Simulation the exponential distribution

vyndk

September 3, 2018

Overview

In this project we shall investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can 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. In this investigation, lambda is always 0.2. The sample mean and and the sample variance will be compared to the theoretical mean and variance of the distribution respectively

Simulations

The following code, we perform 1000 simulations of the exponential distribution. Lambda is set to 0.2 and sample size is 40

```
library(tidyverse)
set.seed(20183)
lambda <- 0.2
sample.size <- 40;
num.of.simulaton <- 1000;
# create a matrix containing simulations
data <- matrix(data=rexp(sample.size*num.of.simulaton,rate=lambda), nrow = num.of.simulaton, ncol = samp
# Calculate the mean value of each simulation
means <- apply(data, 1, mean)
head(means)
```

```
## [1] 6.173655 5.851536 5.339017 4.881206 4.106226 5.022794
```

Sample Mean versus Theoretical Mean

Now, we shall compare the sample mean of simulations with the the theoretical mean. According to The theory, the mean is 1 / lambda

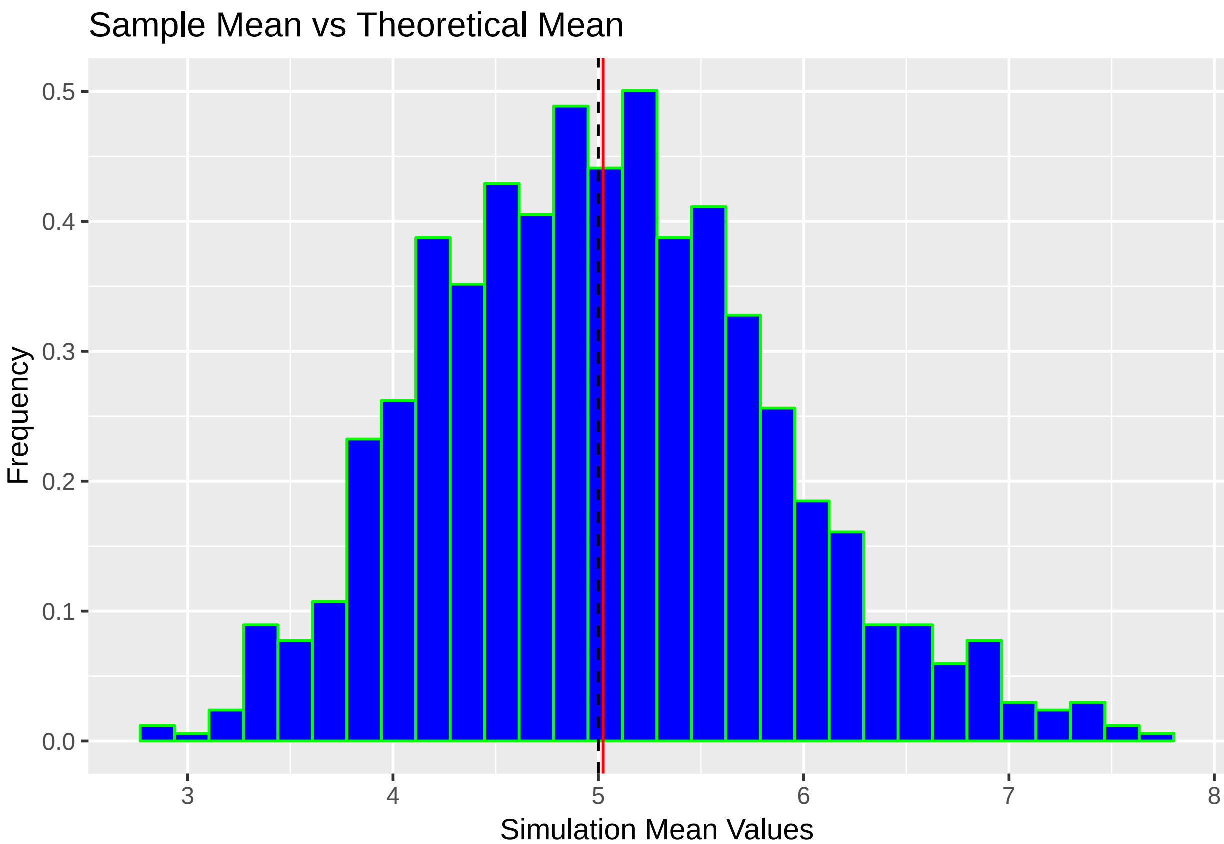
```
theoritcal.mean = 1 / lambda
sample.mean <- mean(means)
sample.mean
```

```
## [1] 5.023344
```

We can see that the sample mean is 5.023, very close to theoretical mean Now, we plot the veritlcal line representing the sample mean and the dashed black vertical line representing the theoretical mean on the histogram of sample mean values

```
ggplot(data.frame(means),aes(x=means)) +
    geom_histogram(fill="blue",colour="green",aes(y = ..density..)) +
    labs(x="Simulation Mean Values",y="Frequency", title="Sample Mean vs Theoretical Mean") +
    geom_vline(xintercept=sample.mean, color="red") +
    geom_vline(xintercept=theoritcal.mean, color="black", linetype="dashed")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



We can see that the read line is very close to the dashed black line. This illustrates the the sample mean is very close to the theoretical mean

Sample Variance versus Theoretical Variance

In this section, we shall compare the sample variance and standard deviation to the theoritical values of them

```
theory.sd <- (1/lambda) / sqrt(sample.size)
sample.sd <- sd(means)
theory.var <- ((1/lambda)^2) / sample.size
sample.var <- sample.sd ^ 2
result <- data_frame("Type" = c("Theory", "Simulation"), "Standard.Deviation" = c(theory.sd, sample.sd)
result
```

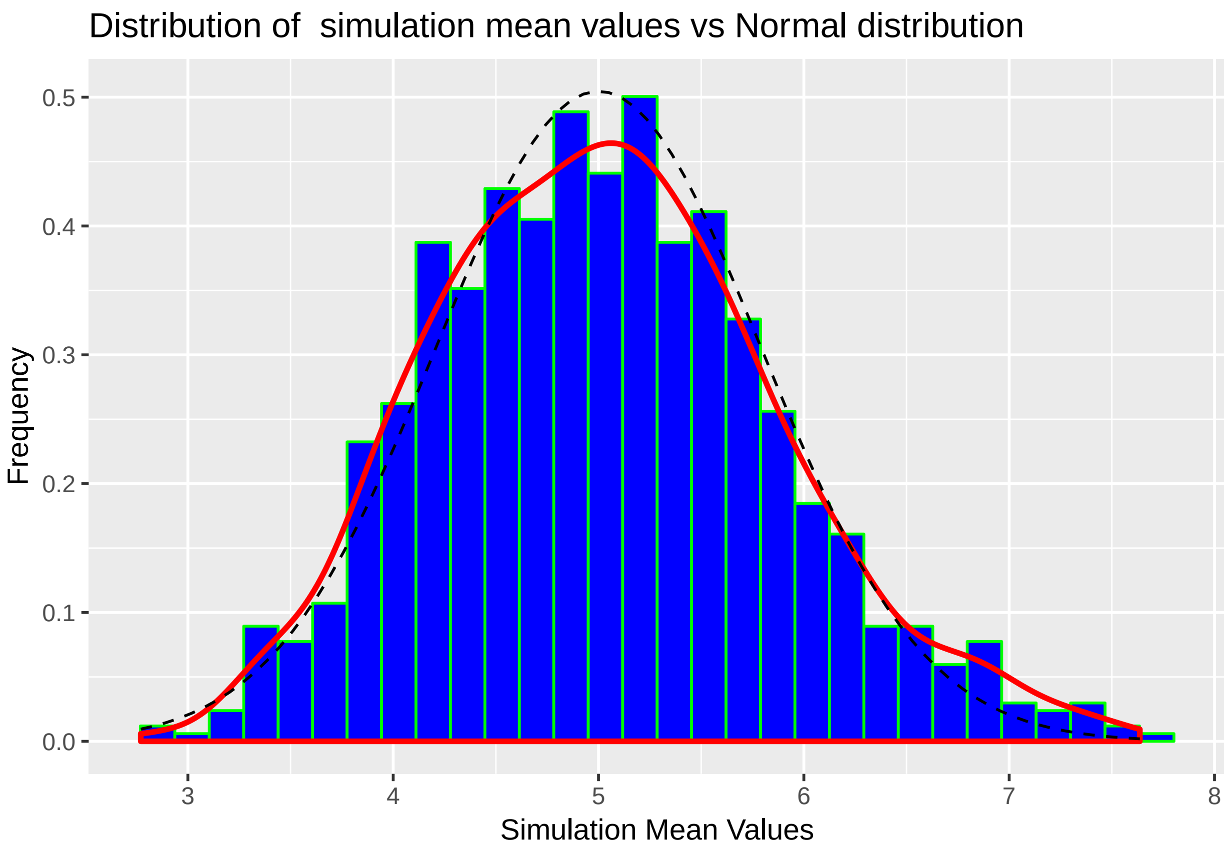
```
## # A tibble: 2 x 3
##   Type       Standard.Deviation Variance
##   <chr>                   <dbl>    <dbl>
## 1 Theory                  0.791    0.625
## 2 Simulation              0.835    0.697
```

The result show us that the variances are very close.

Distribution

Now we plot the distribution of the sample mean to see if it is close to nor mal distribution as the Central Limit Theorem says

```
# create normal distribution line
x <- seq(min(means), max(means), length=2*sample.size)
y <- dnorm(x, mean=1/lambda, sd=sqrt(((1/lambda)/sqrt(sample.size))^2))
nor <- data_frame(x, y)
ggplot(data.frame(means),aes(x=means)) +
    geom_histogram(fill="blue",colour="green",aes(y = ..density..)) +
    geom_density(colour="red",size=1)+
    geom_line(aes(x=x, y=y), data = nor, linetype = "dashed") +
    labs(x="Simulation Mean Values",y="Frequency", title="Distribution of  simulation mean values vs No
```



The distribution of the simulation is present by the read line, and the normal distribution is the dashed line. We can see that the distribution of the simukation mean is approximately the normal distribution as the theory says