# P1

Wang Luyuan

2019/8/24

## Overview:

This course project report looks at a series of exponential distribution iterations, and compares the result set to a normal distribution and the theoretical mean and variance.

```
set.seed(123456789)
```

```
l<-0.2        # lambda
n<-40         # number of exponentials
s<-1000       # number of simulations

# initialize a data frame with a row count of the number of simulations
df<-data.frame(mean=numeric(s))

# iterate 1 to the number of simulations variable
for (i in 1:s) {
  ss<-rexp(n,l)     # simulation set of n exponential with l lambda
  df[i,1]<-mean(ss) # mean of simulation set
}
```

## Sample Mean versus Theoretical Mean:

With our sample data in place, we will compare the sample mean to the theoretical mean. The latter is defined as 1/lambda; we will define `tm` in this manner.

```
tm<-1/l
```

```
format(round(tm,2),nsmall=2)
```

```
## [1] "5.00"
```

We can define the sample mean as the average of the interated set; `am` is defined in the following manner.

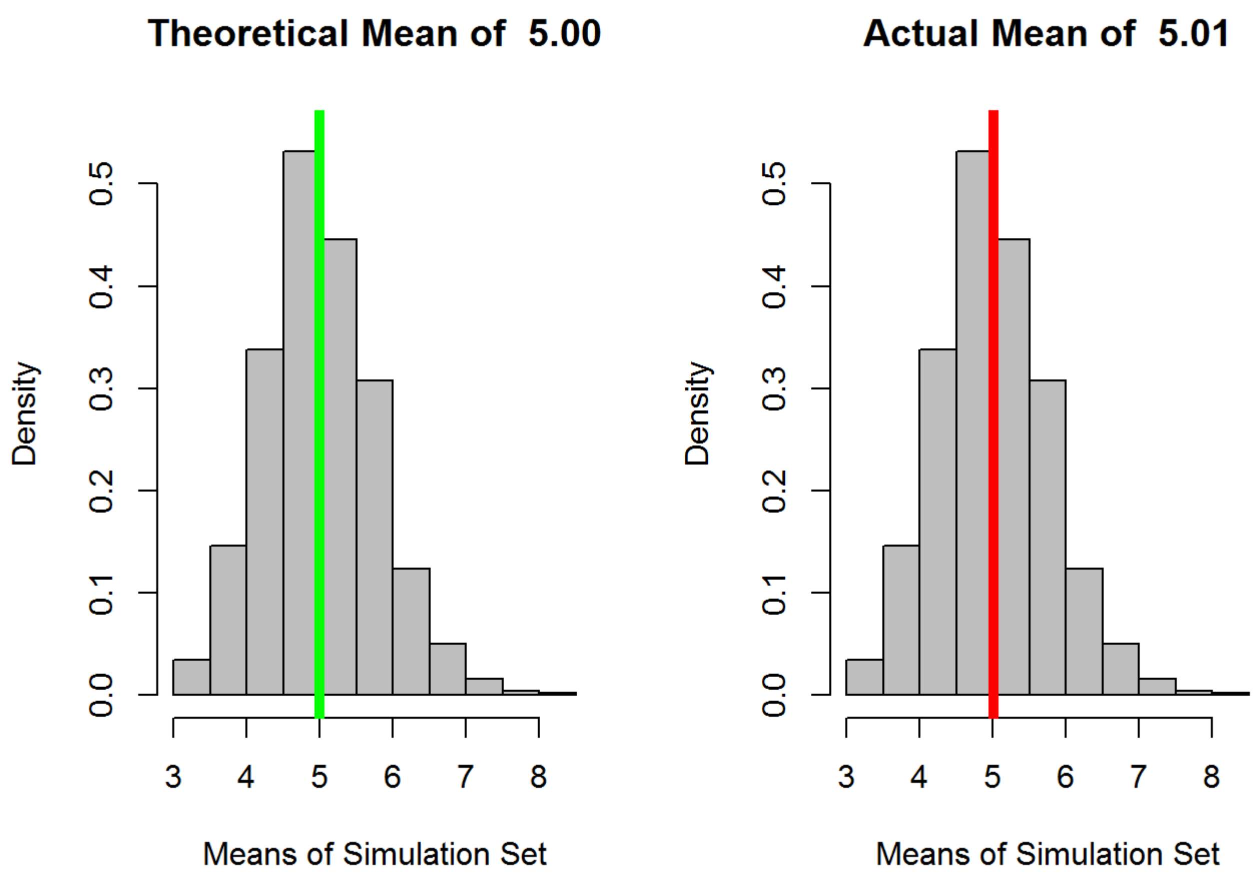
```
am<-mean(df$mean)
```

```
format(round(am,2),nsmall=2)
```

```
## [1] "5.01"
```

As you can see from the above two results, the sample mean (or actual mean) of 5.1 is very close to the theoretical mean of 5.0. When we plot the sample set below and place a vertical line on both mean results, the histograms look almost identical.

```
par(mfrow=c(1,2))
hist(df$mean,probability=T,main=paste('Theoretical Mean of ',format(round(tm,2),nsmall=2)),yl
im=c(0,0.55),col='gray',xlab='Means of Simulation Set')
abline(v=tm,col='green',lwd=5)
hist(df$mean,probability=T,main=paste('Actual Mean of ',format(round(am,2),nsmall=2)),ylim=c(
0,0.55),col='gray',xlab='Means of Simulation Set')
abline(v=am,col='red',lwd=5)
```




## Sample Variance versus Theoretical Variance:

```
tv<-((1/l)^2)/n
```

We can use the R `var` function to determine the actual variance of the observations. We can define `av` as follows.

```
format(round(tv,3),nsmall=3)
```

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

```
av<-var(df$mean)
```

```
format(round(av,3),nsmall=3)
```

```
## [1] "0.601"
```

The results of the theoretical variance as compared to actual variance are very close, as noted above. So, the sample data is ‘spread-out’ in a similar fashion to what is expected.

# Distribution:

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