Statistical_Infer_Part1.Rmd

Joe Okelly

Overview

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.

Simulations

The exponential distribution can be simulated in R with `rexp(n, lambda)`, where `lambda` is the rate parameter and `n` is the number of Simulations. 'lambda' is set to `0.2`. load the `ggplot2` plotting library.

```
library(ggplot2)
library(knitr)
```

Intialiaze the variables

```
num_Simulation <- 1000
sampSize <- 40
lambda <- 0.2
set.seed(300413)
```

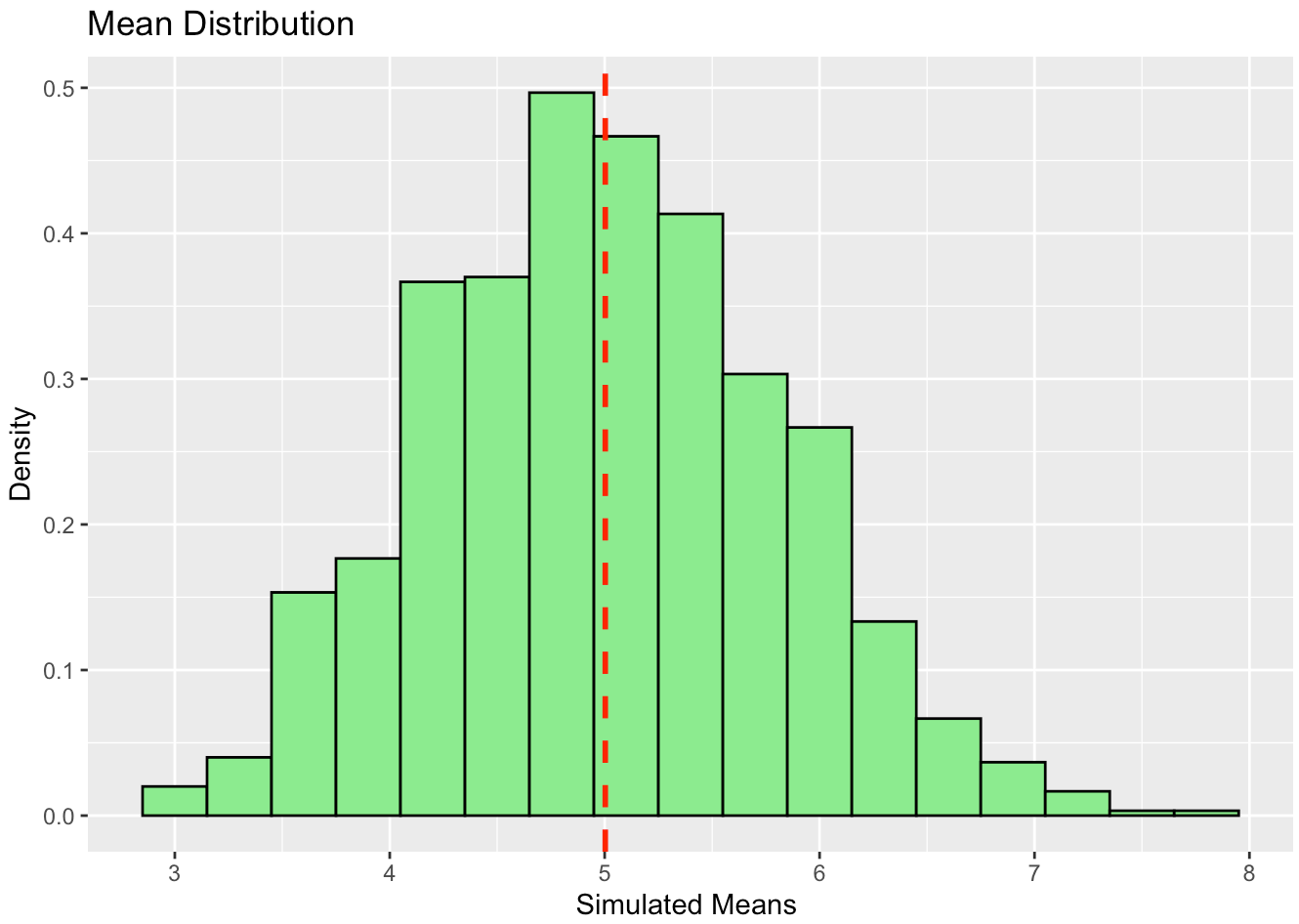
Define a matrix of 1000 rows x 40 columns, corresponds to Number of Sumulations and Sample Size.

```
simu_data <- replicate(num_Simulation, rexp(sampSize, lambda))
mean_simulation_data <- apply(simu_data, 2, mean)

simu_Matrix <- matrix(rexp(n = num_Simulation * sampSize, rate = lambda), num_Simulatio
n, sampSize)
simu_Mean <- rowMeans(simu_Matrix)
simu_Data <- data.frame(cbind(simu_Matrix, simu_Mean))
```

Define a ggplot to visiualise the data.

```
ggplot(data = simu_Data, aes(x = simu_Mean)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 0.3, fill = "lightgreen", colo
r = "black") +
  labs(title = "Mean Distribution", x = "Simulated Means ", y = "Density") +
  geom_vline(aes(xintercept=mean(simu_Mean)), color="red", linetype="dashed", linewidth=
1)
```




```
actual_Mean <- mean(simu_Mean)
theo_Mean <- (1 / lambda)
act_Variance <- var(simu_Mean)
theo_Variance <- ((1 / lambda) ^ 2) / sampSize

print(paste("Actutal Mean :", actual_Mean))
```

```
## [1] "Actutal Mean : 5.00279885440302"
```

```
print(paste("Theo Mean :", theo_Mean))
```

```
## [1] "Theo Mean : 5"
```

```
print(paste("Actual VAriance :", act_Variance))
```

```
## [1] "Actual VAriance : 0.660816925759118"
```

```
print(paste("Theorical VAriance :", theo_Variance))
```

```
## [1] "Theorical VAriance : 0.625"
```

```
qqnorm(mean_simulation_data)
qqline(mean_simulation_data, col = "magenta")
```