Part 1: Simulation Exercise Instructions

Yan He

Overview

This project is to investigate the exponential distribution in R and compare it with the Central Limit Theorem.

Semulations

In this part we will set random seed and prepared parameters. After that we simulate exponential random variables 1000 times. Then we get 1000 results for this simulation.

```
set.seed(2018)

lambda <- 0.2
n <- 40
num <- 1000

simulation_result <- matrix(rexp(n = n*num, rate = lambda), nrow = num)
```

Mean

We calculate the mean for each row:

```
simulation_means <- rowMeans(simulation_result)
```

Now we can draw a histogram plot for our semulation means and compare it to the theoretical mean of the distribution(1/lambda) which is drawn by a red line:

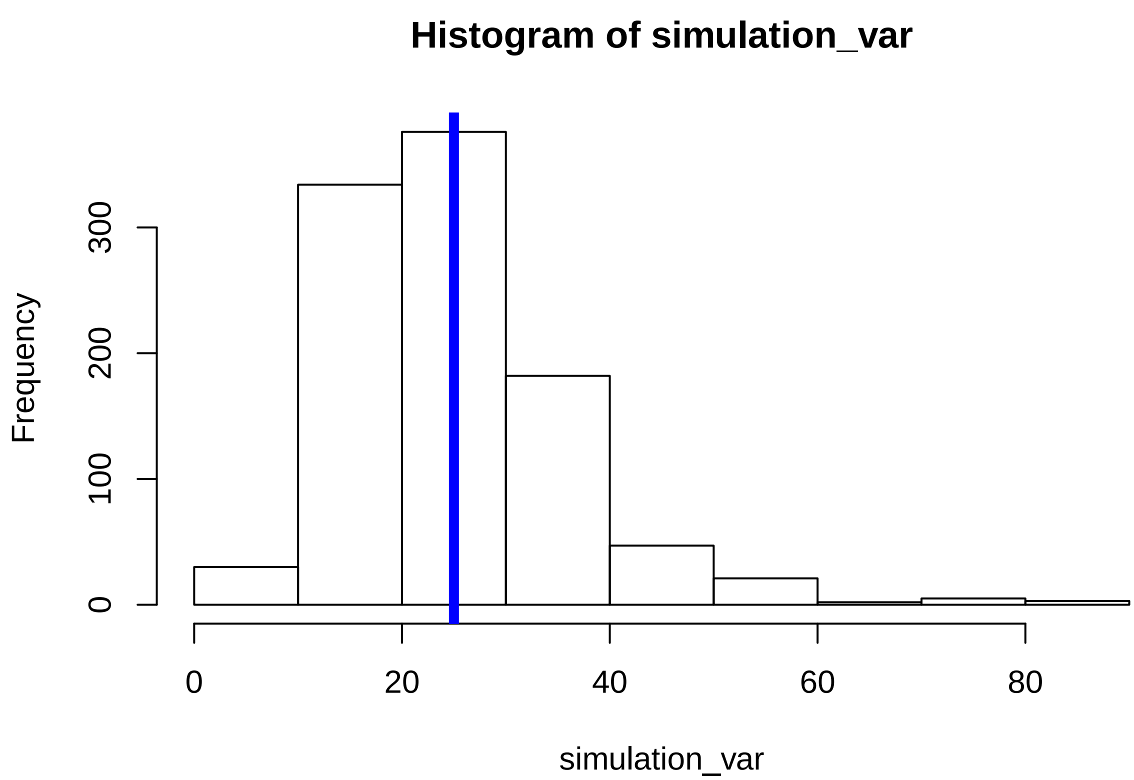
```
hist(simulation_means)
abline(v = (1/lambda), col = 'red', lwd = 5)
```

From the plot we can see that the center of simulation means is very close to the theoretical means.

Variance

Then we calculate the variance and draw histogram to compare the simulation and true variance:

```
simulation_var <- c()
for(i in 1:nrow(simulation_result)){
        s_var <- var(simulation_result[i,])
        simulation_var <- append(simulation_var, s_var)
}
hist(simulation_var)
abline(v = (1/lambda)^2, col = 'blue', lwd = 5)
```



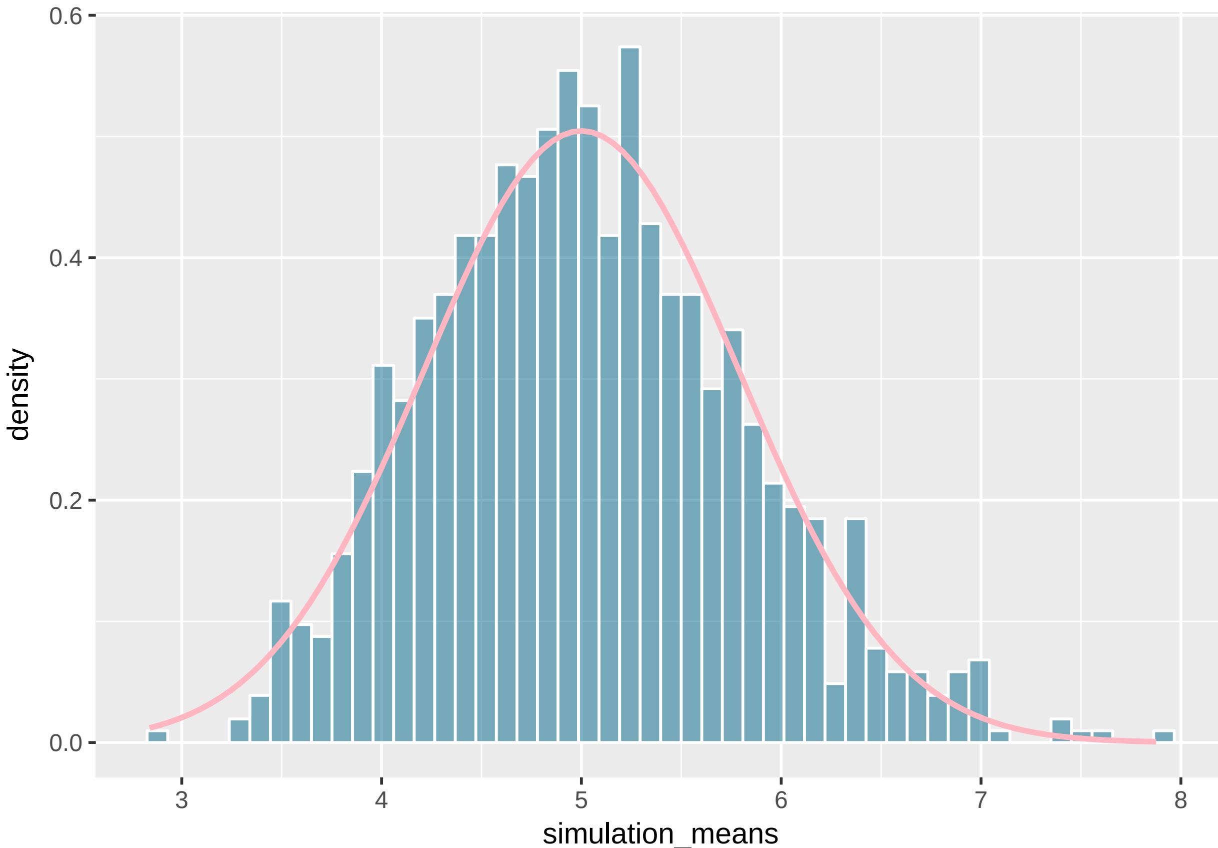
It's obvious to see that the true variance is located in the center of our simulation vars.

Normality

Now we try to explore whether the mean of simulation is close to normal distribution. First, we draw a histogram plot for simulation means.Then we add a normal distribution which mean is 1/lambda and standaed deviation is 1/lambda/sqrt(n).

```
library(ggplot2)
mean_data <- as.data.frame(simulation_means)

ggplot(data = mean_data, aes(x =simulation_means)) +
        geom_histogram(aes(y = ..density..),
                       bins  = 50,
                       color = 'white',
                       alpha = 0.5,
                       fill = 'deepskyblue4') +
        stat_function(fun = dnorm,
                      args = list(mean = 1/lambda,
                                  sd = (1/lambda)/sqrt(40)),
                      color = 'lightpink',
                      size = 1)
```



It's easy to see the distribution of simulate means is very close to theoretical normal distribution from CLT.But we cannot get a precise conclusion. We choose Kolmogorv-Smirnov test for testing the distribution of simulation means and normal distribution:

```
## standard normalize
normal_sim_means <- (simulation_means - mean(simulation_means))/sd(simulation_means)

## KS test
ks.test(normal_sim_means, 'pnorm')
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  normal_sim_means
## D = 0.034677, p-value = 0.1804
## alternative hypothesis: two-sided
```

p value is bigger than 0.05, so we cannot reject the H0 hypothesis: simulation means is normaly. Now we can say this simulate distribution is approximately normal.