Exponential Distribution in R and Comparison to the Central Limit Theorem

Allan Kavuma

22 May 2019

Overview

This project involves investigating the exponential distribution in R and comparing it with the Central Limit Theorem. It illustrates via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials using the steps below:
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.

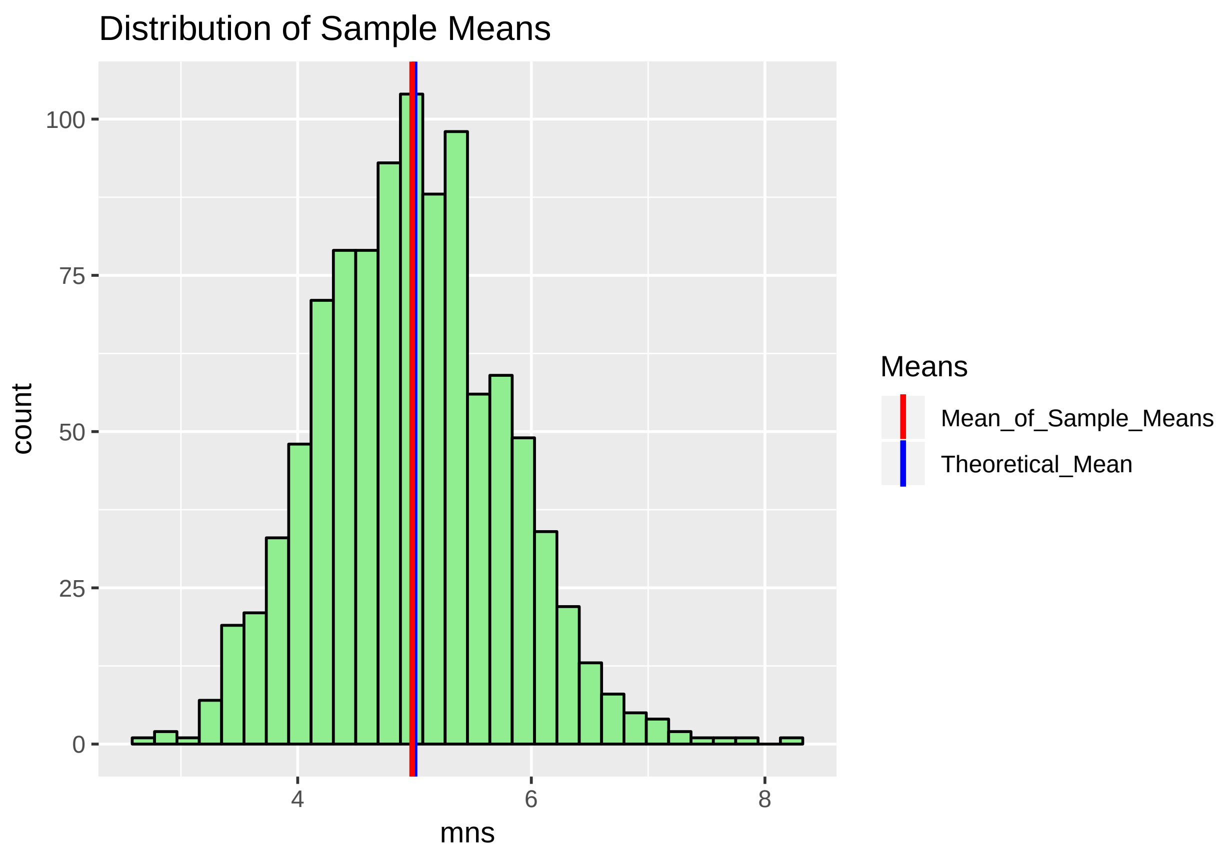
Create simulation distribution of averages of 40 exponentials

```
## Import library for graphs
library(ggplot2)
lbd = 0.2 ## lambda/rate variable
nosim = 1000 ## Number of simulations
obs = 40 ## Number of observations in each simulation
mns = NULL ## Initialize the varible to carry vector of simulated means
## Populate mns with means of 40-element grouped simulations
for(i in 1:1000) mns = c(mns, mean(rexp(n = obs, rate = lbd)))
```

1. Show the sample mean and compare it to the theoretical mean of the distribution.

```
meanSample <- mean(mns) ## Sample mean
meanTheory <- 1/lbd ## Theoretical mean = 1/lambda
meanDiff <- meanSample - meanTheory ## Difference between means

## Graph for the sample means
g <- ggplot(data = data.frame(mns), aes(x = mns))
g <- g + geom_histogram(aes(x = mns), fill = "lightgreen", colour = "black")
## Add lines showing meanSample and meanTheory on the graph
g + geom_vline(aes(xintercept = meanTheory, colour = "Theoretical_Mean"), size = 1) +
        geom_vline(aes(xintercept = meanSample, colour = "Mean_of_Sample_Means"), size = 1) +
        labs(title = "Distribution of Sample Means") +
         scale_color_manual(name = "Means", values = c(Theoretical_Mean = "blue", Mean_of_Sample_Means
```



The sample mean, highlighted in red, is 4.9799964; the theoretical mean, highlighted in blue, is 5; and the difference betweeen the two is -0.0200036. The two means are approximately equal. The Central Limit Theorem (CLT) states that given a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population. This is proved here.

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
varSample <- var(mns) ## Variance of the elements in mns i.e means of 40 simulations
varTheory <- ((1/lbd)^2) / 40 ## Theoretical variance = (SD^2)/n; n = 40
varDiff <- varSample - varTheory ## difference between variances
```

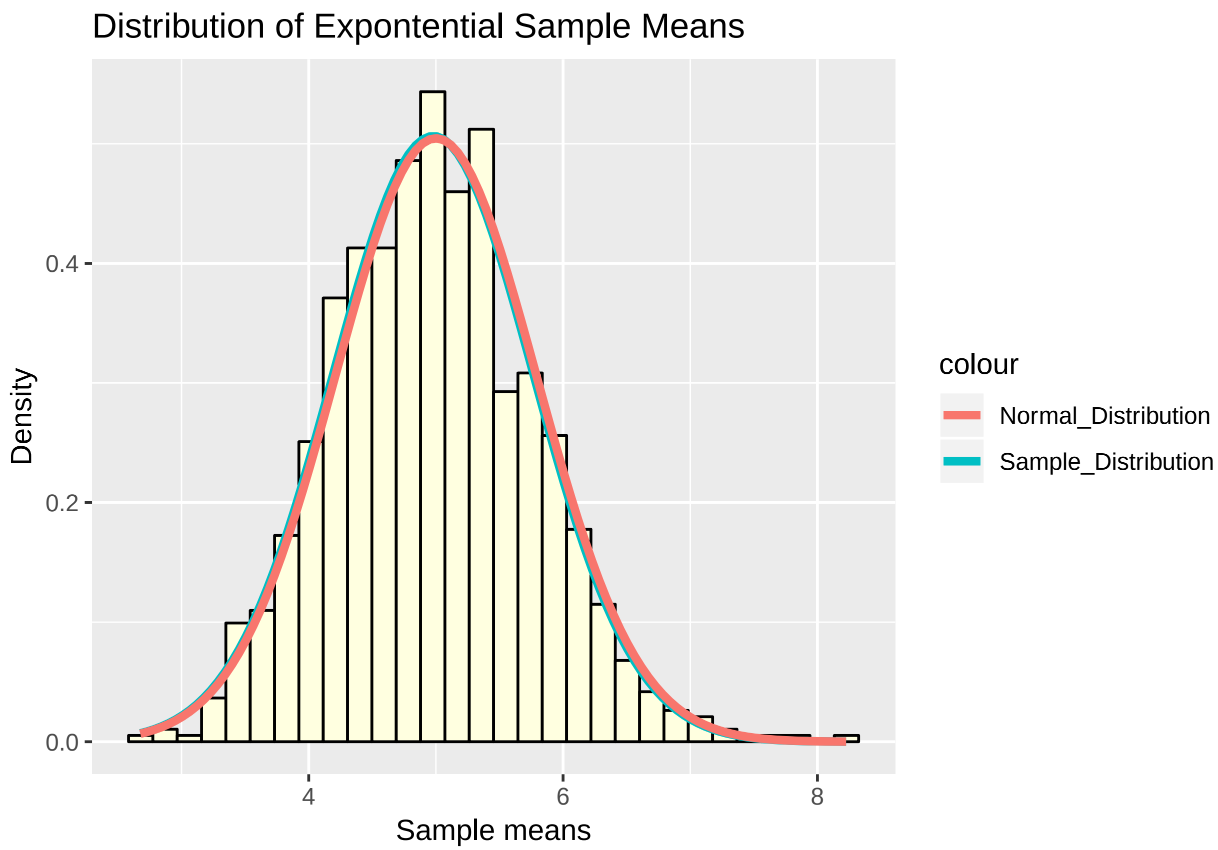
The sample variance is 0.6204997, the theoretical variance is 0.625, and the difference betweeen the two is -0.0045003. The two variances are approximately equal. This implies also that the standard deviation of the sample means is approximately equal to the standard error of the mean of the expontential distribution, hence proving the Central Limit Theorem.

3. Show that the distribution is approximately normal.

```
## Initiate parameters required to show approximation to normal distribution
sdTheory <- sqrt(varTheory) ## Theoretical standard deviation
sdSample <- sqrt(varSample) ## Sample standard deviation

## Plotting the graph for distribution of sample means
g <- ggplot(data = data.frame(mns), aes(x = mns))
```

```
## Construct histogram
g <- g + geom_histogram(aes(y = ..density..), colour = "black", fill = "lightyellow")
## Superimpose normal distribution with mean = meanSample and sd = sdSample
g <- g + stat_function(fun = dnorm, args = list(mean = meanSample, sd = sdSample), aes(colour = "Sample_
## Superimpose normal distribution with mean = meanTheory and sd = sdTheory
g <- g + stat_function(fun = dnorm, args = list(mean = meanTheory, sd = sdTheory), aes(colour = "Normal_
## Add title and axes labels
g + labs(title = "Distribution of Expontential Sample Means", y = "Density", x = "Sample means")
```



From the above graph, the distribution of means in the sample means set, mns, closely follows that of the normal distribution which is superimposed on it.

Conclusions:
- The averages are approximately normal with their distribution centered at the population mean. This proves the Central Limit Theorem
- Also, the standard deviation of the sample means is approximately equal to the actual theoretical standard error of the mean of the population, hence also proving the Central Limit Theorem