Statistical Inference Course Project 1

Eric Scuccimarra

4 January 2018

Overview

In this project I will 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. I will investigate the distribution of averages of 40 exponentials over 1,000 simulations.

Simulating the Data

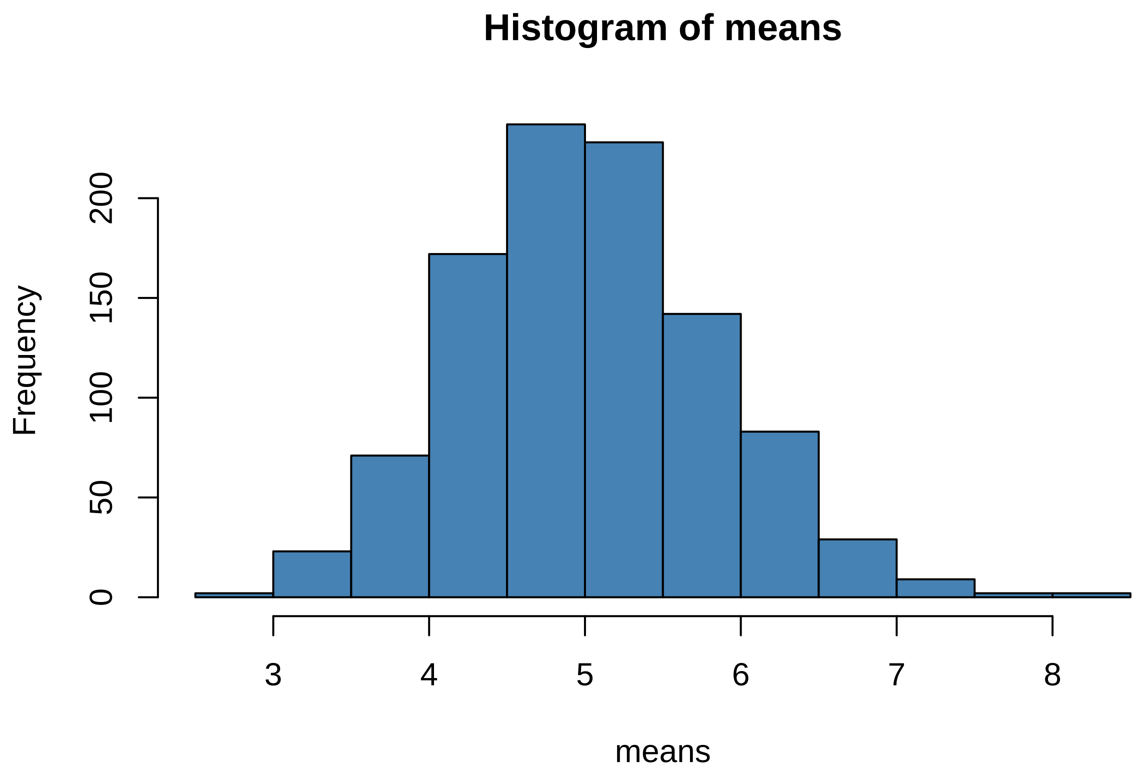
First we simulate the data and create a vector of the mean for each simulation.

```
set.seed(123455)
n <- 40
simulations <- 1000
lambda <- 0.2
data <- matrix(rexp(n * simulations, lambda),nrow=simulations)
means <- apply(data,1,mean)
```

Sample Mean

The theoretical mean is $1/\lambda$, or 5. Looking at this plot of distribution of means we see that the distribution is centered around 5 and appears to be normally distributed.

```
hist(means,col="steelblue")
```



We can calculate the average sample mean:

```
sampleMean <- mean(means)
```

The average sample mean is 5.0229748 which is very close to the theoretical mean.

Sample Variance

The theoretical standard deviation is $\sigma = \frac{1/lambda}{\sqrt(n)}$, and the theoretical variance is σ^2, so we can calculate the theoretical variance:

```
sd <- 1/lambda/sqrt(n)
var <- sd^2
```

The theoretical variance is 0.625.

Now we can calculate the actual variance of the means:

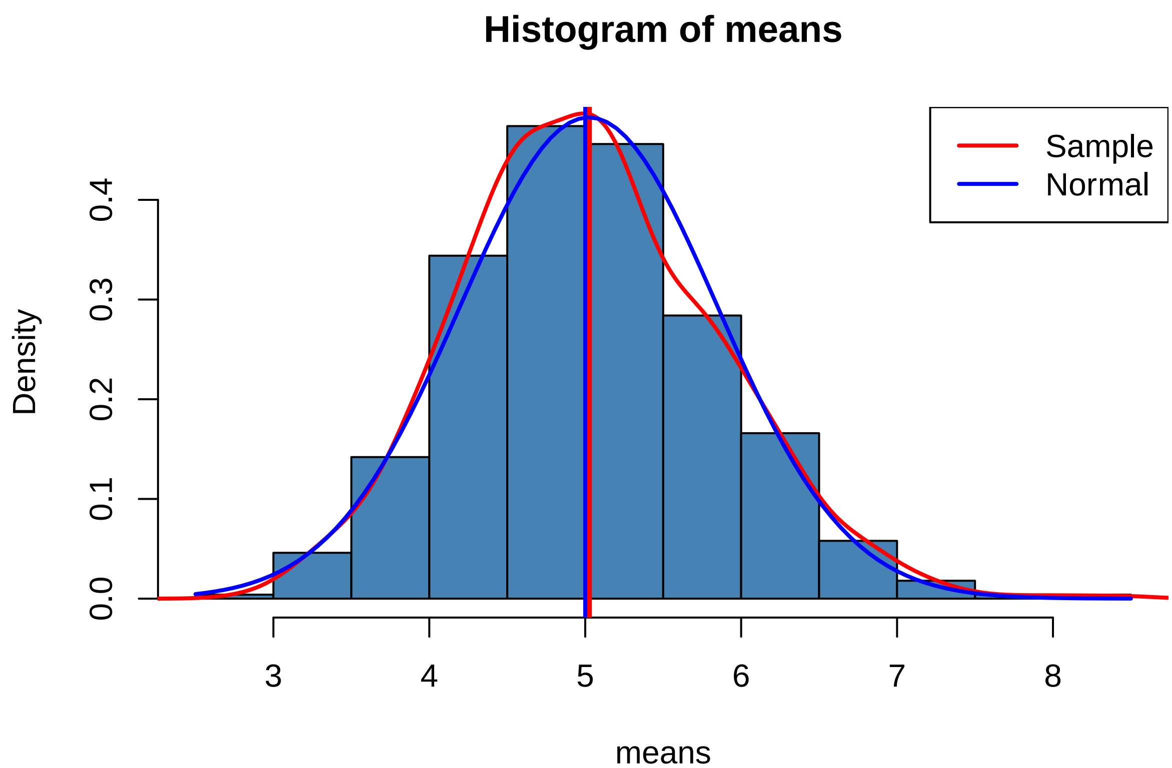
```
sample_var <- var(means)
```

The actual variance of the sample means is 0.6829699 which is relatively close to the theoretical variance.

Distribution

In the histogram above we see that the distribution appears roughly normal, let's take a closer look at the distribution by comparing it to a normal distribution.

```r
hist(means,col="steelblue",freq=FALSE)
lines(density(means),col="red",lwd=2)
abline(v=sampleMean,lwd=3,col="red")
normal <- rnorm(10000,mean(means),sd(means))
curve(dnorm(x, mean=mean(means), sd=sd(means)), add=TRUE,col="blue",lwd=2)
abline(v=5,lwd=2,col="blue")
legend("topright",legend=c("Sample","Normal"),col=c("red","blue"),lwd=2)
```



The blue curve is a normal distribution, with the blue vertical line at the mean. The red curve is a density curve of the sample data with the red vertical line at the sample mean. We have already seen that the means are close, the distribution of the actual data appears to overlap with the normal distribution, leading me to conclude that the sample distribution is also approximately normal.