Investigating the Exponential Distribution

April 25, 2015

Overview

In this project we 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.

Please note, instead of separating code and diagrams into the appendix, I chose to use the "knit and weave" approach of mixing my explanations and diagrams. However, in total this report should meet the page maximum (less than 6 total).

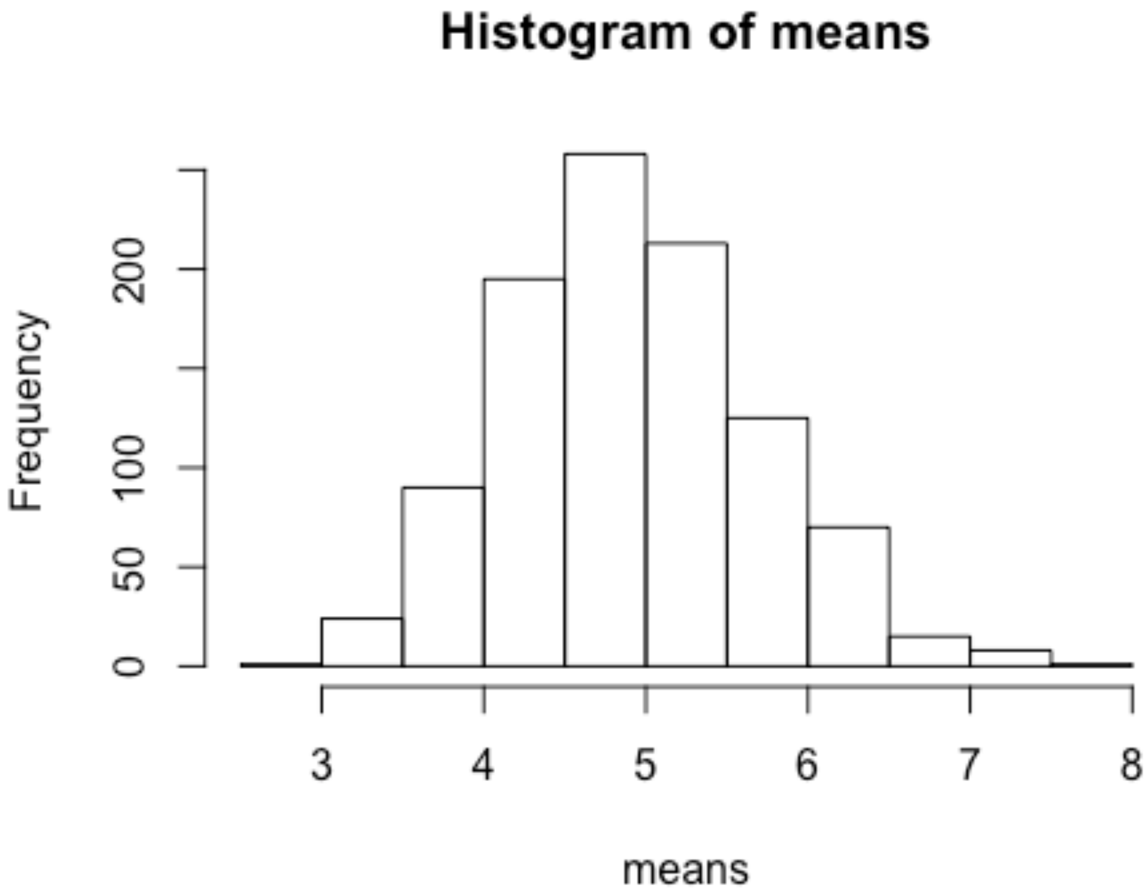
Simulations Code

Here we are running 1000 simulations of taking the mean of 40 samples with lambda = 0.2 and storing those in "means" as well as calculating the variance for each sample to store them in a variable "vars".

```r
means <- NULL
vars <- NULL
for (i in 1:1000) {
    sim <- rexp(40,.2)
    means <- c(means, mean(sim))
    vars <- c(vars, var(sim))
}
```

Sample Mean versus Theoretical Mean

```r
hist(means)
```



As you can see in the "Histogram of means" above, the distribution is uni-modal and appears to be normally distributed around 5. We can get the actual sample mean by taking the mean of the means:

```
mean(means)

## [1] 4.915909
```

The theoretical mean is 1/lambda, and we compute this with .2 as lambda

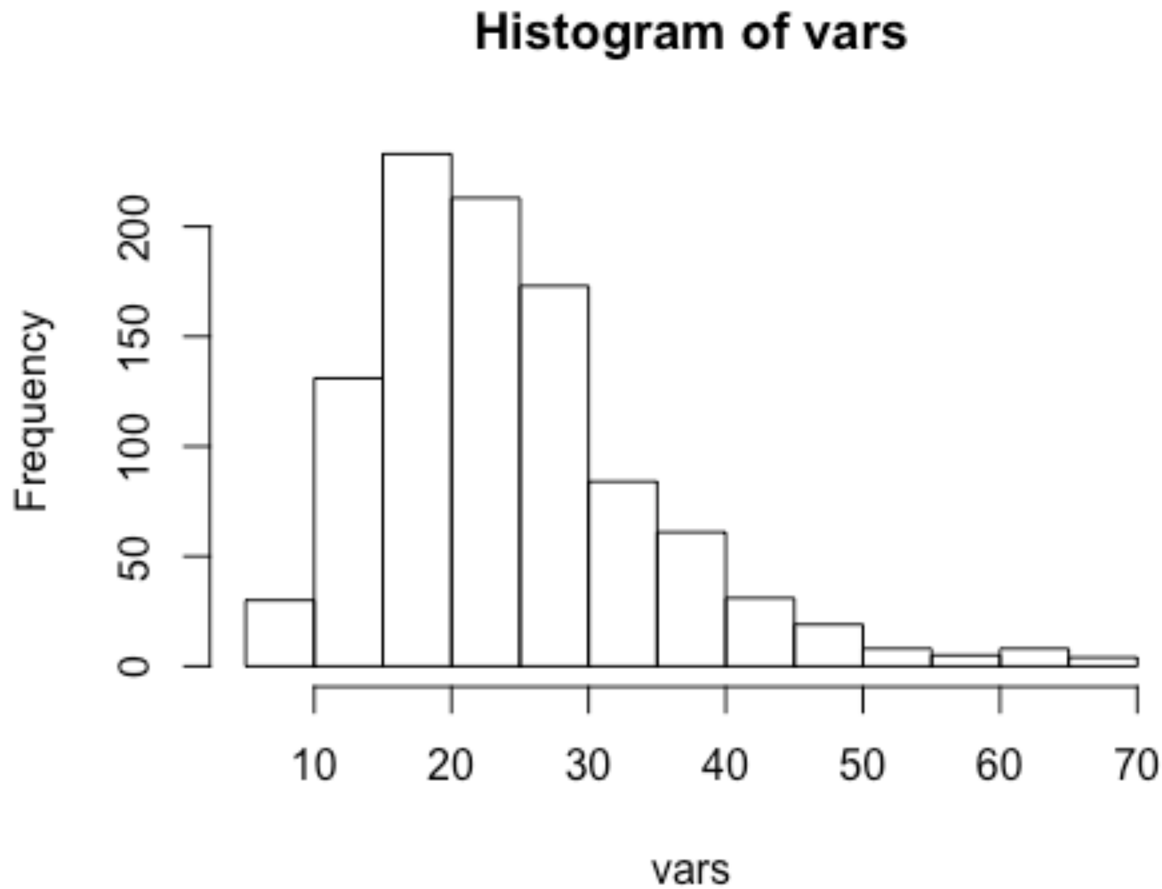
```
1/.2

## [1] 5
```

The theoretical mean is 5 and the mean of our sample means is very, very close to 5.

Sample Variance versus Theoretical Variance

```
hist(vars)
```




```r
mean(vars)

## [1] 24.21085
```

Above we plot a histogram of the variances for each sample of 40 random exponentials. From the plot, you can see that the distribution centers around ~22. Taking the mean of the variances equals about 25. We can check this against the theoretical variance, which is given by (1/lambda)^2 and we see the results are very, very close.

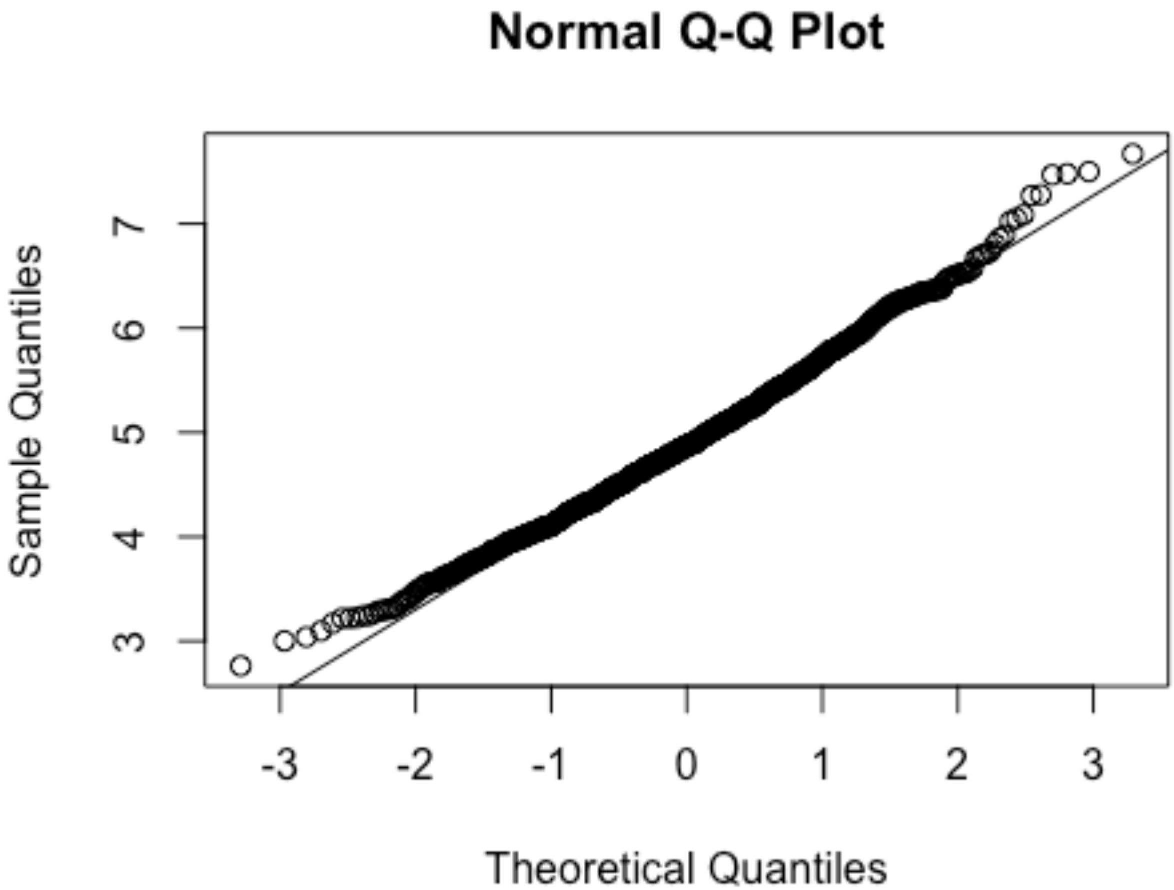
```r
(1/.2)^2

## [1] 25
```

Distribution

As mentioned above, the histogram does appear to be normal. But we can confirm this by using a QQ plot.

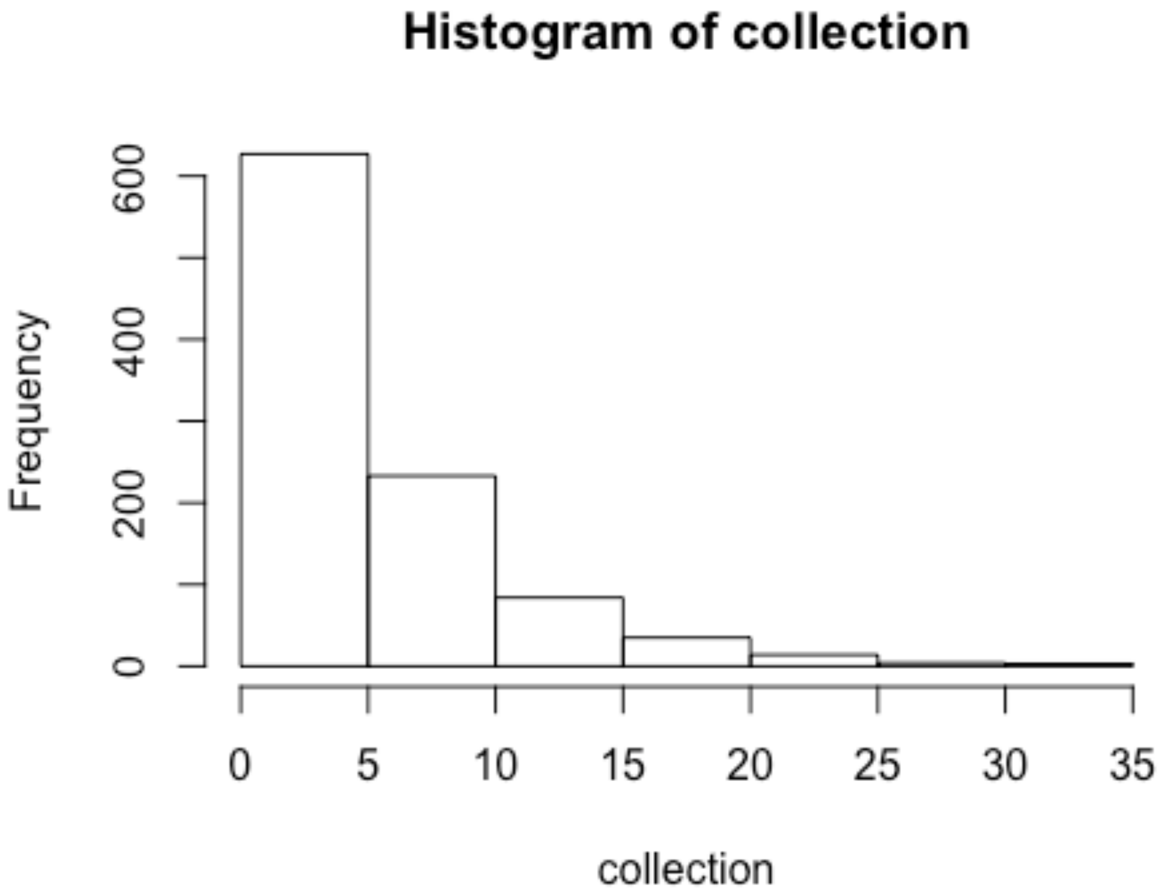
```r
qqnorm(means)
qqline(means)
```



As you can see in the plot above, the data follows the line very closely, which some deviation at the tails. Based on this graph, we can confirm that this data is approximately normal.

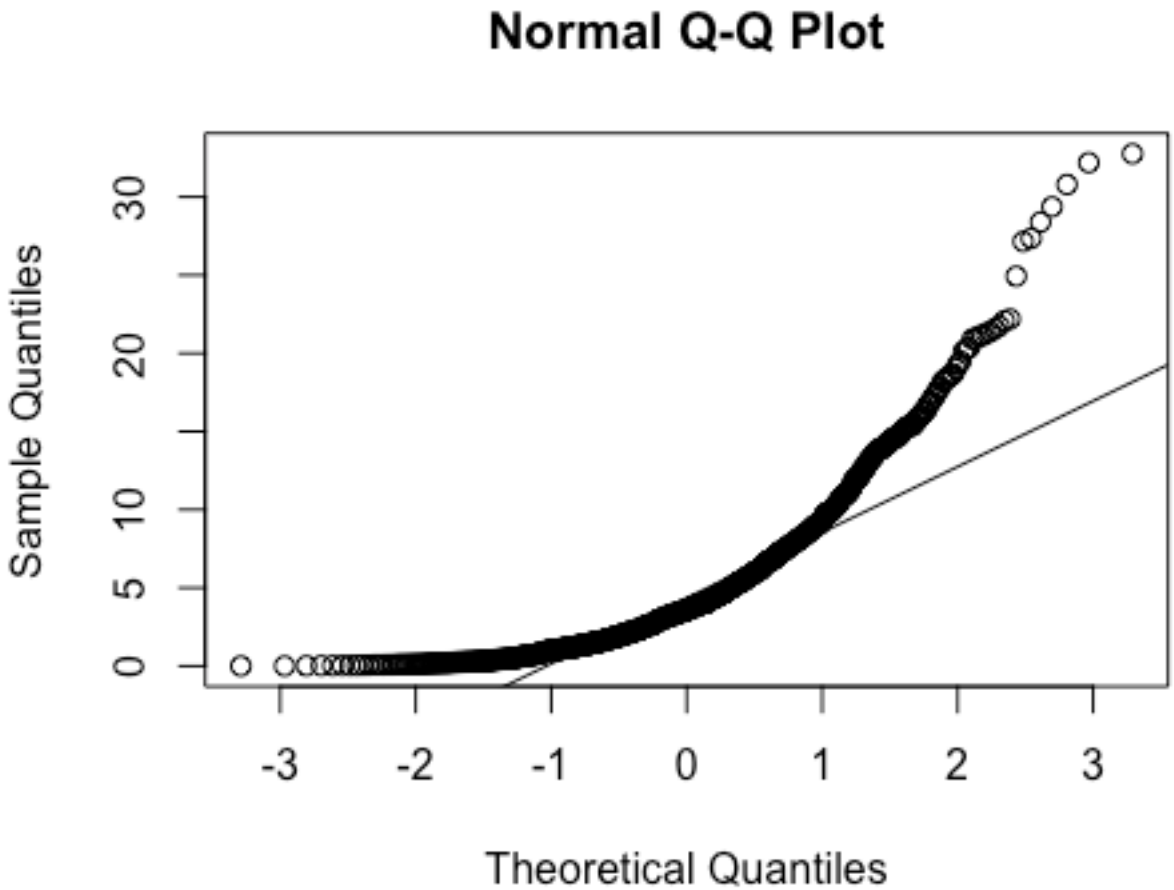
We can further compare this to the distribution of a large collection of random exponentials (as opposed to our data set, which is the distribution of a large collection of averages of 40 exponentials).

```
collection <- rexp(1000,.2)
hist(collection)
```



This distribution of 1000 random exponentials is not at all normally distributed. It's right skewed with data starting a zero. We can further show this distribution's lack of normality with a QQ plot:

```
qqnorm(collection)
qqline(collection)
```



Normal Q-Q Plot
Sample Quantiles
0
5
10
20
30
-3
-2
-1
0
1
2
3
Theoretical Quantiles