---
title: "Statistical Inference course project" Athni K
output: html_document

Overview

This is a project on Statistical Inference. It consists of two parts: 1. A simulation exercise 2. Basic inferential data analysis

This project is to explore inference and do simple inferential data analysis.

Assignment

In this project you 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.

Simulations

```
#set seed for reproducibility
set.seed(10)

#set lambda with given value
lambda <- 0.2

#set n with given value
n=40

#simulations set to 1000
simulations <-1000

#simulate
sim_exponentials <- replicate(simulations, rexp(n,lambda))

#calculate mean of this
mean_exponentials <- apply(sim_exponentials, 2, mean)
```

Sample Mean vs Theoretical Mean

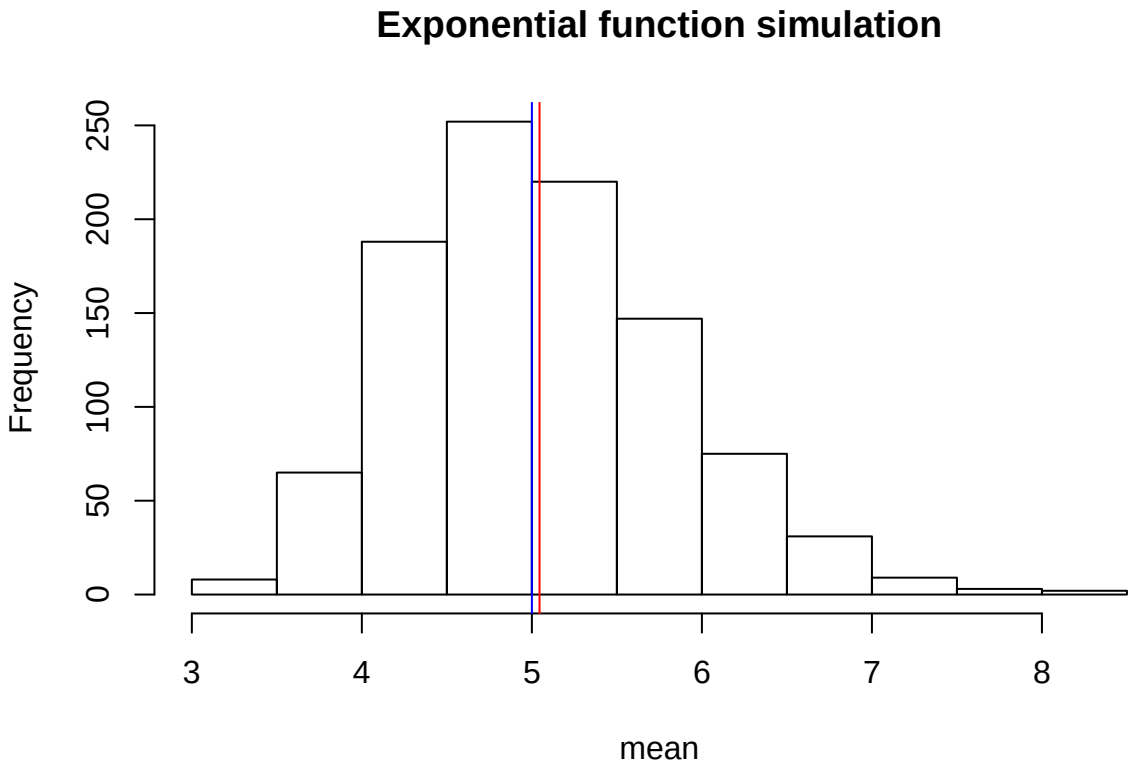
```
#distribution mean
samplemean <- mean(mean_exponentials)
samplemean
```

```
## [1] 5.04506
```

```
#theoretical mean
theoreticalmean <- 1/lambda
theoreticalmean
```

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

```
#visual representation
hist(mean_exponentials, xlab="mean", main="Exponential function simulation")
abline(v=samplemean, col="red")
abline(v=theoreticalmean, col="blue")
```



The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution(5.04 and 5 resp.)

Sample Variance vs Theoretical Variance

```
#standard deviation of the distribution
samplesd <- sd(mean_exponentials)
samplevariance <- samplesd^2
samplevariance
```

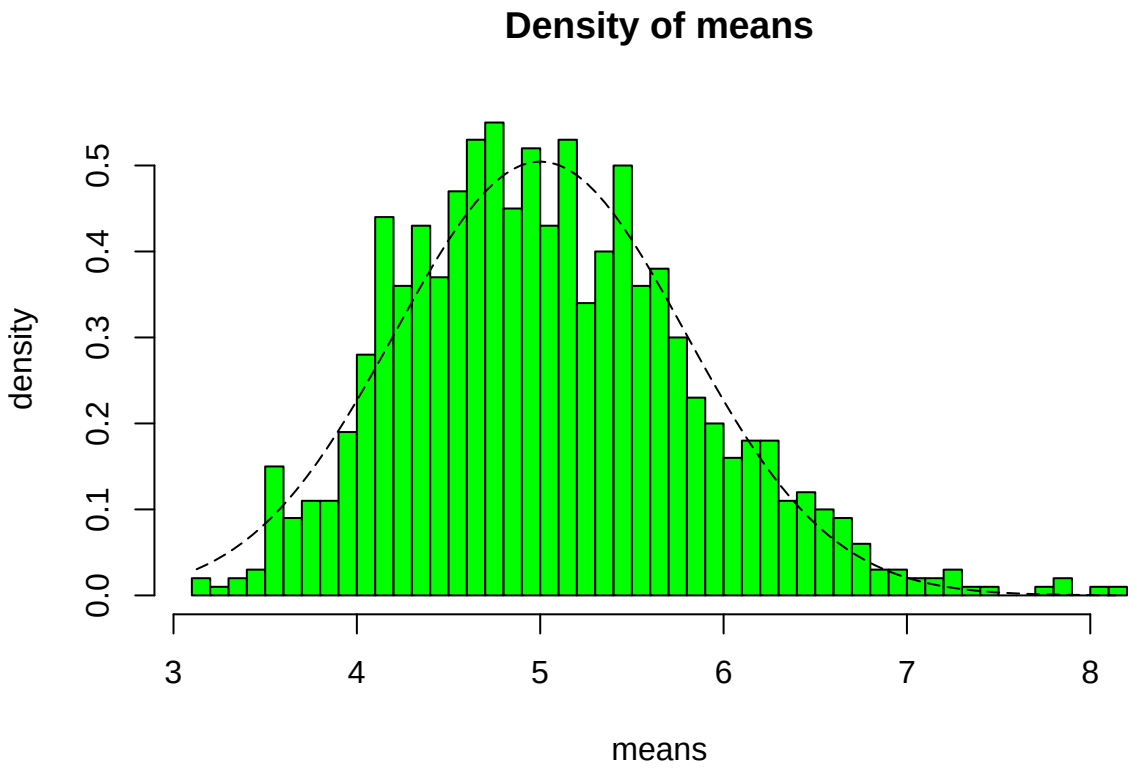
```
## [1] 0.6372544
```

```
#theoretical standard deviation
theoreticalsd <- (1/lambda)/sqrt(n)
theoreticalvariance <- theoreticalsd^2
theoreticalvariance
```

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

The variances by theory and analytically are very close to each other.

```r
xfit <- seq(min(mean_exponentials), max(mean_exponentials), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))
hist(mean_exponentials,breaks=n,prob=T,col="green",xlab = "means",main="Density of means",ylab="density")
lines(xfit, yfit, pch=22, col="black", lty=5)
```



Distribution

Looking at the distribution, it looks to be a normal distribution