Statistical Inference Project by Ola

This is a pdf generated by an R Markdown document for a project in Coursera Statistical Inference Class. In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem.

Before I proceed with the project, I add the library ggplot2.

```
library(ggplot2)
```

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

Set lambda = 0.2 in the exponential distribution for all of the simulations. We will investigate the distribution of averages of 40 exponentials in one thousand simulations. The simulation is saved in the matrix 'simulation' computed with the code below. We also show the distribution of the results of 1000 simulations in the histogram. The vertical line shows the mean for the exponential distribution, 1/lambda.

```
n_sim <- 1000 # number of simulations
n <- 40 # number of draws in a single simulation
lam <- 0.2 #lambda parameter
simulation <- matrix(rexp(n_sim * n , rate=lam), n_sim)
simulation_means <- apply(simulation, 1, mean)

plotdata <- data.frame(simulation_means)
p1 <- ggplot(data=plotdata, aes(plotdata$simulation_means))
p1 <- p1 + geom_histogram(aes(y=..count..), colour="black", fill = "red")
p1 + geom_vline(xintercept = 1/lam)
```
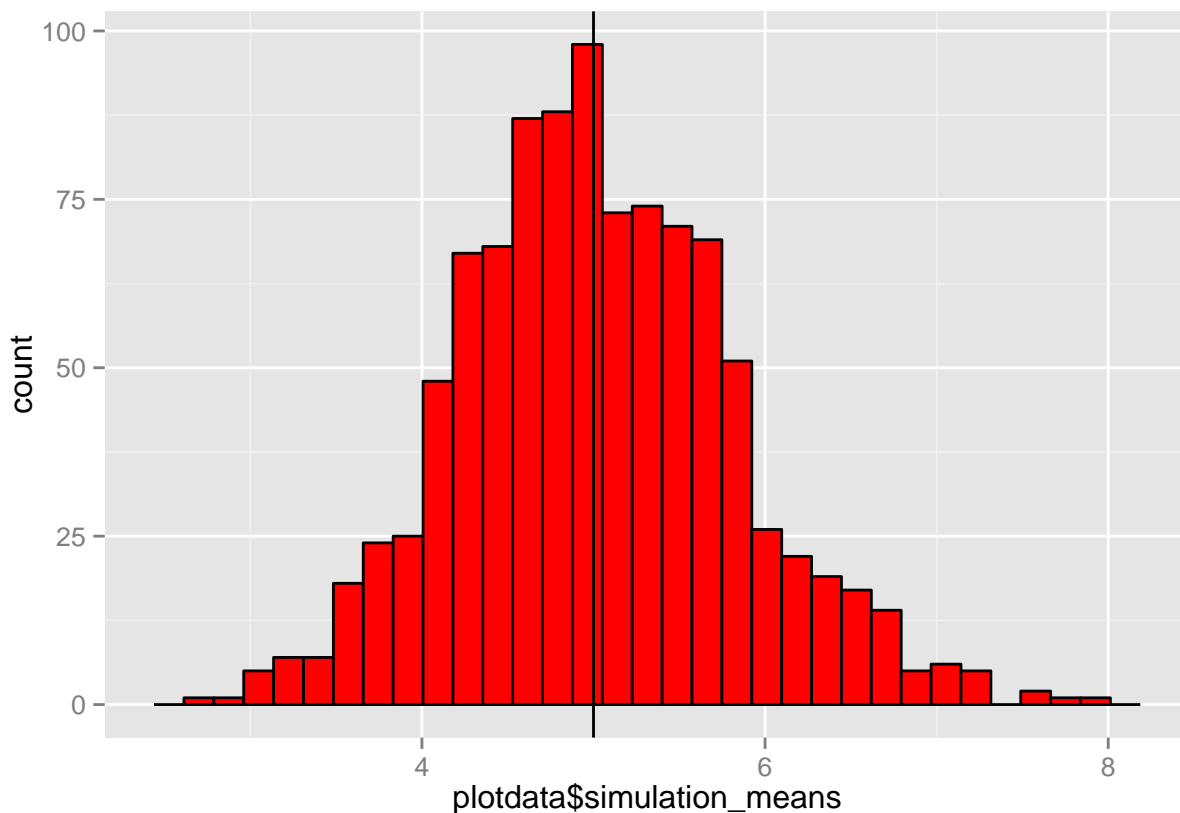
A direct comparison of the mean of the exponential distribution and a sample mean:

```
print (paste("Mean of the exponential distribution = ", 1/lam))
```

```
## [1] "Mean of the exponential distribution =  5"
```

```
print (paste("Sample mean ", round(mean(simulation_means),4)))
```

```
## [1] "Sample mean  5.0224"
```

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

```
print (paste("Variance of the exponential distribution = ", 1/lam/lam/n))
```

```
## [1] "Variance of the exponential distribution =  0.625"
```

```
print (paste("The variance of the sample ", round(var(simulation_means),4)))
```

```
## [1] "The variance of the sample  0.643"
```

3. Show that the distribution is approximately normal.

One can see that the distribution is approximately normal by looking at the histogram of our results. For a better comparison we display a smooth density estimate of the sample data computed automatically within the ggplot 2 (black curve) and add a layer of an actual Gaussian distribution with mean = 1/lambda and sd = 1/lambda/sqrt(n). One can compare those distributions and see they are very close.

```r
p2 <- ggplot(plotdata, aes(x = simulation_means))
p2 <- p2 + geom_histogram(aes(y=..density..), colour="black", fill = "red")
p2 <- p2 + geom_density(colour= "black", size=1)
p2 + stat_function(fun = dnorm, colour="blue", args=list(mean=1/lam, sd=1/lam/sqrt(n)))
```