Statistical Inference Course Project

Ashly Yashchin

August 5, 2017

Part One - Simulation Excercise

The following is an investigation into exponential distributions and their comparison to the Central Limit Theorem (CLT).

For the purpose of this excercise, lambda is assumed to equal 0.2.

Both mean and standard deviation equals 1/lambda.

Let's investigate the distribution of averages of 40 exponentials using 1000 simulations.

Before we begin, let's set a seed for reproducibility. Then let's determine our theoretical mean of 1/lambda.

```
set.seed(2017)

# Number of Exponentials
n=40

# Number of Simulations
sims=1000

lambda <- 0.2
tmean <- 1/lambda
tmean
```

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

The answer we get is 5. We divide this by our sample to first get our standard deviation and then square it to find our variance.

What about our theoretical variance?

```
tsd <- tmean/sqrt(n)
print(paste("Theoretical Standard Deviation:",round(tsd,3)))
```

```
## [1] "Theoretical Standard Deviation: 0.791"
```

```
tvar <- tsd^2
print(paste("Theoretical Variance:",tvar))
```

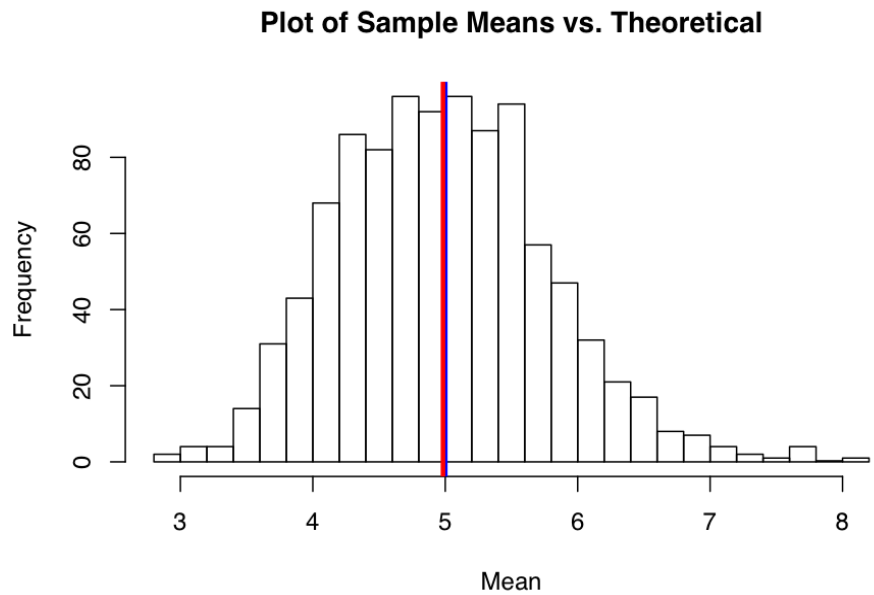
```
## [1] "Theoretical Variance: 0.625"
```

Now, let's calculate our sample mean.

```
simulations <- replicate(sims,rexp(n,lambda))
```

This gives us one row for each exponential (40 total) and one column for each simulation (1000 total). Next, we need to calculate the mean of each simulation and plot. We will be calculating the mean of each column, or each simulation

```
means <- apply(simulations,2,mean)
hist(means, breaks=25, main = "Plot of Sample Means vs. Theoretical", xlab = "Mean")
abline(v=5, col="blue", lwd = 3) #blue line represents the theoretical mean
abline(v=mean(means),col="red",lwd=3) #red line represents the sample mean
```



So what is the sample mean?

```
mean(means)
```

```
## [1] 4.982863
```

The sample mean is very close to our theoretical mean of 5.

This finding supports the Central Limit Theorem in demonstrating that the mean of a significant sample size (1000) will be equivalent to the theoretical (population) mean.

Now what about the variance? We remember that our theoretical variance is 25 (since our standard deviation is 5). Let's calculate our sample variance.

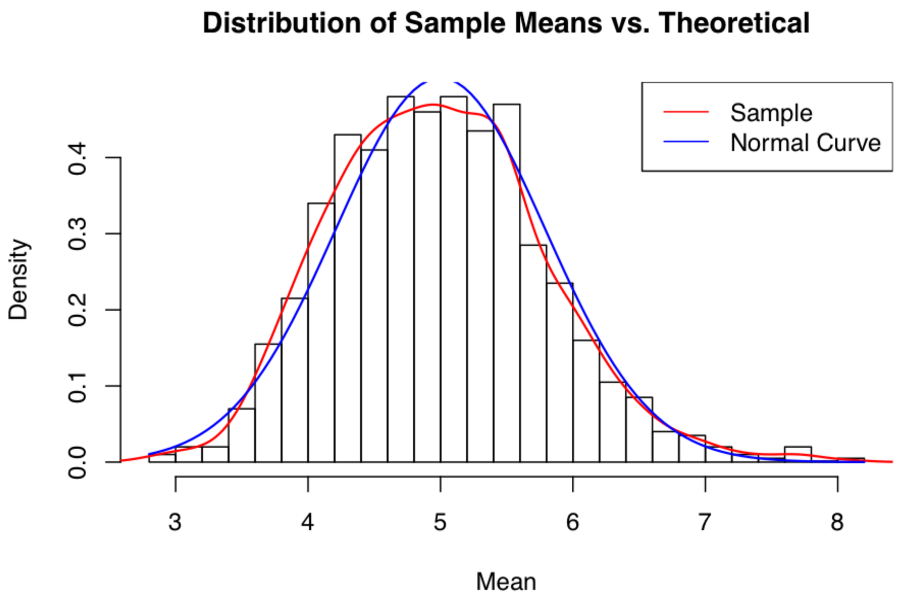
```
svar <- sd(means)^2
print(paste("Sample Variance",round(svar,3)))
```

```
## [1] "Sample Variance 0.627"
```

Our sample variance of 0.62 is very close to our theoretical, above, of 0.625.

Since, our calculations show that this excercise supports the Central Limit Theorem, it follows that the distribution should be normal. To confirm, we'll draw a quick plot over our histogram from earlier.

```
hist(means, prob=TRUE, breaks=25, main = "Distribution of Sample Means vs. Theoretical", xlab = "Mean")
lines(density(means),lwd=1.5, col="red") #red equals sample
curve(dnorm(x,mean=tmean,sd=tsd),lwd=1.5,col="blue",add=TRUE) #blue equals normal distribution
legend("topright",c("Sample","Normal Curve"),lty=c(1,1),col=c("red","blue"))
```



While not identical, our sample means (red) approximate a normal curve (blue).