Statistical Inference Course Project - Part I - Simulation Exercise

Huseyin Emre Balli

December 20, 2018

Overview

In this project we 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**.

We will; 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.

Simulation

Set simulation variables as described in the assignment

```
lambda <- 0.2
n <- 40
sim_count <- 1000
```

Set seed for reproducibility

```
set.seed(88423)
```

Conduct exponential simulation 'sim_count' times with 'n' exponentials and rate parameter equals 'lambda'

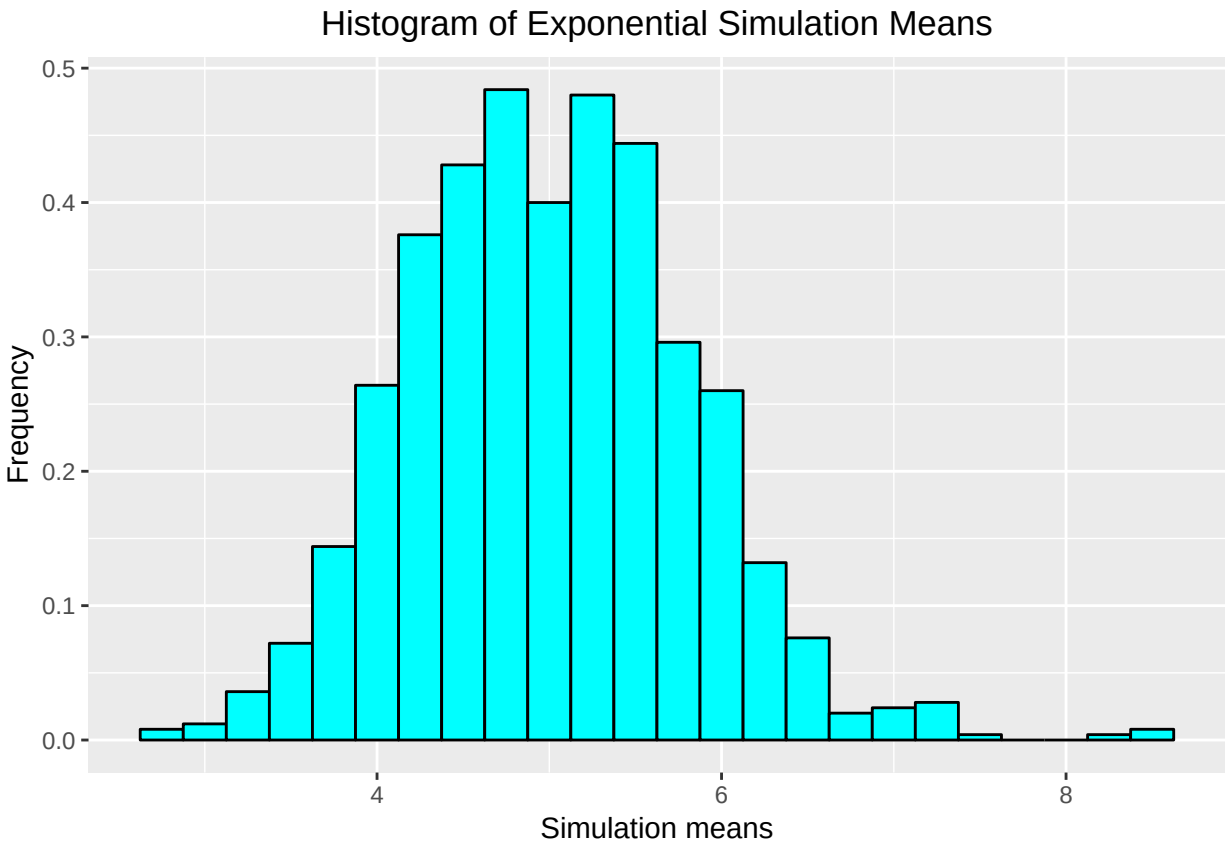
```
simulation <- replicate(sim_count, rexp(n, lambda))
means <- apply(simulation, 2, mean)
```

Take a quick look of means

```
range(means)
```

```
## [1] 2.731732 8.575295
```

```
means <- data.frame(means)
library(ggplot2)
plot_means <- ggplot(means, aes(means)) +
        geom_histogram(aes(y = ..density..), color = "black", fill = "cyan",
                       binwidth = 0.25) +
                labs(x = "Simulation means", y = "Frequency",
                     title = "Histogram of Exponential Simulation Means") +
                        theme(plot.title = element_text(hjust = 0.5))
plot_means
```



Question 1 - Sample Mean vs Theoretical Mean

Calculate theoretical and sample means

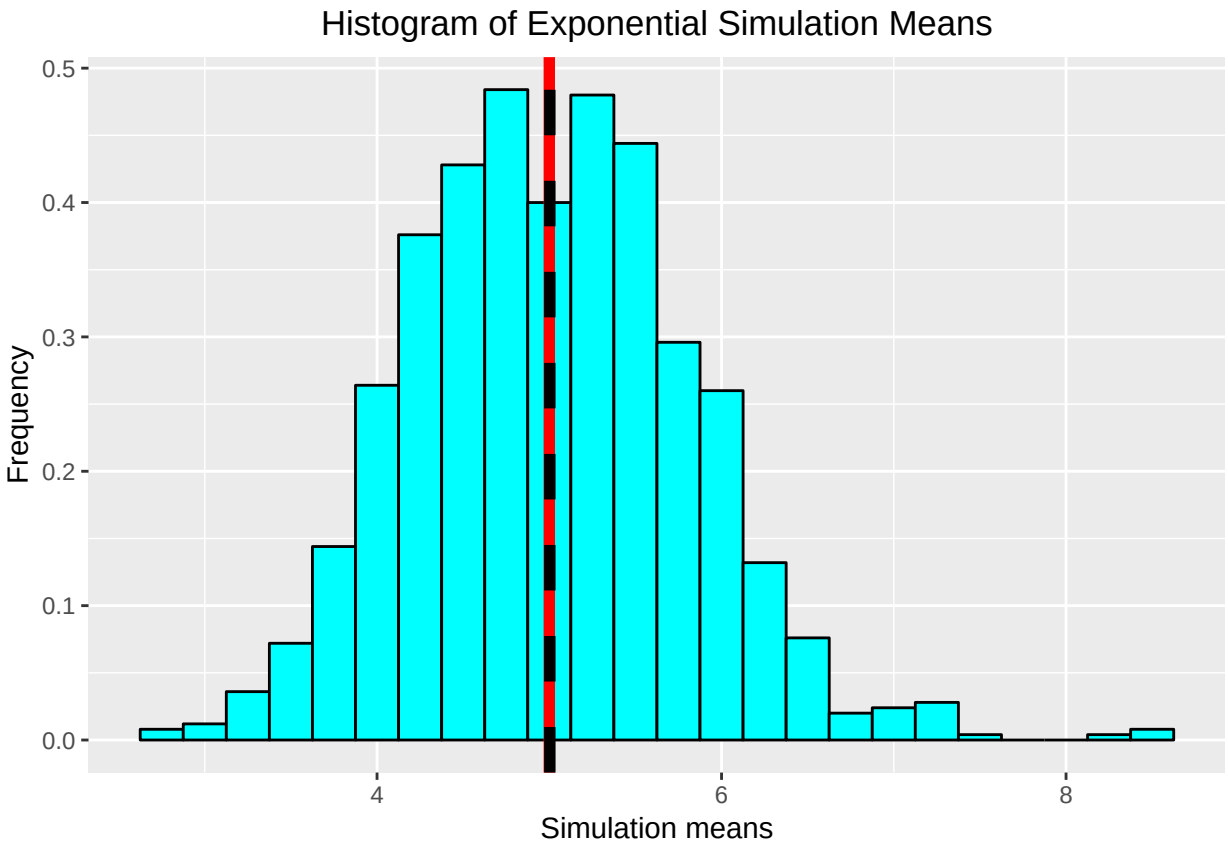
```
theory_mean <- 1 / lambda
sample_mean <- mean(means$means)
```

Compare them and plot

```
abs(theory_mean - sample_mean)
```

```
## [1] 0.003324618
```

```
plot_means +
        geom_vline(xintercept = theory_mean, lwd = 2, color = "red") +
                geom_vline(xintercept = sample_mean, lwd = 2, lty = 2, color = "black")
```



We can see that sample mean and theoretical mean are very close to each other.

Question 2 - Sample Variance vs Theoretical Variance

Calculate theoretical and sample variances

```
theory_variance <- (1 / lambda) ^ 2 / n
sample_variance <- var(means$means)
data.frame(Sample.Variance = sample_variance, Theoretical.Variance = theory_variance,
           Difference = theory_variance - sample_variance)
```

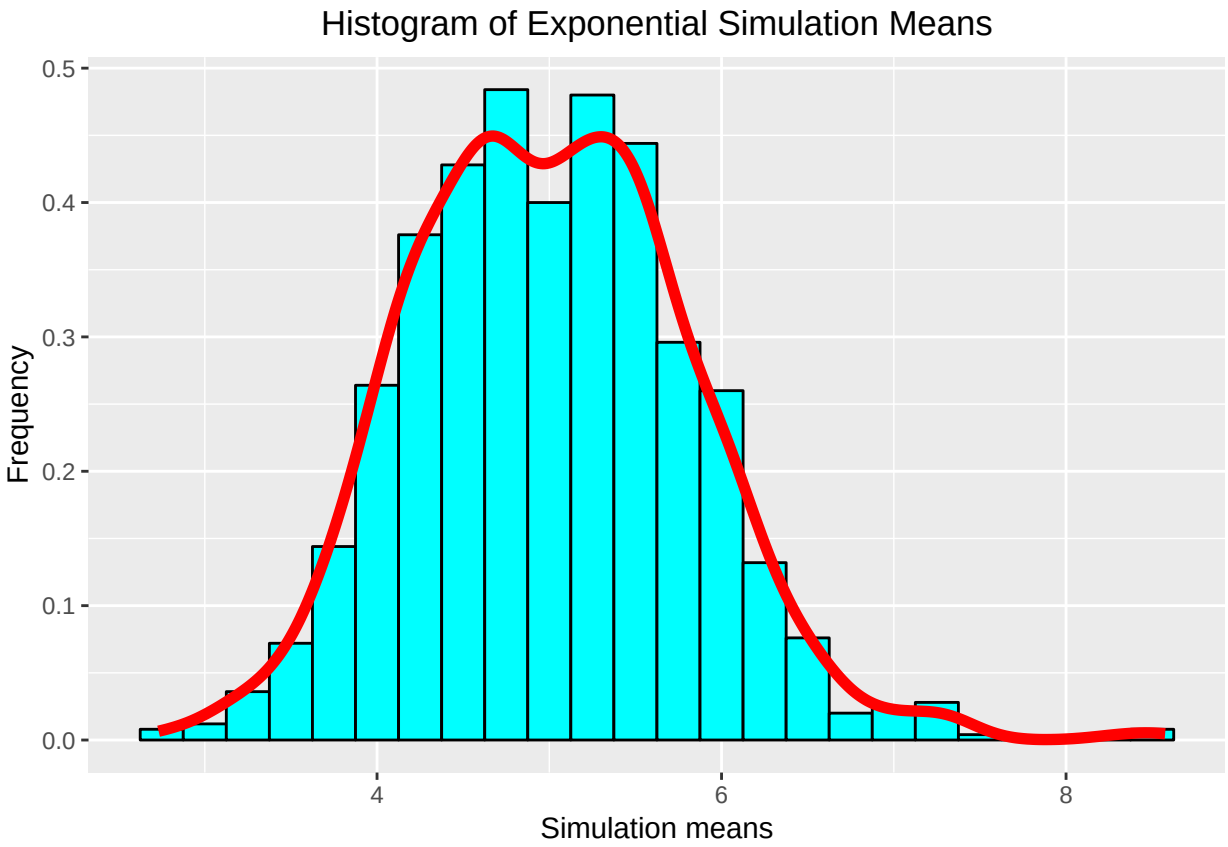
```
##   Sample.Variance Theoretical.Variance Difference
## 1       0.6541354                0.625 -0.0291354
```

We can see that both variance are very close.

Question 3 - Distribution

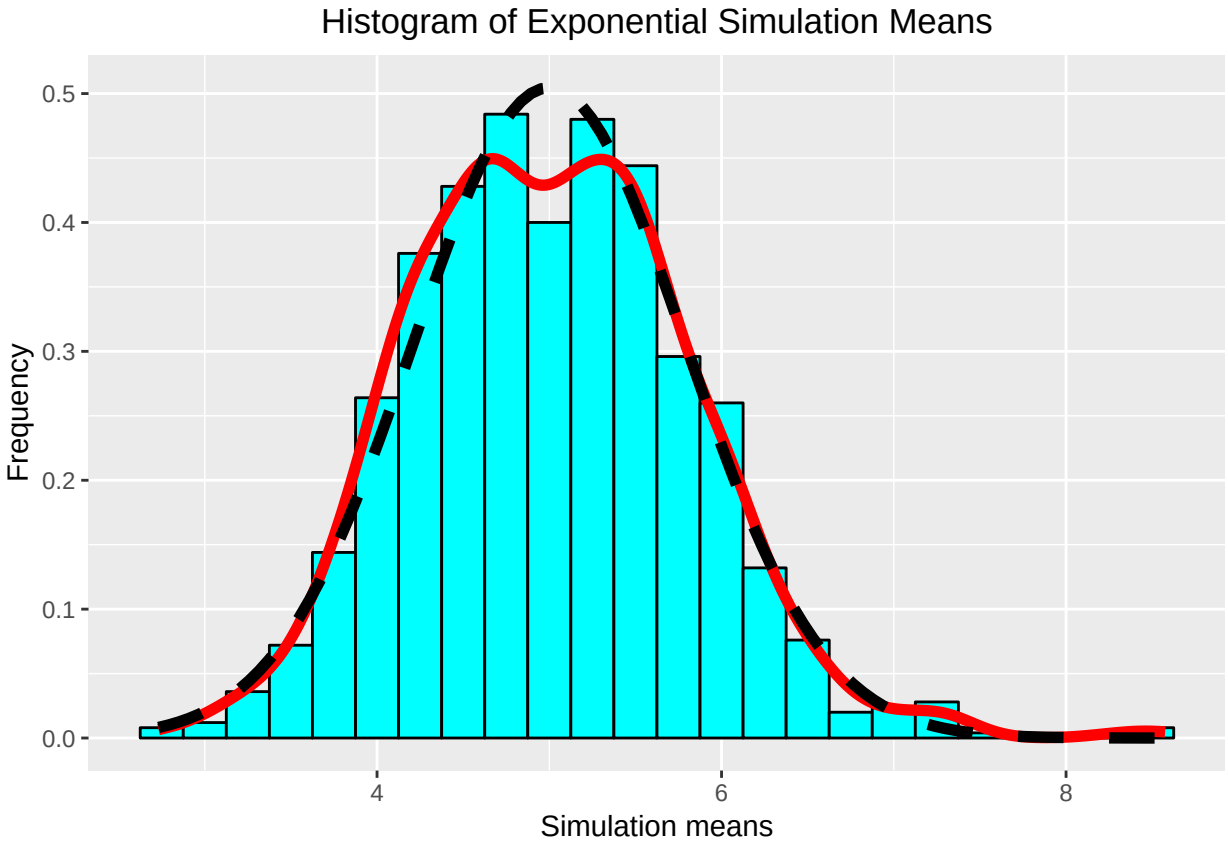
Graph of means' distribution

```
plot_distribution <- plot_means +
        stat_density(geom = "line", color = "red", lwd = 2)
plot_distribution
```



Add normal distribution line for comparison

```
plot_distribution +
        stat_function(fun = dnorm, args = list(mean = theory_mean,
                                               sd = 1 / (lambda * sqrt(n))),
                      color = "black", lwd = 2, lty = 2)
```



Both distribution lines are very close, if we increase number of simulations they should be identical.