Statistical Inference Project

S. Smolenski

Overview:

In this project, I investigate the exponential distribution in R and compare it to the Central Limit Theorem. I consider the distribution of averages of 40 exponentials, performing 1000 simulations to create the data used for this project.

Simulations:

I set the parameters for the exponential function and perform the simulation.

```
n=40
numsim=1000
lambda<-.2
Sim<-matrix(rexp(numsim*n,lambda), numsim)
Means<-apply(Sim, 1, mean)
```

Sample Mean versus Theoretical Mean:

The means are approximately the same:

Sample mean:

```
SimMn<-mean(Means)
SimMn
```

```
## [1] 4.987434
```

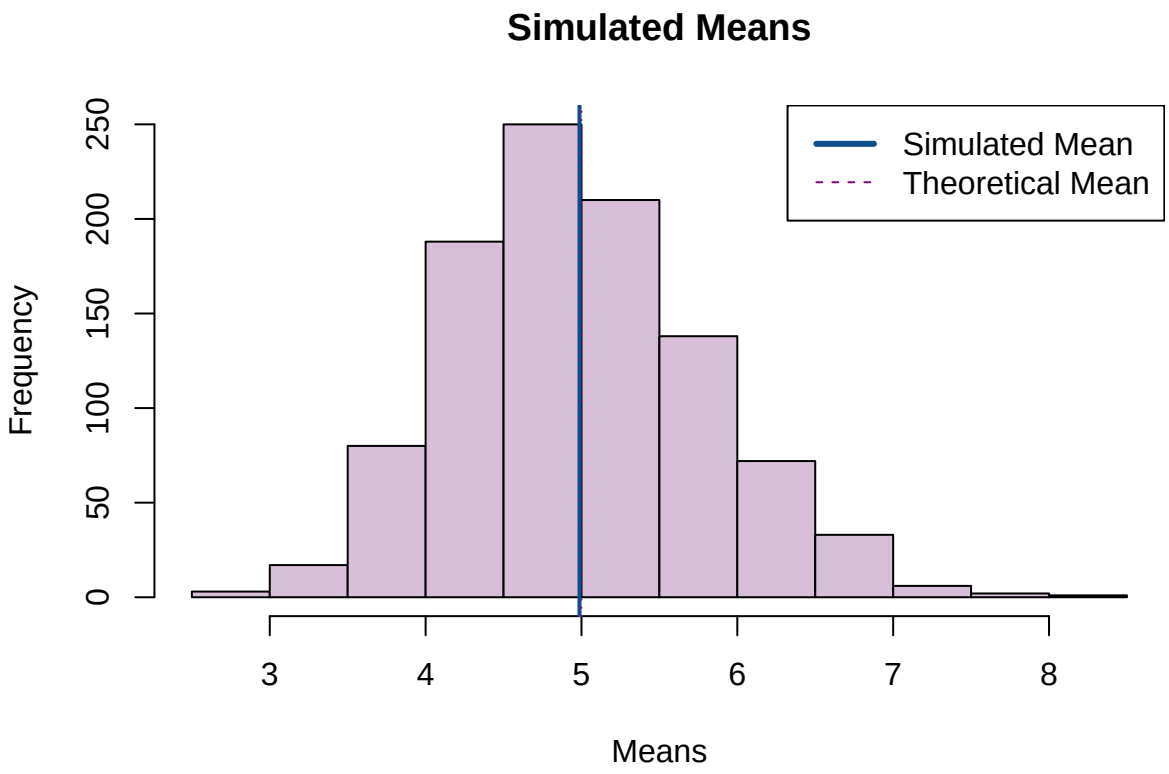
Theoretical Mean:

```
TheoMn<-1/lambda
TheoMn
```

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

Visual Comparison of Means:

```
hist(Means, col="thistle", main="Simulated Means")
abline(v=SimMn, col="dodgerblue4", lwd=2)
abline(v=TheoMn, col="darkmagenta", lty=3)
legend("topright", lwd=c(3,1), lty=c(1,2), col=c("dodgerblue4","darkmagenta"), legend=c("Simulated Mean
```



Sample Variance versus Theoretical Variance:

```
SimVar<-var(Means)
SimVar
```

```
## [1] 0.6460181
```

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

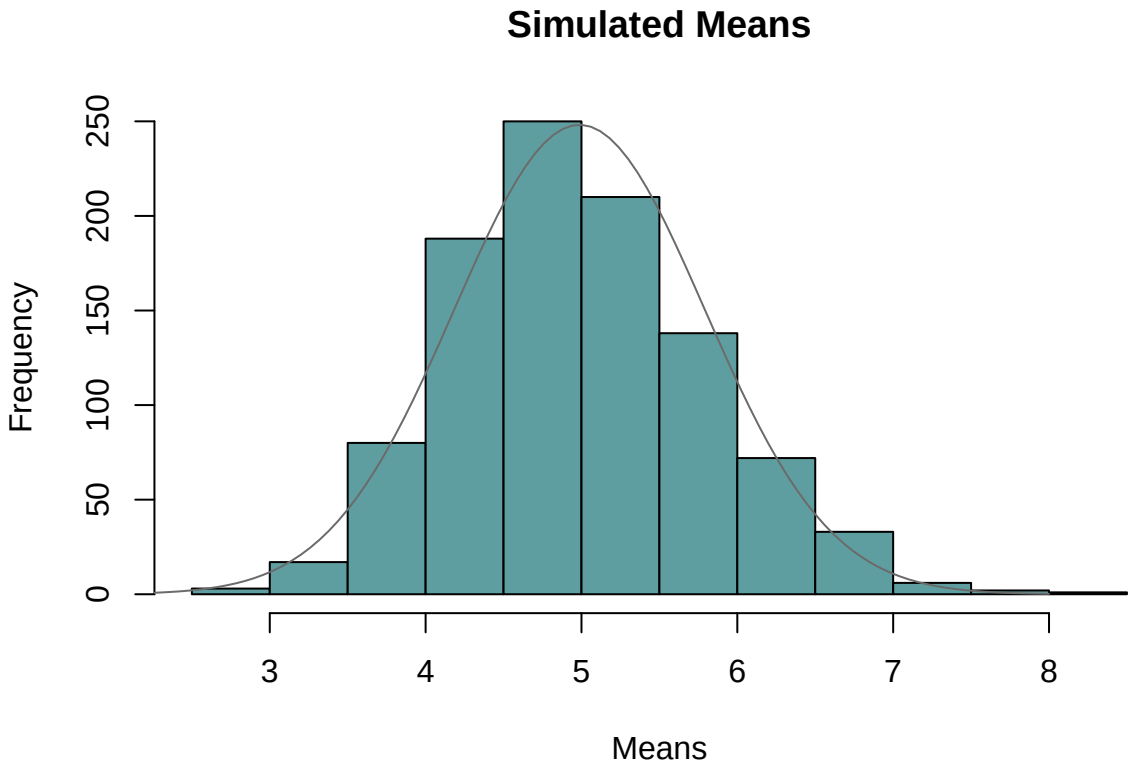
```
## [1] 0.625
```

The theoretical variance is slightly lower than the simulated variance, implying that the simulated distribution will have a slightly larger spread. The larger numsim becomes, the closer to the theoretical variance this value should be.

Distribution

If we overlay a normal density function on top of our simulated data, we find that the distribution is roughly normal:

```
hist(Means, col="cadetblue", main="Simulated Means")
curve( dnorm(x, mean=SimMn,sd=sqrt(SimVar))*500, 2,8, add=TRUE, col="gray42")
```


Simulated Means
Frequency
0
50
100
150
200
250
3
4
5
6
7
8
Means