Statistical Inference

Course Project

Simulation Exercise

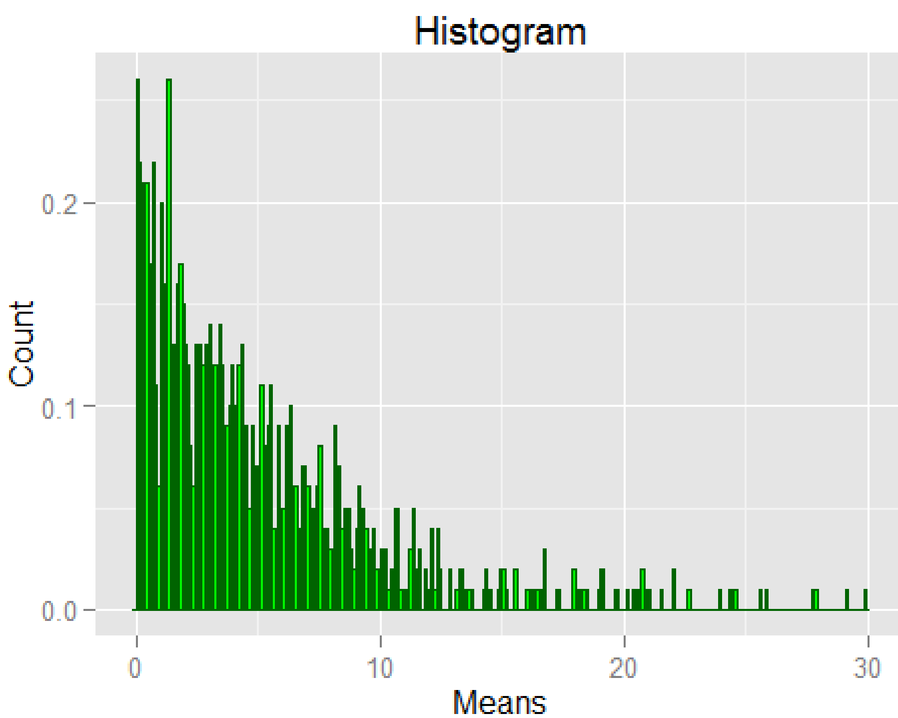
In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

In this simulation, I will set lmabda = 0.2. I will investigate the distribution of averages of 40 exponentials and I will do a thousand simulations.

First, we will show the sample mean and compare it to the theoretical mean of the distribution.

Let's take a look at the exponential distribution

```
## Warning: position_stack requires constant width: output may be incorrect
```



Now let's set the parameters for the simulation.

```
set.seed(526)
lambda <- 0.2
```

```
n <- 40
nsim <- 1000
```

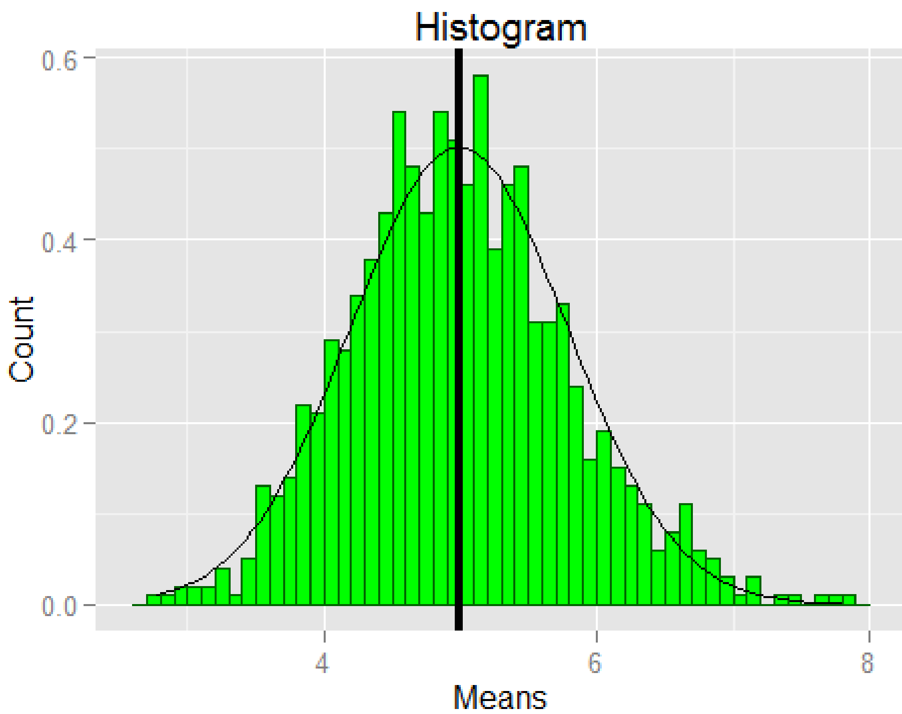
Now we can simulate the distribution.

```
## 
## Attaching package: 'dplyr'
## 
## The following object is masked from 'package:stats':
## 
##     filter
## 
## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

expsim <- as.data.frame(matrix(rexp(nsim*n, lambda),nsim,n))
expsim$Means <- rowMeans(expsim)
```

Now, we will calculate the means of the 40 rows and create a histogram of those averages. We expect that the histogram will show that the averages of 40 exponential distributions are approximately normally distributed.

```
means <- rowMeans(expsim)
histogram <- ggplot(expsim, aes(x=Means))+
    geom_histogram(aes(y=..density..),binwidth=0.1,colour = "darkgreen", fill
= "green")+
    stat_function(fun = dnorm, args = list(mean = mean(expsim$Means),
                                           sd = sd(expsim$Means)))+
    labs(x = "Means", y= "Count", title = "Histogram")+
    geom_vline(aes(xintercept=1/lambda), size=1.5)
print(histogram)
```




```
samplemean <- mean(expsim$Means)
truemean <- 1/lambda
```

The theoretical mean of the exponential distribution is 1/lambda or 5.00. The sample mean of the exponential distributions is 4.99. We can see that the two numbers are very close.

Now I will calculate sample variance and compare it to the theoretical variance of the distribution.

```
truevar = 1/(lambda^2*n)
samplevar = (sd(rowMeans(expsim)))^2
```

The theoretical variance of the distribution is 1/(lambda^2*n) or 0.62. The sample variance of the distribution is 0.63.

According to the Central Limit Theorem, as the sample size increases, the distribution of the means of the exponential distributions will converge to normal distribution even if the exponential distributions themselves are very different.

As is evident from the histogram above, the distribution of means of a large number of exponential distributions is approximately normal.