Exponential Distribution Simulation

Mari Coonley

December 23, 2016

OVERVIEW

This project will investigate the exponential distribution in R and compare it with the Central Liit Theorem. The exponential distribtuion 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. Lambda is set at 0.2 for all simulations. This project investigates the distribution averages of 40 exponentials, simulated 1000 times.

First, establish variables the create data set, called 'data', for exponential distribution simulation.

```
set.seed(1234)
lambda <- .2
n<- 40
sim<- 1000
data<-matrix(rexp(n*sim, lambda), nrow = 40, ncol = 1000)
str(data)
```

```
##  num [1:40, 1:1000] 12.5088 1.2338 0.0329 8.7137 1.9359 ...
```

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

First,calculate mean of each of the 1000 simulations

```
simMeans = NULL   ## simulation means
for (i in 1:1000){
  simMeans[i] = mean(data[ ,i])
}
```

Next, calculate the sample mean from the simulation means as well as the theoretical mean and compare the results.

```
samMean<-mean(simMeans)   ## sample mean
theoMean<-1/lambda        ## theoretical mean
samMean; theoMean         ## display comparison
```

```
## [1] 4.974239
```

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

The results are very similar.

2) Show how variable the sample is (via variance) and compare to theoretical variance

of the distribution

The variance of a sample is the square root of the standard deviation. Therefore, the standard deviation must be calculated first, then the square root of that figure is calculated. Begin by calculating the standard deviation of each of the 1000 simulations.

```
simSD = NULL    ## simulation standard deviations
for (i in 1:1000) {
  simSD[i] = sd(data[ ,i])
}
```

Next, take the average of all the 1000 standard deviations to determine the sample standard deviation. Also calculate the theoretical standard deviation.

```
samSD<-mean(simSD) ## sample standard deviation
theoSD<-1/lambda   ## theoretical standard deviation
```

Lastly, take the square root of both the sample standard deviation and theoretical distribution to calculate and compare the respective variances.

```
samVar<-sqrt(samSD)      ## sample variance
theoVar<- sqrt(theoSD)   ## theoretical variance
samVar; theoVar          ## display comparison
```

```
## [1] 2.198869
```

```
## [1] 2.236068
```

Again, the results are very similar.

3) Show that the distribution is approximately normal.

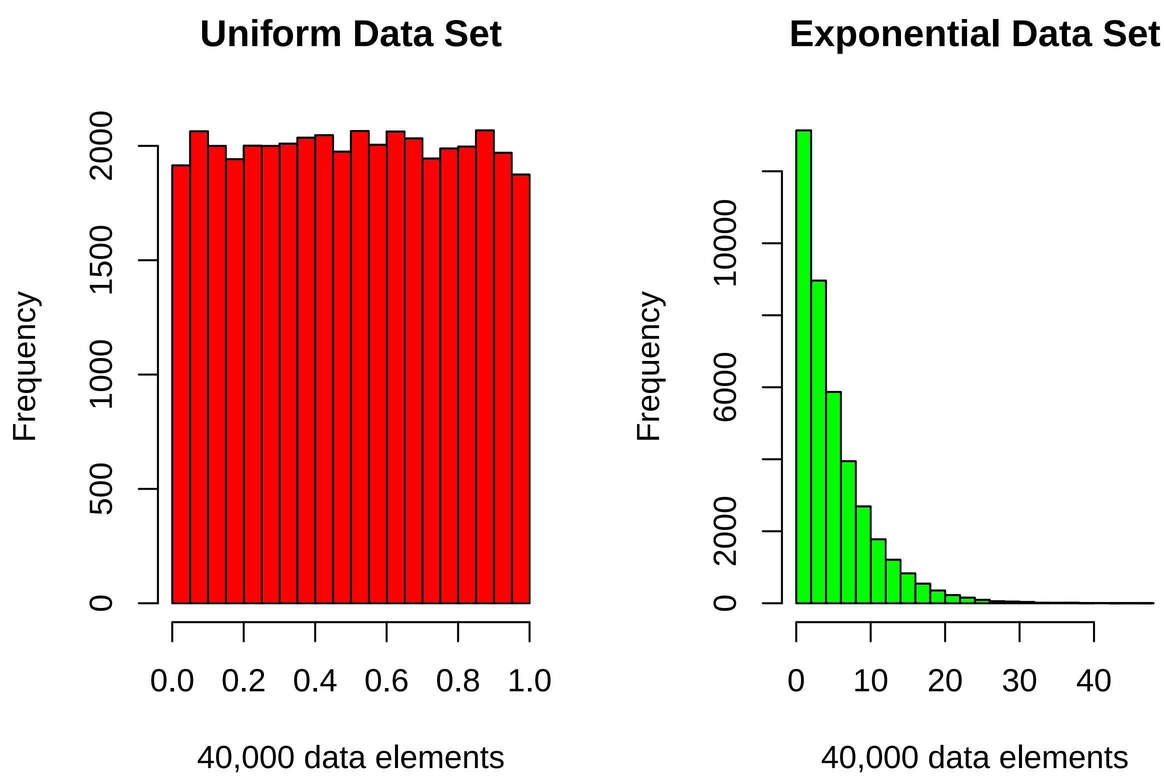
To show the distribution is approximately normal, compare the exponential distribution to a uniform distribution. Use the uniform data suggested as motivating example to create comparison data set, with the exception of increasing to 40,000 elements from 1,000, for consistency purposes.

```
set.seed(4321)
comp<-runif(40000)                ## create uniform distribution data set
mns=NULL                          ## create and calculate the means for each of the 1000 sets of
for(i in 1:1000){                 ##  40 elements uniformly distributed
  mns=c(mns, mean(runif(40)))
}
std=NULL                          ## create and calculate the standard deviation for each of the
for(i in 1:1000){                 ##  1000 sets of 40 elements uniformly distributed, recalling
  std=c(std, sd(runif(40)))       ##  variance is the square root of standard deviation and
}                                 ##  variance is required for proper comparison
```

The most direct method of comparison is graphically, to visually assess each data set and components.

First, compare the original data sets.

```
par(mfrow = c(1,2))
hist(comp, main = "Uniform Data Set", xlab = "40,000 data elements", col="red")
hist(data, main = "Exponential Data Set", xlab = "40,000 data elements", col = "green")
```

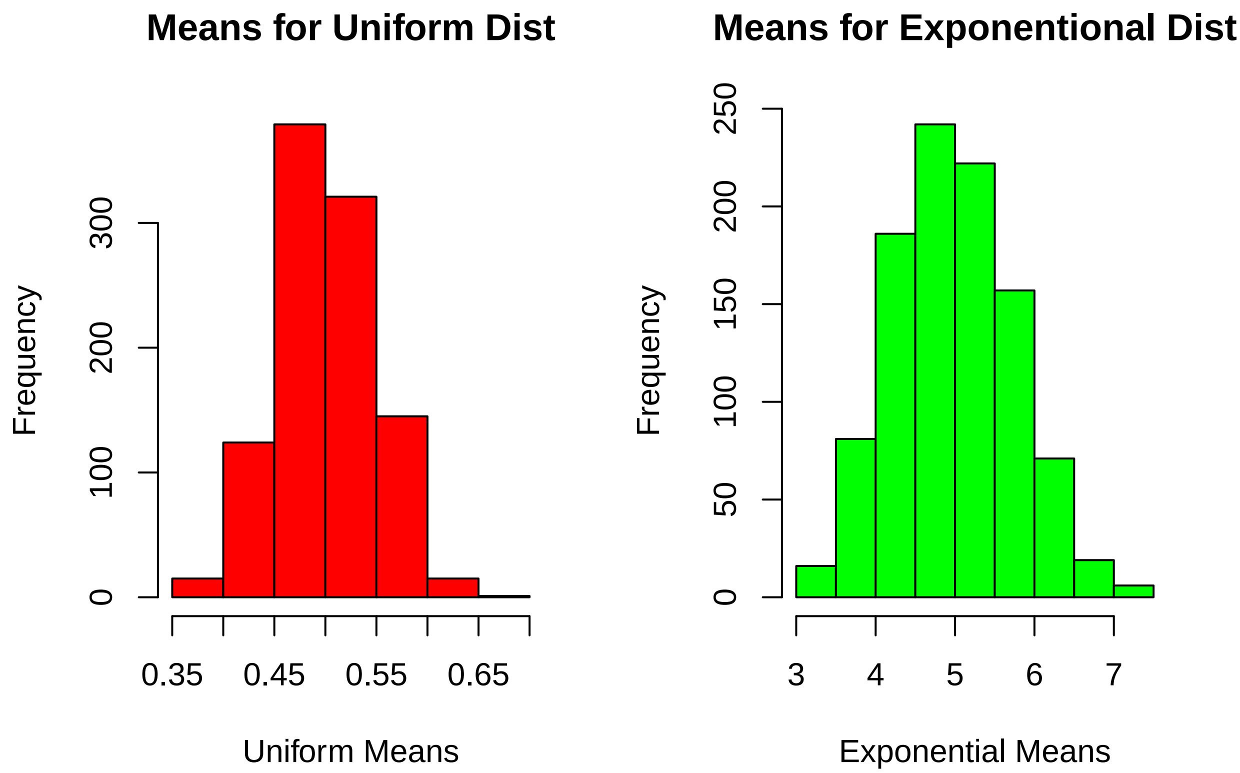


The individual data sets are very different, with charactistics indicative of each type of distribution. The elements of the uniform data set are much closer together or uniform with a smaller range between elements. Whereas the elements of the exponential data set reflect a much wider range between elements, as is expected from a data set using exponents to calculate elements.

However, when comparing statistical components such as mean and standard deviation, they become much more similar.

Looking at the respective means,

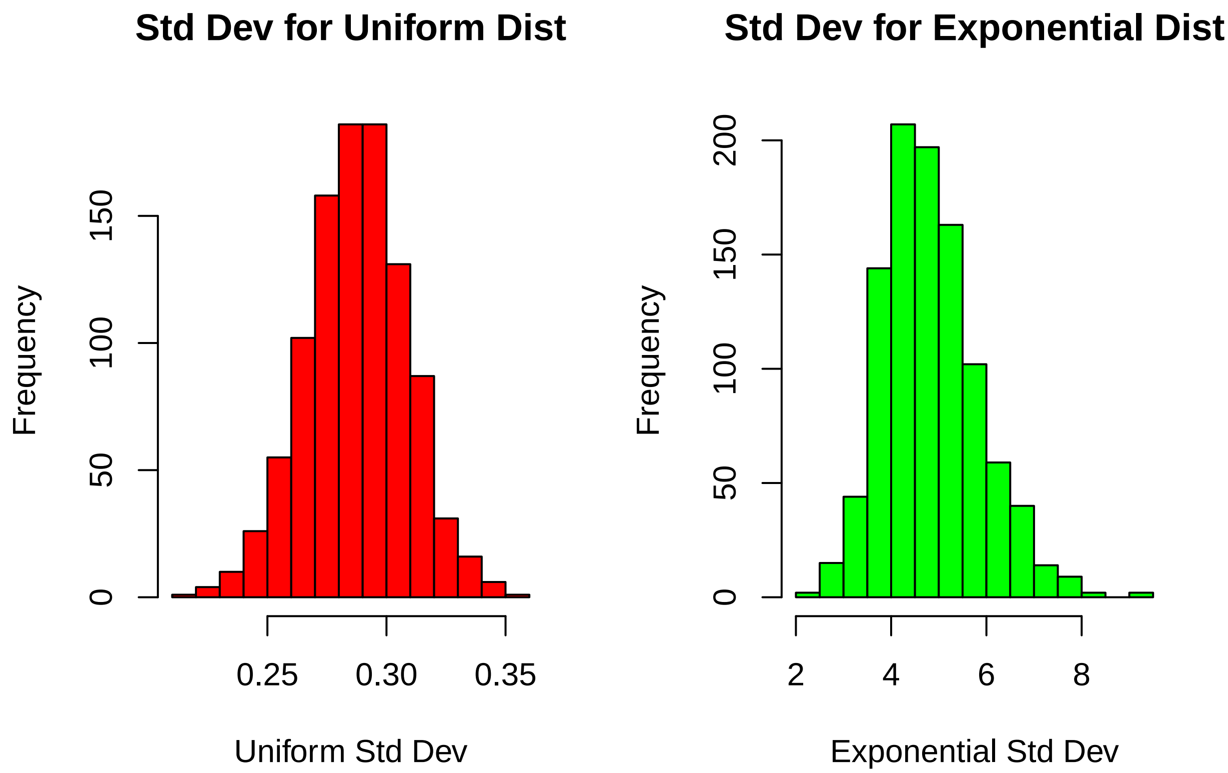
```
par(mfrow = c(1,2))
hist(mns, main = "Means for Uniform Dist", xlab = "Uniform Means",
     col="red")
hist(simMeans, main = "Means for Exponentional Dist",
     xlab = "Exponential Means", col="green")
```



there is a much more obviously similar distribution of means.

Next, compare the standard deviations of the two data sets.

```
par(mfrow = c(1,2))
hist(std, main="Std Dev for Uniform Dist", xlab = "Uniform Std Dev",
     col = "red")
hist(simSD, main = "Std Dev for Exponential Dist",
     xlab = "Exponential Std Dev", col="green")
```



Again, the distribution of standard deviations is also very similar.

These two comparison show the distribution of the simulation of exponential distribution is approximately normal.