Investigate Exponential Distribution and Compare with CLT

Sachin Raje

26 Dec 2016

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.

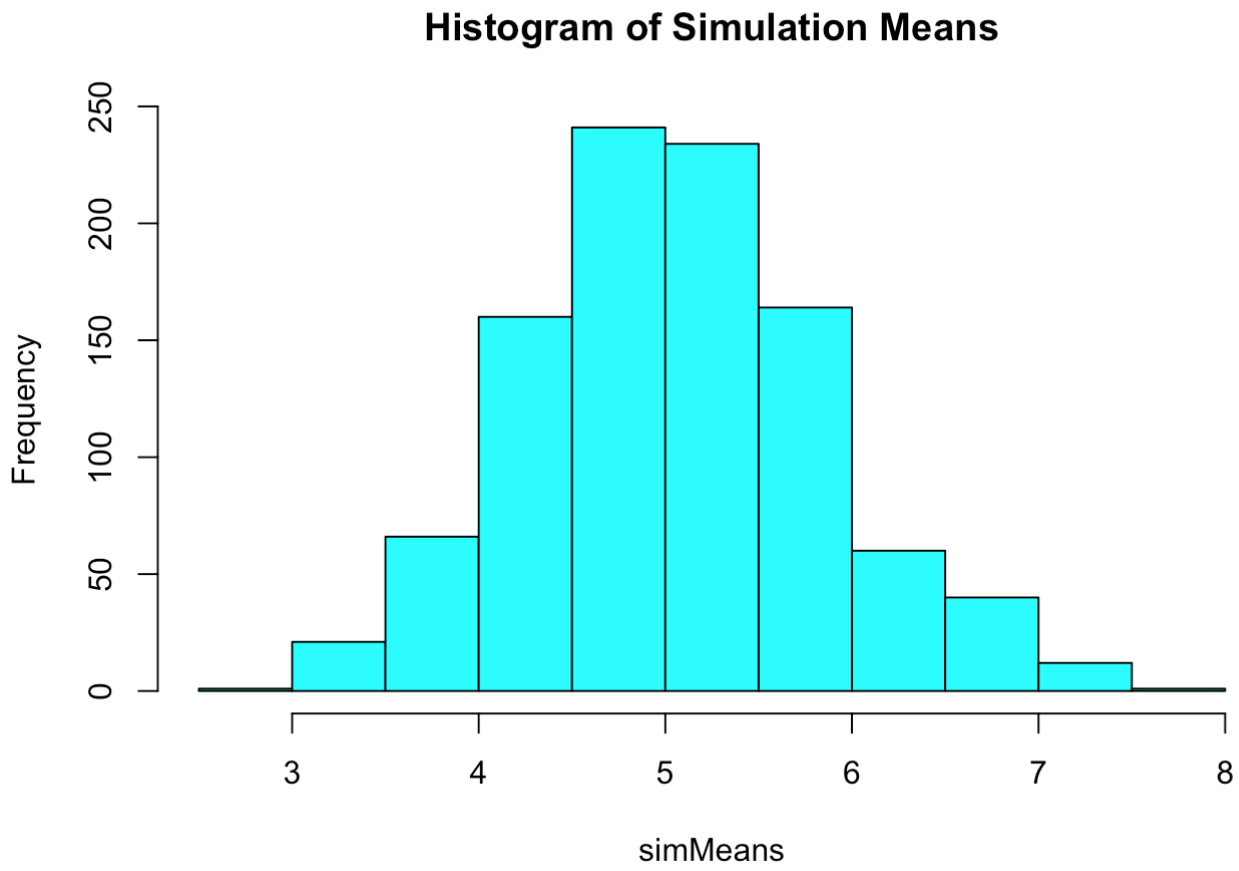
Simulations

Set the simulation variables lambda, exponentials, and seed.

```
set.seed(1337)
lambda = 0.2
exponentials = 40
```

Run Simulations with variables

```
simMeans = NULL
for (i in 1 : 1000) simMeans = c(simMeans, mean(rexp(exponentials, lambda)))
hist(simMeans, breaks = 10, col = "cyan", main="Histogram of Simulation Means")
```



Sample Mean versus Theoretical Mean

Sample Mean

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

```
mean(simMeans)
```

```
## [1] 5.055995
```

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] 0.05599526
```

Sample Variance versus Theoretical Variance

Sample Variance

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

```
var(simMeans)
```

```
## [1] 0.6543703
```

Theoretical Variance

The theoretical variance of an exponential distribution is (lambda * sqrt(n))^-2.

```
(lambda * sqrt(exponentials))^-2
```

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

Comparison

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

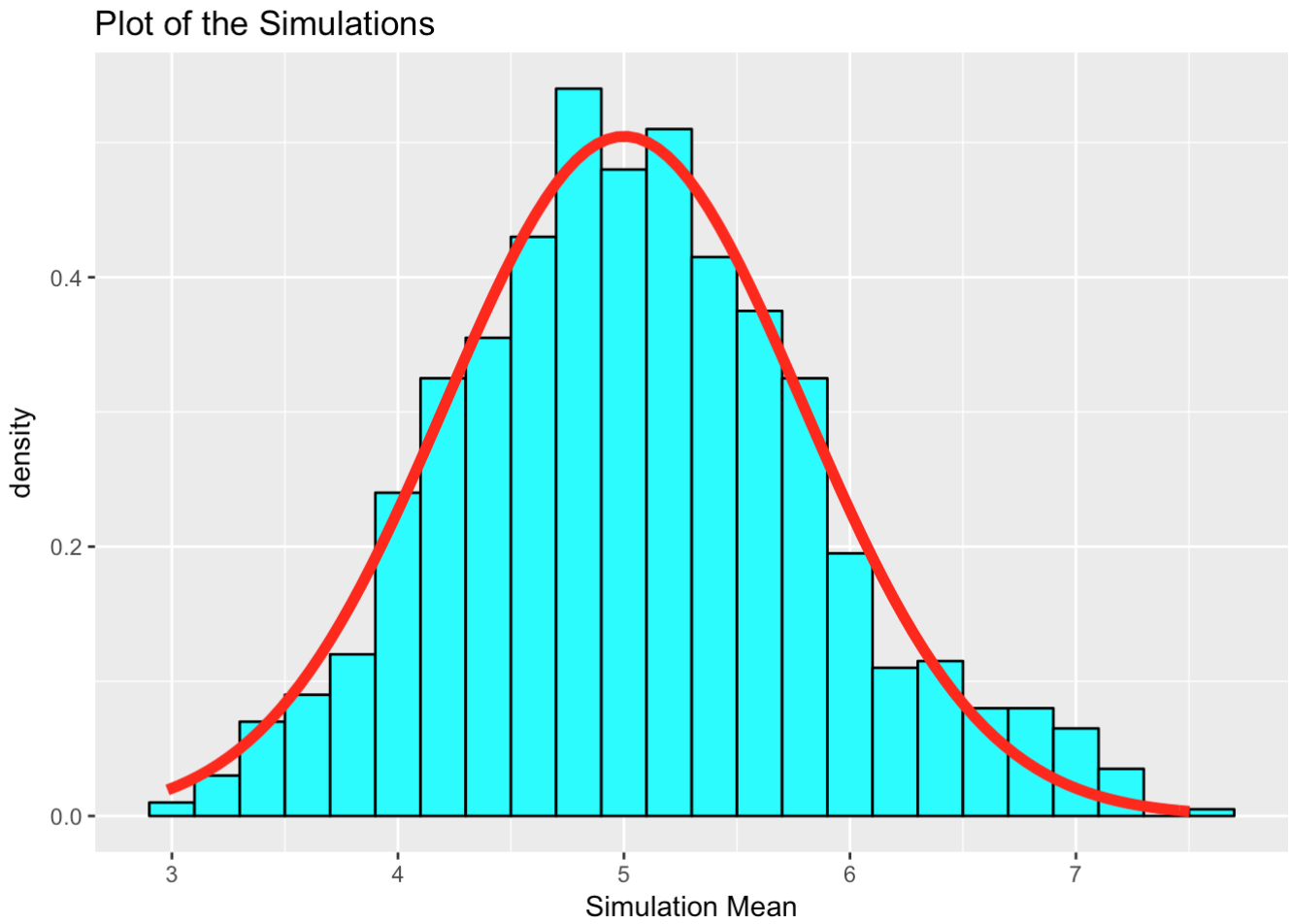
```
abs(var(simMeans)-(lambda * sqrt(exponentials))^-2)
```

```
## [1] 0.0293703
```

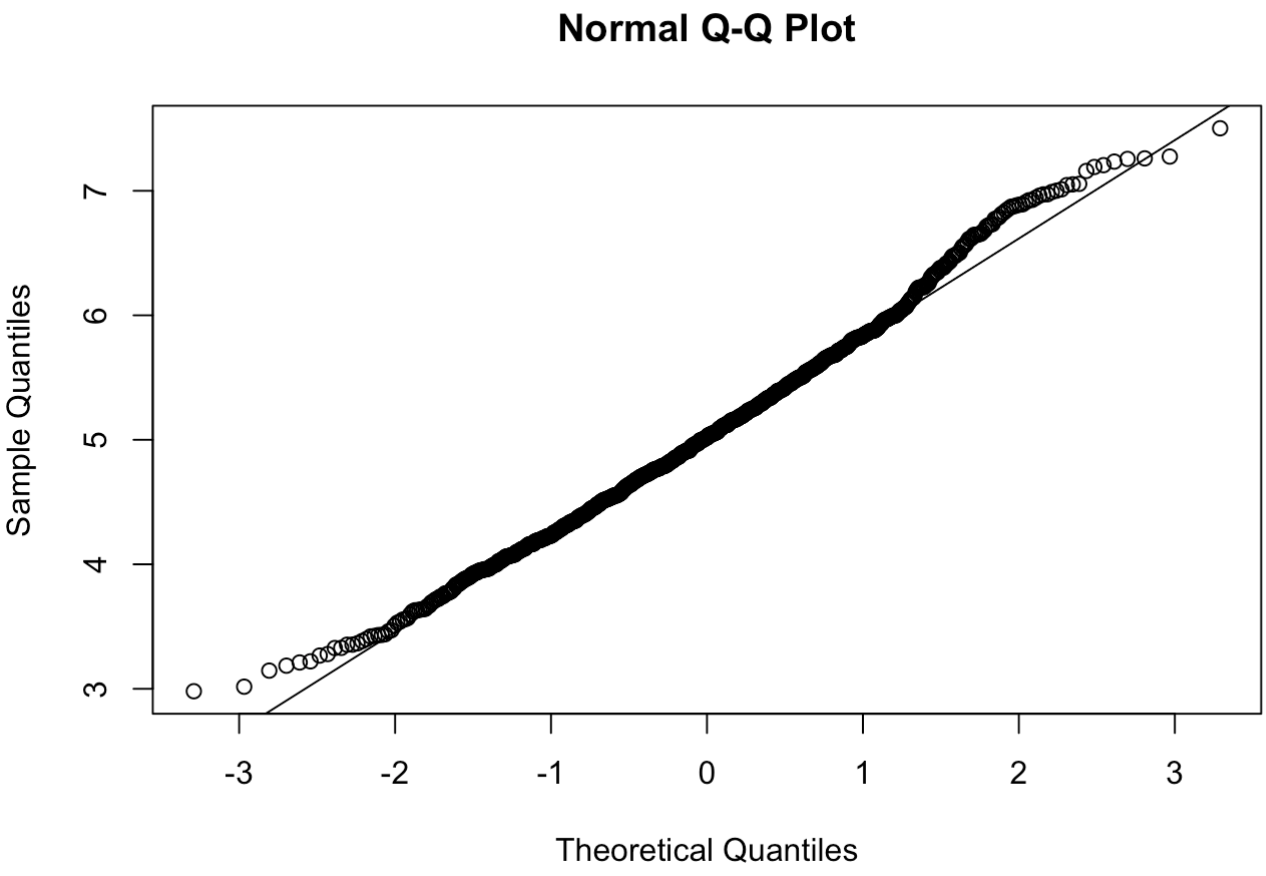
Distribution

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.

```
library(ggplot2)
ggplot(data.frame(y=simMeans), aes(x=y)) +
  geom_histogram(aes(y=..density..), binwidth=0.2, fill="cyan", color="black") +
  stat_function(fun=dnorm, args=list(mean=lambda^-1, sd=
(lambda*sqrt(exponentials))^-1), size=2, colour="red") +
  labs(title="Plot of the Simulations", x="Simulation Mean")
```




```
qqnorm(simMeans); qqline(simMeans)
```



The theoretical quantiles also match closely with the actual quantiles. These evidences prove that the distribution is approximately normal.