Statistical Inference: Project 1 Part 1

JR Sturino

November 17, 2015

Overview

The task is to investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations.

Simulations

Run 1000 simulations using the parameters in the overview section. The following code loops through each of our simulations and obtains the mean:

```
startSimulations = 1
endSimulations = 1000
lambda <- 0.2
n <- 40

set.seed(1)
simulationMeans <- sapply(startSimulations:endSimulations, function(x)
{mean(rexp(n, lambda))})
head(simulationMeans)
```

```
## [1] 4.860372 5.961285 4.279204 4.702298 5.196446 4.397114
```

Sample Mean versus Theoretical Mean

Sample Mean

Obtain the mean for our data set from the simulation

```
mean(simulationMeans)
```

```
## [1] 4.990025
```

Theoretical Mean

We calculate the theoretical mean using 1/lambda

```
(1/lambda)
```

```
## [1] 5
```

Conclusion

In this case the sample and theoretical mean are quite similar.

Sample Variance versus Theoretical Variance

Sample Variance

Obtain the variance from the data set, we do this by obtaining the standard deviation and squaring it:

```
sd(simulationMeans)^2
```

```
## [1] 0.6111165
```

Theoretical Variance

We obtain the theoretical variance by using 1/lambda/sqrt(n) and squaring the result:

```
((1/lambda)/(sqrt(n)))^2
```

```
## [1] 0.625
```

Conclusion

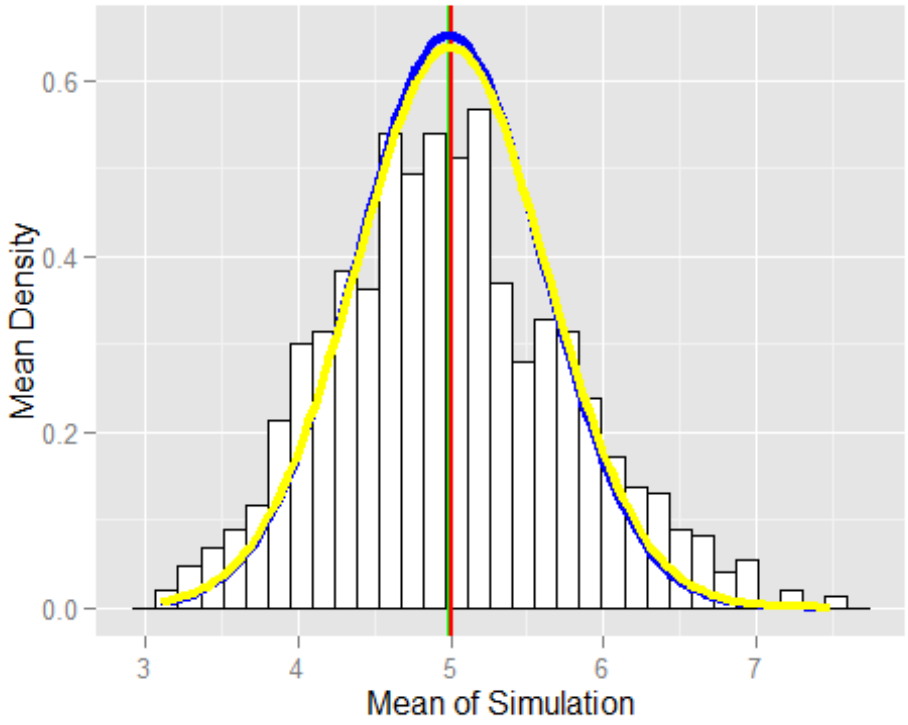
Like the mean, the variance is also very close.

Distribution

In the below graph we plot two lines vertically; the red line plots the theoretical mean while the green plots the sample mean.

```
library(ggplot2)
meanData <- data.frame(x=simulationMeans)
tVar <- ((1/lambda)/(sqrt(n)))^2

ggplot(meanData, aes(x = simulationMeans)) +
  geom_histogram(aes(y=..density..), color="black",fill="white") +
  xlab("Mean of Simulation") +
  ylab("Mean Density") +
  geom_vline(xintercept=mean(meanData$x), color="green", lwd=1) +
  geom_vline(xintercept=(1/lambda), color="red", lwd=1) +
  stat_function(fun=dnorm, args=list(mean=mean(meanData$x),
sd=sd(meanData$x)^2), color="blue", lwd=1.5) +
  stat_function(fun=dnorm, args=list(mean=1/lambda, sd=tVar), color="yellow",
lwd=1.5)
```



Looking at the graph, the blue line represents the sample distrubtion and the yellow line represents the theoretical distrubtion. As you can see they both follow the normal distrubtion.