

# Statistical Inference - Course Project

## Part 1

Anjali Singh

1 October 2017

## Part 1: Simulation Exercise Instructions

In this project we 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  $1/\lambda$ .

We have to 1) Set  $\lambda = 0.2$  for all of the simulations. 2) Investigate the distribution of averages of 40 exponentials. 3) Compare the distribution of 1000 random uniforms

## Simulation

We will first set up the environment and load up the library needed to run the simulation

```
require(knitr)
require(ggplot2)
set.seed(111)
```

Initial Values

```
lambda <- 0.2;
samples <- 40;
simulationCount <- 1000;
simMeans = NULL
for (i in 1 : simulationCount) {
  simMeans = c(simMeans, mean(rexp(samples,lambda)))
}
```

Question 1: Show the sample mean and compare it to the theoretical mean of the distribution. Calculate the theoretical mean and the sample mean.

```
#Theoretical Mean
tMean<-round(1/lambda,3)
tMean
```

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

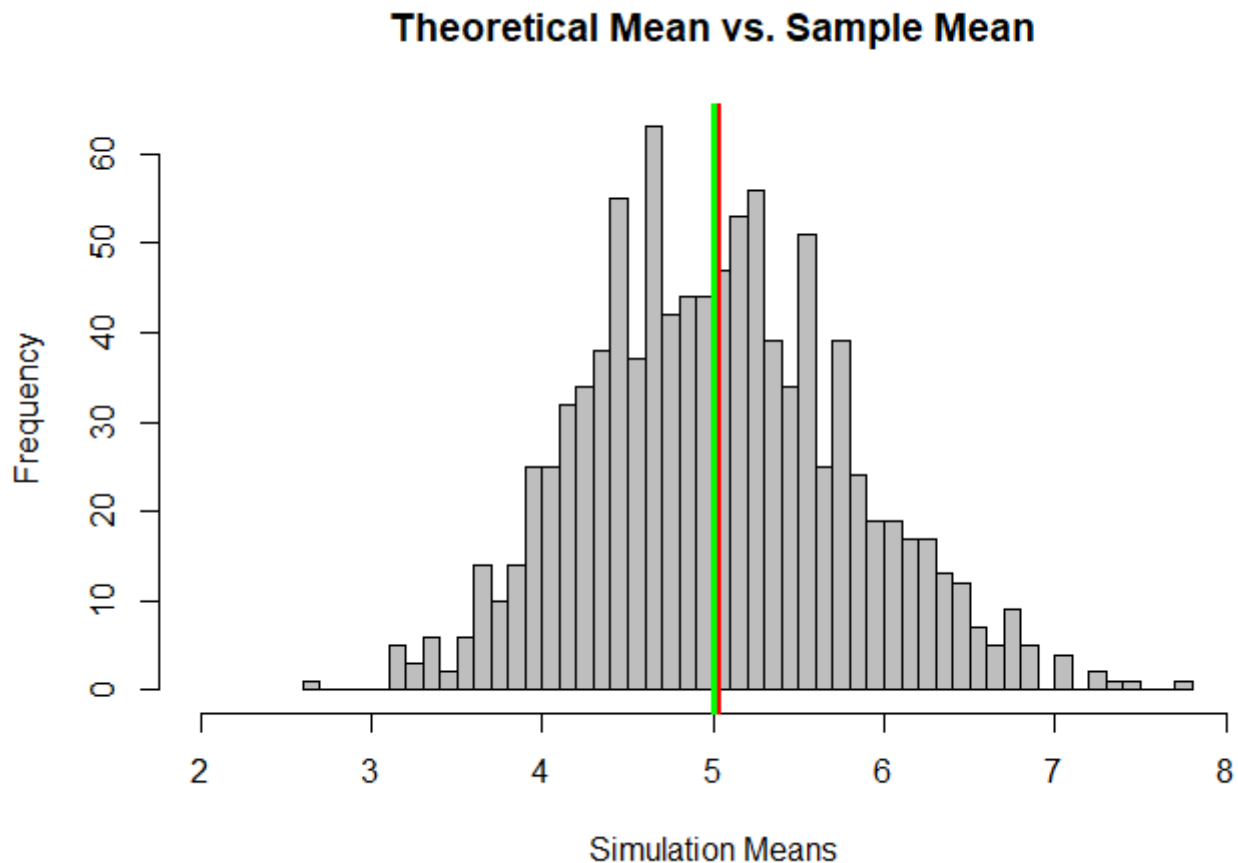
```
#Sample Mean
aMean<-round(mean(simMeans),3)
aMean
```

```
## [1] 5.026
```

Plot the Theoretical Mean and Sample Mean

```
library(ggplot2)
# plot histogram of the sample means
hist(simMeans, col="gray", main="Theoretical Mean vs. Sample Mean", xlim = c(2,8),breaks=40,
xlab = "Simulation Means")

# plot a vertical red line at the mean of the sample means
abline(v=mean(aMean), lwd="3", col="red")
abline(v=mean(tMean), lwd="3", col="green")
```



The green vertical line indicate the theoretical sample mean(5), The red vertical line is the calculated average sample(5.026).The two means are very close.

Question 2: show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
#theoretical variance
tVar <- round((1/lambda)^2/samples,3)
tVar
```

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

```
#Sample variance
aVar <- round(var(simMeans),3)
aVar
```

```
## [1] 0.607
```

```

results <- matrix(c(tMean, aMean,
                    tVar,aVar),
                  ncol = 2, byrow=TRUE)
#Naming the columns
colnames(results) <- c("theroetical","sample mean")
#Naming the row
rownames(results) <- c("mean","variance")
#Display the result as table
as.table(results)

```

```

##          theroetical sample mean
## mean          5.000         5.026
## variance       0.625         0.607

```

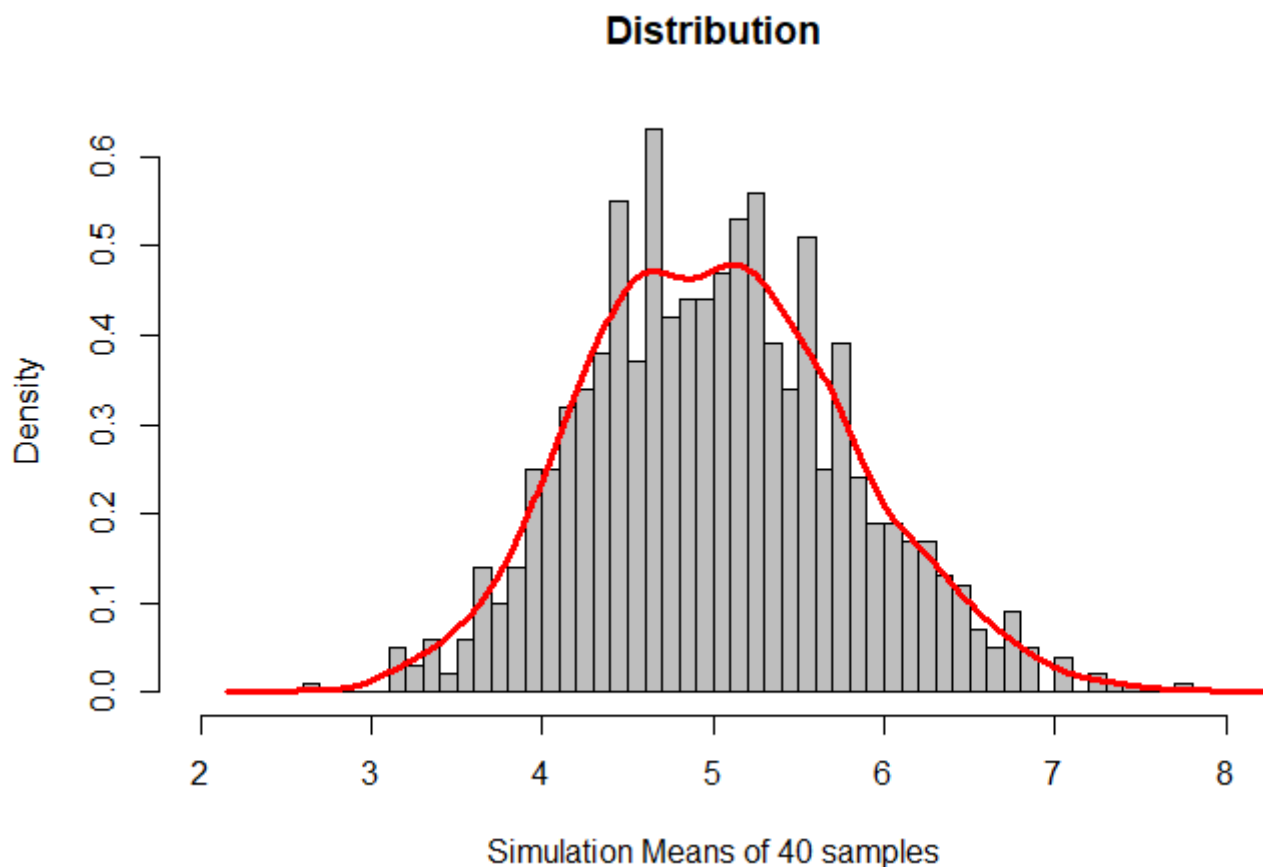
the theoretical variance and sample variance is 0.625 and 0.607 respectively and are quite close

Question 3: Show that the distribution is approximately normal.

```

hist(simMeans, prob=TRUE, col="gray", main="Distribution", breaks=40, xlim=c(2,8), xlab = "Si
mulation Means of 40 samples")
lines(density(simMeans), lwd=3, col="red")

```



The distribution seems to follow the normal distribution and this is due to the central limit theorem normal distribution, due to the Central Limit Theorem (i.e. as we increase the sample size, the distribution follows a normal distribution)