# Statistical Inference Course Project Part 1

Keng Yew Hoe

10/13/2020

## Overview

The purpose is to investigate the exponential distribution in R and comparing it with the Central Limit Theorem. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. with lambda = 0.2 for all the simulations. The investigation compares the distribution of averages of 40 exponentials over a thousand simulations

## Simulations

For the 1000 simulations, lambda is 0.2 and sample size is 40

```
#Set seed to ensure reproducability
set.seed(6666)

#Set lambda
lambda <- 0.2

#Set exponentials
n <- 40

#Simulating the exponential distribution and getting the mean of 1 simulation
simul <- rexp(n,lambda)
simul
```

```
##  [1]  3.54431668  1.53150748  0.32929821  3.27650114  2.69959882  8.74659368
##  [7]  3.56559497  0.62182882  1.61579815  1.29026995  4.73752971  1.46355046
## [13] 13.40054884  2.06984195  0.35475479  4.45554234  8.60265237 12.09602350
## [19]  1.07535294  1.39329189  1.28706599  1.57589110  7.72572449  7.00650160
## [25]  5.11022532  3.02797012  2.19538448  9.61173891  6.33892175  4.12758995
## [31]  0.01988841 11.82788231 18.65521784  1.99198980  8.48981051 14.90360759
## [37]  8.14093665  9.86643793  4.81952719  5.16277506
```

```
expMean <- mean(simul)
expMean
```

```
## [1] 5.218887
```

## Sample Mean versus Theoretical Mean

```
#Getting the means of each simulations
simul1000 <- as.data.frame(replicate(1000,mean(rexp(n, lambda))))
names(simul1000) <- c("sample.mean")

#Calculating the mean of the means of 1000 simulation
mean1000 <- mean(simul1000$sample.mean)
mean1000
```

```
## [1] 5.062025
```

```
#The theoretical mean
theoMean <- 1/lambda
theoMean
```

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

The theoretical mean and the experimental mean have a very similar value therefore the center of distribution of sample means of 40 exponential is close to the theoretical center of the distribution

## Sample Variance versus Theoretical Variance

```
# Calculating the variance of this simulation
var1000 <- var(simul1000$sample.mean)
var1000
```

```
## [1] 0.621038
```

```
# The theoretical variance
theoVariance <- ((1/lambda)^2)/40
theoVariance
```

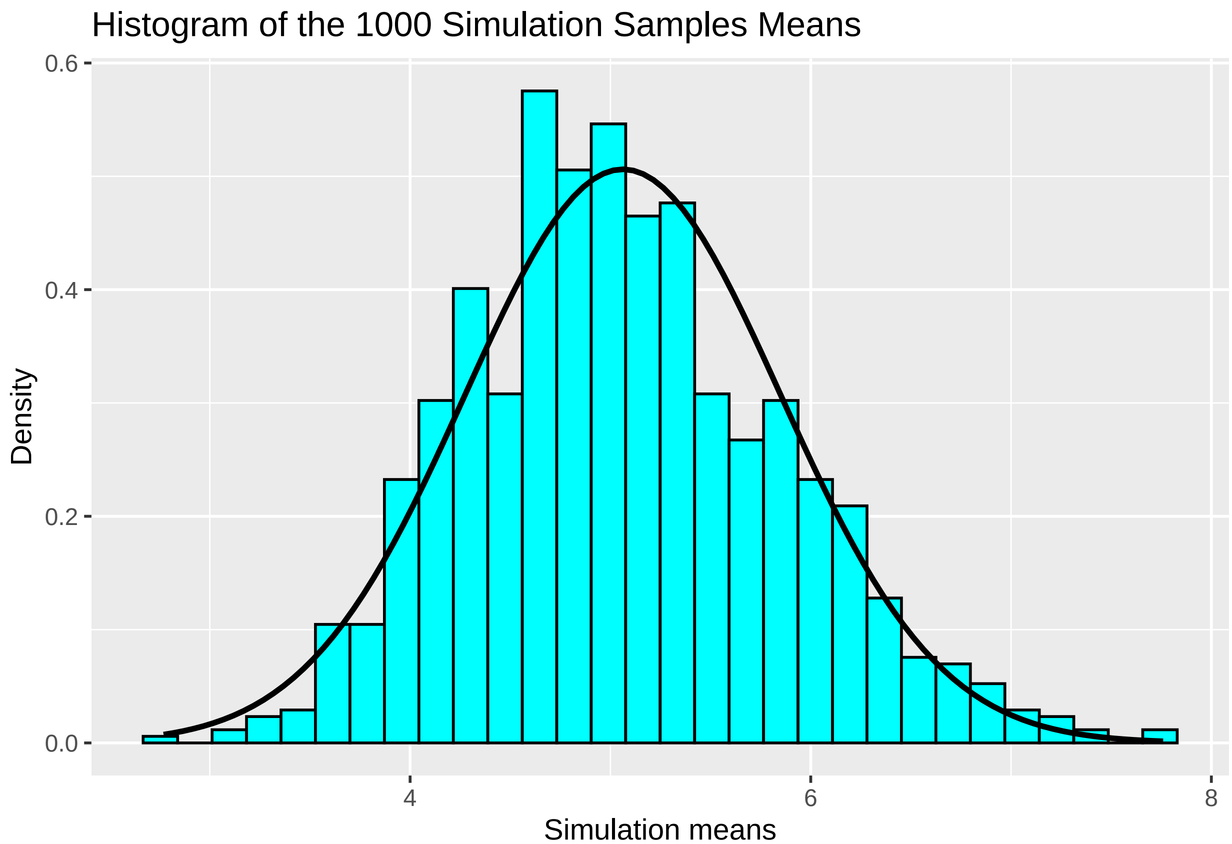
```
## [1] 0.625
```

The variance of the distribution of means have a very similar value to the theoretical variance which is calculated by squaring the standard deviation and dividing by the sample size

## Distribution

```
library(ggplot2)
ggplot(simul1000, aes(x=sample.mean)) +
    geom_histogram(aes(y = ..density..),colour="black",fill="cyan")+
    stat_function(fun=dnorm,args=list( mean=mean1000, sd=sqrt(var1000)),geom="line",color = "black", si
    ggtitle("Histogram of the 1000 Simulation Samples Means") +
    scale_x_continuous("Simulation means")+
    ylab("Density")
```

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




The black curve represents the normal distribution which is used to compare with the histogram. The central limit theorem states that the sample means would become that of a standard normal distribution as the sample size increases whilst meeting the two conditions of independence ( n $<$ 10% ) and normal, or if skewed distribution, that n $>$ 30.

---