Exponential Distribution Course Project

Overview

In this project we investigate the exponential distribution 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. We set lambda = 0.2 for all of the simulations and investigate the distribution of averages of 40 exponentials over 1000 simulations.

Simulations:

The simulation can be built in R using the `rexp(n,l)` function, where `samp` is the number of exponentials per set, and `lambda` is the rate (a constant of these purposes of 0.2). We need to build a single column data frame with the average of each of the sets. We iterate 1k times, or 1k simulations, or `sim`, and use the `mean` function to calculate the average. The code is as follows:

```
set.seed(4523)
lambda <- 0.2
samp <- 40
sim <- 1000
```

initialize a data frame with a row count of the number of simulations

```
df <- matrix(data=NA, nrow=0, ncol = samp)
```

iterate 1 to the number of simulations variable

```
for (i in 1:sim) {
  df <- rbind(df, rexp(samp, lambda))
}
```

Sample Mean versus Theoretical Mean:

With our sample data in place, we will compare the sample mean to the theoretical mean. We start off by calculating the mean and median of the sample means to see how closely it lines up to the theoretical mean.

```
tMean <- 1/lambda             # theoretical mean
sMean <- c(1: sim)            # sample mean
```

```
for(k in 1:sim) {
    sMean[k] <- mean(df[k,])
}
meansMean <- mean(sMean)
medsMean <- median(sMean)
meansMean
```

[1] 5.00569

```
medsMean
```

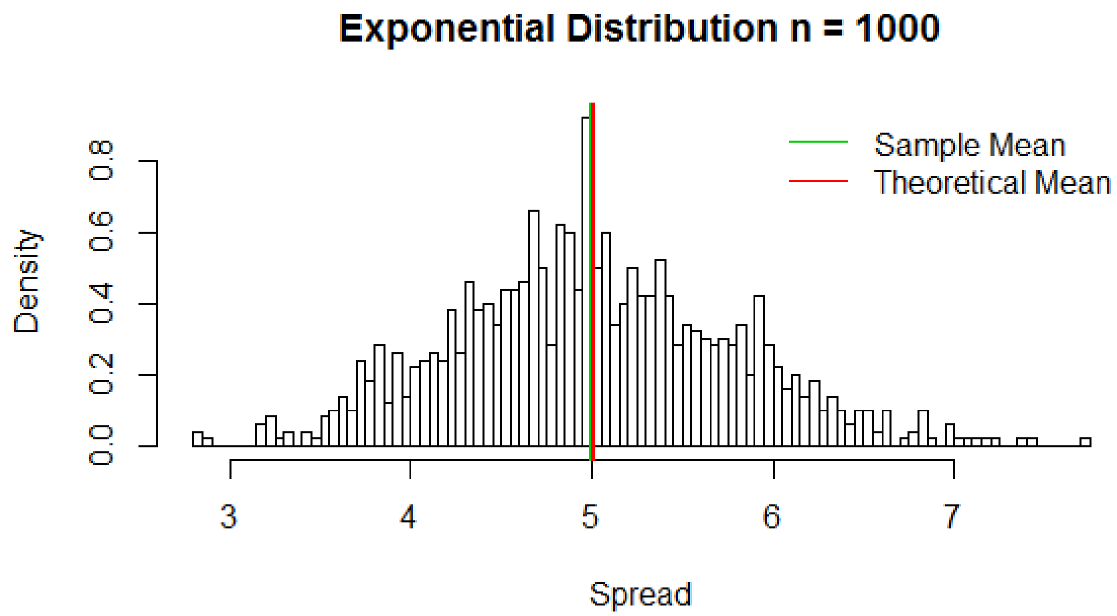
[1] 4.973336

As is evident from the above two results, the sample mean (or actual mean) of 5.00569 is very close to the theoretical mean of 4.973336.

When we plot the sample set below and place a vertical line on both mean results, the histograms look almost identical.

```
 hist(sMean,
                 breaks = 100,
                 prob = TRUE,
                 main="Exponential Distribution n = 1000",
                 xlab="Spread")
                        abline(v = tMean,
                                col= 3,
                                lwd = 2)
                        abline(v = meansMean,
                                col = 2,
                                lwd = 2)

                legend('topright', c("Sample Mean", "Theoretical Mean"),
                                bty = "n",
                                lty = c(1,1),
                                col = c(col = 3, col = 2))
```



Sample Variance versus Theoretical Variance:

We can now examine the variance. The variance present in the sample means of the 1000 simulations will be compared to the theoretical variance of the population.

```
tVar <- ((1/lambda)^2)/samp     # theoretical variance
```

[1] 0.625

We can use the R `var` function to determine the actual variance of the observations.

```
sVar <- var(sMean)     # actual variance
```

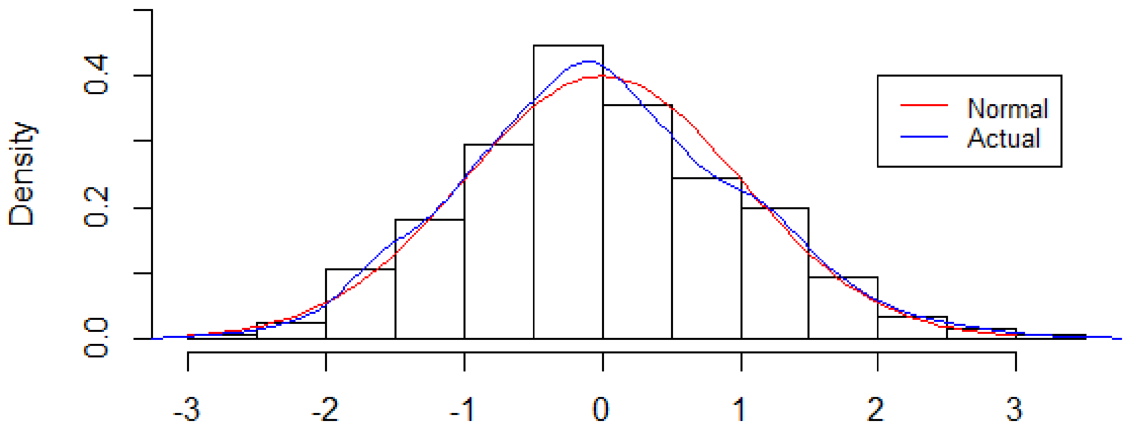
[1] 0.6054884

The results of the theoretical variance as compared to actual variance are very close, as noted above.

Distribution:

Lastly, if a normal curve is plotted and overlayed to the sample set, along with the actual distribution curve, it is evident that they are
very similar, thus, the sample set can be defined as normal, in terms of distribution.

```
par(mfrow=c(1,1))
hist(scale(sMean),probability=T,main='',ylim=c(0,0.5),xlab='')
curve(dnorm(x,0,1),-3,3, col='red',add=T) # normal distribution
lines(density(scale(sMean)),col='blue') # actual distribution
legend(2,0.4,c('Normal','Actual'),cex=0.8,col=c('red','blue'),lty=1)
```



Density
0.0
0.2
0.4
-3
-2
-1
0
1
2
3
Normal
Actual