Statistical Inference Course Project (Part 1)

Jeffrey M. Hunter

09 May, 2019

Contents

Overview

The Central Limit Theorem states that if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population (generally sample sizes greater than 30), then the distribution of the sample means will be approximately normally distributed about the population mean μ - no matter the shape of the population distribution.

This project explores the Central Limit Theorem using the exponential distribution in R. The theoretical normal distribution will be compared to the distribution of calculated means of samples from the exponential distribution.

Simulations

Perform 1000 simulations, each with 40 samples of an exponential distribution. The 40 samples will be used to calculate the arithmetic mean and variance and then compared to the theoretical estimates.

To make the data reproducible, a seed will be set. Also, set the control parameters $\lambda = 0.2$ (the rate) and $n = 40$ (number of samples). A histogram will be provided to show the averages of the 40 exponentials over 1000 simulations.

```
# load libraries
if (!require(ggplot2)) {
    install.packages("ggplot2", repos = "http://cran.us.r-project.org")
    library(ggplot2)
}
```

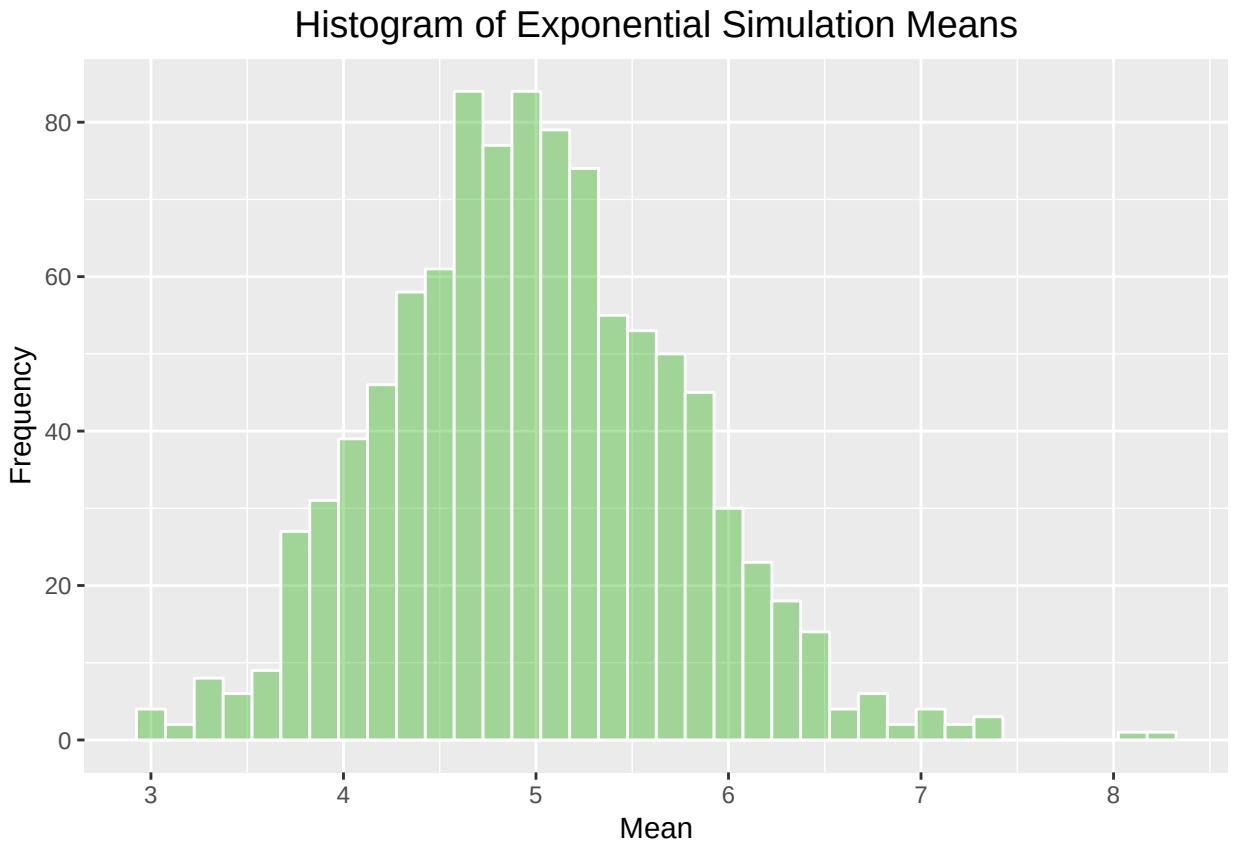
```
## Loading required package: ggplot2
```

```
# set seed for reproducability
set.seed(006766)

# set sampling values:
lambda <- 0.2             # rate parameter
n <- 40                   # number of samples (exponentials) in each simulation
numSimulations <- 1000    # number of simulations

# simulate the population
simMeans <- data.frame(expMean = sapply(1 : numSimulations, function(x) {mean(rexp(n, lambda))}))
```

```r
# plot the distribution
expSimulationMeansChart <- ggplot(simMeans, aes(x = expMean, y = ..count..)) +
                                  geom_histogram(binwidth = 0.15, color = "white", fill = rgb(0.2,0.7,0
                                  xlab("Mean") +
                                  ylab("Frequency") +
                                  theme(plot.title = element_text(size = 14, hjust = 0.5)) +
                                  ggtitle("Histogram of Exponential Simulation Means")
print(expSimulationMeansChart)
```



Sample Mean versus Theoretical Mean

Pending

The mean of the exponential distribution is `1/lambda` and the standard deviation is also `1/lambda`.

Perform 1000 simulations on 40 samples which are exponentially distributed with `lambda = 0.2`, then take the mean of each simulation.

In this project, investigate the exponential distribution in R and compare it with the Central Limit Theorem (CLM). The exponential distribution will be simulated in R with `rexp(n, lambda)` where lambda is the rate parameter.

Lambda will be set to `lambda = 0.2` for all of the simulations. An investigation will be performed on the distribution via simulation of averages of 40 exponentials for one thousand simulations.

sampleMean <- mean(simMeans$mValue) theoMean <- 1/lambda print(sampleMean) print(theoMean)

Sample Variance versus Theoretical Variance Distribution