Investigation of exponential distribution and Central Limit Theorem in R

Erwin Lo

Overview

This article investigates the exponential distribution in R and compare it with the Central Limit Theorem(CLT). The exponential distribution will be simulated using lambda of 0.2 with a thousand simulations. The distribution of averages of 40 exponentials will then be investigated.

Simulations

A thousand simulation of 40 exponentials are simulated below:

```
set.seed(15)
lambda <- 0.2
num <- 40
sim <- 1000

means <- NULL
for (i in 1:sim)
     means <- c(means, mean(rexp(num, lambda)))
```

Sample Mean versus Theoretical Mean

We can find sample mean by taking the average of our simulated means.

```
(samp_mean <- mean(means))
```

```
## [1] 4.980535
```

The theoretical mean of exponential distribution is 1/lambda.

```
(theo_mean <- 1 / lambda)
```

```
## [1] 5
```

We can see that sample mean is approximately the same as theoretical mean.

Sample Variance versus Theoretical Variance

Calculate the sample variance

```
(samp_var <- var(means))
```

```
## [1] 0.6186672
```

compared to theoretical variance of exponential distribution.

```
(theo_var <- 1 / (lambda^2 * num))
```

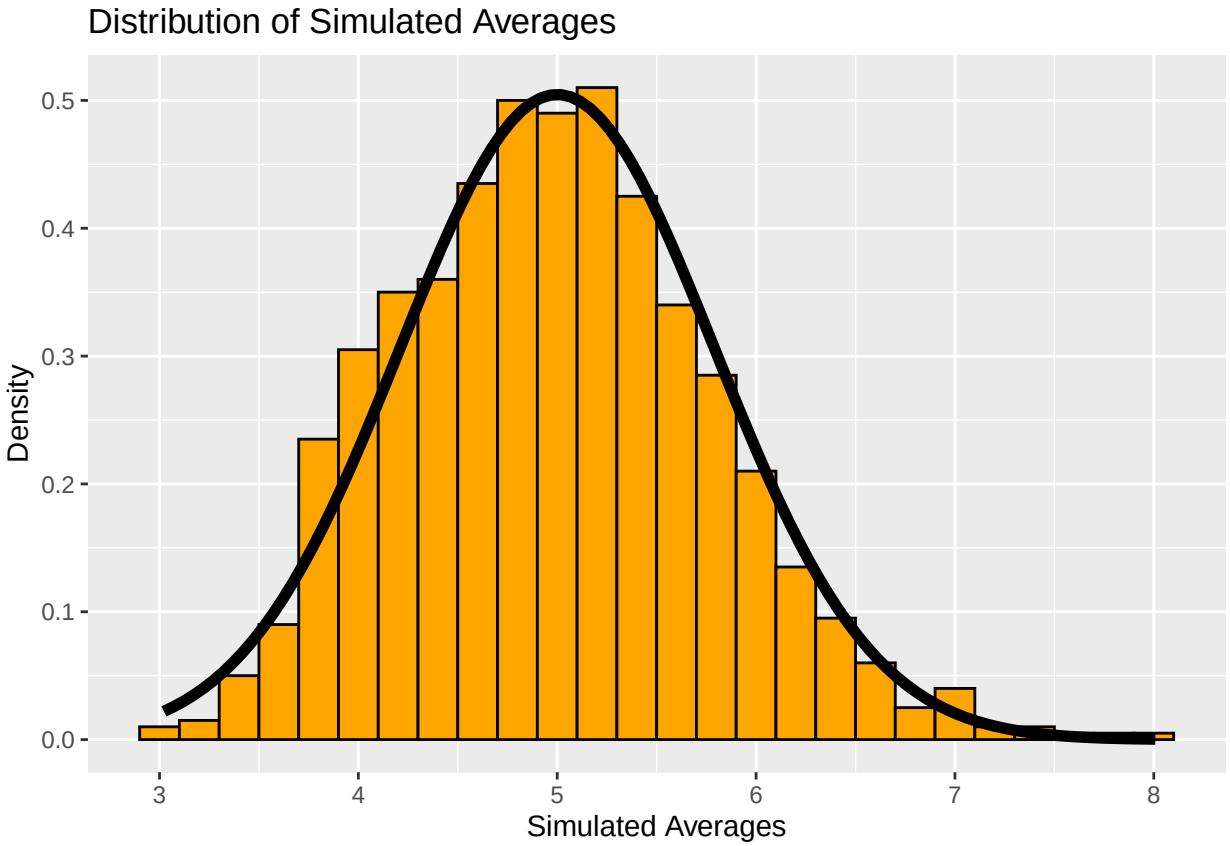
```
## [1] 0.625
```

Sample variance is also approximately the same as theoretical variance.

Distribution

Below is the histogram plot of simulated averages, and normal distribution line in black overlayed on top of the chart. Theoretical mean and standard deviation of exponential distribution were used as the parameters for the normal distribution line.

```
library(ggplot2)
ggplot(data.frame(means), aes(x = means)) +
     geom_histogram(binwidth = lambda, fill = "orange", color = "black",
                    aes(y = ..density..)) +
     stat_function(fun = dnorm, args = list(mean = theo_mean, sd = sqrt(theo_var)),
                   size = 2, color = "black") +
     labs(title = "Distribution of Simulated Averages",
          x = "Simulated Averages",
          y = "Density")
```



From the chart, we can tell that the histogram of simulated averages is approximately normal.