Statistical Inference Project. Part 1

Diego Andres Benitez

24/1/2021

Contents

1 Overview

This is the first part of the Statistical Inference Course Project from Coursera. It is demonstrated that the distribution proves the Central Limit Theorem. This assignment will make calculation and plots and compare confidence intervals, and eventually proves that the distribution is approximately normal.

2 Tasks

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.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

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.

In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

3 Analysis

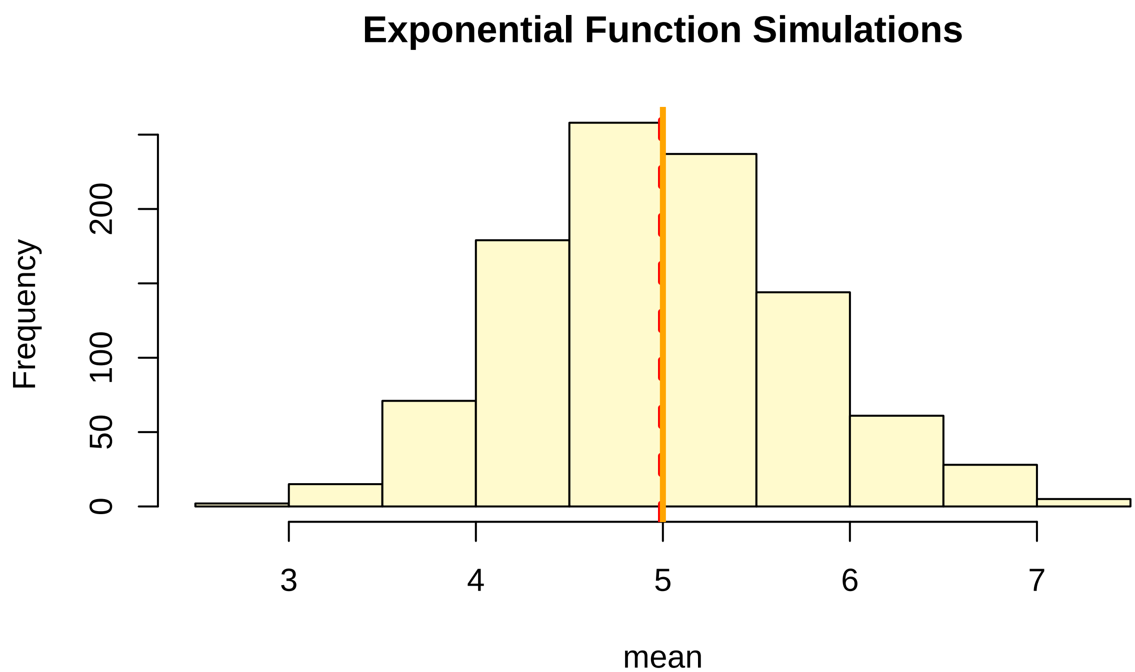
3.1 Simulations

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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

3.2 Mean comparison

Show where the distribution is centered at and compare it to the theoretical center of the distribution.

analytical mean	theorical mean
4.99	5



The analytics mean is 4.99 and the theoretical mean is 5. The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

3.3 Variance comparison

Show how variable it is and compare it to the theoretical variance of the distribution.

	Variance	Deviation
Analytical	0.570	0.75
Theorical	0.625	0.79

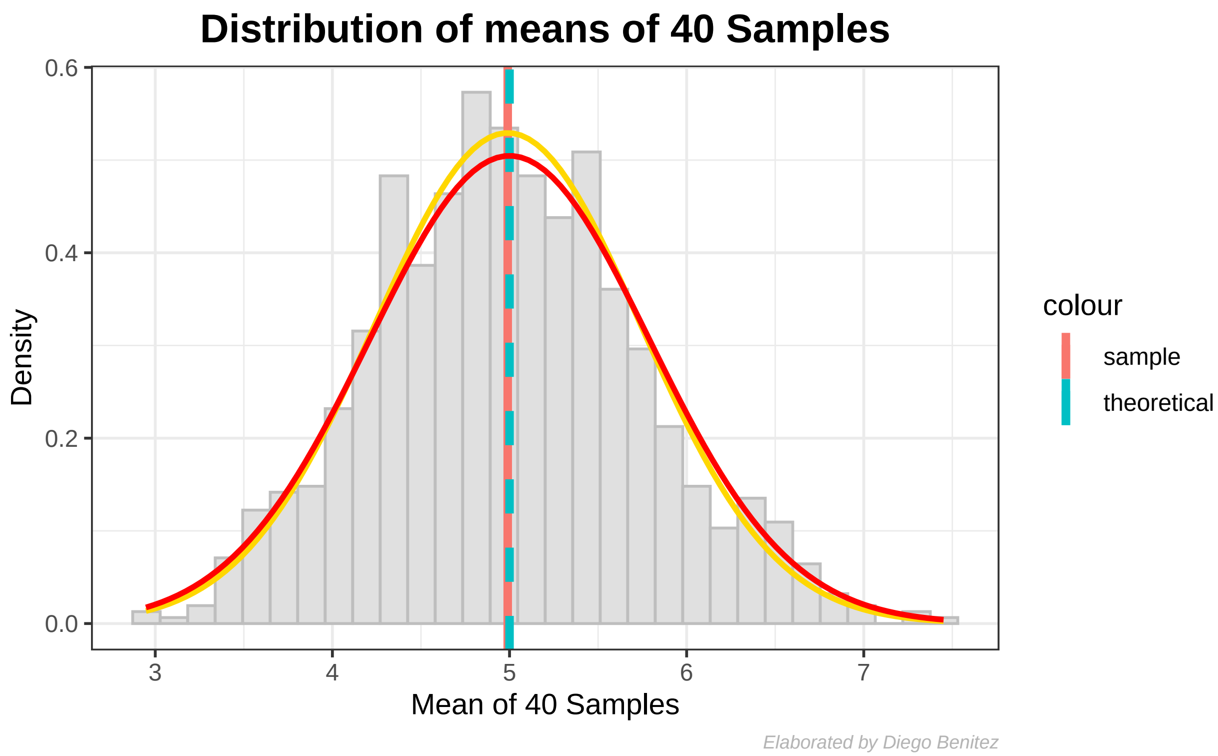
Standard Deviation of the distribution is 0.75 with the theoretical SD calculated as 0.79. The actual variance of the distribution is 0.57 Both standard deviation values are very close to each other and the theoretical variance is calculated as

$$Var = \frac{1}{\lambda^2 n} = 0.625$$

3.4 Distribution

Show that the distribution is approximately normal.

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



The density of the actual data is shown by the light blue bars. The theoretical mean and the sample mean are so close that they nearly overlap. The "red" line shows the normal curve formed by the the theoretical mean and standard deviation. The "gold" line shows the curve formed by the sample mean and standard deviation. As you can see from the graph, the distribution of means of 40 exponential distributions is close to the normal distribution with the expected theoretical values based on the given lambda. Due to Due to the central limit theorem (CLT), the distribution of averages of 40 exponentials is very close to a normal distribution.

4 Appendix

The Rmd code, generated by `knitr`, are available clicking here.