The Central Limit Theorem Applied to the Exponential Distribution in R

Kaushal Yadav

13 December 2018

Overview

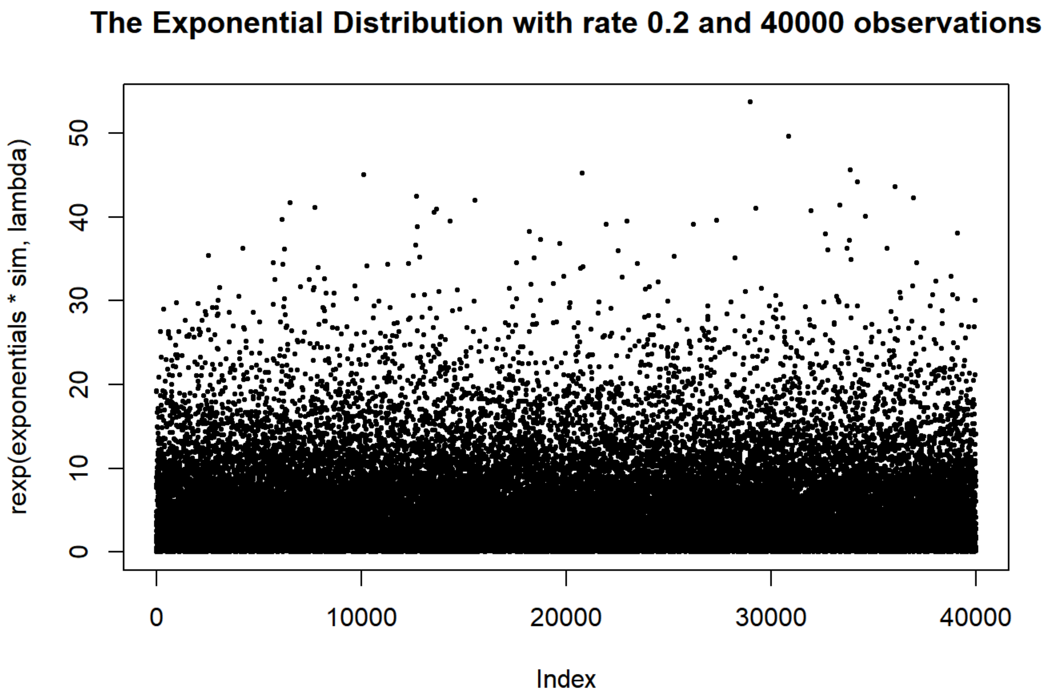
The purpose of this data analysis is to investigate the exponential distribution and compare it to the Central Limit Theorem. For this analysis, the lambda will be set to 0.2 for all of the simulations. This investigation will compare the distribution of averages of 40 exponentials over 1000 simulations.

Simulations

Set the simulation variables lambda, exponentials, number of simulations and seed.

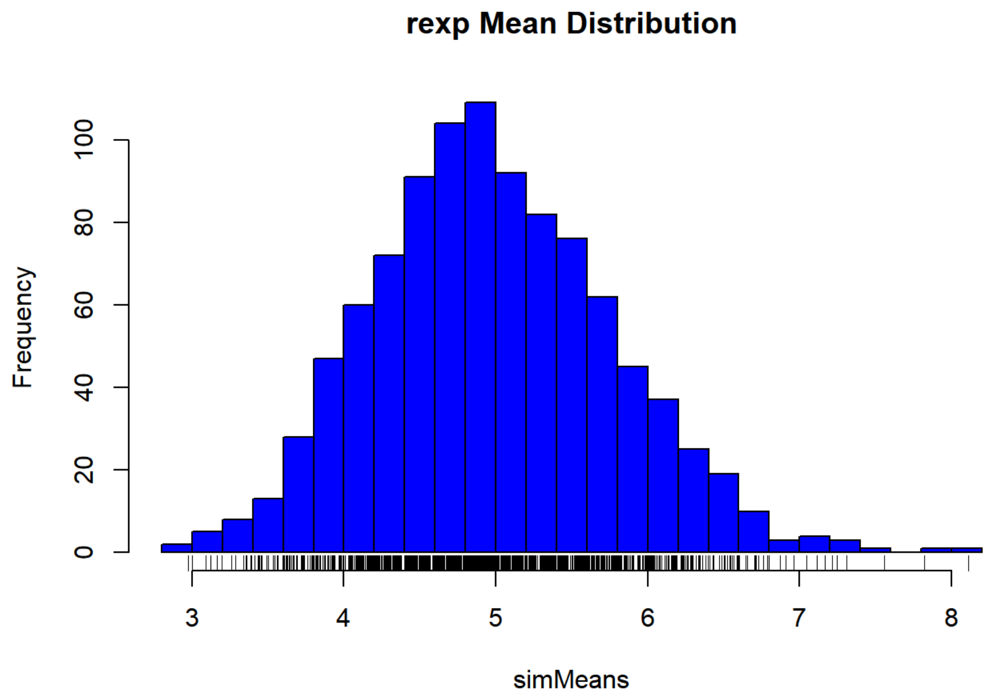
```
ECHO=TRUE
set.seed(157)
lambda = 0.2
exponentials = 40
sim = 1000

#the exponential distribution
plot(rexp(exponentials*sim, lambda), pch = 20, cex = 0.6, main = "The Exponential Distribution with rate 0.2 
and 40000 observations")
```



Run Simulations with variables and plots the distribution

```
simMeans = NULL
for (i in 1 : 1000) simMeans = c(simMeans, mean(rexp(exponentials, lambda)))
hist(simMeans, col = "blue", main = "rexp Mean Distribution", breaks = 20)
rug(simMeans)
```



Sample Mean versus Theoretical Mean

Sample Mean

Calculating the mean from the simulations with give the sample mean.

```
mean(simMeans)
```

```
## [1] 4.978361
```

Theoretical Mean

The theoretical mean of an exponential distribution is lambda^-1.

```
lambda^-1
```

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

Comparison

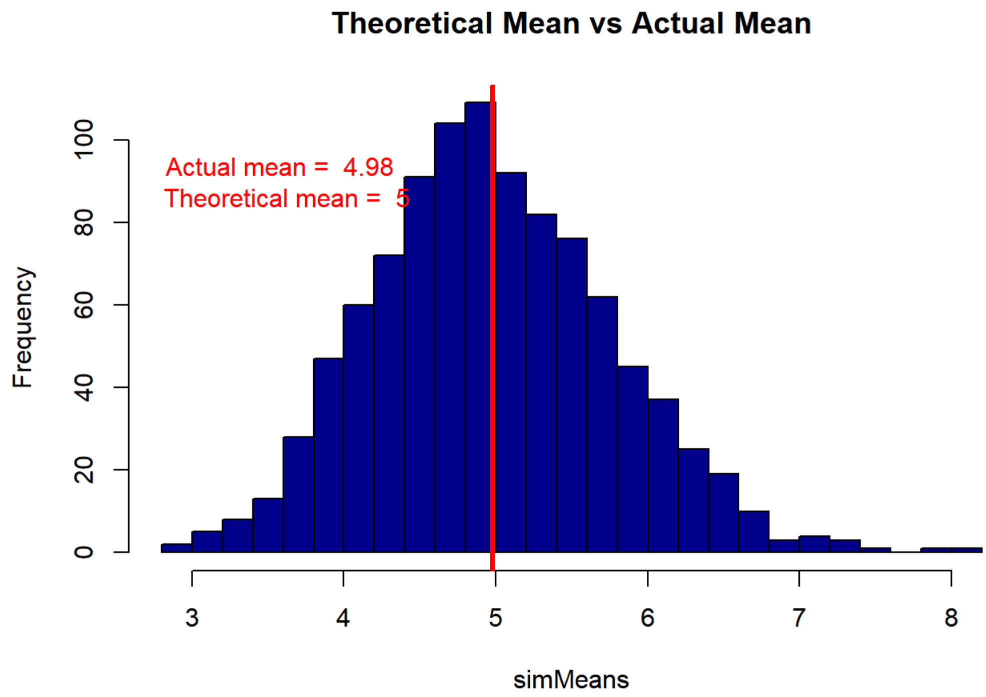
There is only a slight difference between the simulations sample mean and the exponential distribution theoretical mean.

```
abs(mean(simMeans)-lambda^-1)
```

```
## [1] 0.02163885
```

Plot Comparing Theoretical Mean vs Sample Mean

```
hist(simMeans, col="darkblue", main="Theoretical Mean vs Actual Mean", breaks=20)
abline(v=mean(simMeans), lwd="3", col="red")
text(3.6, 90, paste("Actual mean = ", round(mean(simMeans),2), "\n Theoretical mean = ", round(lambda^-1,2)
), col="red")
```



Sample Variance versus Theoretical Variance

Sample Variance

Calculating the variance from the simulation means with given sample variance.

```
var(simMeans)
```

```
## [1] 0.619628
```

Theoretical Variance

The theoretical variance of an exponential distribution is (1/lambda)^2/exponentials.

```
(1/lambda)^2/exponentials
```

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

Sample Standard Deviation versus Theoretical Standard Deviation

Sample Standard Deviation

Calculating the standard deviation for the simulation means.

```
sd(simMeans)
```

```
## [1] 0.7871646
```

Theortical Standard Deviation

The theoretical standard deviation of an exponential distribution is sqrt((1/lambda)^2/exponentials).

```
sqrt((1/lambda)^2/exponentials)
```

```
## [1] 0.7905694
```

Comparison

There is only a slight difference between the simulations sample variance and the exponential distribution theoretical variance.

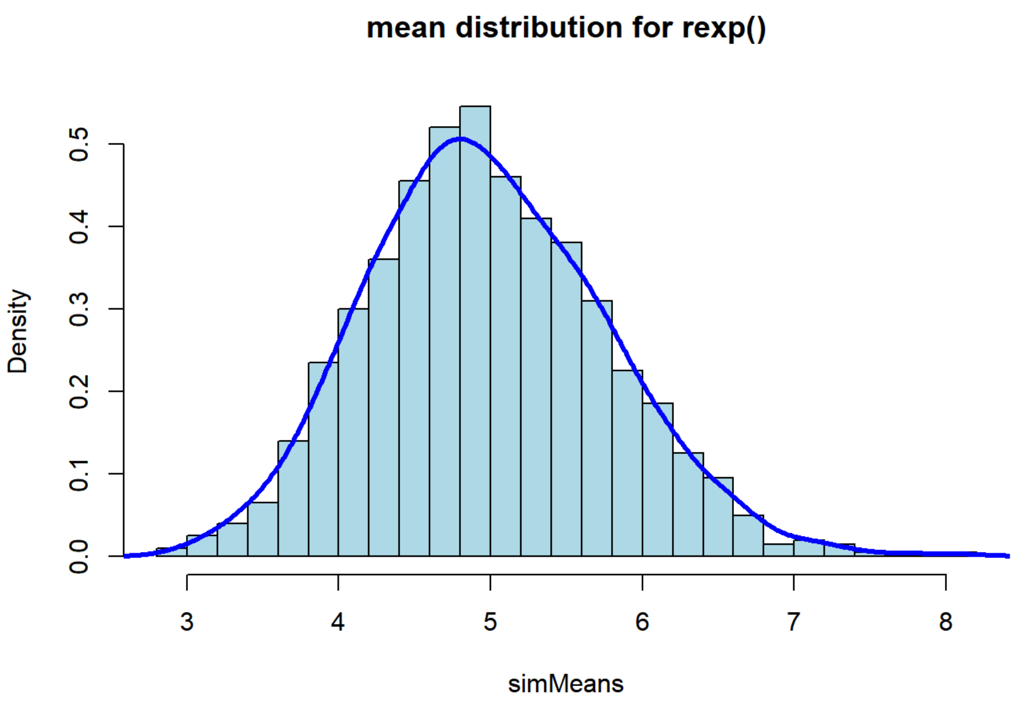
```
abs(sd(simMeans)-sqrt((1/lambda)^2/exponentials))
```

```
## [1] 0.003404856
```

Distribution

This is a density histogram of the 1000 simulations. There is an overlay with a normal distribution that has a mean of lambda^-1 and standard deviation of (lambda*sqrt(n))^-1, the theoretical normal distribution for the simulations.

```
hist(simMeans, prob=TRUE, col="lightblue", main="mean distribution for rexp()", breaks=20)
lines(density(simMeans), lwd=3, col="blue")
```



As shown in the graph, the calculated distribution of means of random sampled exponantial distributions overlaps with the normal distribution, due to the Central Limit Theorem. The more samples we would get (now 1000), the closer will the density distribution be to the normal distribution bell curve.

Confidence Interval Comparison

Check the Confidence Interval levels to see how they compare.

Sample CI

Calculate the sample confidence interval; sampleCI = mean of x plus or minus the .975th normal quantile times the standard error of the mean standard deviation of x divided by the square root of n (the length of the vector x).

```
sampleConfInterval <- round (mean(simMeans) + c(-1,1)*1.96*sd(simMeans)/sqrt(exponentials),3)
sampleConfInterval
```

```
## [1] 4.734 5.222
```

Theoretical CI

Calculate the theoretical confidence interval; theoCI = theoMean of x plus or minus the .975th normal quantile times the standard error of the mean standard deviation of x divided by the square root of n (the length of the vector x).

```
theoConfInterval <- lambda^-1 + c(-1,1) * 1.96 * sqrt(lambda^-1)/sqrt(exponentials)
theoConfInterval
```

```
## [1] 4.307035 5.692965
```

The sample confidence interval is 4.749 5.254 and the theoretical confidence level is 4.755 5.245. The confidence levels also match closely. Again, proving the distribution is approximately normal.

Conclusion

It is determined that the distribution does indeed demonstrate the Central Limit Theorem; a bell curve. After graphing all the values above and comparing the confidence intervals the distribution is approximately normal.