Statistical Inference Course Project Part 1

Ankitha Giridhar

18/10/2021

Overview

This project investigates the exponential distribution in R and compares it with the Central Limit Theorem alongside a basic inferential analysis. The simulation, variance comparison and distributions are explored in separate sections.

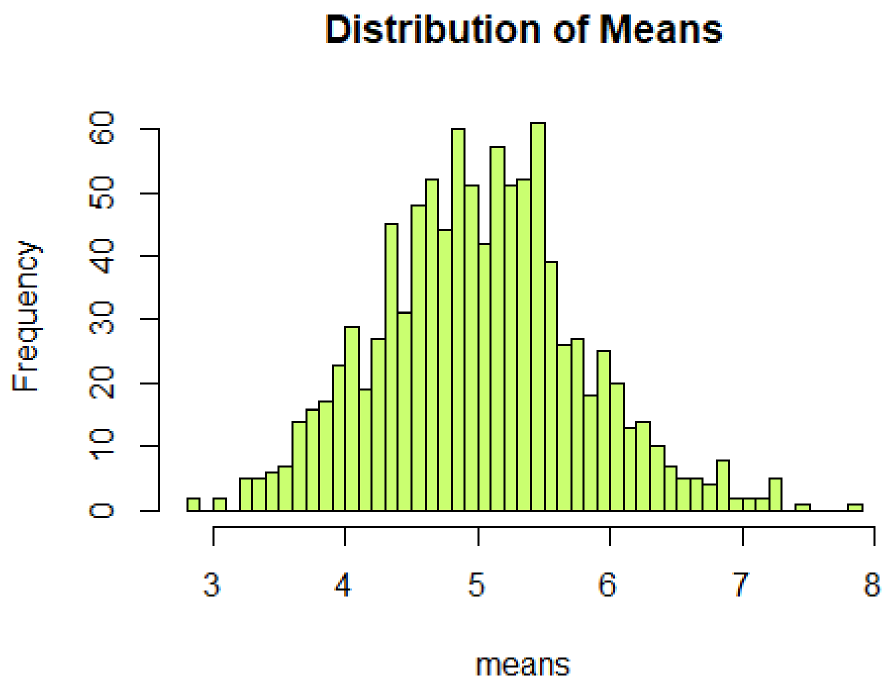
Simulation

The exponential distribution is simulated in R for 40 exponentials, with a lambda of 0.2.There are a thousand simulations. Therefore, the variables lambda, n, and sims are set to 0.2, 40 and 1000 respectively.

```
set.seed(12)
lambda <- 0.2
n <- 40
sims <- 1000
```

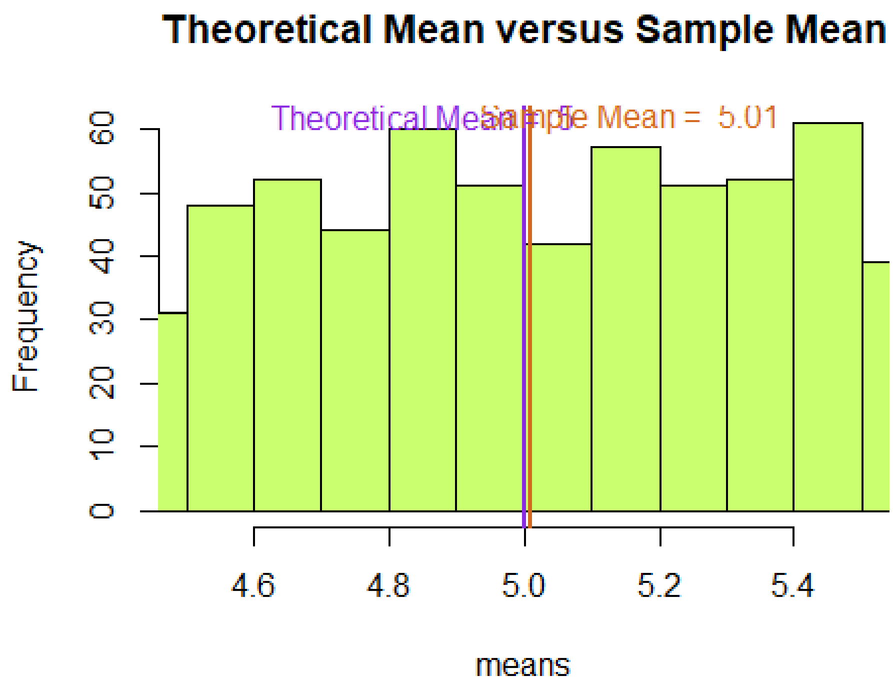
The replicate and rexp functions are used for the simulations. A distribution with all the means is visualised.

```
reps <- replicate(sims, rexp(n, lambda))
means <- apply(reps, 2, mean)
hist(means, breaks = 40, main="Distribution of Means", col =
"darkolivegreen1")
```



A blue line indicates the theoretical mean, which is 1/lambda, i.e. 5.A red line shows the sample mean.The limit of the X axis is set such that these two lines can clearly indicate the means.

```
hist(means, xlim = c(4.5,5.5), col="darkolivegreen1", main="Theoretical Mean versus Sample Mean", breaks=40)
abline(v=mean(means), lwd="2", col="chocolate")
abline(v = 1/lambda, lwd="2", col="blueviolet")
text(4.85, 62,  paste("Theoretical Mean = ",1/lambda), col="blueviolet")
text(5.16, 62, paste("Sample Mean = ",round(mean(means),4)), col="chocolate")
```



As observed, the sample mean is quite close to the theoretical mean.

Variance

The theoretical variance

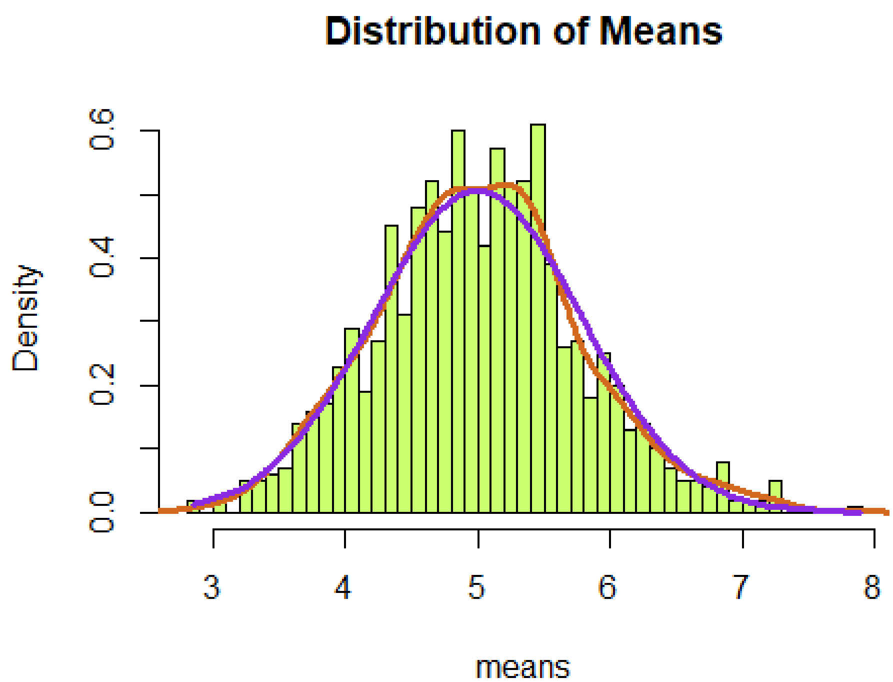
```
print(paste("Theoretical Standard Deviation:", (1/lambda)/sqrt(n)))

## [1] "Theoretical Standard Deviation: 0.790569415042095"

print(paste("Sample Standard Deviation:", sd(means)))

## [1] "Sample Standard Deviation: 0.774059441890319"

print(paste("Theoretical Variance:",((1/lambda)/sqrt(n))^2))

## [1] "Theoretical Variance: 0.625"

print(paste("Sample Variance:", sd(means)^2))

## [1] "Sample Variance: 0.599168019579552"
```

The sample standard deviation and sample variance are both quite close to their theoretical values.

Distribution

The distribution of means is compared with the normal distribution plotted with theoretical values.

```
hist(means, prob=TRUE, col="darkolivegreen1", main="Distribution of Means",
breaks=40)
lines(density(means), lwd=3, col="chocolate")
lines(means[order(means)],dnorm(means[order(means)],5,0.79), lwd = 3, col =
"blueviolet")
```



The blue bell curve shows the normal distribution taken from the theoretical values, while the red curve shows the actual distribution. Since both of them are quite close, it is implied that, due to the Central Limit Theorem, more samples would lead to better convergence.

Conclusion

The data distribution, the sample mean and variance as well as the bell curve all prove the Central Limit Theorem, as they are close enough to the theoretical values to conclude that they can converge to their ideal values.