

Comparison of the Exponential Distribution vs. the Central Limit Theorem

S Fraser

May 29, 2016

Overview

This project will investigate the distribution of 1000 random simulations of 40 exponentials. The averages of exponential distribution will then be compared to the Central Limit Theorem. It will attempt to prove that when the mean of the simulations are taken, the resulting plot of the data will form a normal bell curve or something approaching it. The project will also compare Sample mean vs. theoretical mean and sample variance vs. theoretical variance.

Setup

```
# set up variables for the Exponential Distribution
# based on the information given in the assignment
set.seed(12345678)    ## for reproducibility
lm <- 0.2             ## lamda
n <- 40               ## size of the exponentials
sims <- 1000          ## number of simulations
```

Simulations

Create a matrix of exponential distributions for 40000 occurrences (40 exponentials * 1000 simulations).

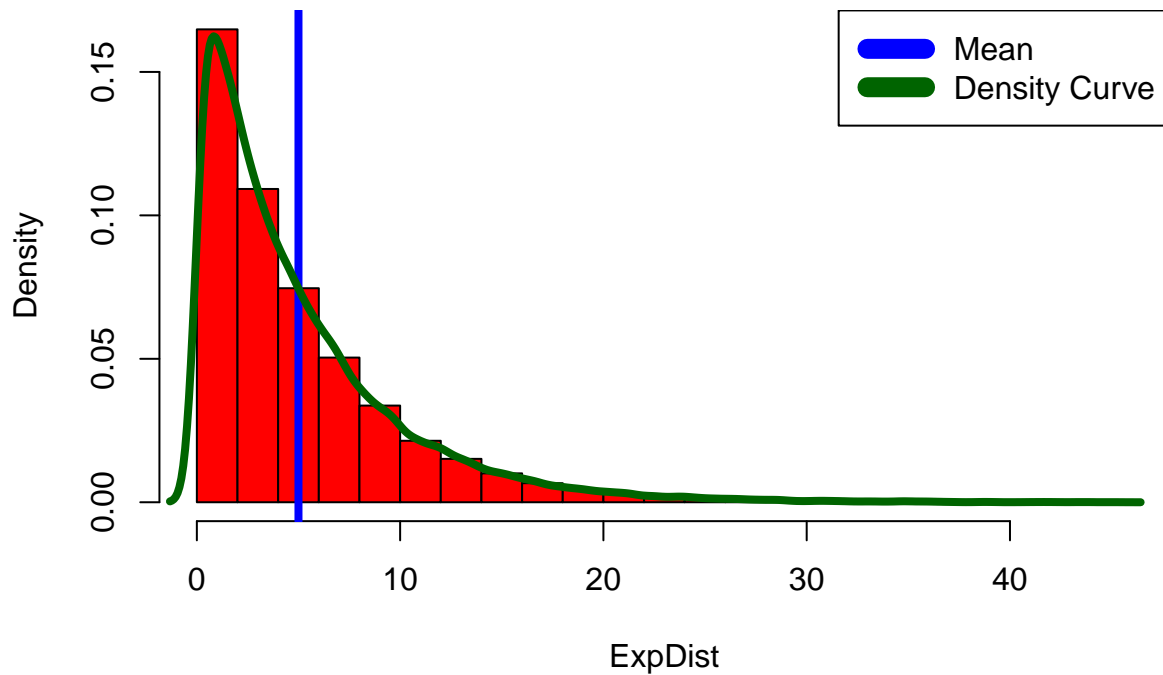
```
ExpDist <- matrix(rexp(sims * n, rate=lm), sims)
```

Various Plots

Plot a histogram of the Exponential Distribution of 40 exponentials with 1000 simulations

```
hist(ExpDist,prob=TRUE,col="red",border="black",breaks=20,cex.main=.8,
main="GRAPH A - Exponential Distribution of 40 exponentials of 1000 simulations" )
abline(v = 1/lm, col = "blue", lwd = 4)
lines(density(ExpDist),col="darkgreen",lwd=4)
legend("topright", c("Mean","Density Curve"), col=c("blue","darkgreen"), lwd=10)
```

GRAPH A – Exponential Distribution of 40 exponentials of 1000 simulations



Simulate and plot a histogram showing 40 random samples for an exponential distribution.

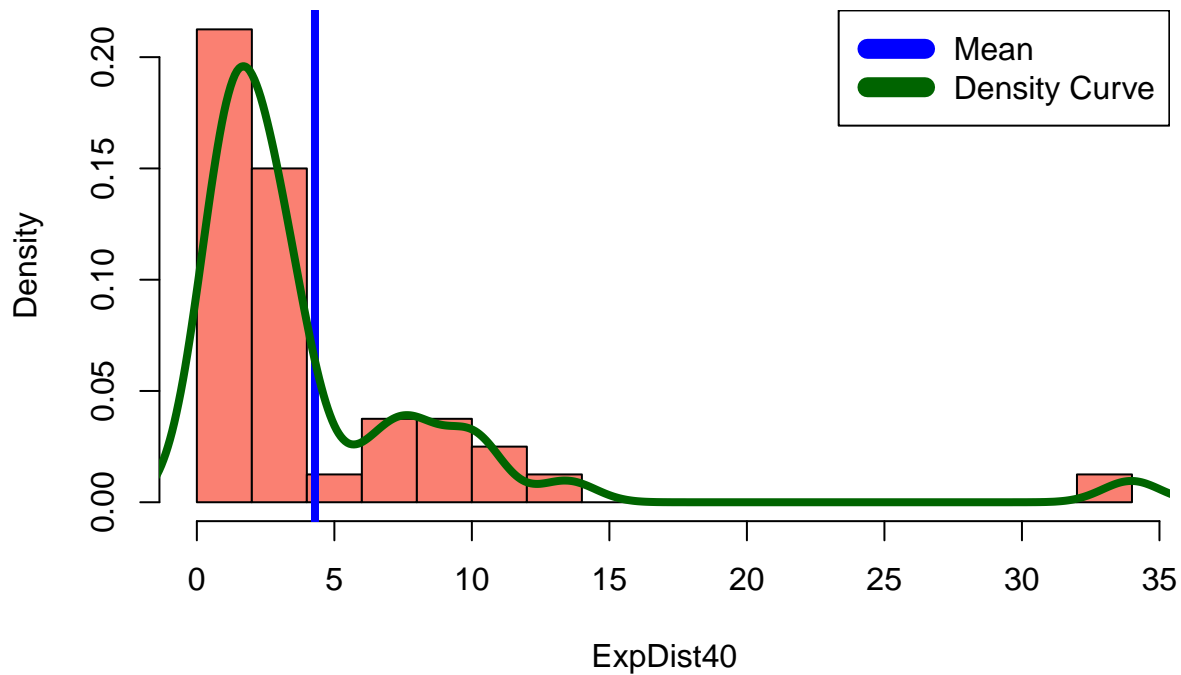
```
ExpDist40 <- rexp(n,lm)
```

```
ExpDist40Avg <- mean(ExpDist40)  
ExpDist40Avg
```

```
## [1] 4.295142
```

```
## plot the Exponential Distribution of 40 exponentials  
hist(ExpDist40,prob=TRUE,col="salmon",border="black",breaks=20,cex.main=.8,  
main="GRAPH B - Exponential Distribution of 40 exponentials" )  
abline(v = ExpDist40Avg, col = "blue", lwd = 4)  
lines(density(ExpDist40),col="darkgreen",lwd=4)  
legend("topright", c("Mean","Density Curve"), col=c("blue","darkgreen"), lwd=10)
```

GRAPH B – Exponential Distribution of 40 exponentials



We can see that for one set of 40 exponentials the rule of $1/\lambda$ (λ) does not necessarily hold true.

Sample Mean vs. Theoretical Mean

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

We expect the theoretical mean to be $1/\lambda$ (λ) for an exponential distribution.

```
## mean of exponential distribution
ExpMn <- 1/lambda
ExpMn
```

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

For the sample mean, compute the average of all the exponential distribution simulations and then taken the mean of these.

```
## calculate the average of all of the exponential distribution simulations
ExpDistAvg <- apply(ExpDist,1,mean)

## mean of the average of the exponential distribution
ExpMnAvg <- mean(ExpDistAvg)
ExpMnAvg

## [1] 5.017885
```

We can see the sample mean and theoretical mean are very close to each other.

Sample Variance vs. Theoretical Variance

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

We can calculate the theoretical variance as the standard deviation squared. We were given the fact the the standard deviation can be computed as $1/\lambda$ (lamda). Divide this by \sqrt{n} .

```
## standard deviation of exponential distribution
ExpSD <- 1/lm/sqrt(n)
ExpSD
```

```
## [1] 0.7905694
```

```
## variance of the exponential distribution
ExpVar <- ((1/lm)/sqrt(n))^2
ExpVar
```

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

We can calculate the sample variance as standard deviation of the averages of all simulations of the exponential distribution squared.

```
## standard deviation of the average of the exponential distribution
ExpSDAvg <- sd(ExpDistAvg)
ExpSDAvg
```

```
## [1] 0.7759495
```

```
ExpVarAvg <- var(ExpDistAvg)
ExpVarAvg
```

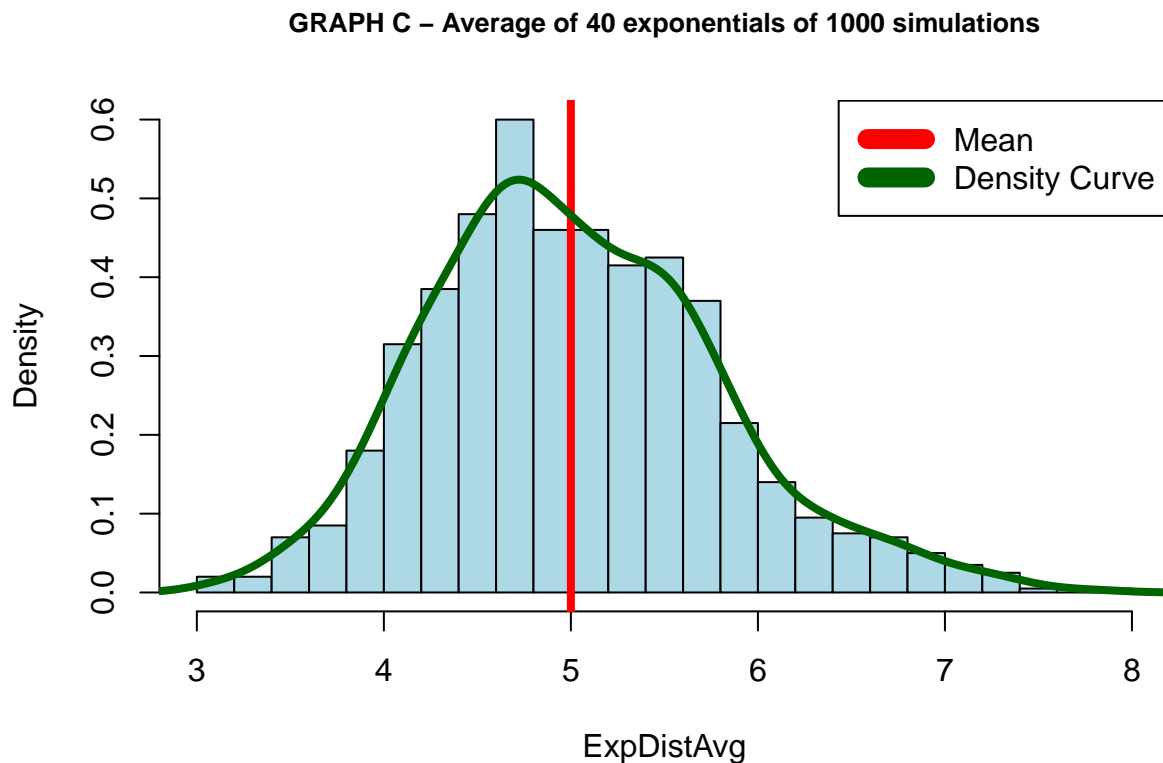
```
## [1] 0.6020976
```

The Sample variance and theoretical variance are also close to each other.

Distribution

3. Show that the distribution is approximately normal.

```
## plot a histogram of the average of the exponential distribution
hist(ExpDistAvg,prob=TRUE,col="lightblue",border="black",breaks=20,cex.main=.8,
main="GRAPH C - Average of 40 exponentials of 1000 simulations" )
abline(v = ExpMn, col = "red", lwd = 4)
lines(density(ExpDistAvg),col="darkgreen",lwd=4)
legend("topright", c("Mean", "Density Curve"), col=c("red", "darkgreen"), lwd=10)
```



Histogram of the distribution of the averages of the 1000 simulations of 40 exponentials showing the density curve and mean.

As can be seen by the dark green line in the histogram, we have a nearly normal curve for the averages of the exponential distribution.

We can also see that GRAPH A, which plots the raw data for the exponential distribution values, is very different from GRAPH C which plots the averages of the exponential distribution data. GRAPH C follows the Central Limit Theorem which says that taking the averages (mean) of a large enough group of random variables with a well-defined variance will be approximately normally distributed regardless of the underlying distribution.