Confirming the Central Limit Theorem by Simulation of an Exponential Distribution

Desiré De Waele

10 maart 2016

Contents

Synopsis

If we were to randomly sample n values of an exponention distribution, take their mean, and repeat this process an infinite number of times, the Central Limit Theory states that the distribution of these means will approximate a normal distribution, regardless of the original distribution. The higher our n-value, the more a normal distribution would be approximated. This assignment likes to see the Central Limit Theorem confirmed in a simulation of the exponential distribution.

Simulation

I will simulate 1000 trials of sampling 40 samples of the exponential distribution with rate (λ) 0.2. The rexp function will generate these samples for me, the sapply function lets us take the mean for each of samples. But firstly, the seed is set for reproducability.

```
set.seed(6)
lambda = 0.2
n = 40
means <- data.frame(x = sapply(1:1000, function(x) mean(rexp(n, lambda))))
head(means)
```

```
##          x
## 1 3.500695
## 2 5.036769
## 3 5.223307
## 4 4.030061
## 5 5.243362
## 6 4.346943
```

To clarify, each of these means (6 of 1000 are given as an example) is the mean of a random sample of 40 values of the exponential distribution with lambda 0.2. Of course, we would like to see the Central Limit Theorem confirmed by our simulation. We can compare the mean, variance and shape of our generated sampling distribution with the mean, variance and shape proposed by the Central Limit Theorem

Comparing sampling mean with theoretical mean

The Central Limit Theorem states that the mean of any sampling distribution converges to the true mean of its population. Here we can calculate this true mean of the exponantion distribution, by simply taking the reverse of lambda:

```
theoreticalMean <- 1 / lambda
print(theoreticalMean)
```

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

If we calculate the mean of our sampling distribution (i.e. the mean of means), we'll see that it converges to the theoretical mean.

```
mean(means$x)
```

```
## [1] 4.950171
```

Comparing sampling variance with theoretical variance

The Central Limit Theorem states that the variance of a sampling distribution is simply the variance of the original (exponential) distribution, divided by the sample size. Our original standard deviation is the same as the mean, 1 / lambda, so we just have to square it to get the variance. We get:

```
theoreticalVariance <- (1/lambda)^2 / n
print(theoreticalVariance)
```

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

If we then calculate the variance of our own sampling data, we get a close approximation.

```
var(means$x)
```

```
## [1] 0.6639611
```

Comparing sampling shape with theoretical shape

If one were to explain the Central Limit Theorem, the theoretical normality of a sampling distribution would be the first thing to address. To compare the shape of our sampling data with the gaussian curve, I plot the data in a histogram and overlay it with a normal curve with our theoretical mean and variance calculated above. Note that the standard deviation is just the square root of the variance.
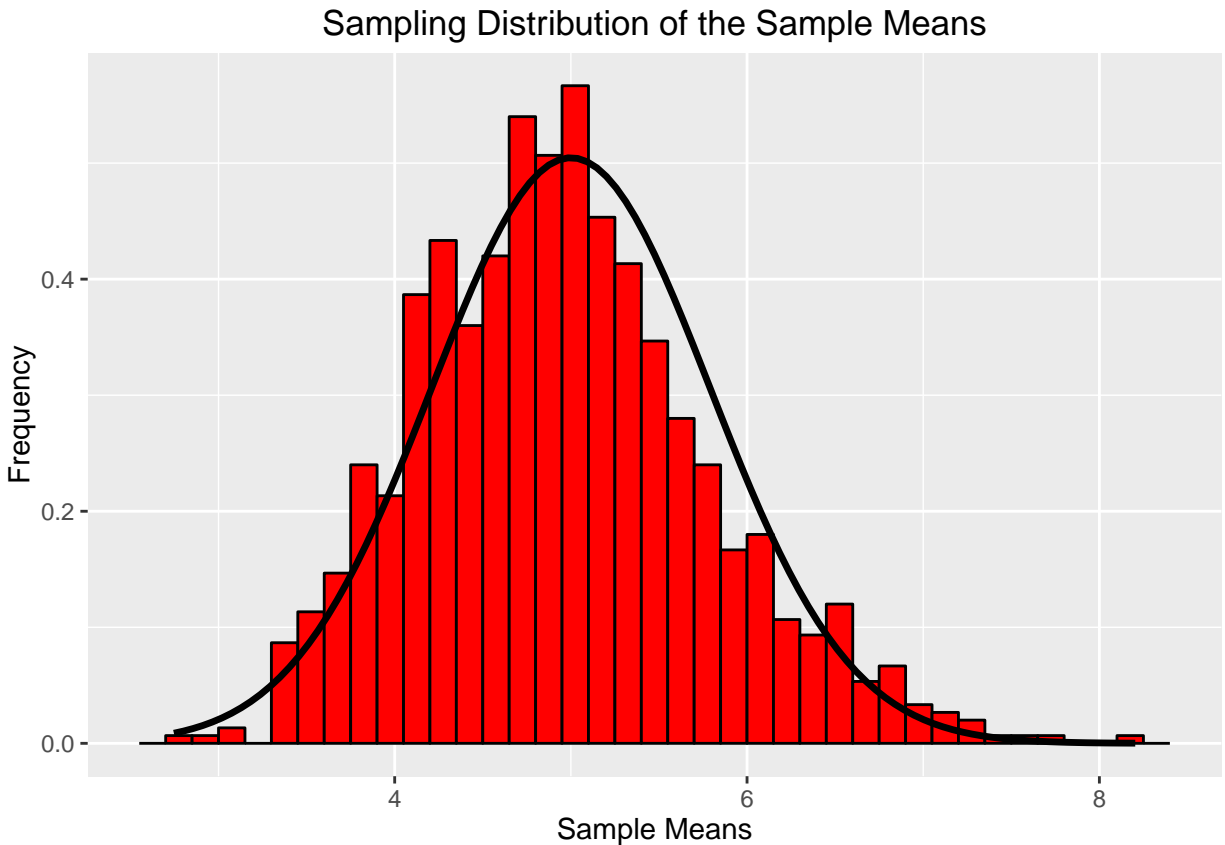
I use the ggplot package for plotting.

```
library(ggplot2)
ggplot(data = means, aes(x = x)) +
        geom_histogram(
                aes(y = ..density..),
                binwidth = 0.15,
                color = I("black"),
```

```
            fill = I("red")) +
    stat_function(
            fun = dnorm,
            arg = list(mean = theoreticalMean, sd = sqrt(theoreticalVariance)),
            size = 1.2) +
    labs(
            title = "Sampling Distribution of the Sample Means",
            x = "Sample Means",
            y = "Frequency")
```

Sampling Distribution of the Sample Means

We can see that our sampling distribution approximates the normal distribution reasonably well for a sample size of 40. It is important to note that a greater sample size would result in an even closer approximation of the normal distribution, whereas a smaller sample size would result the opposite. This makes sense, as a hypothetical sampling distribution with sample size 1 would just be the same as the original - in this case exponential - distribution.

In conclusion, the distribution of means of 40 exponentials behave as predicted by the Central Limit Theorem.