

# Statistical Inference Simulation Project

## Overview:

To investigate the relationship between the exponential distribution in R and the Central Limit Theorem (CLT). There are 3 tasks:

- 1) Show the sample mean and compare it to the theoretical mean of the distribution.
- 2) Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
- 3) Show that the distribution is approximately normal. Which I hope to answer in order.

First, to provide reproducibility for the random variables used. Then we set the base characteristics that was provided by the project guideline.

```
set.seed(3032)
lambda <- 0.2
mean=1/lambda
n<-40
nosims<-1000
data=rexp(n*nosims, lambda)
```

To simulate `rexp(n, lambda)` 1000 times, we will keep the data in a neat matrix. Then, to answer the first question, we will find the sample mean of this simulation and compare it with the theoretical mean ( $1/\lambda$ ).

```
exD<- matrix(data,nrow=nosims)
exDmean<-apply(exD,1,mean)
##The sample mean of the simulated mean will be:
mean(exDmean)
```

```
## [1] 5.014347
```

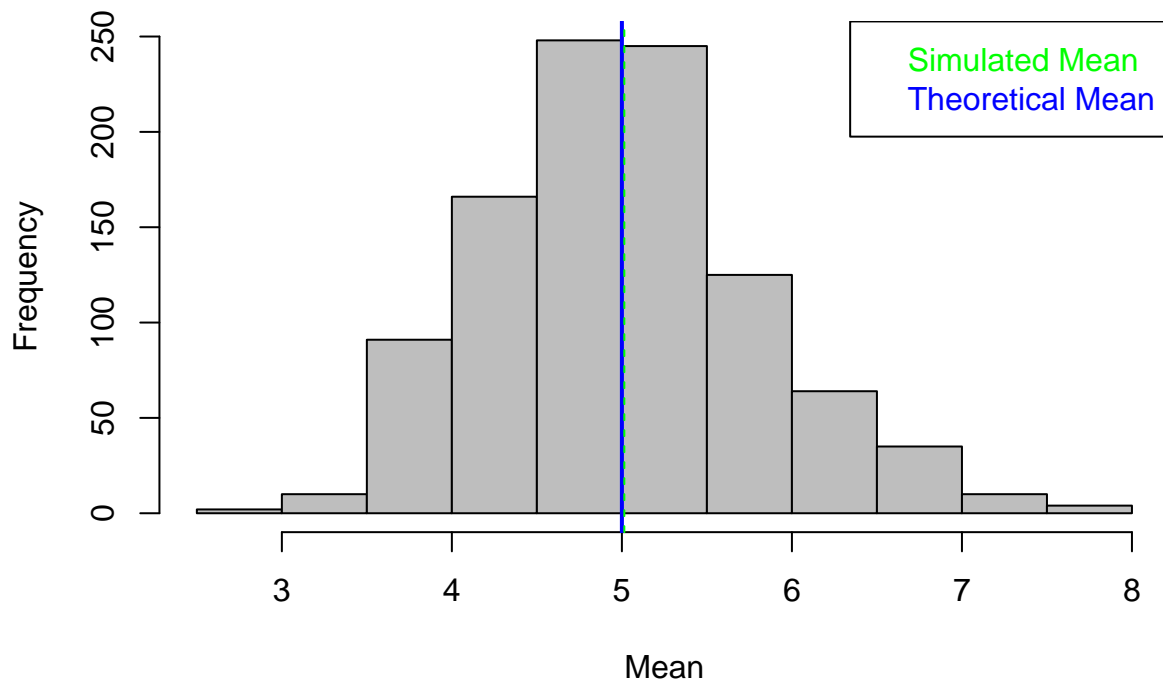
```
##while the theoretical mean is 1 over lambda, which is:
print(mean)
```

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

To show it graphically, we will use a histogram, where the green dotted line is the simulated mean and the blue is the theoretical. They are almost on top of each other and might be a bit difficult to see separately, which is why I plotted the dotted line on top of the blue one.

```
hist(exDmean, xlab="Mean", main="Histogram of Simulated Exponential Means", col="grey")
legend("topright", c("Simulated Mean", "Theoretical Mean"), text.col=c("green", "blue"))
abline(v=mean, col="blue", lwd=2)
abline(v=mean(exDmean),col="green", lty=2)
```

## Histogram of Simulated Exponential Means



#Question 2 The variance of a probability is the square of the standard deviation (s), which also measures how far from the mean a random sample can fall. We use s to measure the variability of the data as it uses the same unit of measurement as the original variable, and can be applied to the mean.

```
exDsd<- sd(exDmean)
##The sample standard distribution of the simulation will be:
print(exDsd)
```

```
## [1] 0.8011365
```

```
##while the standard distribution of the theoretical value will be
sd<-(1/lambda)/sqrt(n)
##To answer the question, we will have to find variance by applying the formula v=sd^2
var<- exDsd^2
print(var)
```

```
## [1] 0.6418196
```

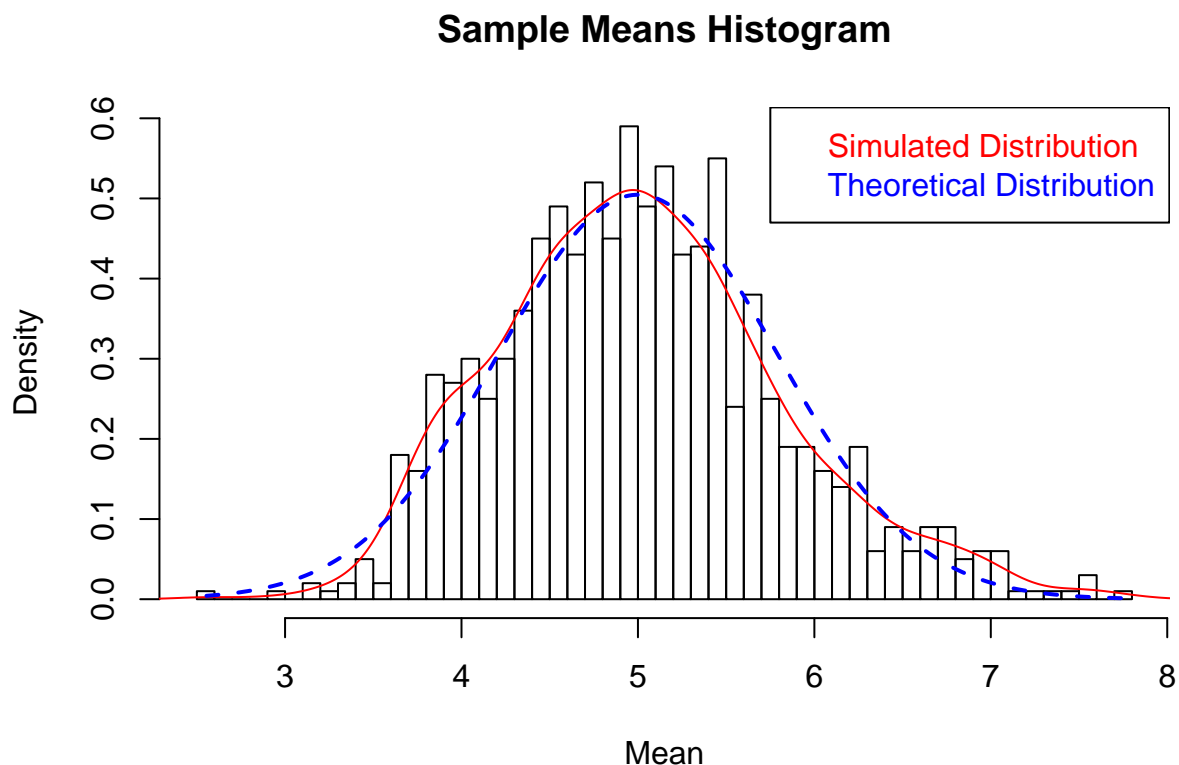
```
tvar<- sd^2
print(tvar)
```

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

So, the theoretical variance is smaller at 0.625 while the actual sample variance is smaller at approximately 0.642. This means that this sample will vary more than theoretical version, with one standard error as 0.801.

**Question 3: Show that the distribution is approximately normal.**

```
hist(exDmean, breaks=40, prob=T,  
     main="Sample Means Histogram", xlab="Mean")  
  
##After drawing the histogram of the mean, showing more counts of the means than the above histogram, w  
  
lines(density(exDmean), col="red")  
  
xfit<- seq(min(exDmean), max(exDmean), length=250)  
yfit<- dnorm(xfit, mean, sd)  
lines(xfit, yfit, pch=30,col="blue",lty=2, lwd=2)  
legend("topright", c("Simulated Distribution", "Theoretical Distribution"), text.col=c("red", "blue"))
```



The shape of the theoretical mean distribution is a perfect bell shape, also known as the standard normal distribution, albeit centered around 5 rather than 0. The simulated data is a similar bell shape, with kinks close to 4 and 7, which is typical of true samples to not be perfect. Even so, the distribution is centered at the same point and will therefore be standard normal in nature.