Project: Statistical Inference - Part I

Part 1: Simulation Exercise Instructionsmenos

In this project you 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. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

As a motivating example, compare the distribution of 1000 random uniforms

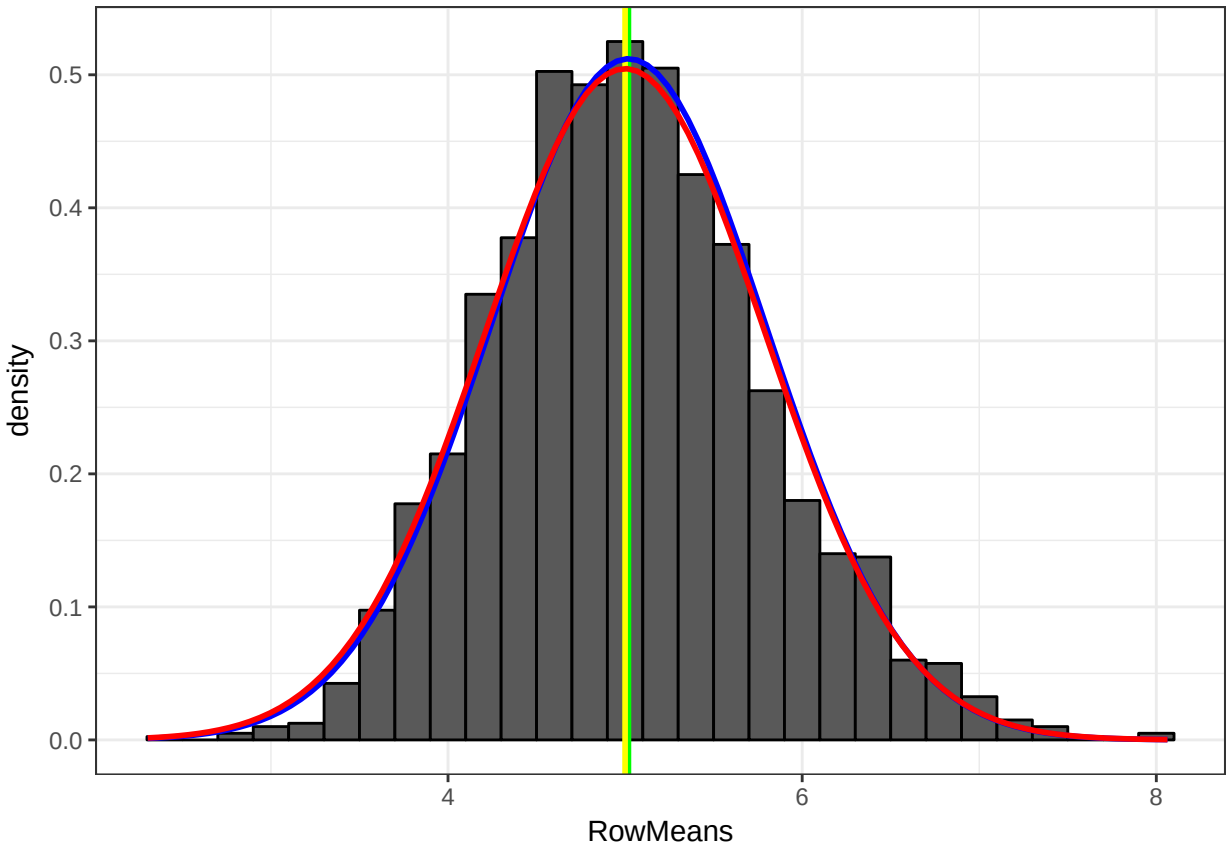
```
# Number of values (n) = 40 lambda = 0.2 number of iterations, as least 1000. numsim=2000
n <-40
lambda <-0.2
numsim <-2000
dataset <- matrix(rexp(n*numsim,lambda),numsim)


theoretical_mean<-1/lambda
RowMeans<-apply(dataset,1,mean)
actual_mean<-mean(RowMeans)
theoretical_sd<-((1/lambda) * (1/sqrt(n)))
actual_sd<-sd(RowMeans)

theoretical_var<-theoretical_sd^2
actual_var<-var(RowMeans)

# - Plot
dfRowMeans<-data.frame(RowMeans)

ggplot(dfRowMeans,aes(x=RowMeans)) +
    geom_histogram(binwidth = lambda, color="black", aes(y = ..density..)) +
    # Show the sample mean and compare it to the theoretical mean of the distribution.
    geom_vline(xintercept=actual_mean,size=1.0, color="green") +
    geom_vline(xintercept=theoretical_mean,size=1.0, color="yellow") +
    # Show how variable the sample is (via variance) and compare it to the theoretical
    # variance of the distribution.
    stat_function(fun=dnorm,args=list(mean=actual_mean, sd=actual_sd),
                  color = "blue", size = 1.0) +
    stat_function(fun=dnorm,args=list(mean=theoretical_mean, sd=theoretical_sd),
                  color = "red", size = 1.0) +
    theme_bw()
```


density
0.0
0.1
0.2
0.3
0.4
0.5
4
6
8
RowMeans