

Statistical Inference Course Project, Part 1

Tiffany Stoeke

September 19, 2015

Overview:

In this project we investigate the exponential distribution in R and compare it with the Central Limit Theorem. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$, and we have set $\lambda = 0.2$ for all simulations. We investigate the distribution of averages of 40 exponentials with a thousand simulations.

Simulations

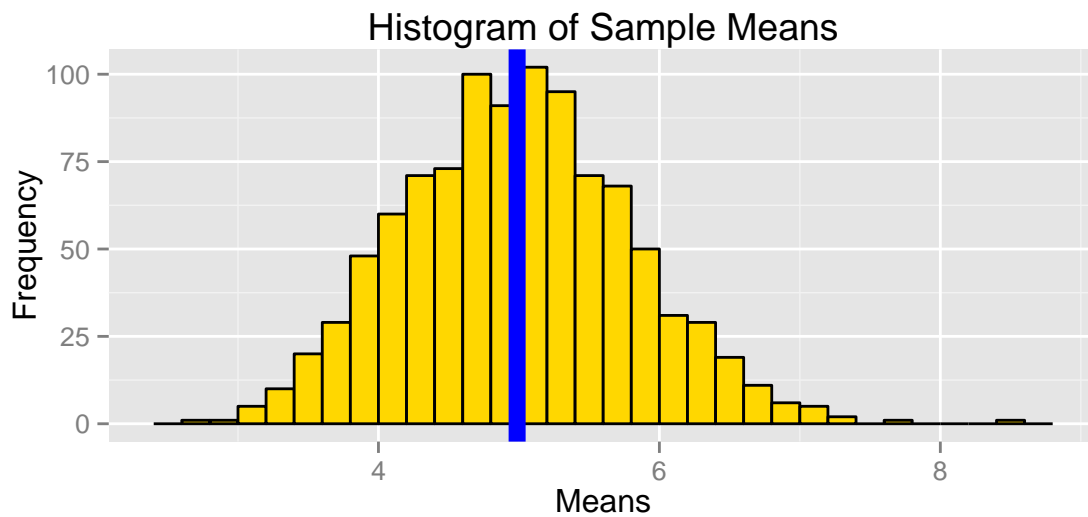
The below code gathers the averages of the 40 exponentials and plots them in a histogram, adding a blue vertical line to show the mean of the means. This will be the starting point of our investigation. Ensure loading ggplot before attempting to run the code.

```
library(ggplot2)

##Use set.seed() command so that the set of random numbers generated in R is reproduceable.
set.seed(11)
lambda<-0.2
n<-40
nosim<-1000

##Creates a matrix of the means of the 40 exponentials
simmeans<-apply(matrix(rexp(nosim * n, lambda), nosim), 1, mean)

##Plots a histogram showing the density of the means of the 40 exponentials,
###with vertical blue line showing the mean of the data.
simmeansdf<-as.data.frame(simmeans)
g<-ggplot(simmeansdf,aes(simmeans))+geom_histogram(fill="gold",color="black",binwidth=.2)
g<-g+geom_vline(xintercept=mean(simmeans),size=3,color="blue")
g<-g+xlab("Means")+ylab("Frequency")+ggtitle("Histogram of Sample Means")
g
```



As we can see from the above plot, the distribution appears to be approximately normal (consistent with the Central Limit Theorem) with a mean of around 5, but we will continue our exploration of the data to prove the point.

Question 1: Sample Mean vs. Theoretical Mean

The following code calculates the actual and theoretical means of our data. According to theory, the mean of our data should be $1/\lambda = 1/0.2 = 5$.

```
actmean<-mean(simmeans)
actmean
```

```
## [1] 4.987157
```

```
theomean<-1/lambda
theomean
```

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

Success! The actual mean of our data is approximately 5, which matches our theoretical mean.

Question 2: Sample Variance vs. Theoretical Variance

Let's see if the equivalency holds true for the actual and theoretical variances. According to theory, the variance of our data should be $(1/\lambda)^2/n = (1/.2)^2/40 = 0.625$.

```
varsim<-var(simmeans)
varsim
```

```
## [1] 0.6509313
```

```
vartheo<-(1/lambda)^2/n
vartheo
```

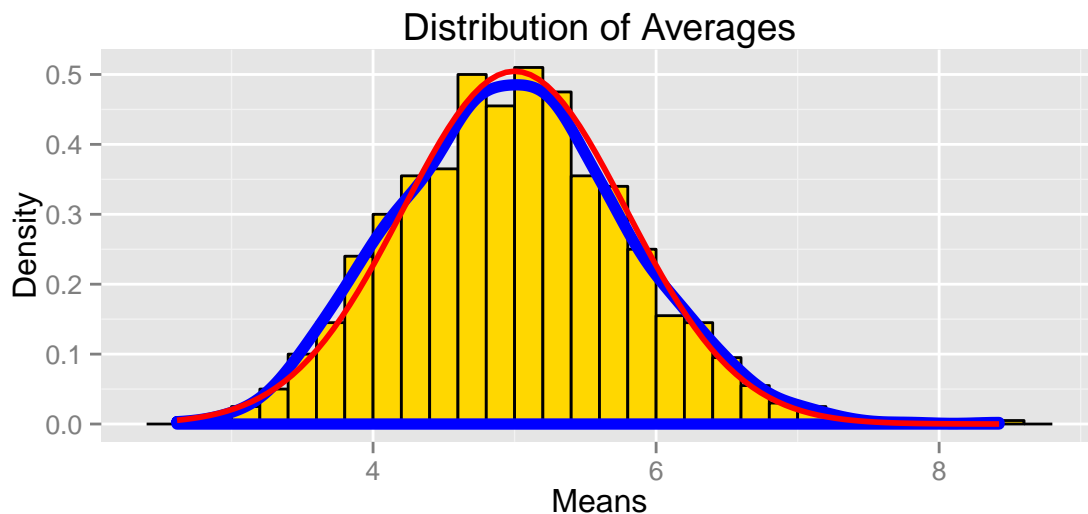
```
## [1] 0.625
```

Another success! The actual variance of our data is close to 0.625, which is our theoretical variance.

Question 3: Distribution

The below plot will finalize our investigation of the exponential distribution in R compared with the Central Limit Theorem. Here we plot another histogram showing the density computed using the simulated data and the normal density plotted with our theoretical mean and variance values.

```
##Plots a histogram showing that the distribution is approximately normal.
##Blue line = simulated distribution, red line = normal distribution.
g2<-ggplot(simmeansdf, aes(x=simmeans))
g2<-g2+geom_histogram(aes(y=..density..),fill="gold",color="black",binwidth=0.2)
g2<-g2 + xlab("Means") + ylab("Density") + ggtitle("Distribution of Averages")
g2<-g2+geom_density(color="blue",size=2)
g2<-g2+stat_function(fun = dnorm, colour = "red",size=1,arg=list(mean=theomean,sd=sqrt(vartheo)))
g2<-g2+scale_colour_manual("Legend",breaks=c("Simulation","Normal"),values=c("blue","red"))
g2
```



In this plot, the blue line shows the distribution of our simulated data and the red line shows the normal distribution. As we can see from the plot, our simulated data is approximately normally distributed, which is consistent with the Central Limit Theorem.