Peer-graded Assignment: Statistical Inference Course Project PART1

Gabriela Ochoa

Instructions

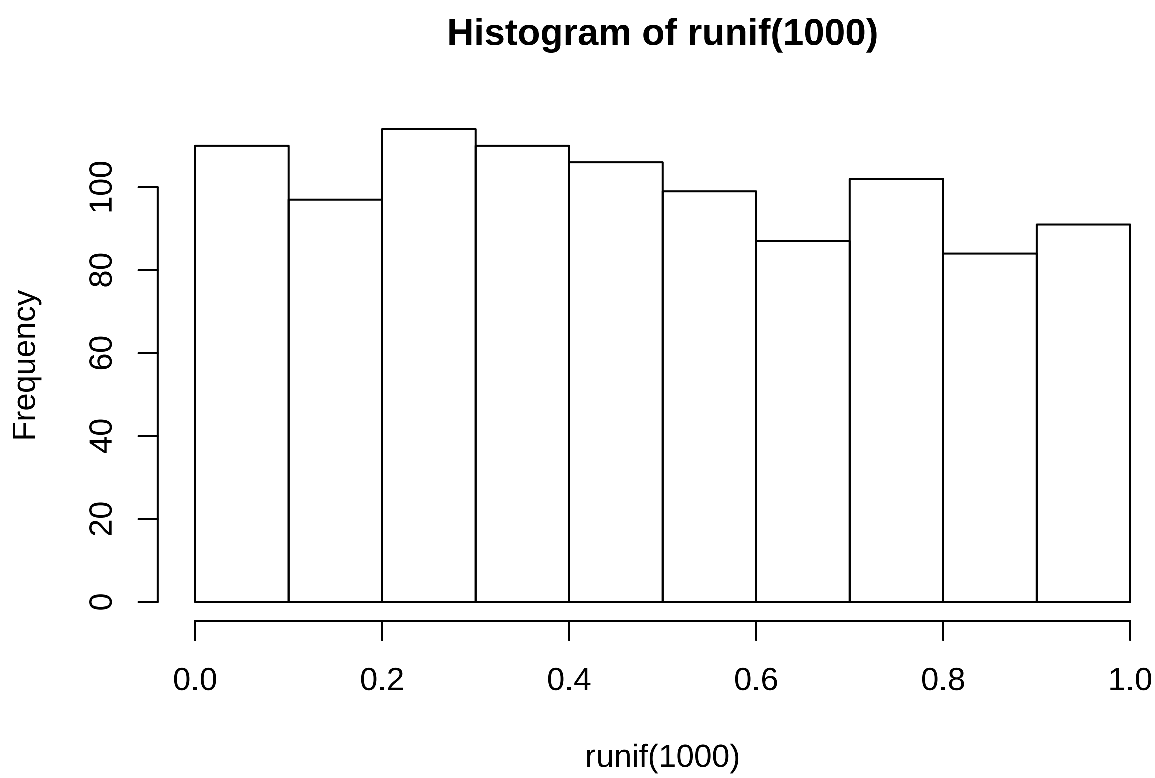
The project consists of two parts:

- A simulation exercise.
- Basic inferential data analysis.

Part 1: Simulation Exercise instructions

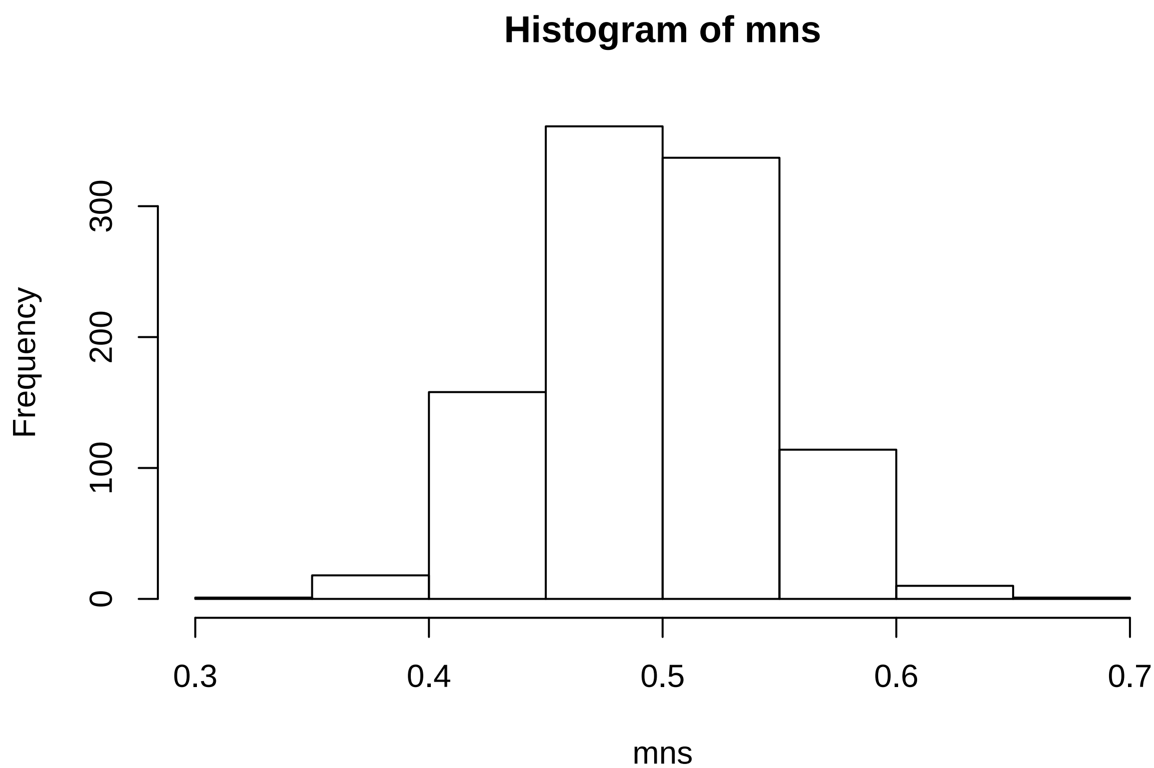
Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

```
hist(runif(1000))
```



and the distribution of 1000 averages of 40 random uniforms

```r
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(runif(40)))
hist(mns)
```



This distribution looks far more Gaussian than the original uniform distribution!

This exercise is asking you to use your knowledge of the theory given in class to relate the two distributions.

Resolution Part 1

We start by running a 1000 simulations of 40 exponentials.

```r
## Required libraries
library(ggplot2)
library(knitr)
## setting seed
set.seed(1)
```

```r
#No. of values (n) = 40, lambda = 0.2, No.of iterations, at least 1000, numsim=2000,Theoretical mean =1
lambda <- 0.2
nosim <- 1:1000 # Number of Simulations/rows
n <- 40
```

Generating data using rexp

Use the rexp function to develop a dataset with the mean and lambda specified above.

```r
#sd(apply(matrix(rnorm(nosim*n), nosim), 1, mean))
```

```r
#Create a matrix of simulated values:
e_matrix <- data.frame(x = sapply(nosim, function(x) {mean(rexp(n, lambda))}))
```

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

```r
sim_mean <- apply(e_matrix, 2, mean)
sim_mean
```

```
##        x
## 4.990025
```

Which is very close to the expected theoretical center of the distribution:

```r
th_mean <- 1/lambda
th_mean
```

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

2.Show how variable it is and compare it to the theoretical variance of the distribution. .

```r
sim_SD <- sd((e_matrix$x))
sim_SD
```

```
## [1] 0.7817394
```

```r
sim_Var <- var(e_matrix$x)
sim_Var
```

```
## [1] 0.6111165
```

Let's compare, the expected theretical SD and Variance are:

```r
th_SD <- (1/lambda)/sqrt(n)
th_SD
```

```
## [1] 0.7905694
```

```r
th_Var <- th_SD^2
th_Var
```

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

Comparing Theoretical and actual Values of mean,Standard deviation and variance Table

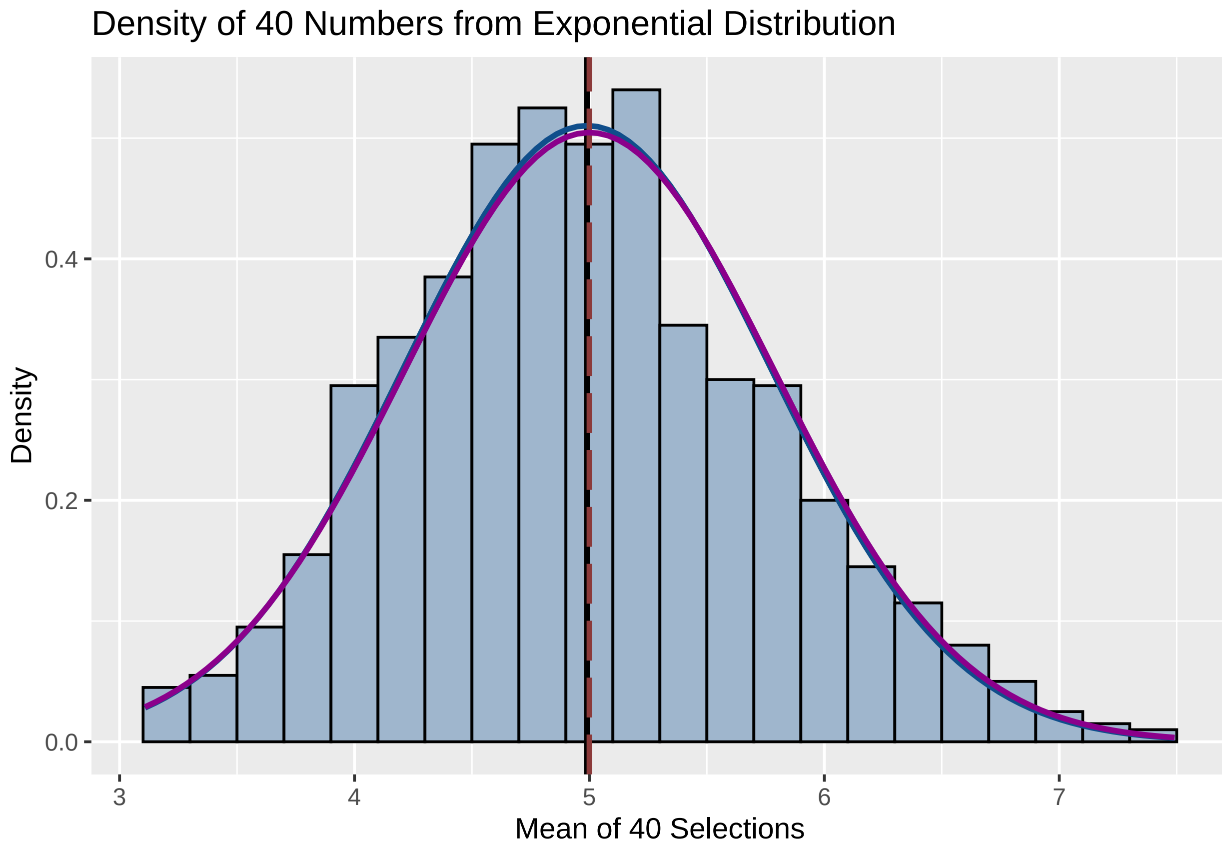
Variable	Theoretical val	Actual Val
Mean	5	5.048
SD	0.791	0.796
Var	0.625	0.634

We can verify that the differences are minimal, as expected.

3. Show that the distribution is approximately normal.

```r
plot <- ggplot(data = e_matrix, aes(x = x)) +
    geom_histogram(aes(y=..density..), binwidth = 0.20, fill="slategray3", col="black")
plot <- plot + labs(title="Density of 40 Numbers from Exponential Distribution", x="Mean of 40 Selectio
plot <- plot + geom_vline(xintercept=sim_mean,size=1.0, color="black")
plot <- plot + stat_function(fun=dnorm,args=list(mean=sim_mean, sd=sim_SD),color = "dodgerblue4", size =
plot <- plot+ geom_vline(xintercept=th_mean,size=1.0,color="indianred4",linetype = "longdash")
plot <- plot + stat_function(fun=dnorm,args=list(mean=th_mean, sd=th_SD),color = "darkmagenta", size =
```

```
plot
```



we conclude that the function appears to aproximate to nearly Normal.