Statistical Inference- Project Part 1

Jeff Sheremata

October 24, 2015

Overview

In this project the exponential distribution is studied and compared to the Central Limit Theorem. 1000 simulations of the distributions of 40 exponentials were performed. Values of the variance and mean calculated from the simulations were compared to values calculated from the normal distribution.

Simulations

Generation of Simulated Exponential Values

Step 1 - Define the Distribution and Simulation Parameters

```
library(ggplot2)
lambda =0.2
n=40
numSamples = 1000
set.seed=300

#These are the theoretical values of the mean, standard deviation, and variance
theoretical_mean = 1/lambda
theoretical_standarddeviation= ((1/lambda) * (1/sqrt(n)))
theoretical_variance = (1/lambda)^2/n
```

Step 2 - Generate the Simulated Data and Calculate the Mean, Standard Deviation, and Variance

The replicate function is used to run the simulations. n samples from an exponential distribution are randomly generated and stored in a matrix. The mean, standard deviation, and variance of the simulations are then determined.

```
rep<-replicate(numSamples, rexp(n, lambda))
#this calculates the mean of each simulation
means<-colMeans(rep)
#this calculates the mean of all of the means
samples_mean = mean(means)
#this calculates the standard deviation
samples_standarddeviation<-sd(means)
#this calculates the variance
samples_variance = var(means)
```

Step 3. Compare the Mean and Variance of the Simulations To Theoretical Values.

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

The mean of the simulations is 5. The theoretical mean is 5.007112. Thus, there is very close agreement between the center of the simulated values and the center of the theoretical distribution.

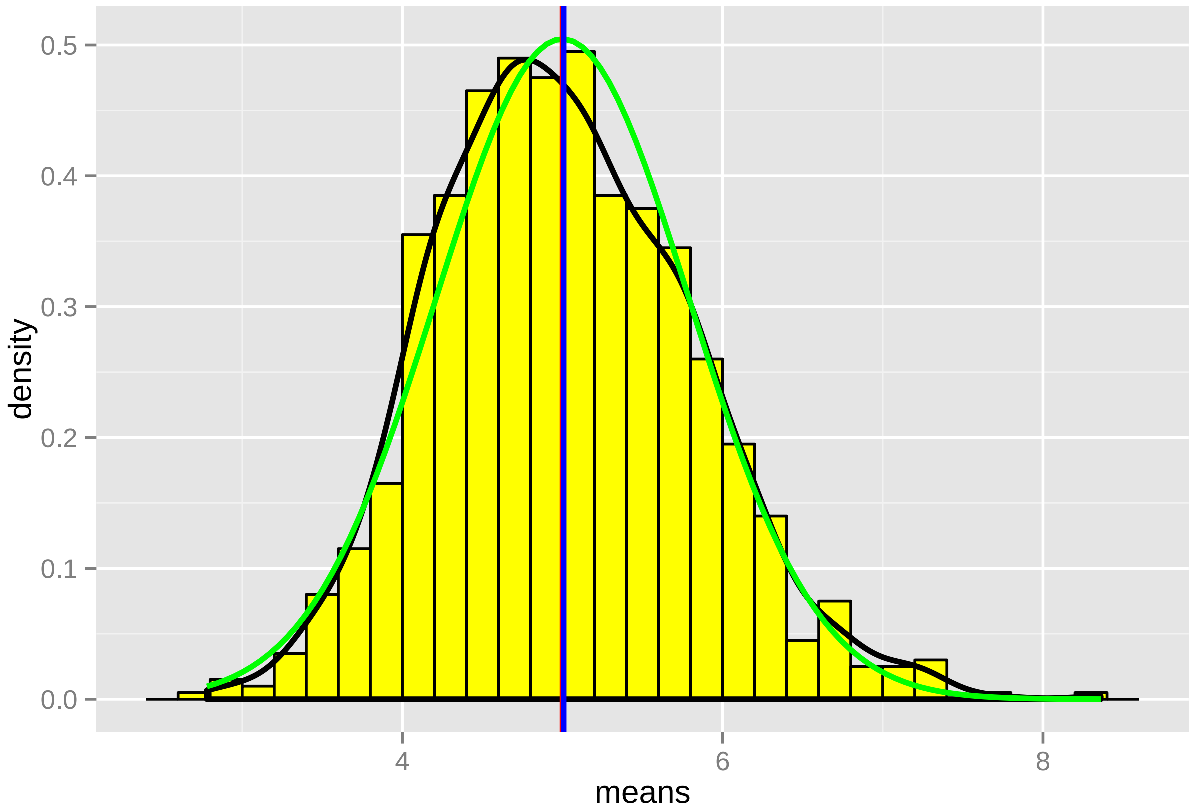
b. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

The variance of the simulations is 0.6389222. The theoretical variance is 0.625. Thus, there is very close agreement between the variance of the simulated values and the variance of the theoretical distribution.

c. Show that the distribution is approximately normal.

Let's start by looking at a histogram of the data and comparing it to a normal distribution.

```
simdata <- data.frame(means);
simd <- ggplot(simdata, aes(x =means))
simd <- simd + geom_histogram(binwidth=lambda, aes(y=..density..), colour="black", fill = "yellow");
simd <- simd + geom_density(colour="black", size=1);
simd <- simd + stat_function(fun=dnorm,args=list(mean= theoretical_mean, sd=theoretical_standarddeviati
simd<-simd + geom_vline(xintercept=theoretical_mean,size=1.0, color="red")
simd<-simd + geom_vline(xintercept=samples_mean,size=1.0, color="blue")
simd
```



The histogram of the exponential simulated data is represented by the plot in yellow, while a smooth density estimate is in black. The center of the simulated data is the blue line. The normal distribution is the green plot, while the center of the normal data is the red line.

This plot clearly shows that the centers of the exponential simulated and normal data are very closely matched. Likewise, the plot also clearly shows that the distrubiton of the exponentials is approximately normal. From looking at these plots, it is apparent that the variances of the simulated and normal data are similar.

Step 4. Compare Confidence Intervals of the Simulated and Normal Values

```
simulated_conf_interval <- round (mean(samples_mean) + c(-1,1)*1.96* samples_standarddeviation/sqrt(n),4
theoretical_conf_interval <- theoretical_mean + c(-1,1)*1.96* theoretical_standarddeviation/sqrt(n);
simulated_conf_interval
```

```
## [1] 4.7594 5.2548
```

```
theoretical_conf_interval
```

```
## [1] 4.755 5.245
```

The 95% confidence interval of the exponential simulated values is 4.7594, 5.2548. This is very similar to the 95% confidence interval of the normal distribution of 4.755, 5.245.

Conclusion. The data in this study clearly indicate that the simulated exponentials are aproximately normal.