

Peer-graded Assignment: Statistical Inference Course Project

Gilbert Toroitich Tarus

27/01/2021

Overview

This submission is a report on the investigation of exponential distribution in R comparing it with the Central Limit Theorem. This project is done using a sample of 40 exponential for thousand simulations.

Simulation Example

The Exponential Distribution

In order to simulate 40 observations, we utilize the `rexp()` function in R with arguments `n = 40` and `lambda = .2`. However, to generate 1000 such simulations, we use the `matrix()` function with dimensions `1000 * 40`

```
nosim <- 1000
n <- 40
lamda <- 0.2
expSimulation <- matrix(rexp(n*nosim,lamda),nosim,n)
theoreticalMean <- 1/lamda
```

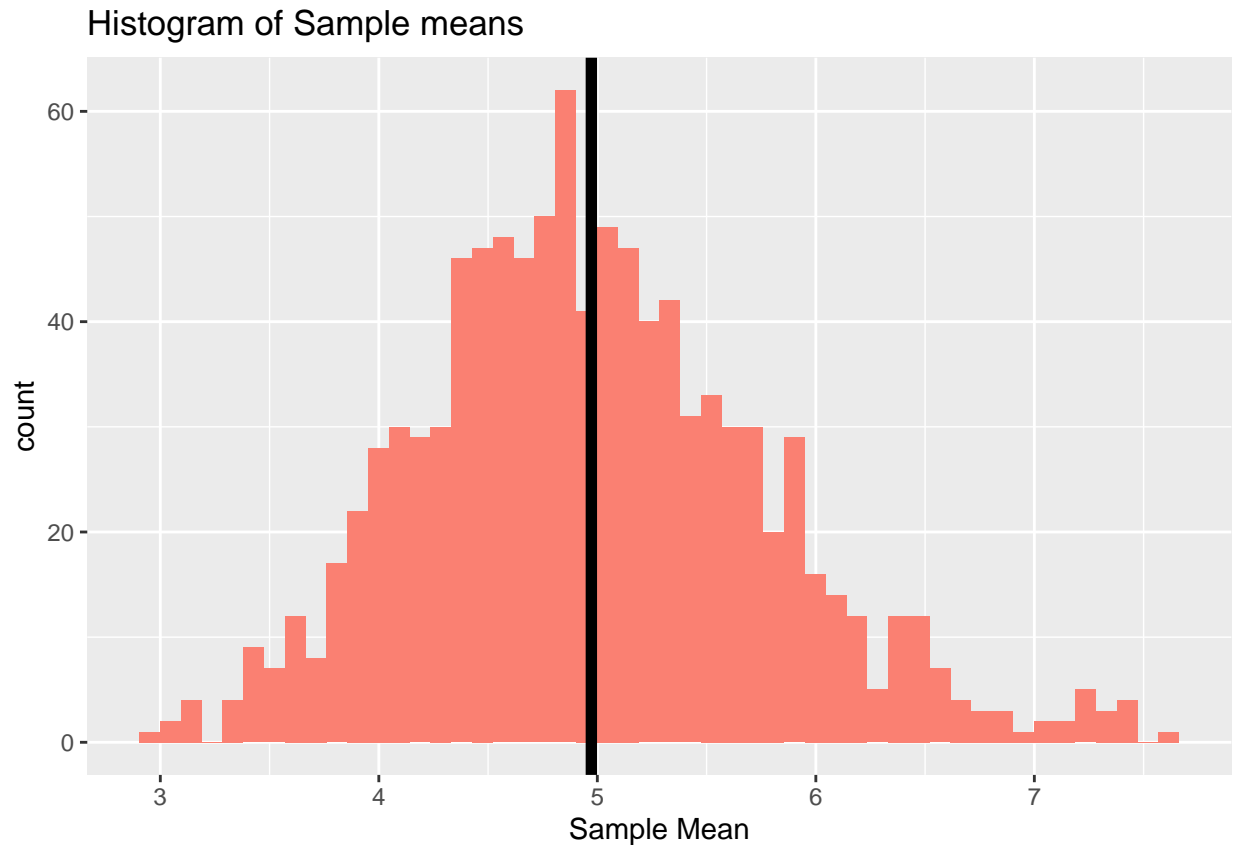
Calculate the mean of each sample. We will have 1000 averages.

```
sampledMean <- apply(expSimulation, 1, mean)
```

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

Plot

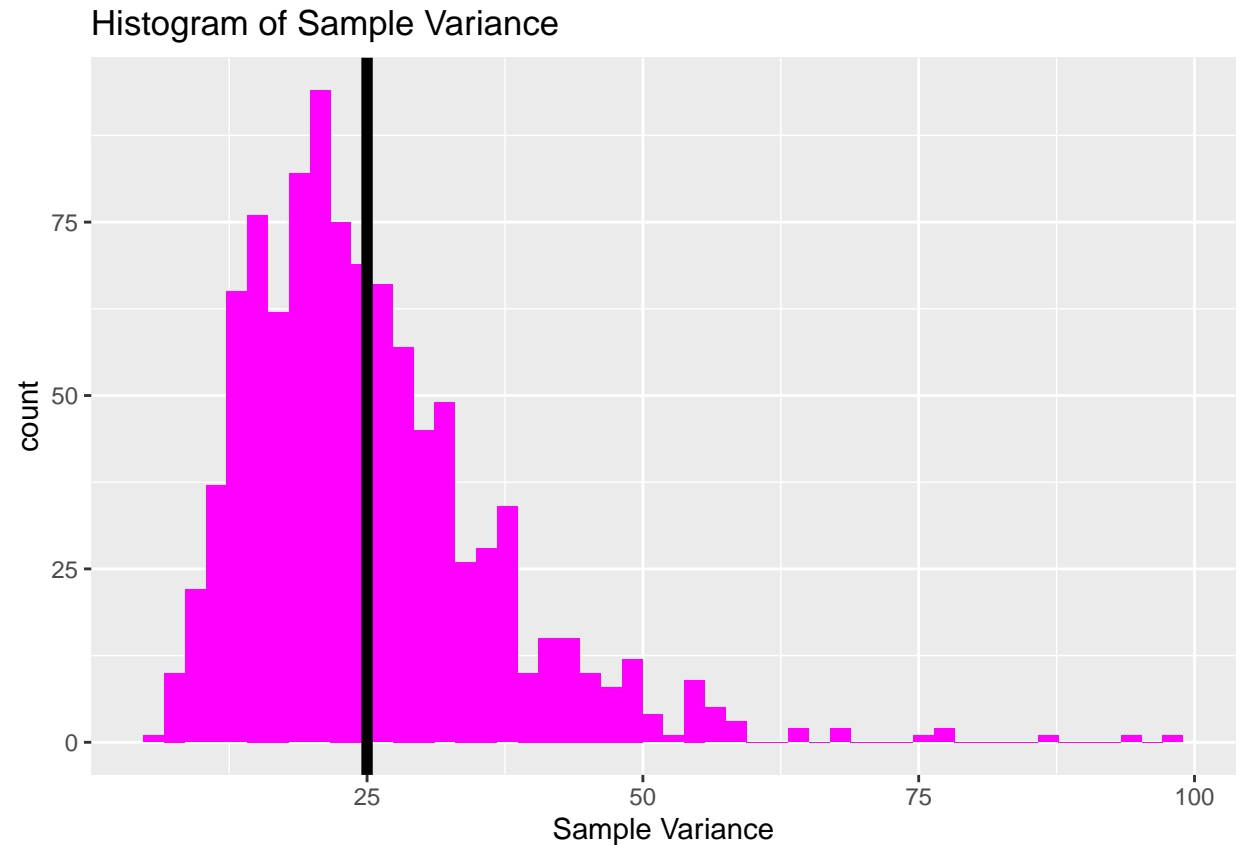
```
library(ggplot2)
ggplot(data.frame(x <- sampledMean),aes(x = x))+
  geom_histogram(fill = "salmon",bins = 50)+
  geom_vline(xintercept = mean(x),size = 2, color = "black")+
  labs(title = "Histogram of Sample means",x = "Sample Mean")
```



The sample means are centered around 5 shown by the vertical line in the plot is equal to theoretical mean 5.

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

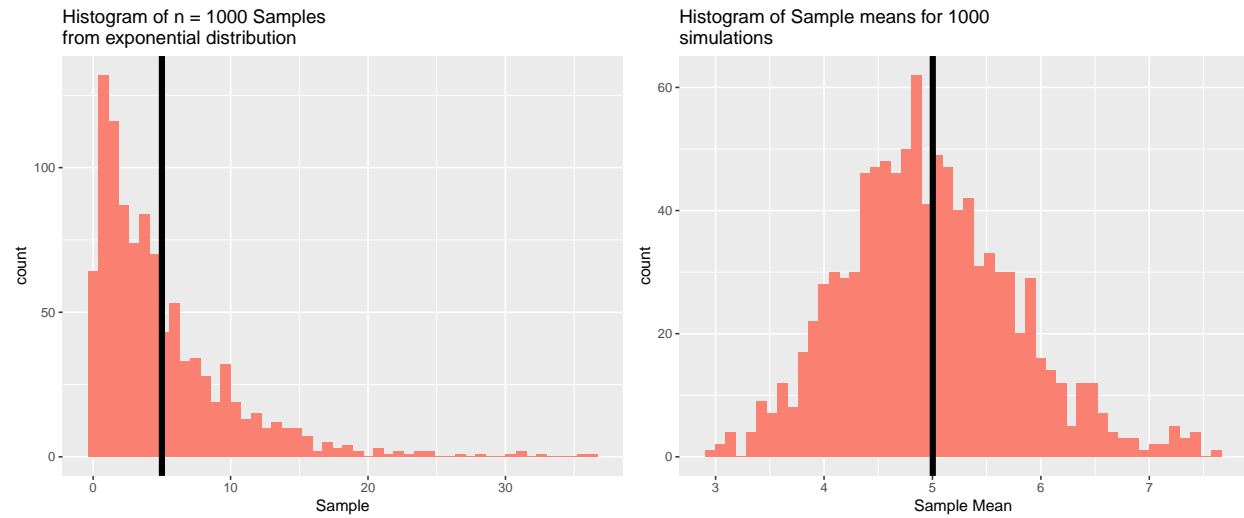
```
sampledVariance <- apply(expSimulation, 1, var)
ggplot(data.frame(x <- sampledVariance), aes(x = x)) +
  geom_histogram(fill = "magenta", bins = 50) +
  geom_vline(xintercept = (1/lamda)^2, size = 2, color = "black") +
  labs(title = "Histogram of Sample Variance", x = "Sample Variance")
```



The distribution of sample variance is centered around the theoritical variance 25

3. Show that the distribution is approximately normal.

```
library(gridExtra)
set.seed(1234)
x <- rexp(nosim, lamda)
g1 <- ggplot(data.frame(x), aes(x = x)) +
  geom_histogram(fill = "salmon", bins = 50) +
  geom_vline(xintercept = mean(x), size = 2, color = "black") +
  labs(title = "Histogram of n = 1000 Samples\nfrom exponential distribution", x = "Sample")
g2 <- ggplot(data.frame(x = sampledMean), aes(x = x)) +
  geom_histogram(fill = "salmon", bins = 50) +
  geom_vline(xintercept = mean(x), size = 2, color = "black") +
  labs(title = "Histogram of Sample means for 1000 \nsimulations", x = "Sample Mean")
grid.arrange(g1, g2, ncol = 2)
```



The plot on the left side follows an exponential distribution with mean 5.

The plot on the right side is bell-shaped. This follows a normal distributin with mean 5 and standard error 0.0251094