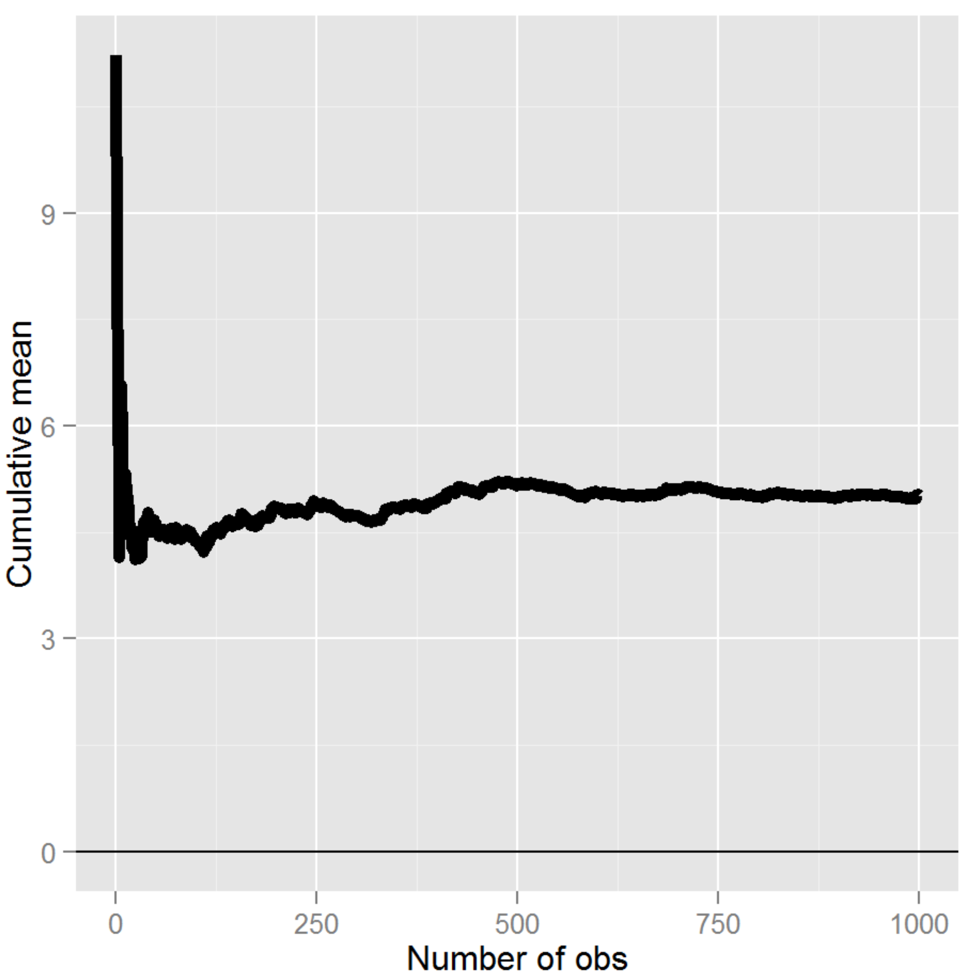
Statistical Inference Project Exponential distribution vs CLT

Alfonso Gonzalez Damian

Overview

- In this project the exponential distribution is investigated in R and compared with the Central Limit Theorem.
- The exponential distribution is 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.
- lambda was set to 0.2 for all of the simulations of the distribution of averages of 40 exponentials. ## Simulations
- We can calculate the mean of random values from an exponential distribution with R, using the next code, where n= 1000 is the number of observations, lambda = 0.2 that is the rate of the exponential distribution and plot the result of the means.

```
n <- 1000; lambda <- 0.2; means <- cumsum(rexp(n, lambda)) / (1  : n); library(ggplot2)
g <- ggplot(data.frame(x = 1 : n, y = means), aes(x = x, y = y))
g <- g + geom_hline(yintercept = 0) + geom_line(size = 2)
g <- g + labs(x = "Number of obs", y = "Cumulative mean")
g
```

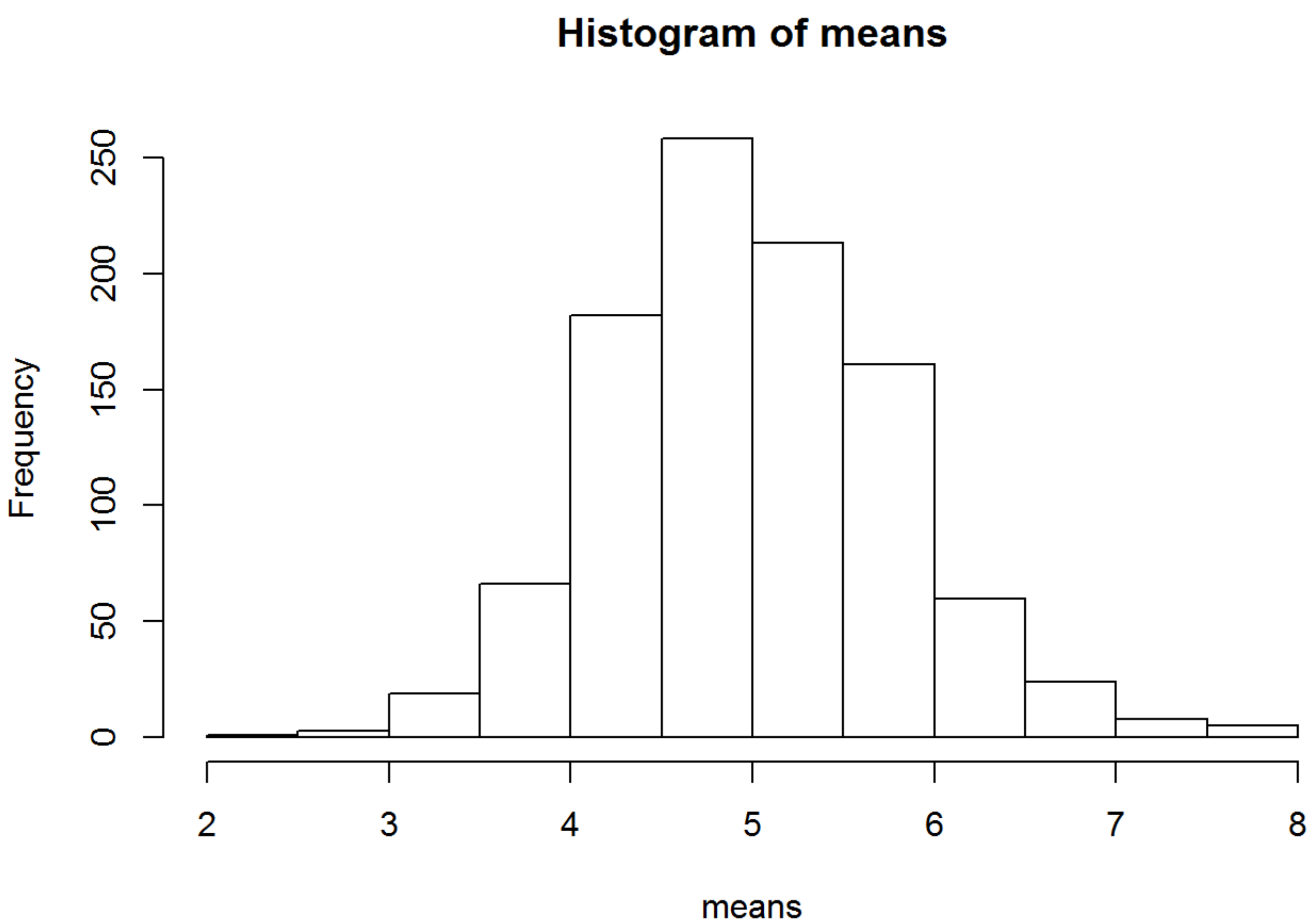


- In this graph you can see the graph of simuled data means, and how they are close to 5 (theoretical mean = 1/lamda) when we have more observations.

Sample Mean versus Theoretical Mean

- As we saw in the last graph, theoretical mean for this distribution is 5, and the Law of Large numbers says that when we have a large number of repetitions of the experiment, the sample mean that we obtain will be closer to it. The CLT states that the distribution of averages of iid, of 1000 repetitions of 40 observations each.

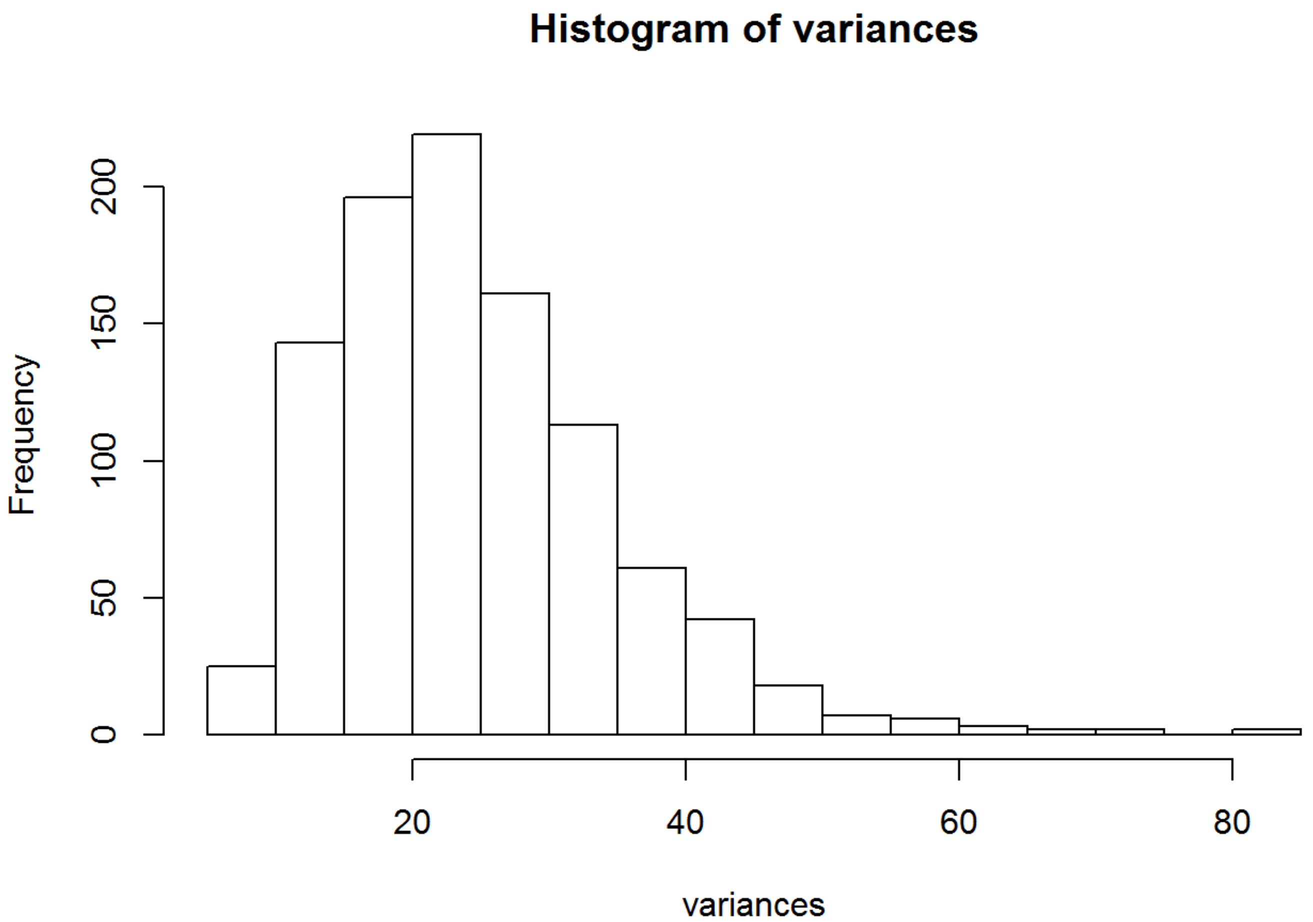
```
means = NULL
lambda = 0.2
for (i in 1 : 1000) means = c(means, mean(rexp(40, lambda)))
hist(means)
```



Sample Variance versus Theoretical Variance

- We can calculate now the variance of 40 random values from an exponential distribution with R, using the next code, where n= 1000 is the number of means, lambda = 0.2 that is the rate of the exponential distribution and plot the result of the variances.

```
variances = NULL
lambda = 0.2
for (i in 1 : 1000) variances = c(variances, var(rexp(40, lambda)))
hist(variances)
```



Distribution

- The final distribution of the means is approximately normal when compared with the distribution of a large collection of random exponentials as can be seen in the next graph

```
hist(rexp(10000, 0.2))
```