Statistical Inference Course Project Part 1

Nur Seto Dimas

9 September 2019

Contents

Overview

The first part of the project will try to investigate the exponential distribution and compare it with the Central Limit Theorem. Investigation will be done through calculation with R and comparison between plots.

Analysis

Simulations

The exponential distribution will 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. There will 1000 simulations for distribution of averages of 40 exponentials.

```
set.seed(2019)
n <- 40         # number of exponentials
sims <- 1000    # number of simulations
lambda <- 0.2

# Runs 1000 simulations and calculating mean
simulation_sims <- replicate(sims, rexp(n, lambda))
means_sims <- colMeans(simulation_sims)
```

Sample Mean versus Theoretical Mean

```
sample_mean <- mean(means_sims)
theoretical_mean <- 1 / lambda

cbind(sample_mean, theoretical_mean)
```

```
##      sample_mean theoretical_mean
## [1,]    5.038755                5
```

From calculation, sample mean is **5.038755** and the theoretical mean is **5**. This shows that sample mean is very close to theoretical mean thus confirming the CLT.

Sample Variance versus Theoretical Variance

```
sample_variance <- var(means_sims)
theoretical_variance <- (1/lambda)^2 / n

cbind(sample_variance, theoretical_variance)
```

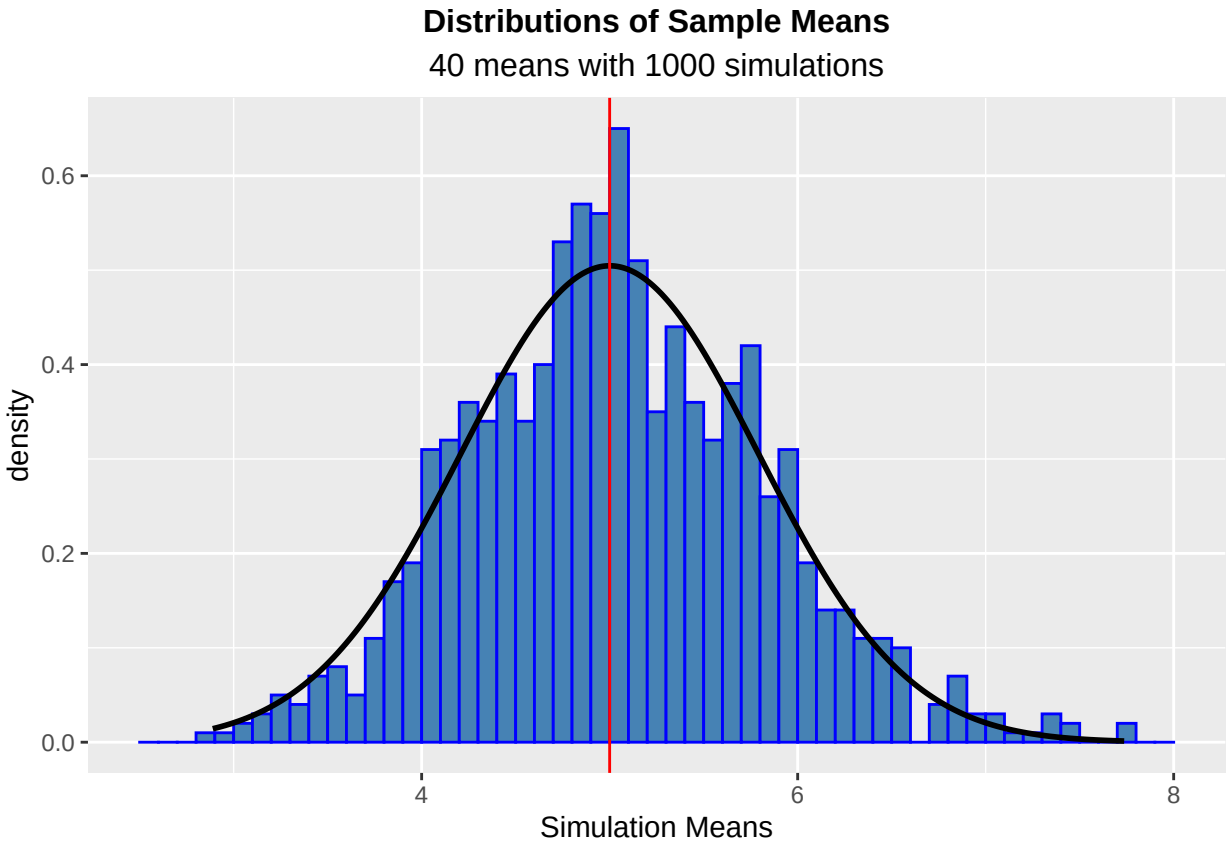
```
##      sample_variance theoretical_variance
## [1,]       0.6253269                0.625
```

The variance of sample is **0.6253269** and the theoretical is **0.625**. Both of them are very close, further confirming the CLT.

Distribution

```
library(ggplot2)
plotdata <- data.frame(means_sims)
x_norm <- seq(min(plotdata$means_sims), max(plotdata$means_sims),
              length.out = 1000)
y_norm <- dnorm(x_norm, mean = 1/lambda, sd = (1/lambda)/sqrt(n))

mean_plot <- ggplot(data = plotdata, aes(x = means_sims))
mean_plot +
        geom_histogram(aes(y = ..density..),
                       breaks = seq(2.5,8, by = 0.1),
                       fill = "steelblue",
                       color = "blue") +
        ggtitle(label = "Distributions of Sample Means",
             subtitle = "40 means with 1000 simulations") +
        theme(plot.title = element_text(size = 12, hjust = 0.5, face = "bold"),
              plot.subtitle = element_text(size = 12, hjust = 0.5)) +
        xlab("Simulation Means") +
        geom_line(aes(x_norm, y_norm), size = 1) +
        geom_vline(aes(xintercept = theoretical_mean), color = "red")
```



The plot shows that the histogram plot of means distribution of samples overlays normal distribution plot (the bell curve plot). From CLT, mean from sample should follows normal distribution if number of n(simulations) increases, it also can be said that if the simulations is done more than 1000 times it will even become closely fits.