Investigate the Exponential Distribution and Compare it with the Central Limit Theorem

Simin Talebpour

Overview

In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The mean of the exponential distribution is 1/lambda and the standard deviation is also 1/lambda. I will investigate the distribution of averages of 40 exponentials in 1000 simulations, and set lambda = 0.2 for all of the simulations.

```
# Load required packages
library(ggplot2)
```

Simulations

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter.

```
#Set the seed
set.seed(77777)

#Set Parameters
mns <- NULL
lambda <- .2
n <- 40
simulation <- 1000
#Simulate the exponential
for (i in 1 : simulation) mns = c(mns, mean(rexp(n , lambda)))
```

Comparison of mean

```
#Claculate the Simulation mean
simulation_mean <- mean(mns)
simulation_mean
```

```
## [1] 4.999655
```

```
#Calculate the theoretical mean
dist_mean <- 1/lambda
dist_mean
```

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

The calculated mean of my simulation is 4.9996, and the mean of the exponential distribution for lambda = 0.2 is 5. The sample mean and the theoretical mean values are close.

Comparison of Variance

```
#Claculate the simulation variance
simulated_var <- var(mns)
simulated_var
```

```
## [1] 0.607749
```

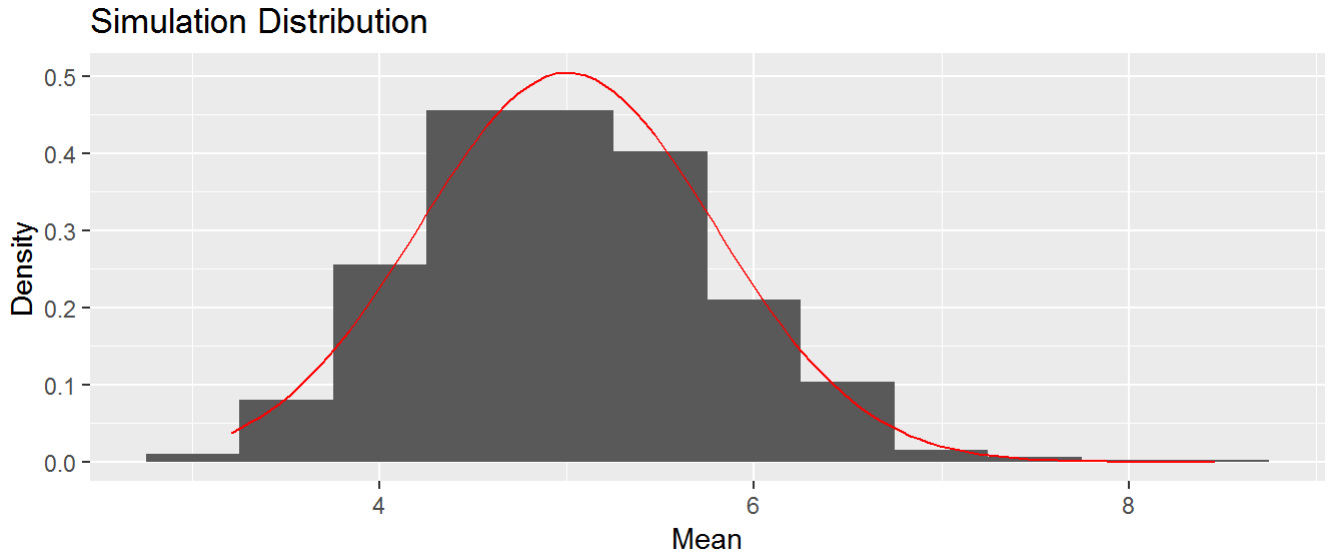
```
#Calculate the theoretical variance
dist_var <- ((1/lambda)/sqrt(n))^2
dist_var
```

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

The calculated variance of my simulation is 0.608, and the variance of the exponential distribution for lambda = 0.2 is 0.625. The sample variance and the theoretical variance values are close.

Distribution of Samples

```
#Display the histogram for the averag of 40 exponentials in 1000 simulations
df <- as.data.frame(mns)
ggplot(aes(x = mns), data = df)+
  geom_histogram(binwidth = .5 , aes( y = ..density..))+
  stat_function(fun = dnorm, colour = 'red', args = list(mean = 5, sd = sqrt(dist_var)))+
  xlab('Mean')+
  ylab('Density')+
  ggtitle('Simulation Distribution')
```



The central Limit Theorem says that the averages of iid variables become standard normal as the sample size increases. Here the histogram shows that for a sample size of 40 exponentials, the distribution of averages is approximately normal, and is centered around 5 which is the mean of my simulation distribution.