Exploration of the Exponential Distribution

Sharon Francisco

July 20, 2016

Synopsis

This report shows how the distribution of averages of an exponential distribution approximates a normal distribution as the sample size increases because of the Central Limit Theorem (CLT).

Installing Packages and Loading Libraries

The packages used in the analysis are installed below.

```
if(!require(ggplot2, quietly = TRUE)) install.packages("ggplot2")
library(ggplot2)
```

Exponential Distribution

To explore the exponential distribution the assignment says to use lambda = 0.2 as the rate parameter.

The code below creates the exponential distribution with lambda = 0.2 and puts it in a data frame. I will use this later to create a plot of the distribution.

```
lambda = 0.2
x <- seq(0, 40, length.out = 1000)
px <- dexp(x, rate = lambda)
dat <- data.frame(x = x, px)
```

The mean, mu, of an exponential distribution is 1/lambda, and the standard deviation, sigma, is also 1/lambda.

```
mu <- 1 / lambda
mu
```

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

The variance, according to Wikipedia, is 1/lambdaˆ2, which is sigmaˆ2.

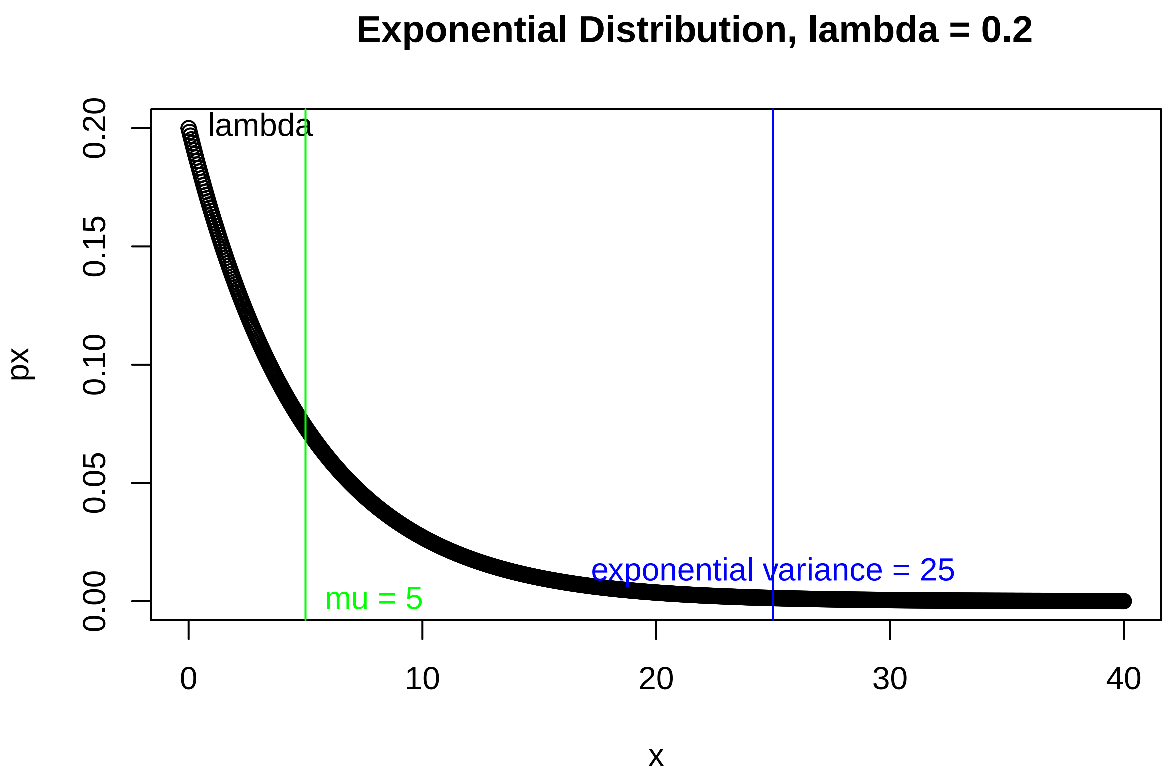
```
sigma <- 1 / lambda

xpDistrVar <- 1 / lambda^2
xpDistrVar
```

```
## [1] 25
```

The code below plots the exponential distribution with lambda, the rate, = 0.2, which is the y-intercept. It also shows the theoretical mean, mu, and the variance.

```r
with(dat, plot(x, px))
par(mar = c(3, 4, 4, 0))
title(main = "Exponential Distribution, lambda = 0.2")
text(0, lambda, labels = "lambda", pos = 4)
abline(v = mu, col = "green")
muLabel <- paste("mu =", mu)
text(mu, 0, labels = muLabel, pos = 4, col = "green")
abline(v = xpDistrVar, col = "blue")
xpDistVarLabel <- paste("exponential variance =", xpDistrVar)
text(xpDistrVar, 0.025, labels = xpDistVarLabel, pos = 1,
     col = "blue")
```



So for an exponential distribution with lambda = 0.2, the mean and standard deviation are 5, and the variance is 25.

Distribution of 1000 Simulations of the Means of 40 Exponential Distributions

Now I'll look at the distribution of the *averages* of 40 exponentials simulated 1000 times.

```r
set.seed(77)
n <- 40
simulations <- 1000
cLim <- replicate(simulations, mean(rexp(n, lambda)))
```

I'll calculate the sample mean of this disribution.

```r
sampleMean <- mean(cLim)
sampleMean
```

```
## [1] 4.996833
```

The sample mean, 4.9968333, is very close to the theoretical mean, 5.

Next I'll calculate the theoretical variance using the formula, sigmaˆ2 / n, and then the sample variance using the **var** function in R.

```r
theoreticalVarn <- sigma^2 / n
theoreticalVarn
```

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

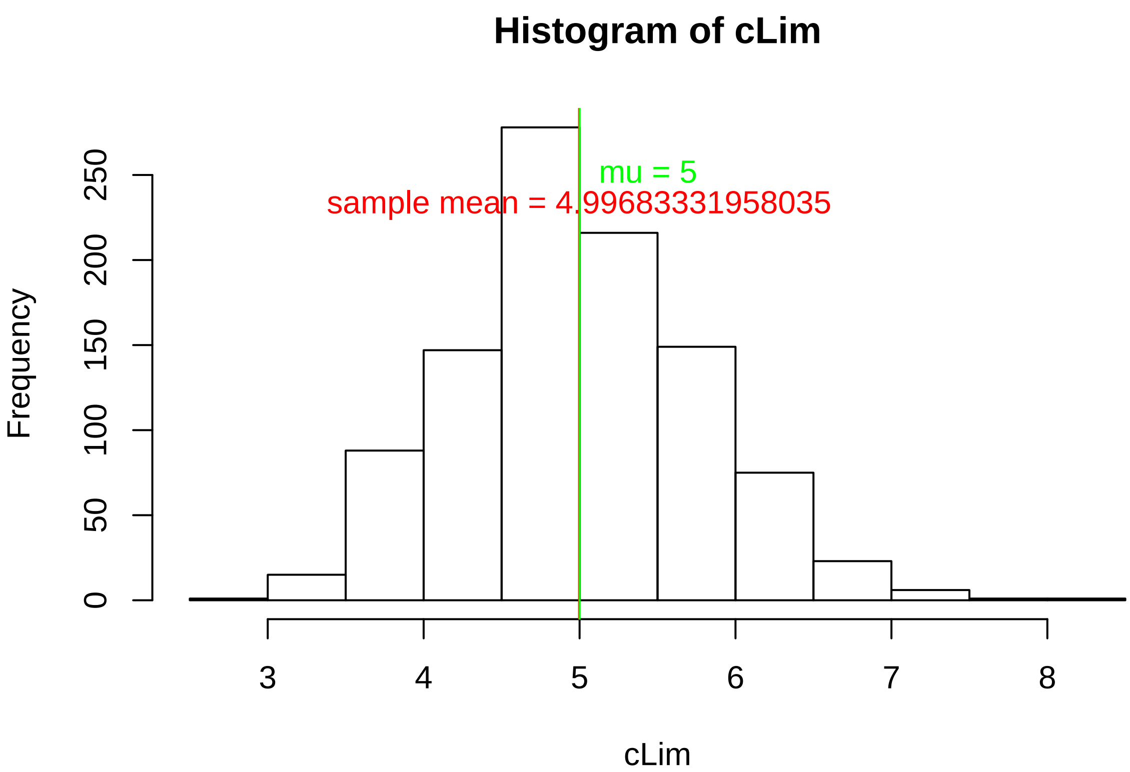
```r
varnSample <- var(cLim)
varnSample
```

```
## [1] 0.6162158
```

The results of the two variance calculations are close. Because we ran 1000 simulations, the variance is relatively small. Because this is a distribution of averages, the variance is very different than the variance of the original exponential distribution.

I'll plot a histogram of this distribution of the *averages* of 40 exponentials simulated 1000 times.

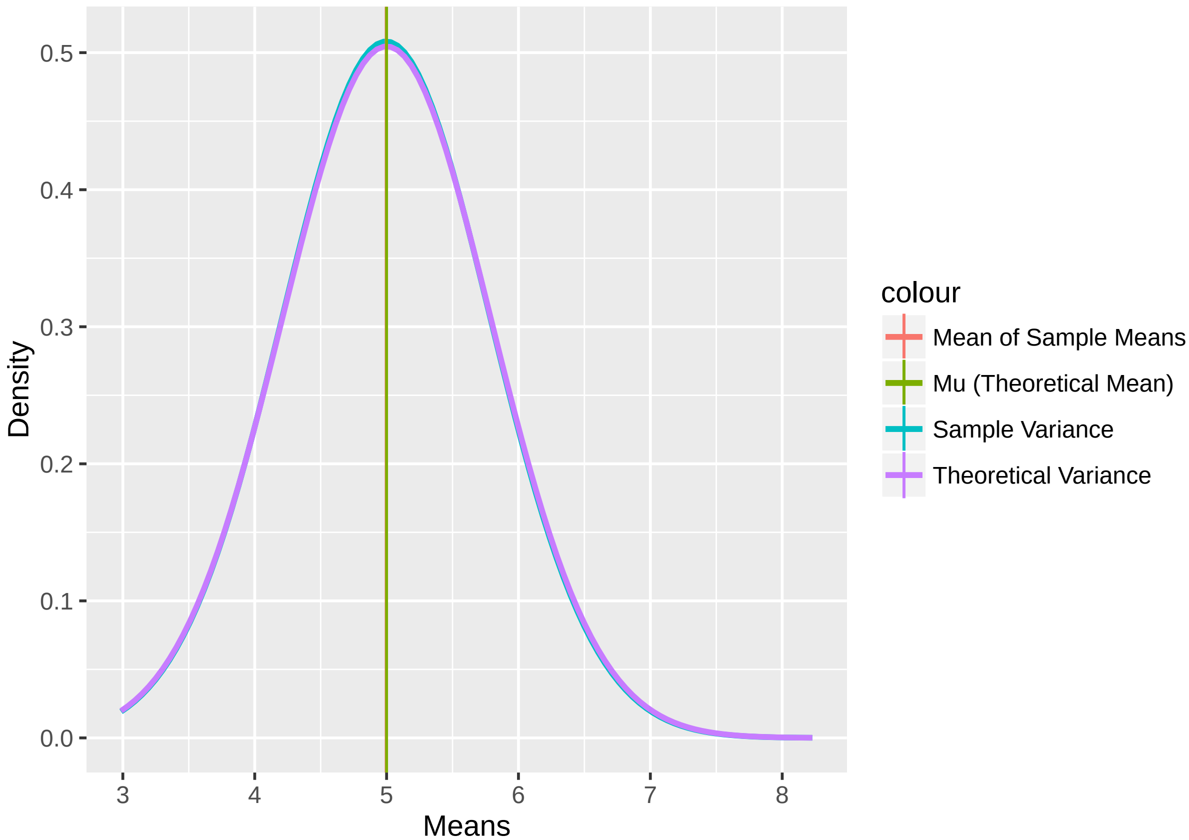
```r
hist(cLim)
abline(v = sampleMean, col = "red")
mnClimLabel <- paste("sample mean =", sampleMean)
text(sampleMean, 250, labels = mnClimLabel, pos = 1,
     col = "red")
abline(v = mu, col = "green")
text(mu, 250, labels = muLabel, pos = 4, col = "green")
```



Since this is a distribution of the averages of 40 distributions, the distribution looks Gaussian because of the Central Limit Theorem, which states that the distribution of averages of iid variables becomes that of a standard normal as the sample size increases.

Here are the density plots of the distribution of averages using the sample mean (4.9968333) and sample variance (0.6162158) and the theoretical mean, mu (5), and the theoretical variance (0.625).

```
g <- ggplot(as.data.frame(cLim), aes(cLim)) +
        xlab("Means") +
        ylab("Density")
g <- g + stat_function(fun = dnorm,
                       args = list(mean = sampleMean,
                                   sd = sqrt(varnSample)),
                       aes(color = "Sample Variance"),
                       size = 1) +
        stat_function(fun = dnorm,
                      args = list(mean = mu,
                                  sd = sqrt(theoreticalVarn)),
                      aes(color = "Theoretical Variance"),
                       size = 1)
g <- g + geom_vline(aes(xintercept = sampleMean, color =
                                "Mean of Sample Means")) +
        geom_vline(aes(xintercept = mu, color =
                               "Mu (Theoretical Mean)"))
g
```



This demonstrates the characteristics of the Central Limit Theorem:

1. The mean of the sample means is approximately equal to the theoretical mean;
2. The sample variance is approximately equal to the theoretical variance divided by population size;
3. The distribution of the sample means is approximately normal.