Exponential Simulation Distribution with R

Jesus A. Gonzalez

Saturday, May 16, 2015

Overview

The exponential distribution is used to answer questions related to how much time we need to wait for an event to happen. When we do not know when such an event will happen, we have a random variable that follows an exponential distribution. In this report we show, through simulation, how this distribution behaves according to the Central Limit Theorem (CLT).

Simulations

In this assignment we simulate the exponential distribution in R. In our simulation we run 1,000 sets of 40 exponential experiments. For each set of 40 exponentials, we calculate the mean and variance and create a data frame containing our 1,000 values for means and variances. For all the experiments we used a value of lambda of 0.2 for the exponential calculations. The code used to run our simulations is shown in the following chunk of code. The result of our simulations is stored in the *simdata* variable.

```
    nosim <- 1000
    rate <- 0.2
    simMean = NULL
    simVar = NULL
    for (i in 1 : 1000)
    {
      d <- rexp(40, rate)
      simMean <- c(simMean, mean(d))
      simVar <- c(simVar, var(d))
    }
    simdata <- data.frame(simMean, simVar, size = factor(rep(c(40),
rep(nosim, 1))))
```

Sample Mean versus Theoretical Mean

Now we compare the sample mean of our simulations with the theoretical value of the mean of the distribution. The theoretical mean is defined as 1/lambda. Since lambda $= 0.2$, the theoretical mean is equal to $1/0.2 = 5$.

The mean of our sample of 1,000 averages of 40 exponentials using lambda $= 0.2$ is equal to 5.0164927. If we compare the theoretical mean to the simulation mean, it is very similar as we expected because the sample size was large. This means that with a large enough sample, the mean of our simulated exponential distribution has a *normal* behavior as the CLT states. We can see this in the plot shown in figure 1, where the mean of our sample mean estimate follows a normal distribution with a mean of approximately 5. As we can see, the mean is centered at the value of **5**. We can also appreciate a normal distribution over the plot created for the sample mean, which adjusts very well to the distribution of the sample mean plot.

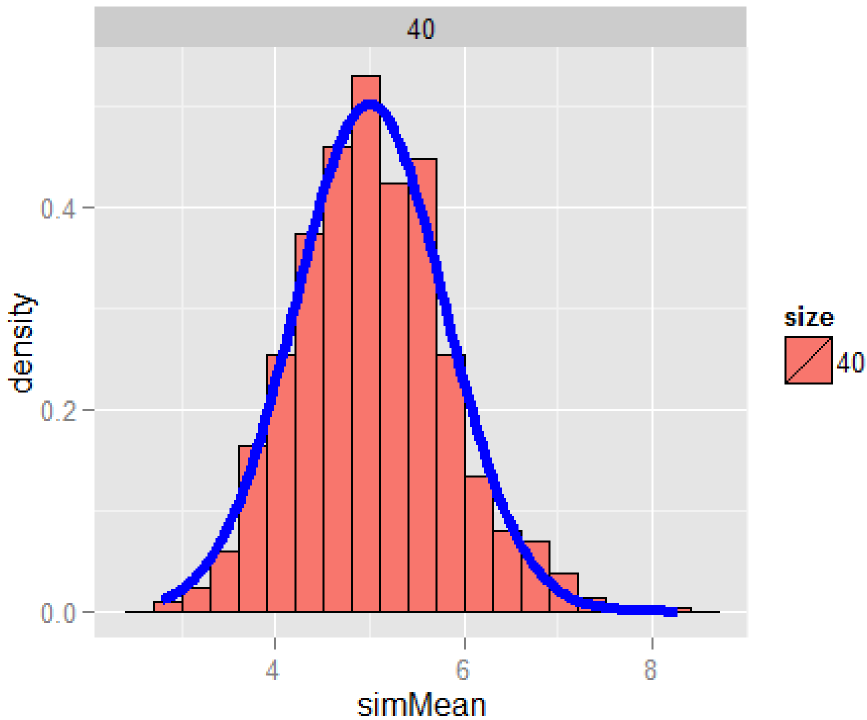

Figure 1. Plot of the means of 1,000 simulations of the exponential distribution with lambda = 0.2. Each simulation consists of the mean of 40 experiments.

Sample Variance versus Theoretical Variance

The theoretical variance of our exponential distribution with lambda $= 0.2$ is equal to $1/(0.2^2) = 25$ but the variance for our experiments is 119.3747197, which is far from the theoretical value. Then, the value of the sample variance is much higher than the value of the theoretical variance. We can see this in figure 2, in which we can also observe a normal distribution with mean $= 5$ and variance $= 25$. However in this same plot we can appreciate that the mean of the sample variance for our 1,000 experiments is centered very close to the value of 25, which is the theoretical variance.

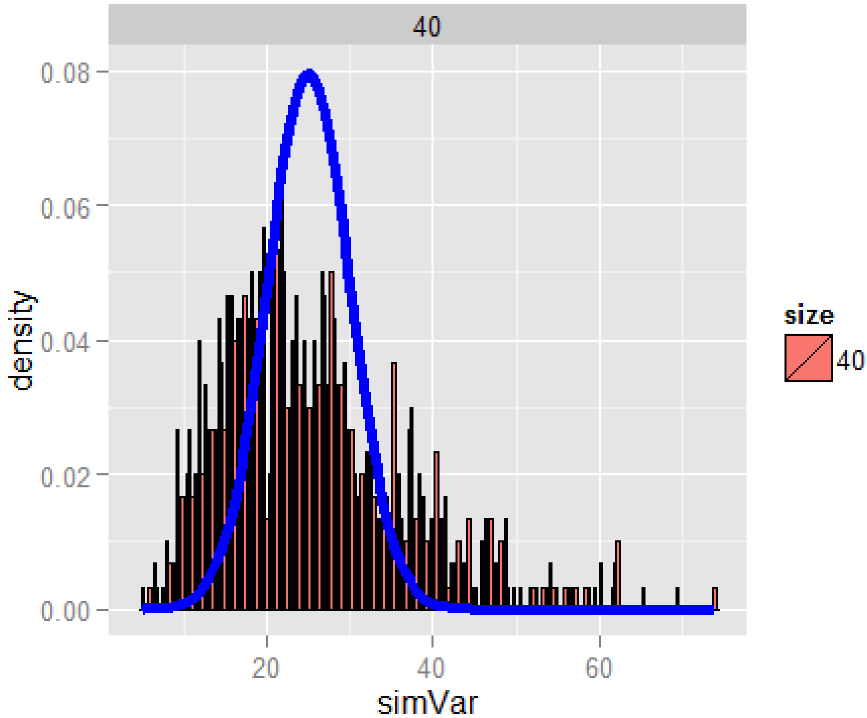

Figure 2. Plot of the variance of 1,000 simulations of the exponential distribution with lambda = 0.2. Each simulation consists of the variance of 40 experiments.

Distribution

As we can see from figures 1 and 2, the distribution of the sample mean and sample variance approximate a normal distribution. In these figures we appreciate a normal distribution curve with mean and standard deviation equal to the theoretical mean and standard deviation of the exponential distribution. This is better appretiated in figure 1 created for the sample mean. On the other hand, the distribution of the sample variance also approximates a normal distribution (the blue curve) and is centered very close to the theoretical variance with a value of 25.