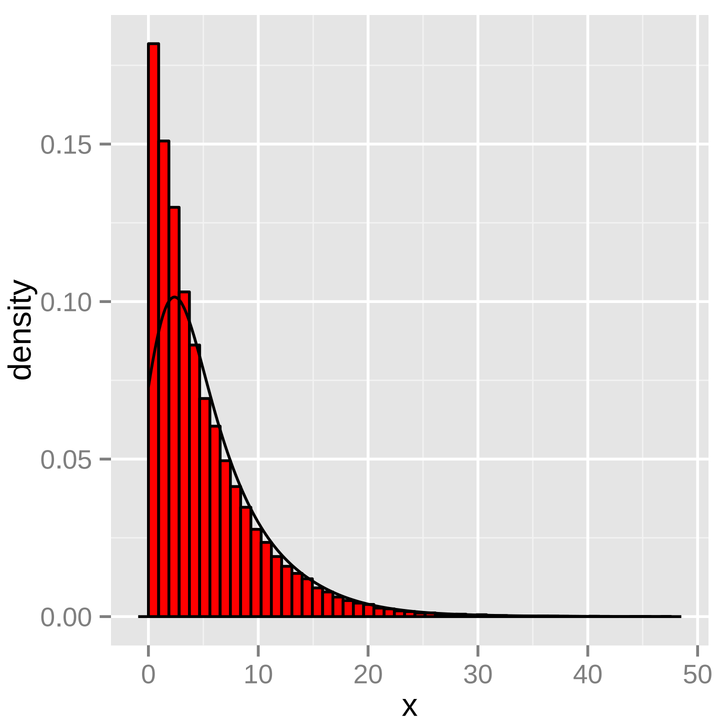
Statistical Inference Project - Part 1

```
lambda=.2

nosim<-1000

mean<-1/lambda

sd<-1/lambda

set.seed(1234)

data<-rexp(nosim * 40,lambda)

n1<-data.frame(x=data)
n10<-data.frame(x=apply(matrix(data,nrow =  nosim,ncol = 10),1,mean))
n20<-data.frame(x=apply(matrix(data,nrow =  nosim,ncol = 20),1,mean))
n25<-data.frame(x=apply(matrix(data,nrow =  nosim,ncol = 25),1,mean))
n40<-data.frame(x=apply(matrix(data,nrow =  nosim,ncol = 40),1,mean))
```

The exponential density is strongly skewed, we need a large value of n before we see the bell curve emerge.

```
ggplot(n1,aes(x = x)) +
        geom_histogram(binwidth=diff(range(data))/50,fill="red",colour = "black",
                       aes(y = ..density..)) +
        stat_density(geom = "line",adjust=5)
```



- 1. Show the sample mean and compare it to the theoretical mean of the distribution.
     Both almost the same as we knew for the LLN theorem.

```r
#Sample mean
mean(n40$x)
```

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

```r
#Theoretical mean of the distribution
mean
```

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

- 2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. For the CLT theorem we knew that Standar deviation of the sample Sx~sd/sqrt(n).

```r
#Standar deviation of the sample
sd(n10$x)
```

```
## [1] 1.522724
```

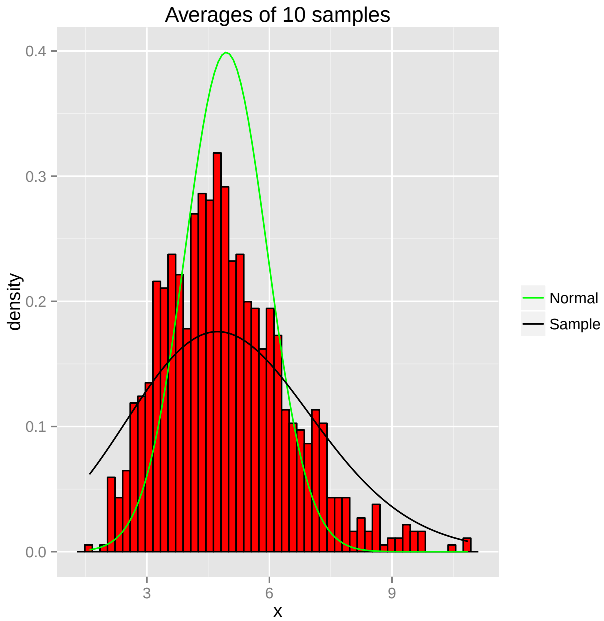
```r
#The theoretical standard deviation of the distribution
sd/sqrt(10)
```

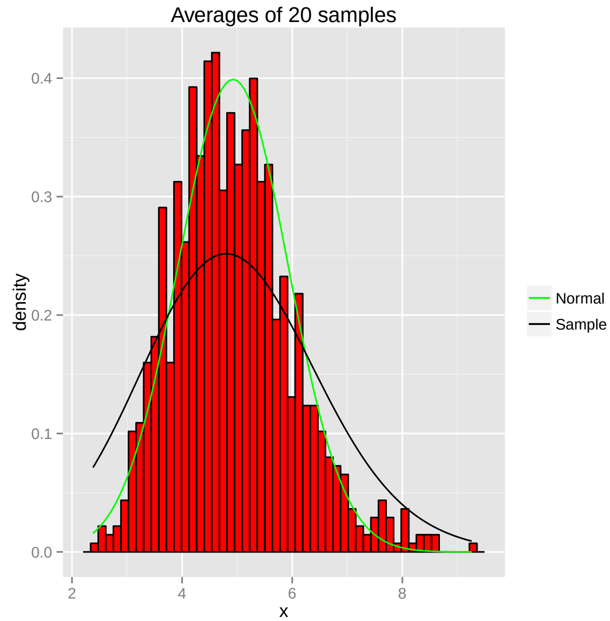
```
## [1] 1.581139
```

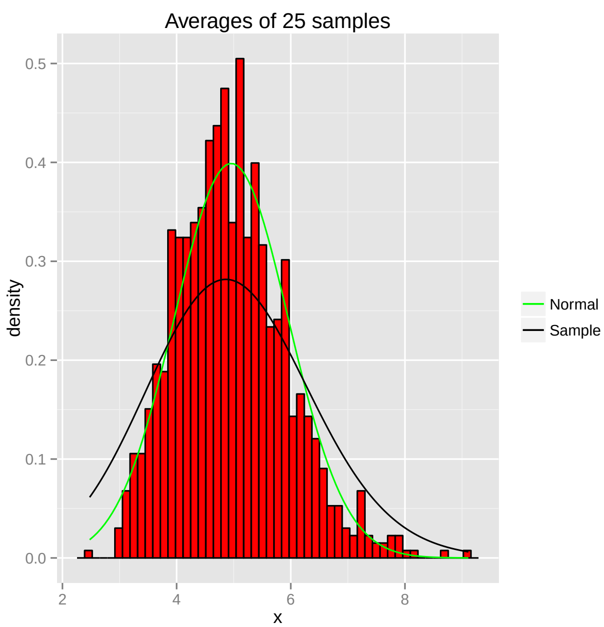
- 3. Show that the distribution is approximately normal.

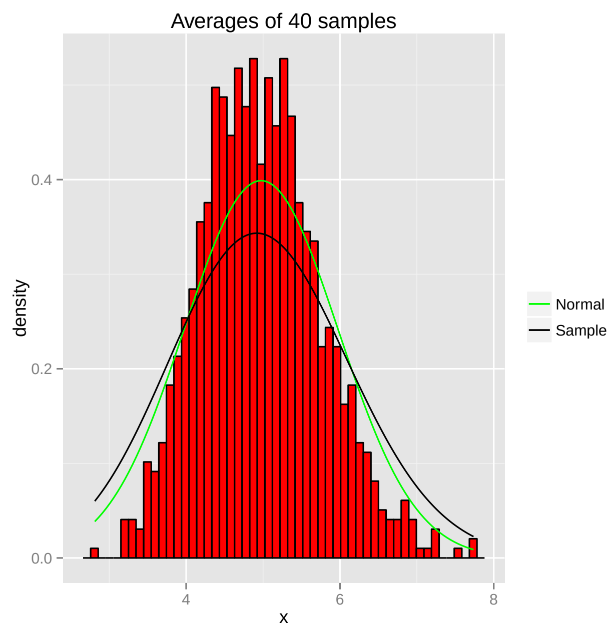
The larger is n the more the closer the exponetial distribution approximate to the normal. I have use the $stat_density$ function that stimates the density of the sample so viasually it is easier to see how the shape is moving to a normal distribution.

```
g1 <- ggplot(n10, aes(x = x)) +
        geom_histogram(binwidth=diff(range(n10$x))/50,
                       fill="red",colour = "black", aes(y = ..density..)) +
        stat_function(fun=dnorm, arg = list(mean=mean(n10$x)),aes(colour = "Normal")) +
        stat_density(geom = "line",adjust=5, aes(colour = "Sample")) +
        scale_colour_manual("",values = c("green","black")) +
        ggtitle("Averages of 10 samples")

g2 <- ggplot(n20, aes(x = x)) +
        geom_histogram(binwidth=diff(range(n20$x))/50,
                        fill="red",colour = "black", aes(y = ..density..)) +
        stat_function(fun=dnorm, arg = list(mean=mean(n20$x)),aes(colour = "Normal")) +
        stat_density(geom = "line",adjust=5, aes(colour = "Sample")) +
        scale_colour_manual("",values = c("green","black")) +
        ggtitle("Averages of 20 samples")

g3 <- ggplot(n25, aes(x = x)) +
        geom_histogram(binwidth=diff(range(n25$x))/50,
                        fill="red",colour = "black", aes(y = ..density..)) +
        stat_function(fun=dnorm, arg = list(mean=mean(n25$x)),aes(colour = "Normal")) +
        stat_density(geom = "line",adjust=5, aes(colour = "Sample")) +
        scale_colour_manual("",values = c("green","black")) +
        ggtitle("Averages of 25 samples")

g4 <- ggplot(n40, aes(x = x)) +
        geom_histogram(binwidth=diff(range(n40$x))/50,
                        fill="red",colour = "black", aes(y = ..density..)) +
        stat_function(fun=dnorm, arg = list(mean=mean(n40$x)),aes(colour = "Normal")) +
        stat_density(geom = "line",adjust=5, aes(colour = "Sample")) +
        scale_colour_manual("",values = c("green","black")) +
        ggtitle("Averages of 40 samples")

grid.arrange(g1, g2, g3,g4, ncol = 2, nrow=2)
```


Averages of 10 samples
density
0.0
0.1
0.2
0.3
0.4
3
6
9
x
Normal
Sample


Averages of 20 samples
density
0.0
0.1
0.2
0.3
0.4
2
4
6
8
x
Normal
Sample


Averages of 25 samples
density
0.0
0.1
0.2
0.3
0.4
0.5
2
4
6
8
x
Normal
Sample


Averages of 40 samples
density
0.0
0.2
0.4
4
6
8
x
Normal
Sample