Comparison of the Exponential Distribution to the Central Limit Theorum

S Duffy

March 9, 2016

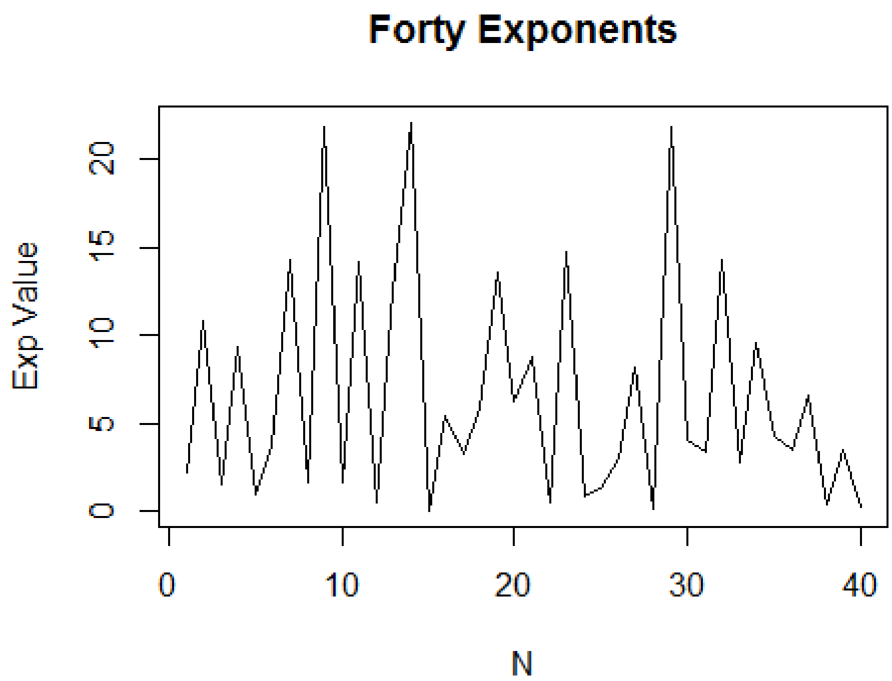
Overview

This project investigates the exponential distribution in R and compares it with the Central Limit Theorem. I will investigate the distribution of averages of 40 exponentials performing 1,000 simulations. I will then compare the mean and variance to the theoretical values of each, and show how the distribution relates to the Central Limit Theorum.

Simulations

First, create the exponential distribution using 40 exponentials and a rate (or lambda) of 0.2. For this distribution, the mean and the standard deviation are 1/rate. This code chunk will also set our variables and seed.

```
set.seed(2317)
lambda <- 0.2;n <- 40;iterations <- 1000
exp_dist <- rexp(n, lambda)
m <- mean(exp_dist);v <- var(exp_dist)
plot(exp_dist, type="l", main = "Forty Exponents",xlab = "N",
     ylab = "Exp Value")
```



From the graph above, we can see that a exponential distribution of 40 values has almost no relation to a normal distribution. It also has a mean of 6.5848754 and a variance of 40.828777. However, if one runs the simulation of 40 exponents and takes the mean of the values, and does so many times, the Law of Large Numbers states that the distribution of those means will approach normal. Below is code taking of 1,000 samples of the mean of the 40 exponents.

```
##create an array of 1000 means of the 40 exponents
exp_mns <- NULL
for (i in 1 : iterations) exp_mns = c(exp_mns, mean(rexp(n, lambda)))
```

The following code provides descriptive statistics on the array of 1000 means. We will use the output of this to compare our sampled mean to the theoretical mean.

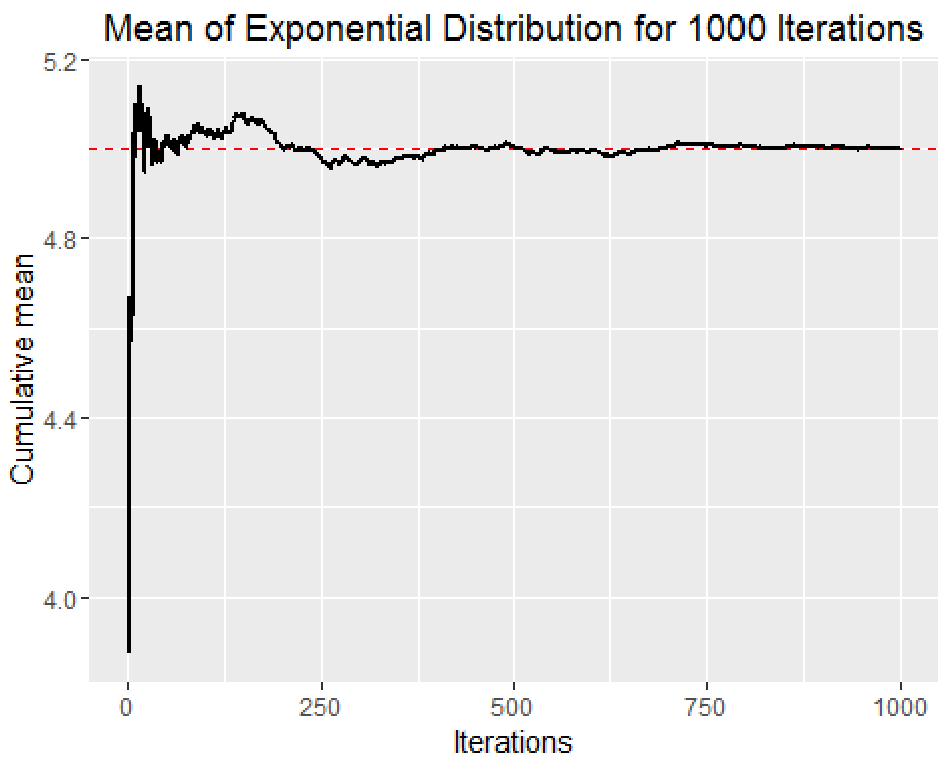
```
library(pastecs)

## Loading required package: boot

stats <- stat.desc(exp_mns);sample_mean <- round(stats[[9]],4)
sample_var <- round(stats[[12]],4)
```

Sample Mean v. Theoretical Mean

The theoretical mean for the distribution should be 1/lambda, which calculates to a value of 5 (1/0.2).

The sample mean, taken from a distribution of 1,000 samples of 40 exponential values equals 5.0034. This is very close to the theoretical mean. The graph below shows that as the number of iterations increases, the sample mean (black line) does indeed approach the theoretical mean (red dotted line).

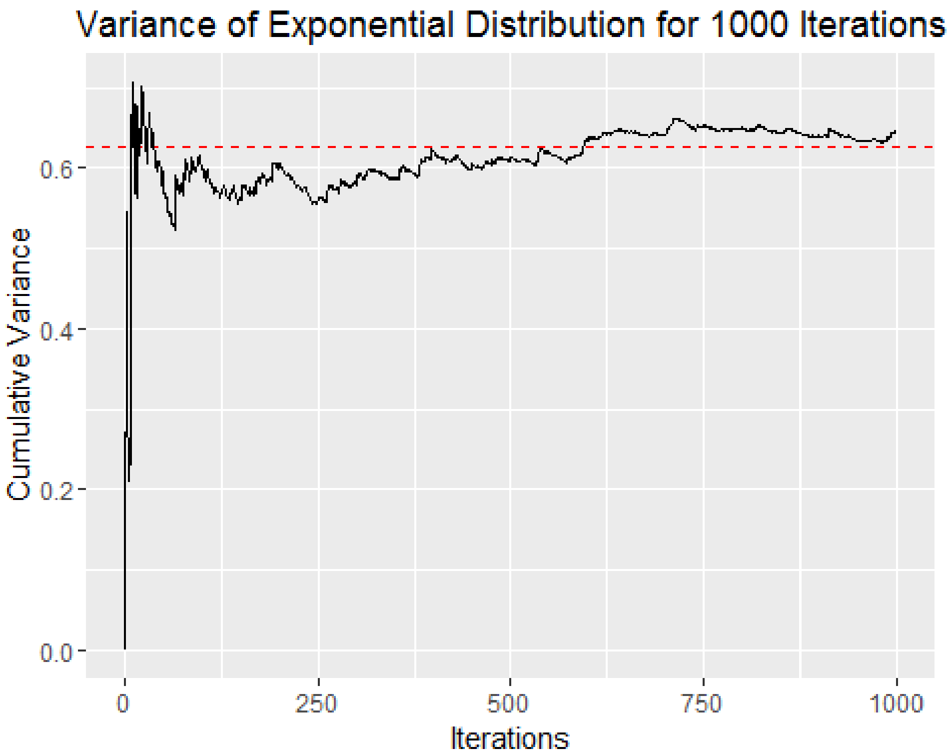
```
##Using the array of sample means created above, plot the cummulative sum of
the means across iterations
means <- cumsum(exp_mns) / (1  : iterations); library(ggplot2)
g <- ggplot(data.frame(x = 1 : iterations, y = means), aes(x = x, y = y))
g <- g + geom_hline(aes(yintercept = 5), color = 'red', linetype = 'dashed')
+ geom_line(size = 1)
g <- g + labs(x = "Iterations", y = "Cumulative mean")
g <- g +  ggtitle("Mean of Exponential Distribution for 1000 Iterations")
g
```



##Sample Variance v. Theoretical Varriance Similarly, the sample variance should approach the theoretical variance as the number of samples grows larger. The theoretical variance is equal to ((1/rate)/sqrt(n))^2, or ((1/0.2)/sqrt(40))^2 which equals 0.625. In our simulation, we retrieved a sample variance of 0.6474. This can be seen in graph below, as iterations increase, the sample variance (black line) moves to meet the theoretical variance (red dotted line).

```
##Using the array of sample means created above, plot the cummulative
variance of the means across iterations
vars <- cumsum(exp_mns^2) / (1  : iterations) - means^2
g <- ggplot(data.frame(x = 1 : iterations, y = vars), aes(x = x, y = y))
g <- g + geom_hline(aes(yintercept = .625), color = 'red', linetype =
'dashed') + geom_line()
g <- g + labs(x = "Iterations", y = "Cumulative Variance")
```

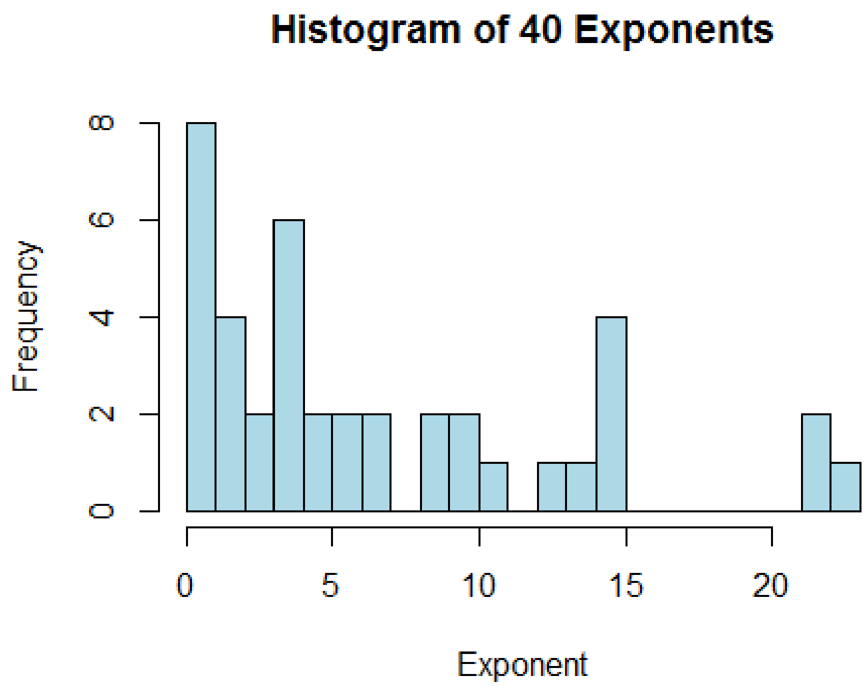
```
g <- g +  ggtitle("Variance of Exponential Distribution for 1000 Iterations")
g
```



Distribution

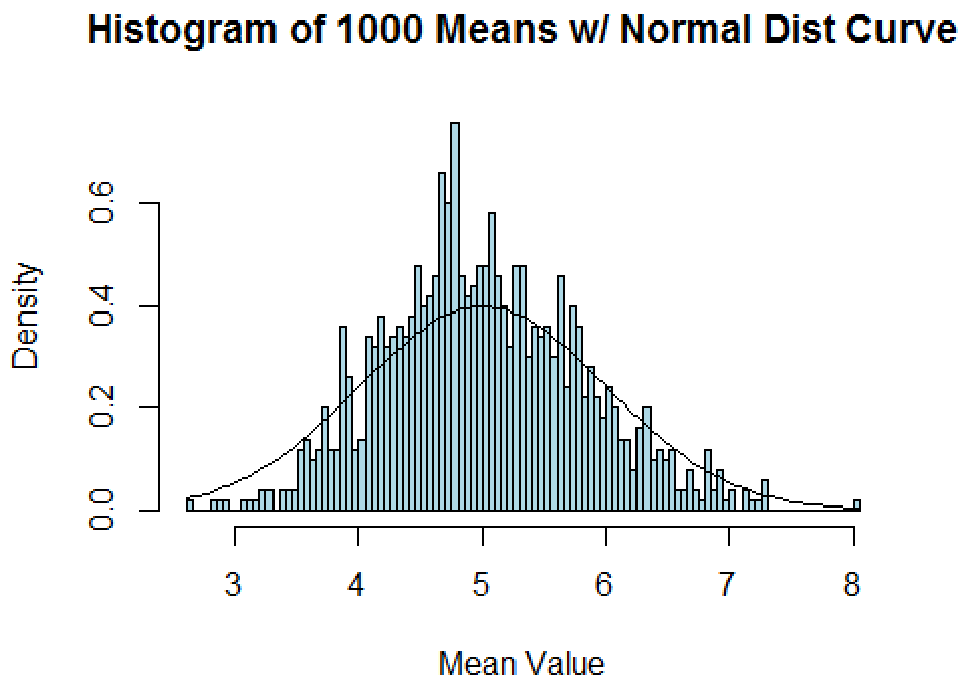
The histogram below shows the exponential distribution of 40 values taken at random.

```
hist(exp_dist, breaks = 20, xlab = 'Exponent', main = 'Histogram of 40
Exponents', col = 'lightblue')
```



You can see that there is little relation to a normal distribution. The next histogram plots the 1,000 sample means used above against a normal distribution curve. Now you can see that with a high number of iterations, taking samples of the mean from a population, the distribution of those means approaches normal.

```
x <- seq(0,1000,0.01)
hist(exp_mns, breaks = 100, prob = TRUE,xlab = 'Mean Value', main =
'Histogram of 1000 Means w/ Normal Dist Curve',col = 'lightblue')
curve(dnorm(x, mean=5, sd=1), add=TRUE)
```



Conclusion

The Law of Large Numbers wins again. With a large number of iterations of sampling the mean of exponential values, we were able to see that both the mean and variance of the samples approach the theoretical mean and variance for the distribution.