Simulation of Exponential Distribution Approaching Normal Distribution Using R

Tyler Peterson

10/25/2016

Overview

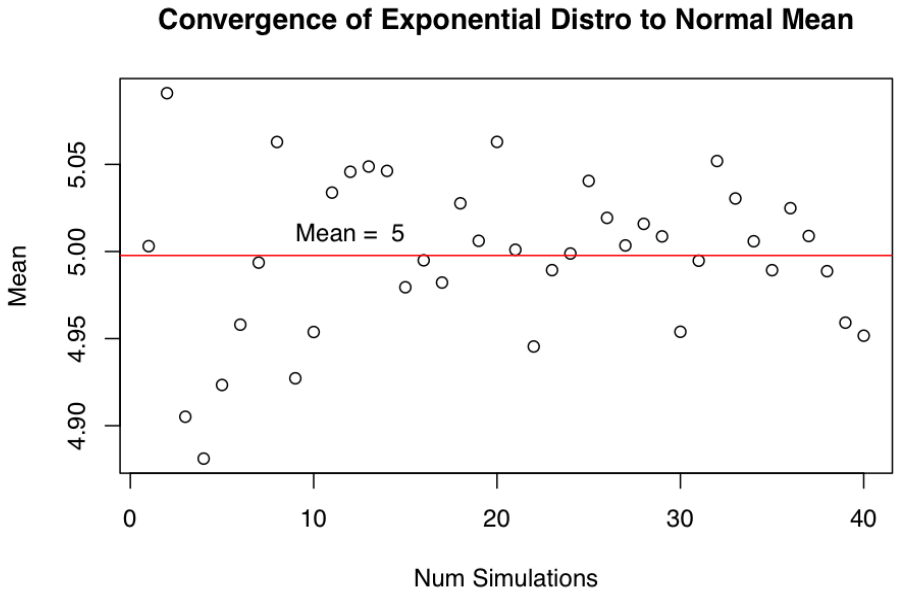
In this project, we investigate the exponential distribution and, using the Central Limit Theorem, show that it can be simulated with a normal distribution for a large sample size. We will use the mean and standard deviation of 5 for the exponential distribution. We will also use a standard normal distribution with a mean and standard deviation of 5 as a baseline.

Simulate The Exponential Distribution

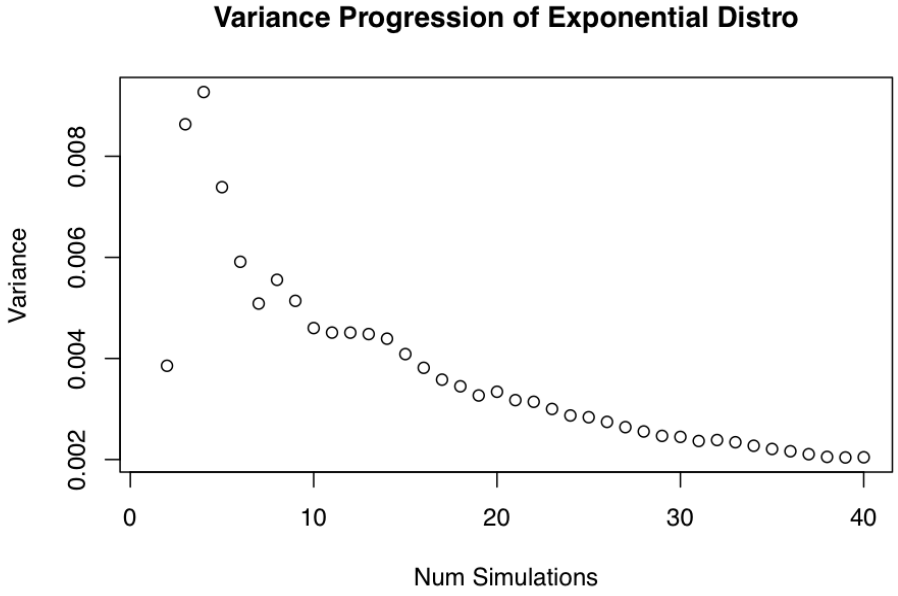
For our simulation we will show the progression of the sample mean as we generate 1000 simulations over 40 exponentials. Following we simulate the distributions of means from 1 to 40 exponentials.

```
# Show convergence progression to the normal distribution over 40 simulations
  overallMean <- c()
  overallSD <- c()
  overallVar <- c()
  for(i in 1:n){
    expData <- rexp(n = i*numSims, rate = lambda)
    expDataMatrix <- matrix(expData, nrow = numSims, ncol = i)
    expDataMeanArray <- apply(expDataMatrix, 1, mean)
    overallMean <- c(overallMean, mean(expDataMeanArray))
    overallSD <- c(overallSD, sd(overallMean))
    overallVar <- c(overallVar, var(overallMean))
  }
```

We plot the simulation here along with the overall mean value to illustrate the ultimate convergence of the exponential distribution to 5.



Notice the value for the mean of the overall progression is 5. Also notice that the distribution of the points moving left to right seem to be converging to the mean as the number of samples increases. In other words, as we progress from left to right, the distance of the plot points above and below the mean line decreases. Indeed, as we examine the variance of the distribution through the following plot



Notice that the variance of the distribution decreases through each simulation and appears to converge as we

approach 40 simulations. This indicates that the distribution becomes more centered around the mean as the number of simulations grows - consistent with the Central Limit Theorem.

```
# Show distribution of means over final simulation:
  finalExpData <- rexp(n = n*numSims, rate = lambda)
  finalExpDataMatrix <- matrix(finalExpData, nrow = numSims, ncol = n)
  finalExpMeanArray <- apply(finalExpDataMatrix, 1, mean)
  finalExpVarArray <- apply(finalExpDataMatrix, 1, var)
  finalExpMean <- mean(finalExpMeanArray)
  finalExpSD <- sd(finalExpMeanArray)
  finalExpVar <- var(finalExpMeanArray)
  standError <- finalExpSD/sqrt(n)
  upperCI <- finalExpMean + 1.96*standError
  lowerCI <- finalExpMean - 1.96*standError
```

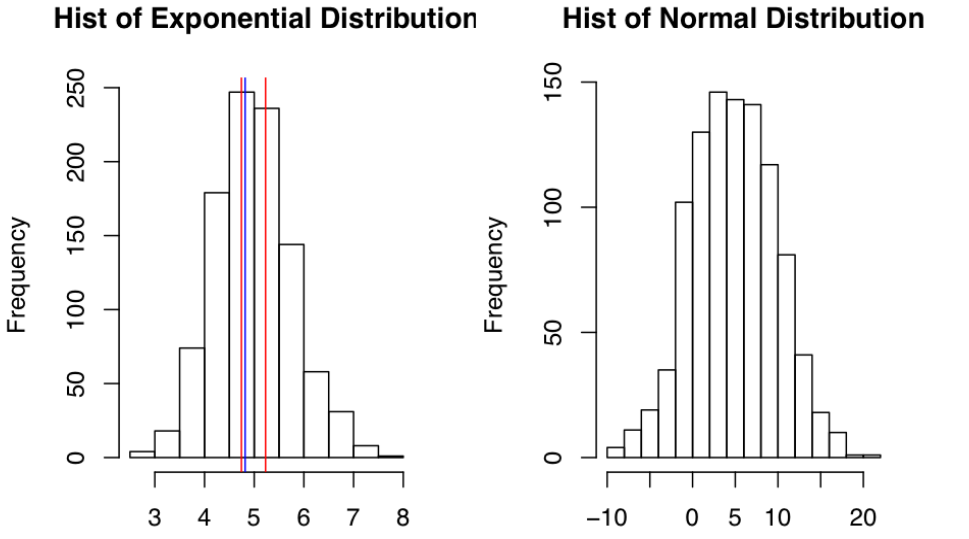
As we model the 40th exponential, we can see that the mean closely approximates the true mean of 5.0 with a value of 4.9858178.

Simulate The Normal Distribution

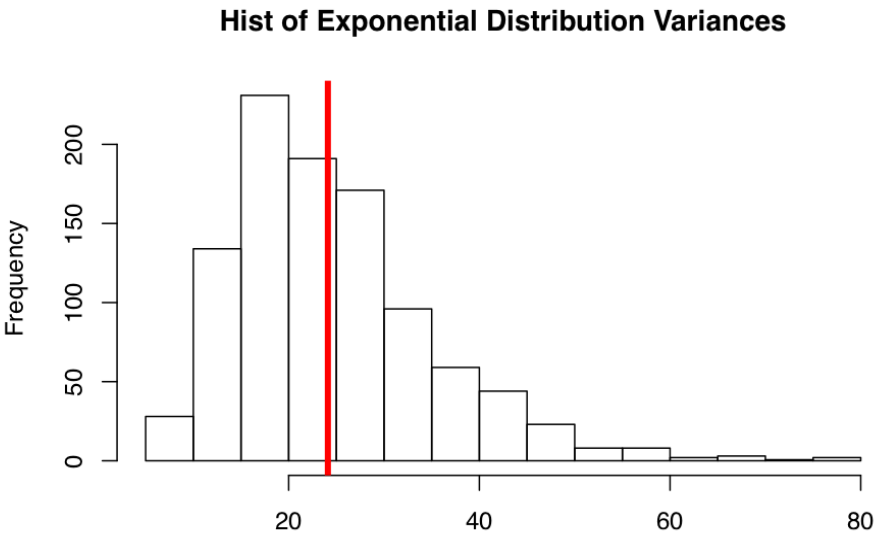
As a baseline, we simulate the normal distribution over 1000 random uniforms:

```
# Normal distribution
  normDist <- rnorm(n = numSims, mean = 1/lambda, sd = 1/lambda)
  normDistMean <- mean(normDist)
  normDistSD <- sd(normDist)
  normDistVar <- var(normDist)
```

Notice that the mean of this distribution is 4.8175763. This mean is very close to the mean simulated in the exponential simulation. We plot these distributions side by side to show that they both appear to be distributed normally around their respective means.



Furthermore, as we explore the variance between the exponential distribution and the normal distribution, we can see that the variance of the normal distribution 24.1106563 lies very close to the expected value of the exponential variance simulated distribution 24.6566176.



Conclusion

Notice that the distributions are very similar in shape. Also notice that the red and blue lines in the Exponential Distribution. The red lines indicate a 95% upper and lower confidence interval. The blue line indicates the true mean of the normal distribution. Notice that the normal distribution's mean lies within the 95% confidence interval of the exponential mean's distribution. Our conclusion is therefore that the means of the exponential distribution begin to resemble that of the normal distribution as the number of simulations grows.