Statistical Inference Course Project (Part 1)

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. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

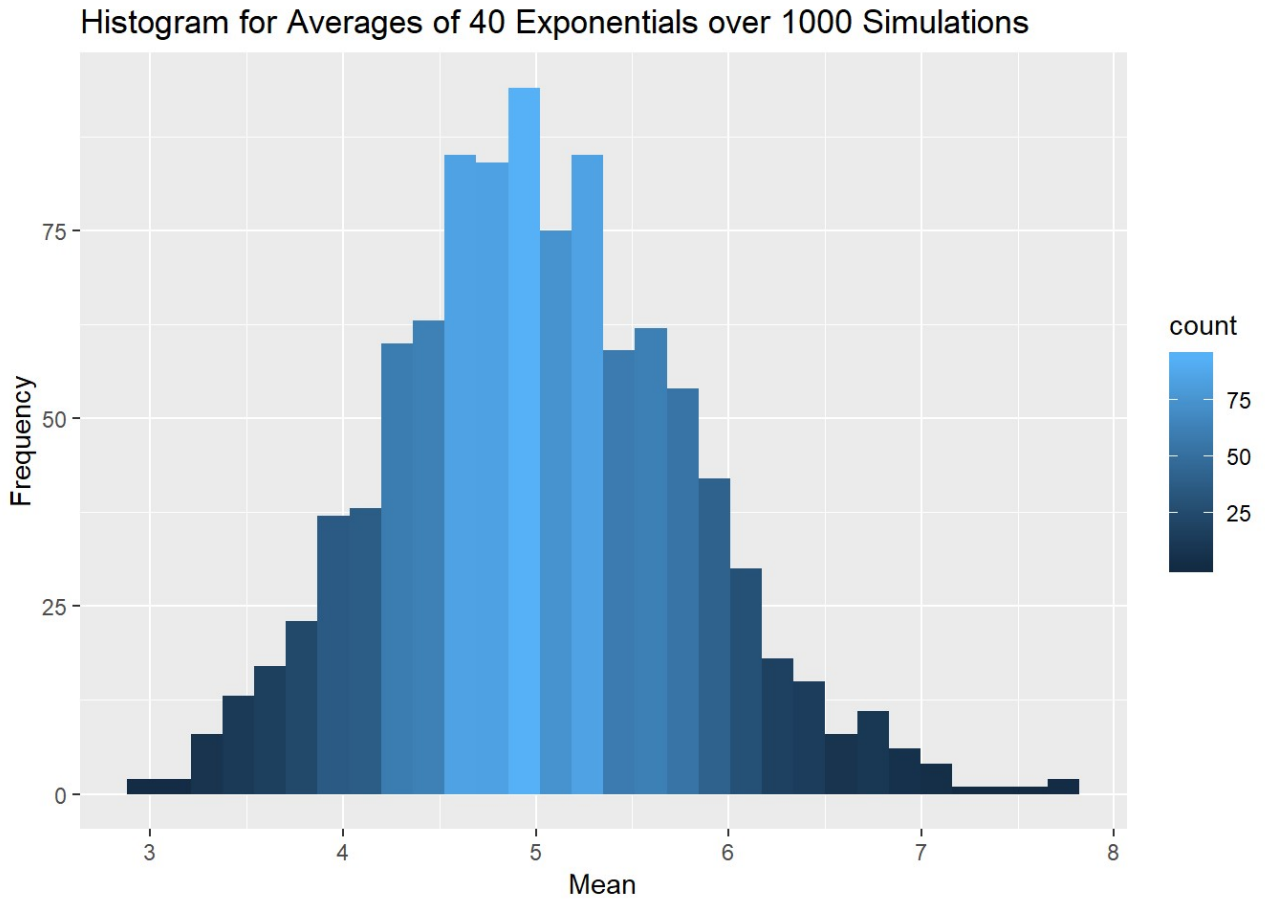
We show the sample mean and compare it to the theoretical mean of the distribution. We show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. And we show that the distribution is approximately normal. In point 3, we focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials. #Simulations

```
# Using pre-defined parameters
lambda <- 0.2
n <- 40
sims <- 1:1000
set.seed(123)

# Check for missing dependencies and load necessary R packages
if(!require(ggplot2)){install.packages('ggplot2')}; library(ggplot2)
```

```
## Loading required package: ggplot2
```

```
# Simulate the population
population <- data.frame(x=sapply(sims, function(x) {mean(rexp(n, lambda))}))

# Plot the histogram
hist.pop <- ggplot(population, aes(x=x)) +
  geom_histogram(aes(y=..count.., fill=..count..)) +
  labs(title="Histogram for Averages of 40 Exponentials over 1000 Simulations", y="Fre
quency", x="Mean")
hist.pop
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



#Sample Mean versus Theoretical Mean

```
# Tabulating the Sample Mean & Theoretical Mean
sample.mean <- mean(population$x)
theoretical.mean <- 1/lambda
cbind(sample.mean, theoretical.mean)
```

```
##      sample.mean theoretical.mean
## [1,]    5.011911                5
```

```
# Checking 95% confidence interval for Sample Mean
t.test(population$x)[4]
```

```
## $conf.int
## [1] 4.963824 5.059998
## attr(,"conf.level")
## [1] 0.95
```

Sample Variance versus Theoretical Variance

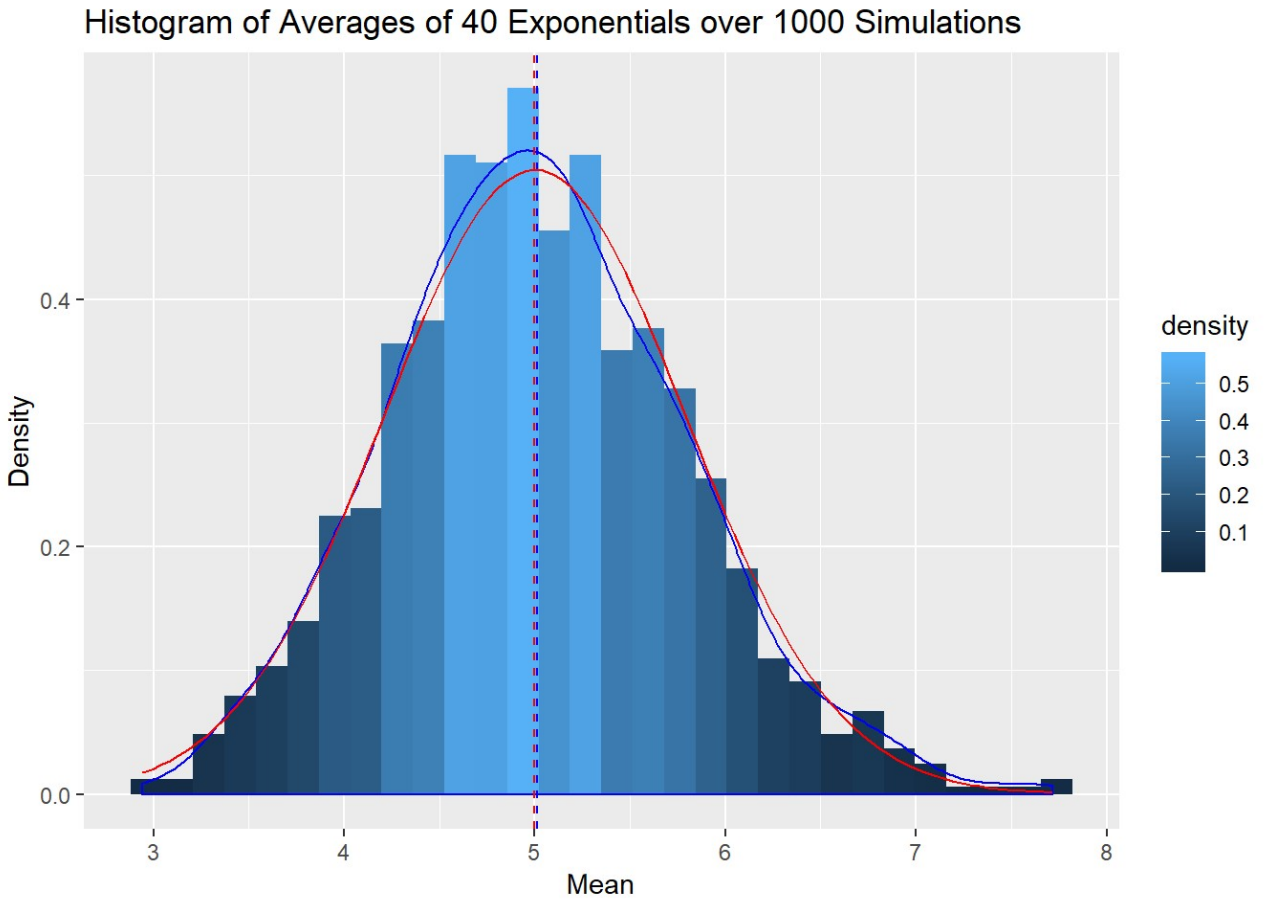
```
sample.variance <- var(population$x)
theoretical.variance <- ((1/lambda)^2)/n
cbind(sample.variance, theoretical.variance)
```

```
##      sample.variance theoretical.variance
## [1,]       0.6004928                0.625
```

Distribution

```
# Plotting Sample Mean & Varience vs Theoretical Mean & Varience
gg <- ggplot(population, aes(x=x)) +
  geom_histogram(aes(y=..density.., fill=..density..)) +
  labs(title="Histogram of Averages of 40 Exponentials over 1000 Simulations", y="Dens
ity", x="Mean") +
  geom_density(colour="blue") +
  geom_vline(xintercept=sample.mean, colour="blue", linetype="dashed") +
  stat_function(fun=dnorm,args=list( mean=1/lambda, sd=sqrt(theoretical.variance)),col
or = "red") +
  geom_vline(xintercept=theoretical.mean, colour="red", linetype="dashed")
gg
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
#Results: the Sampled mean for 40 exponentials simulated 1000 times are very close to
the Theoretical mean for a normal distribution.
```