Distribution of Exponential Function Simulations

Jon Taylor

September 20, 2016

Overview

In this experiment we will show that the exponential distribution follows the central limit theorem by showing that the mean of sets of 40 independent random variables from an exponential function is normally distributed and converges on a value equal to that of the theoretical mean of the exponential function.

Simulations

First we'll set the rate parameter of the exponential function, lambda, to 0.2. Then we'll create variables to store the number of variables (40) in each exponential function and the total number of simulations (1000) to be run. We'll also set the mean and standard deviation of the exponential distribution to 1/lambda.

```
#Set lambda = 0.2 for all of the simulations
lambda <- 0.2

#You will investigate the distribution of averages of 40 exponentials.
#Note that you will need to do a thousand simulations.
n.exp <- 40
n.sim <- 1000

#The mean of exponential distribution is 1/lambda and the
#standard deviation #is also 1/lambda.

mean.dist.exp <- 1/lambda
sd.dist.exp <- 1/lambda
```

Let's take a look at a simulation exponential function. It's mean and standard deviation have an expected value of 1/lambda or 5. And it should result in a histogram skewed to the right.

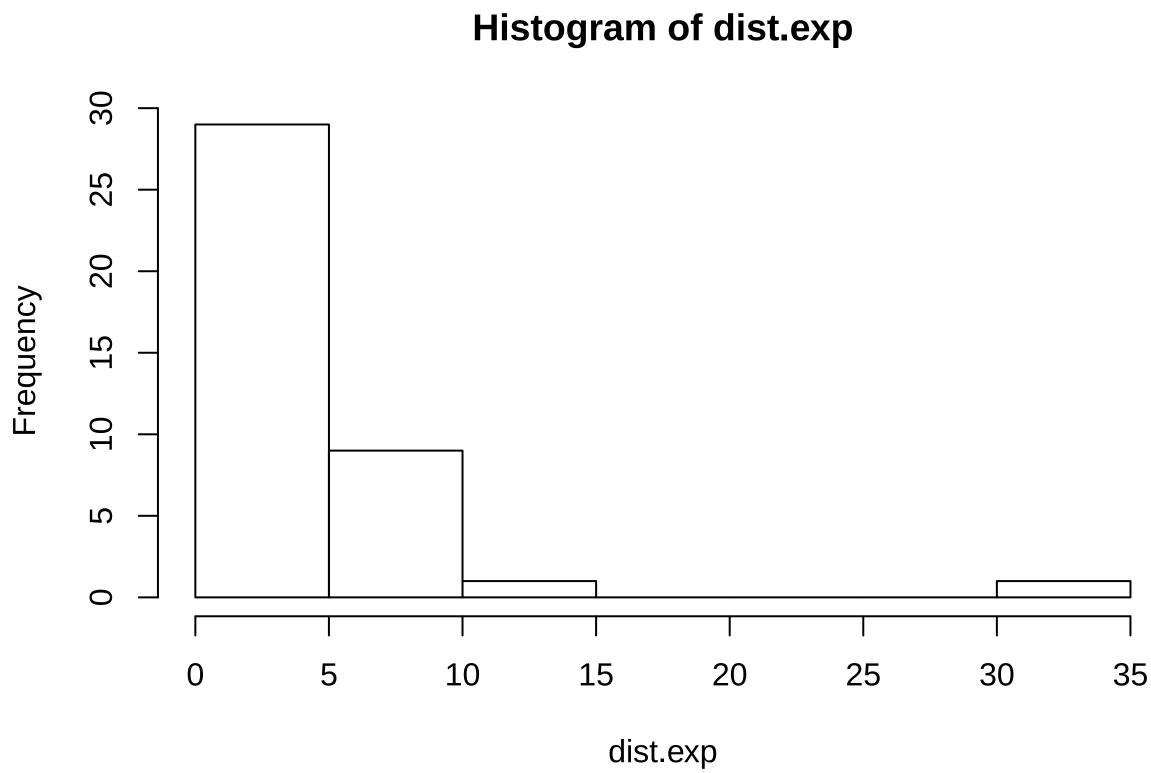
```
#The exponential distribution can be simulated in R with rexp(n, lambda)
#where lambda is the rate parameter.

dist.exp <- rexp(n.exp,lambda)

#The sample mean is
mean(dist.exp) # 1/lambda exopected value
```

```
## [1] 4.370812
```

```
#The sd is
sd(dist.exp) #1/lambda exopected value
```

```
## [1] 5.843897
```

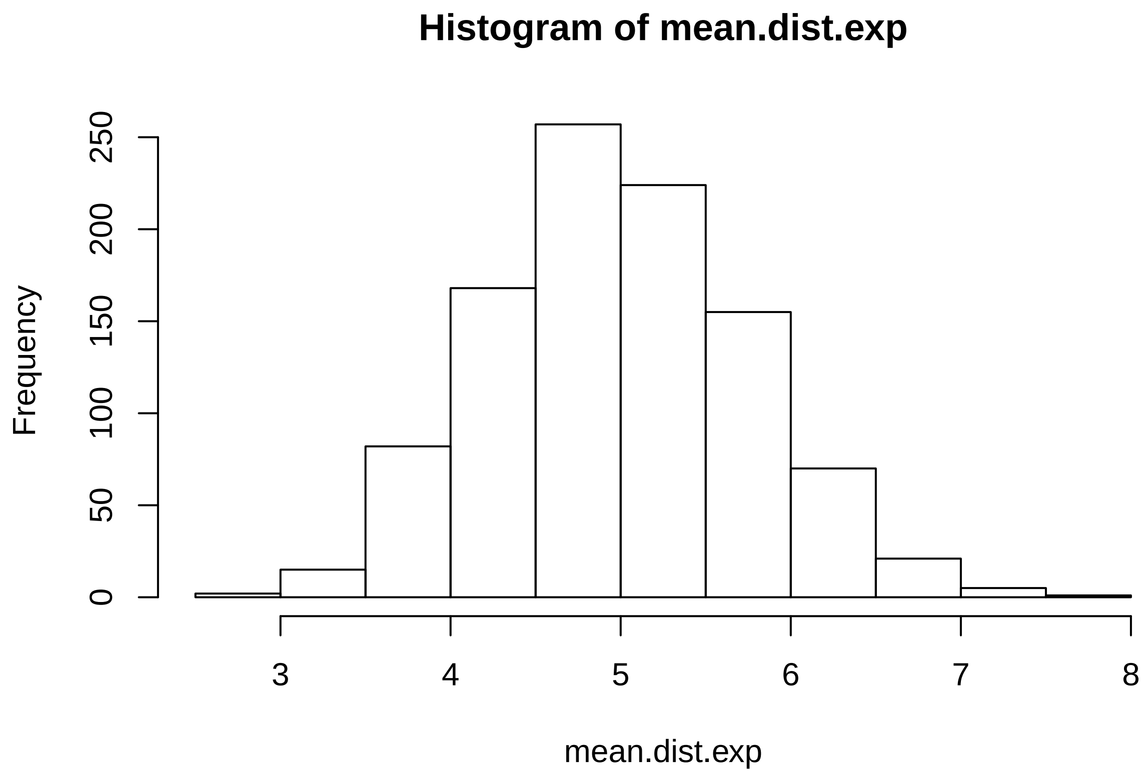
```r
#and the distribution looks like
hist(dist.exp)
```




```r
#Skewed right histogram expected value
```

Now we'll replicate the simulation 1000 times and store the mean of each distribution. Then we'll create a histogram out of the means to view the distibution. The mean should be quite close to 5. We know the standard deviation of a distribution following the central limit theorem should be s/sqrt(n) which is about 0.79.

```r
#Now lets do it 1000 times and take the mean.
mean.dist.exp = NULL
for (i in 1 : 1000) mean.dist.exp = c(mean.dist.exp,
                                     mean(rexp(n.exp,lambda)))

hist(mean.dist.exp)
```




```r
mean(mean.dist.exp) #5 expected
```

```
## [1] 4.986504
```

```r
sd(mean.dist.exp) #0.79 expected
```

```
## [1] 0.7639468
```

The mean and standard deviation fit the theoretical values and the histogram visually shows a somewhat normal distribution which can be identified by the "bell curve" shape centered around the mean.

Sample Mean versus Theoretical Mean

Let's calculate and plot the samples and theoretical means. They should be very close to each other after 1000 simulations.

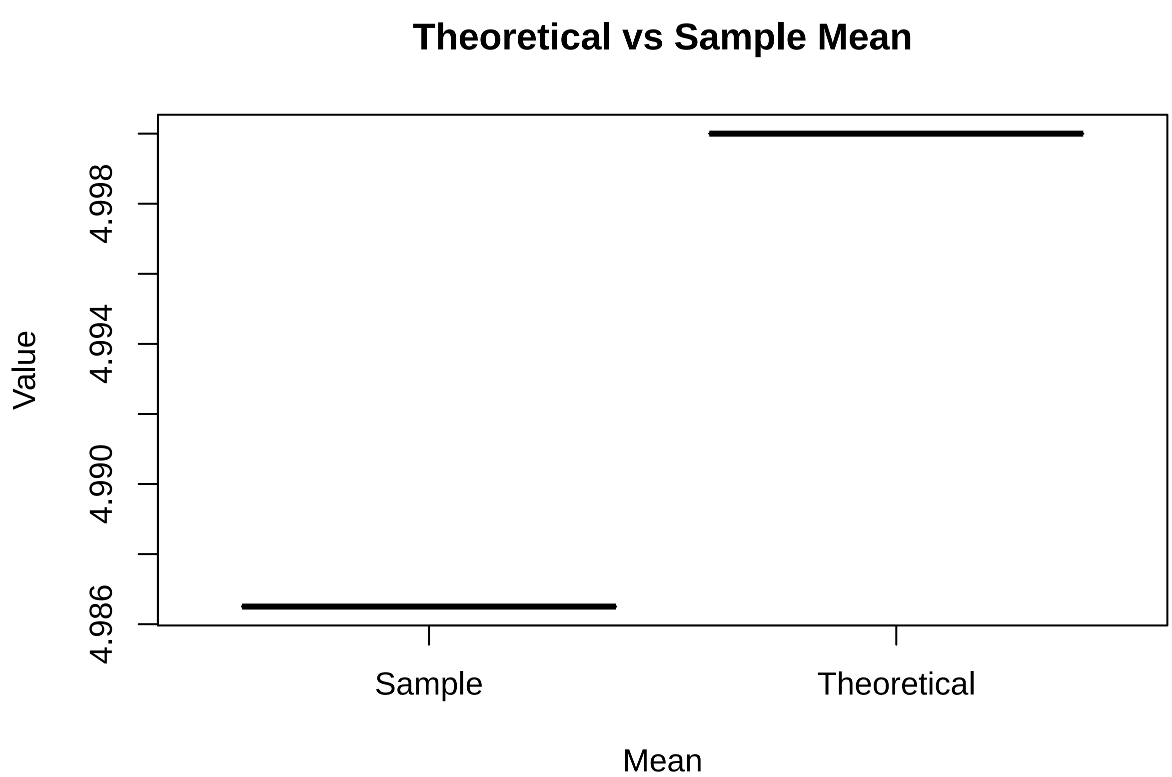
```r
#1. Show the sample mean and compare it to the theoretical mean of the #distribution.

#The mean is
mean(mean.dist.exp)
```

```
## [1] 4.986504
```

```r
#And the difference between it and the theoretical mean is
mean(mean.dist.exp)- (1/lambda)
```

```
## [1] -0.01349612
```

```r
mean.df <- data.frame(Sample =  mean(mean.dist.exp), Theoretical = (1/lambda))

boxplot(mean.df$Sample, mean.df$Theoretical, main="Theoretical vs Sample Mean",
        xlab="Mean", ylab="Value", names=names(mean.df))
```



Sample Variance vs Theoretical Variance

Now lets do the same for the variance.

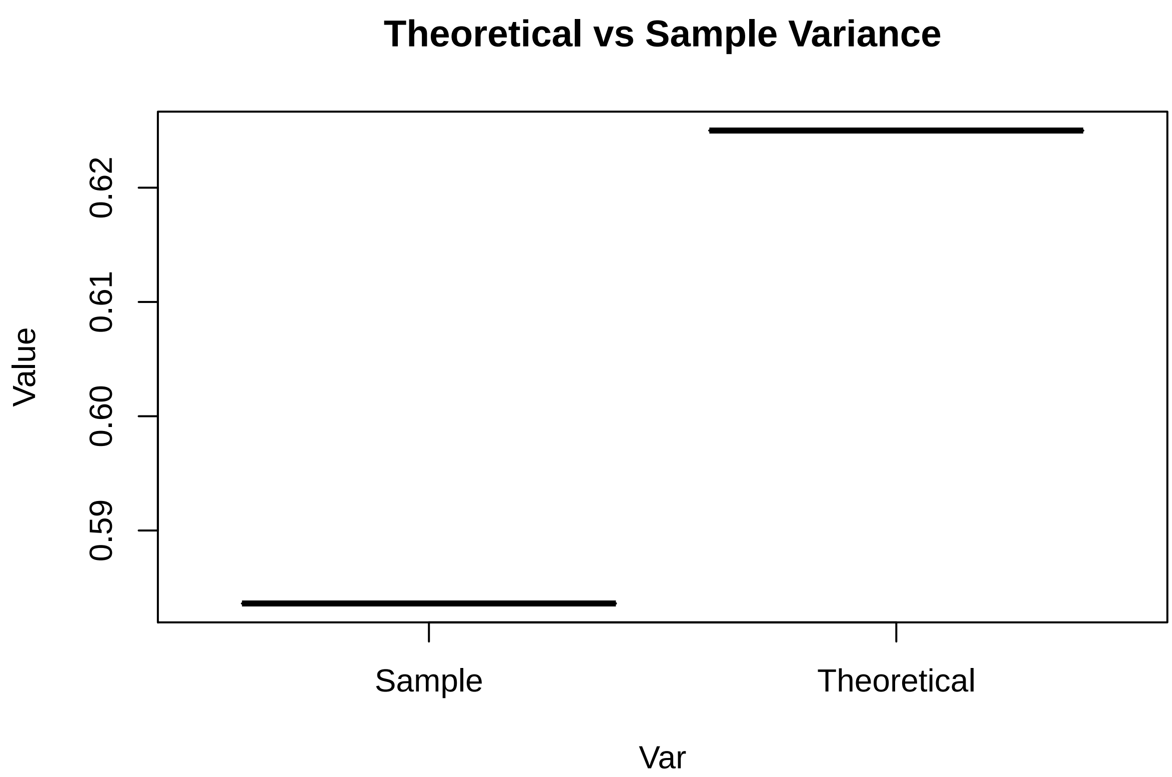
```r
#2. Show how variable the sample is (via variance) and compare it to the
#theoretical variance of the distribution.

#The variance is
var(mean.dist.exp)
```

```
## [1] 0.5836148
```

```r
#The difference between the theoretical variance and sample variance is

var(mean.dist.exp) - (sd.dist.exp^2/n.exp)
```

```
## [1] -0.04138525
```

```r
var.df <- data.frame(Sample =  var(mean.dist.exp), Theoretical = (sd.dist.exp^2/n.exp))

boxplot(var.df$Sample, var.df$Theoretical, main="Theoretical vs Sample Variance",
        xlab="Var", ylab="Value", names=names(var.df))
```



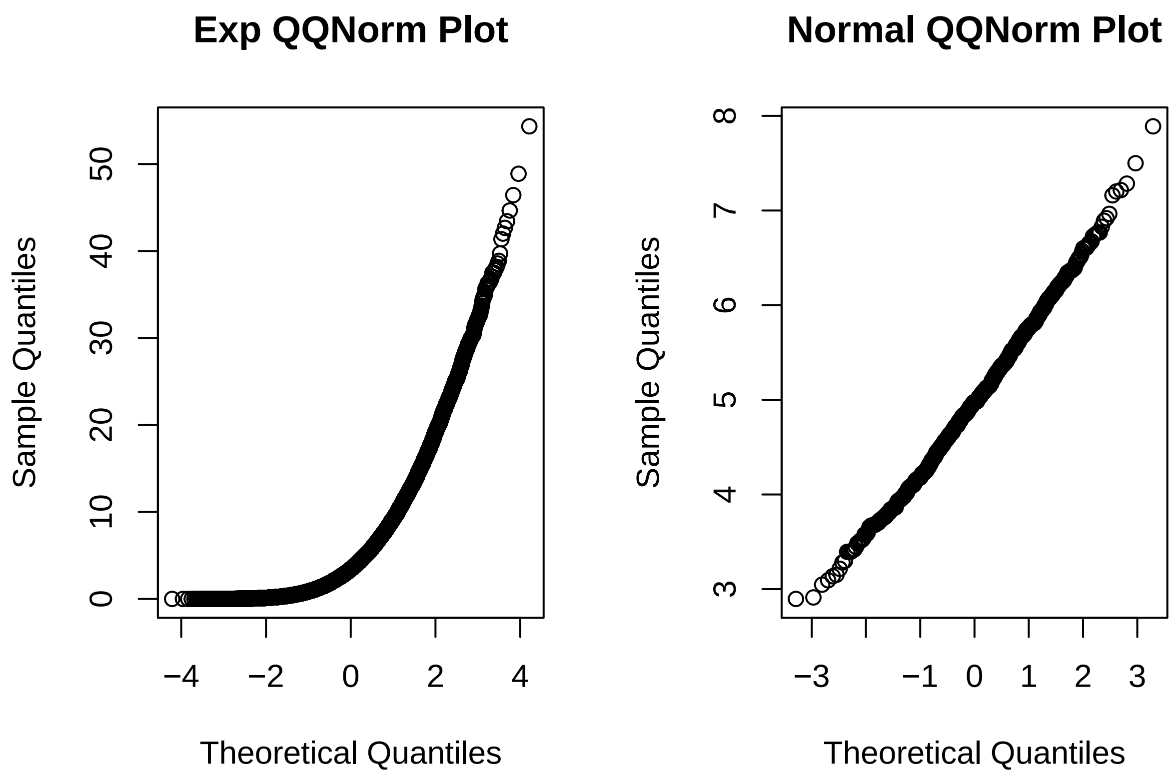
Distribution

Now let's graphically test if the distribution is normal. As mentioned above, the histogram of the distribution of means shows the typical "bell shape" curve of a normal distribution centered around the mean. Here we'll show another graphical test.

The qqnorm function plots the theoretical vs sample quantiles. This means that the plot should be linear if our distribution is normal. Let's plot a collection of 1000 exponential function simulations on the left and the distribution of means of 1000 exponential functions on the right.

```r
#3.Show that the distribution is approximately normal.
collection.dist.exp = NULL
```

```r
for (i in 1 : 1000) collection.dist.exp = c(collection.dist.exp,
                                              rexp(n.exp,lambda))

#formatting
par("mfcol"=c(1,2))


#By usin qqnorm we can test if the distribution is normal
#We expect normal distributions to be linear.
qqnorm(collection.dist.exp, main="Exp QQNorm Plot")
qqnorm(mean.dist.exp, main="Normal QQNorm Plot")
```




```r
#Here we see that the mean of exp functions is normal and the
#collection of exp functions is not normal
```

Conclusion

In summation we have shown that the distribution of means of 1000 simulated exponential functions follows the central limit theorem.