Exponential distribution in R versus the Central Limit Theorem

James Portman

July 16, 2016

Overview

In this project we 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. We will investigate the distribution of averages of 40 exponentials over a thousand simulations.

We will:

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.

Simulation

```
library(ggplot2)
set.seed(1969)

lambda = 0.2 ## rate parameter.
number_of_Distributions = 40
number_of_Simulations = 1000

# Create a matrix with number_of_Simulations rows and number_of_Distributions columns.
simulation_matrix <- matrix(
                            rexp(number_of_Simulations * number_of_Distributions, rate=l
ambda),
                            number_of_Simulations,
                            number_of_Distributions)
simulation_means <- rowMeans(simulation_matrix)
```

See figure 1 for a histogram of simulation means frequency.

Sample Mean versus Theoretical Mean

The actual mean for the sample data and theoretical mean are calculated below:

```
actual_mean  <- mean(simulation_means)
theoretical_mean <- 1/lambda

actual_mean
```

```
## [1] 4.984071
```

```
theoretical_mean
```

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

Actual center of the distribution based on the simulations is 4.984071 while the theoretical mean for lambda = 0.2 is 5. The actual mean for the sample data is very close to the theoretical mean of normal data.

Sample Variance versus Theoretical Variance

The actual variance for the sample data and theoretical variance are calculated below:

```
actual_variance <- var(simulation_means)
theoretical_variance <- (1/lambda)^2/number_of_Distributions

actual_variance
```

```
## [1] 0.5947001
```

```
theoretical_variance
```

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

Actual variance for the sample data is 0.59470011 while the theoretical variance is 0.625. The actual variance for the sample data is very close to the theoretical variance of normal data.

Distribution

To explain how we can tell the distribution is approximately normal, we will do the following steps:

1. Create an approximate normal distribution and see how our sample data aligns with it.
2. Compare the confidence interval along with the mean and variance with normal distribution.
3. Create a q-q plot for quantiles.

Step 1

```
df_simulation_means <- data.frame(simulation_means);
the_plot <- ggplot(df_simulation_means, aes(x =simulation_means))
the_plot <- the_plot + geom_histogram(aes(y=..density..), colour="black", fill =
"green")
```

The plot shown in Figure 2 indicates that the histogram can be approximated with the normal distribution.

Step 2

```
standard_normal_deviate <- 1.96
actual_conf_interval <- round (mean(simulation_means) +
                        c(-1,1)*standard_normal_deviate*sd(simulation_means)/sqrt(number
_of_Distributions),3)
theoretical_conf_interval <- theoretical_mean +

c(-1,1)*standard_normal_deviate*sqrt(theoretical_variance)/sqrt(number_of_Distributions)

actual_conf_interval
```

```
## [1] 4.745 5.223
```

```
theoretical_conf_interval
```

```
## [1] 4.755 5.245
```

The actual 95% confidence interval of [4.755, 5.223] is very close to the theoretical 95% confidence interval of [4.755, 5.245]

Step 3

Figure 3 shows that the theoretical quantiles closely match the actual quantiles.

Appendix

Figure 1: Histogram of simulation means frequency.

```
hist(simulation_means, col = "green", main = "Histogram of Simulation Means", xlab = "Si
mulation Means")
```

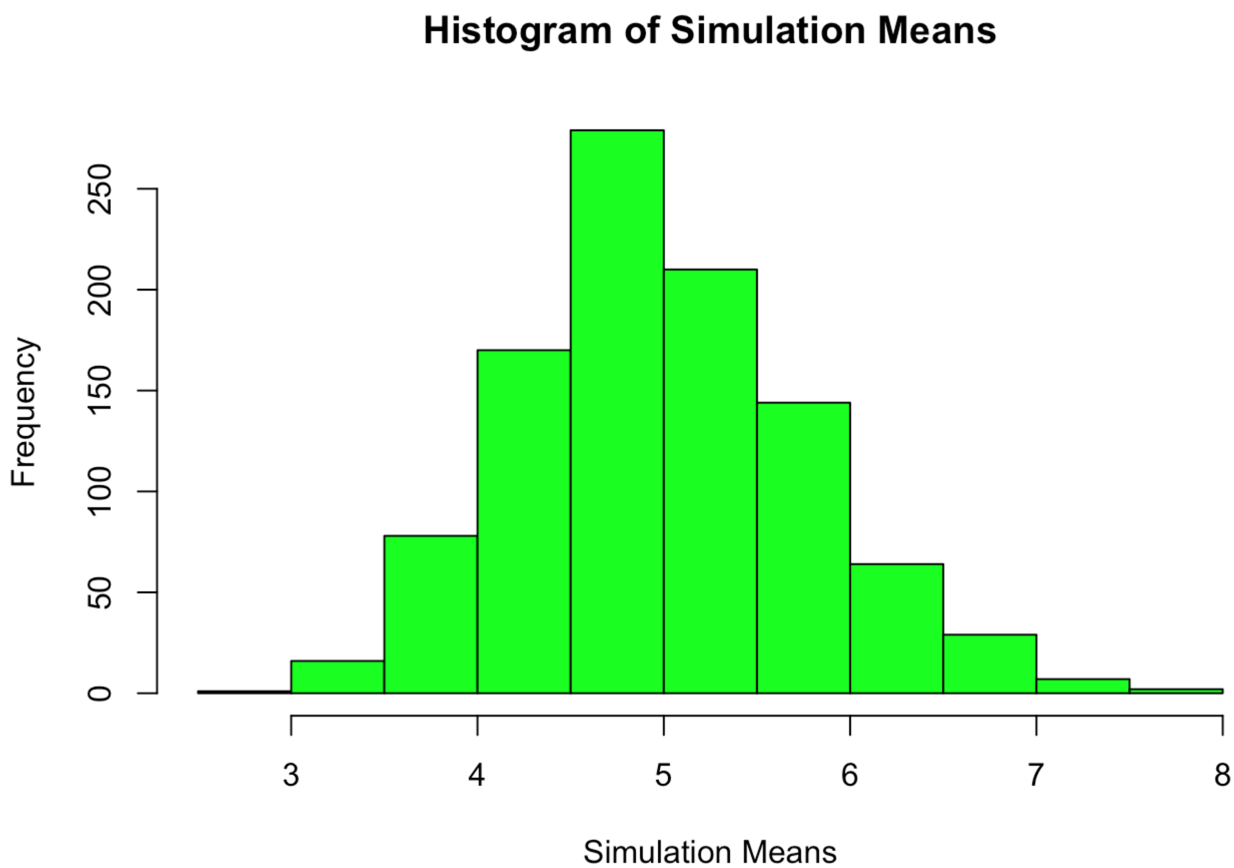


Figure 2: Approximate normal distribution superimposed over sample data.

```
the_plot + geom_density(colour="blue", size=1)
```

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

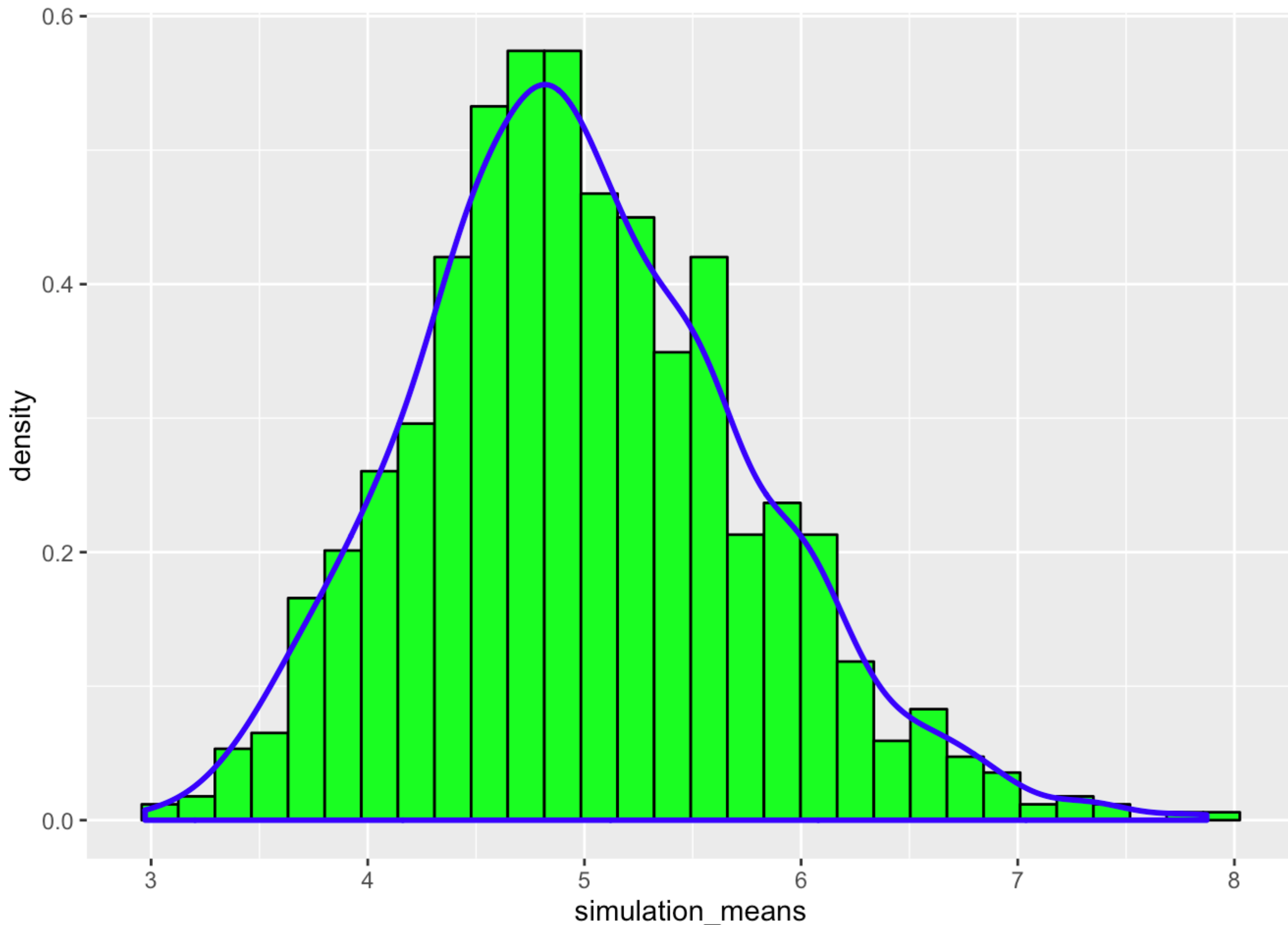
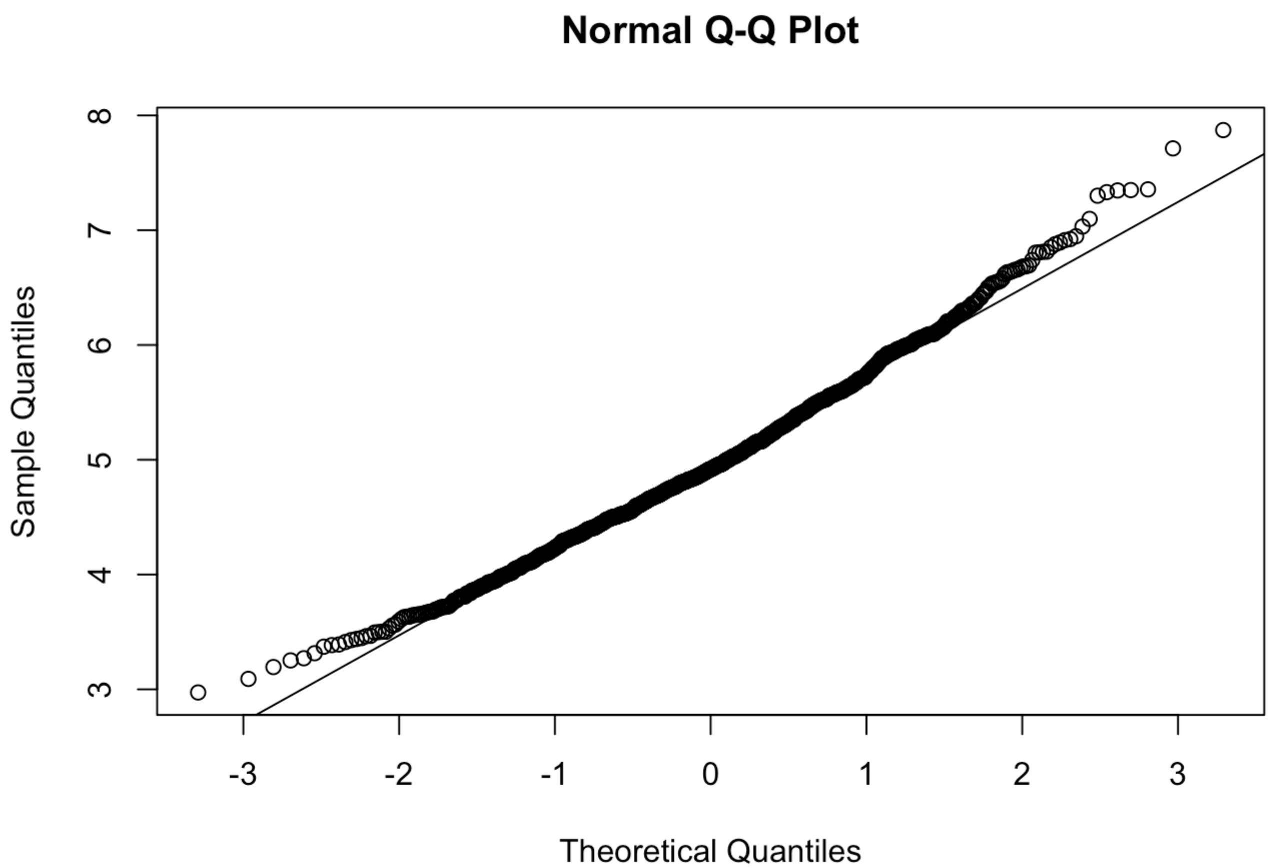


Figure 3: Theoretical quantiles closely match the actual quantiles

```
qqnorm(simulation_means)
qqline(simulation_means)
```


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
3
4
5
6
7
8
-3
-2
-1
0
1
2
3