent printing of the R code unless absolutely necessary. To view the details of the code that has generated the simulated variables, please view R Markdown file that can be found here: https://github.com/Chaendryn/StatsInf_Project/blob/master/simQ1.Rmd

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

The theoretical mean σ (center of the distribution) for an exponential distribution is $\frac{1}{\lambda}$. For the purposes of this simulation $\lambda = 0.2$. The sample population (n) for calculation per simulation is set to 40, and the initial number of simulations will be set to 100. The theoretical variance for this simulation is calculated as $\frac{\sigma^2}{n}$.

Given the above assumptions, the theoretical variables for this simulation are:

Mean - 5

Standard Deviation - 5

Variance - 0.625

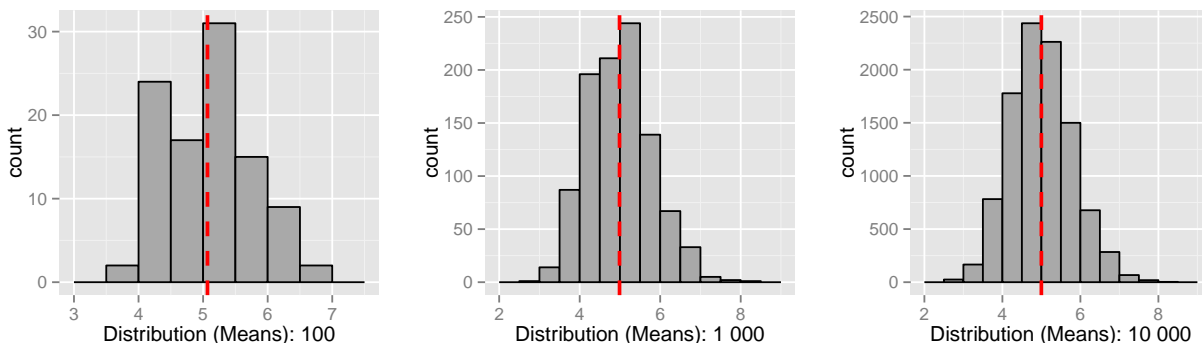
Running a simulation of only 100 iterations, the observed center of the distribution is **5.0676** which is slightly higher than the expected **5**. Running the simulations for 1000 iterations produces **4.9897** and for 10 000 iterations, **4.999**. As the number of iterations increases, the observed centre of the distribution of means approaches the expected theoretical center of the distribution of 5.

2. Show how variable it is and compare it to the theoretical variance of the distribution.

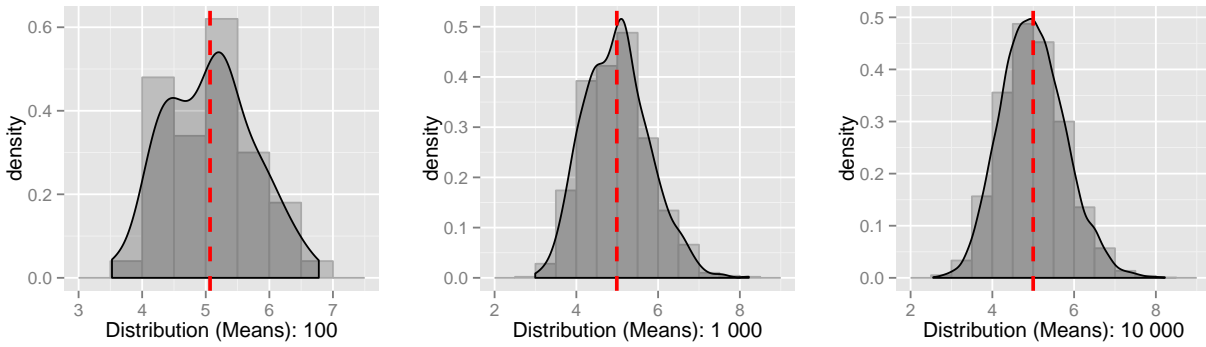
The theoretical variance of the exponential distribution is **0.625**. The 100 iteration variance is **0.4774**, which is well below the theoretical variance. But increasing the number of sample iterations to 1000 and 10000, and running the simulation, the sample variance becomes **0.6182** and **0.6189** respectively. One can see that as the number of iterations increase, the observed variance approaches the theoretical variance.

3. Show that the distribution is approximately normal.

Plotting the means of the three simulations of 100, 1 000 and 10 000 iterations, the frequency distribution progressively resembles a normal distribution.

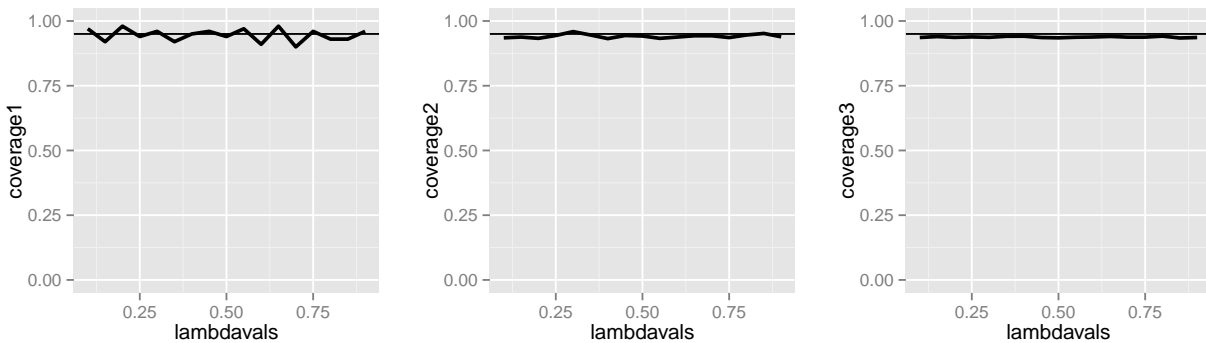


Plotting the distribution density of the simulations illustrates that a sufficient number of samples will eventually result in a distribution of means that approximates the normal distribution more closely.



4. Evaluate the coverage of the confidence interval for $\frac{1}{\lambda}$: $\bar{X} \pm 1.96 \frac{S}{\sqrt{n}}$.

When simulating samples from a known distribution, a good test is to see whether your sample means for your simulated sample populations fall within the 95% confidence interval for the actual population mean. Below, I plot the confidence interval coverage for 100, 1 000 and 10 000 simulations.



For 100 simulations, there is a more than 95% chance that the expected population mean μ of 5 will be found in the the distrubition of sample means, though there is a lot of variability. On simulating 1 000 and 10 000 samples, the variability is reduced and the confidence that the simulated (estimated) mean of your sample population will be close to the true population mean approaches 95%.