Final Project Part I

Table of Contents

Part 1 - Simulation

Author: Davut Emrah Ayan

Abstract: In this project, first, I will investigate the exponential distribution and compare it with Central Limit theorem.

Data generating process

I generated 1000 samples of size 40 from exponential distribution of rate = 0.2

```
rm(list = ls())
set.seed(12345)
simN = 1000
n = 40
lambda = 0.2

X = rexp(n = n*simN, rate = lambda)
sampleX = matrix(X, nrow = simN)
```

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

Distribution of simulated means

I put 1000 simulated means in meanX vector. I find that mean of sample means is 4.971972. On the other hand, theoretical mean is $1/\lambda = 1/0.2 = 5$.

```
meanX = as.data.frame(apply(sampleX, 1, mean))
names(meanX) = "means"
mean(meanX$means)

## [1] 4.971972
```

Below I show the distribution of simulated sample means in the histogram and compare means with the theoretical Mean represented by dashed line.

```
library(ggplot2)
ggplot(meanX, aes(means)) +
    geom_histogram(binwidth = 0.2) +
    geom_vline(xintercept = 1/lambda, linetype = "dashed") +
    labs(title = 'Histogram of Simulated Means',
        subtitle = "Theoretical mean (1/lambda) = 5") +
    theme(plot.title = element_text(hjust = 0.5)) +
    xlab(element_blank())
```

Histogram of Simulated Means
Theoretical mean (1/lambda) = 5
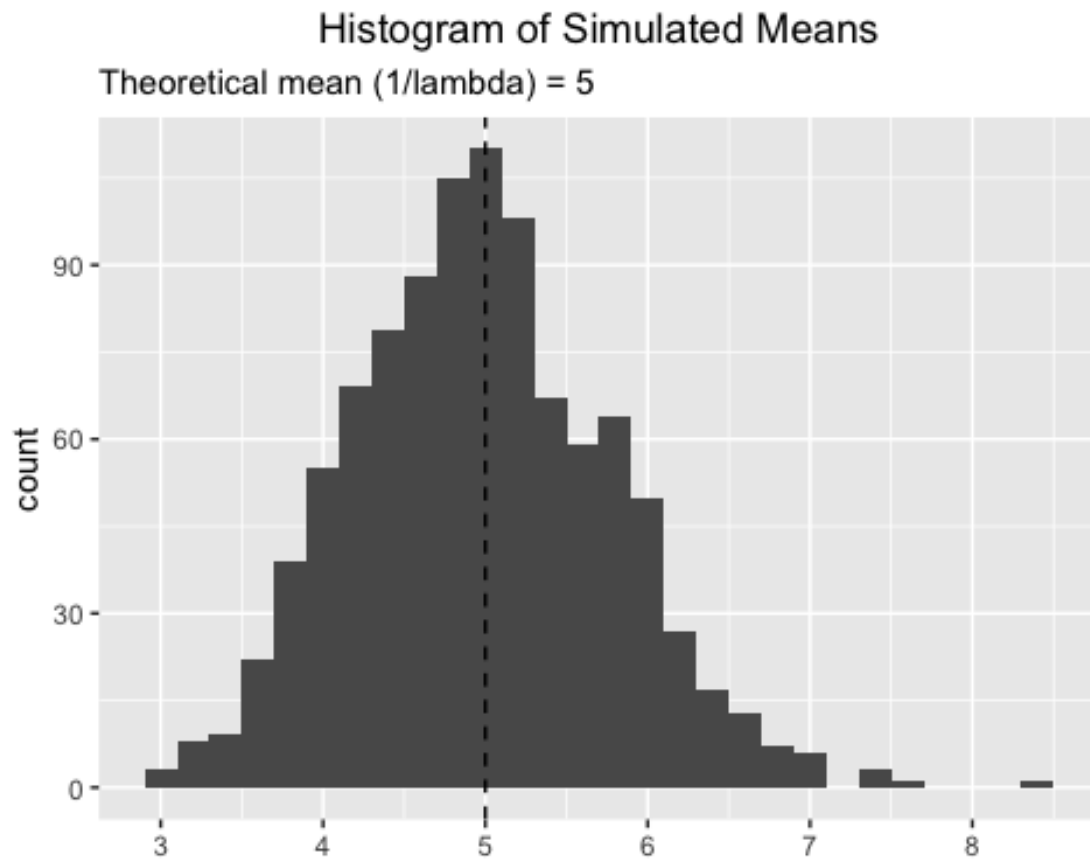
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

Below I calculated variance of 1000 samples and put them in varianceX vector.

```
varianceX = as.data.frame(apply(sampleX, 1, var))
names(varianceX) = "variances"
```

Variance (Standard error) of the simulated sample is $s_{\bar{x}}^2 = 0.6157926$

```
var(meanX$means)
```

```
## [1] 0.6157926
```

and theoretical variance of the sample is $var(\bar{x}) = \sigma_{\bar{x}}^2/n = 0.625$

```
(1/lambda)^2/40

## [1] 0.625
```

Below I show distribution of simulated sample variances and compare with the theoretical Variance. Theoretical variance is $(1/\lambda)^2 = (1/0.2)^2 = 25$. In the histogram, dashed line represents the theoretical variance of 5.

Distribution of simulated variances

```
library(ggplot2)
g2 <- ggplot(varianceX, aes(variances)) +
    geom_histogram(binwidth = 2.5) +
    geom_vline(xintercept = ((1/lambda)/sqrt(40))^2, linetype = "dashed") +
    labs(title = 'Histogram of Simulated Variances',
         subtitle = paste("Theoretical variance is (1/lambda)^2) = 25")) +
    theme(plot.title = element_text(hjust = 0.5)) +
    xlab(element_blank())
g2
```


Histogram of Simulated Variances
Theoretical variance is (1/lambda)^2) = 25
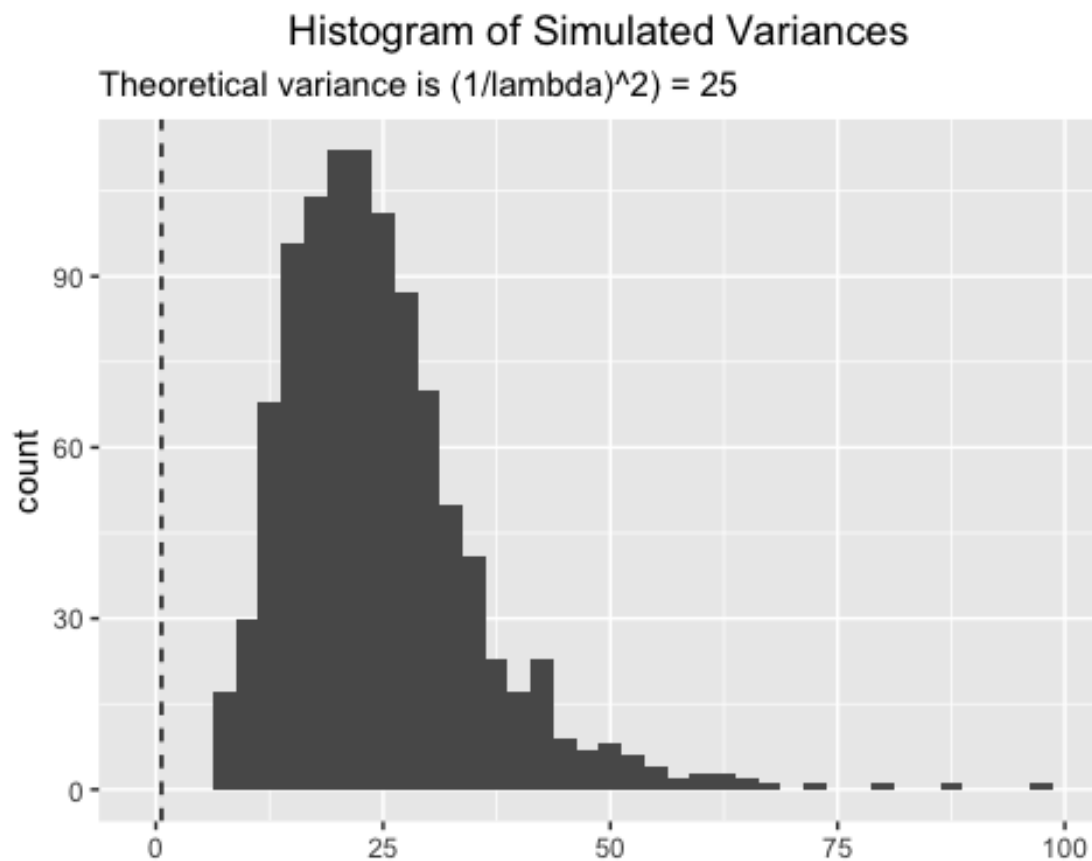
3. Show that the distribution is approximately normal.

Central limit theorem suggests that distribution of samples from an unknown distribution are approximated to normal distribution as sample sizes increase. In this project, I draw 1000 samples of size 40 from exponential distribution where rate = 0.2.

By CLT, I expect that distribution of simulated data approximates the Normal Distribution where $\mu = E(\bar{x})$, and $Var = s^2/n$ in this project.

Both distributions closely overlap.

```
ggplot(data = meanX, aes(x = means)) +
    geom_density() +
    stat_function(fun = dnorm,
                    args = list(mean = (1/lambda), sd = (1/lambda)/sqrt(40)),
                    colour = "red", size = 1) +
    geom_density(colour="blue", size=1, alpha = .2) +
    geom_vline(xintercept = c(1/lambda, mean(meanX$means)), linetype =
"dashed", color = c("red", "blue")) +
    labs(title = 'Density Plots of Simulated Means and Normal distribution',
            subtitle = "Theoretical mean (1/lambda) = 5") +
    theme(plot.title = element_text(hjust = 0.5)) +
    xlab(element_blank())
```


Density Plots of Simulated Means and Normal distribution
Theoretical mean (1/lambda) = 5