Investigation of the exponential distribution in R

Jeevithiesh Duggani

06/05/2020

Contents

Abstract

This project investigates the `exponential distribution` in R and compare it with the `Central Limit Theorem`. It investigates the distribution of **averages** of **forty exponentials**. **Thousand** simulations are performed.

The `aims` of the project are listed below :

1. Show the `sample mean` and compare it to the `theoretical mean` of the distribution.
2. Show how variable the sample is (via `variance`) and compare it to the `theoretical variance` of the distribution.
3. Show that the distribution is approximately normal, i.e. the difference between the distribution of a `large collection of random exponentials` and the distribution of a `large collection of averages of 40 exponentials` is **negligible**.

Import the necessary packages

The following packages are to be imported.

- ggplot2

```
library(ggplot2)
```

Initialization of important Variables

The `seed` is initialized with `100`. The `lambda (rate parameter)` is set as `0.2`. The `sample size` is set to `40`. The `number of simulations` is `1000`.

```
    set.seed(100)

    lambda      <- 0.2
    nsim        <- 1000
    sampleSize  <- 40
```

Run the simulations

The simulations are run and stored in the matrix `sims`.

```
    sims <- matrix(rexp(nsim*sampleSize, lambda), nsim, sampleSize)
```

Comparision between Sample Mean and Theoritical Mean

The value of the Theoritical mean is `1/lambda = 5`.

The Sample mean is calculated below.

```
    tMean <- 1/lambda # Theoritical Mean
    sMean <- mean(rowMeans(sims)) #Sample Mean
```

The value of the Sample mean is 4.999702.

The difference between the two means is -0.000298, which is -0.00006% of the Theoritical mean.

Comparision between Sample Variance and Theoritical Variance

The value of the Theoritical variance is `(1/lambda^2)*(1/sampleSize) = 0.625`.

The Sample mean is calculated below.

```
    tVar <- (1/lambda^2)*(1/sampleSize) # Theoritical Variance
    sVar <- var(rowMeans(sims)) #Sample Variance
```

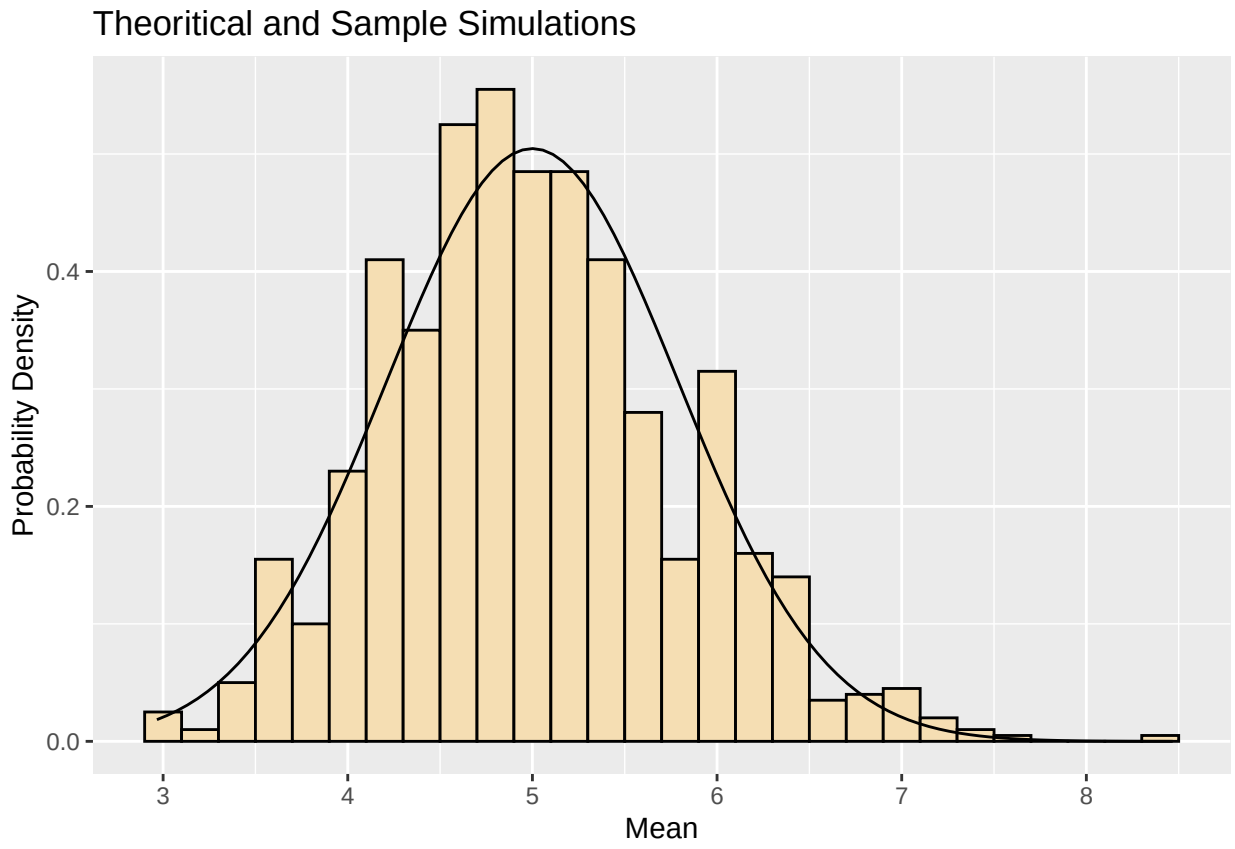
The value of the Sample variance is 0.63353.

The difference between the two variances is 0.00853, which is 0.013648% of the Theoritical variance.

Comparisions of the Distributions

Plot the Theoritical as well as the sample Distributions.

```
graph <- ggplot(data.frame(mean = rowMeans(sims)), aes(mean))
graph <- graph + geom_histogram(aes(y=..density..), binwidth = 0.2,
                                fill = "wheat", color = "black")
graph <- graph + stat_function(fun = dnorm, args = list(mean = tMean, sd = sqrt(tVar)))
graph <- graph + labs(title = "Theoritical and Sample Simulations",
                      x = "Mean", y = "Probability Density")
graph
```



Both the distributions follow the same pattern.

Results

1. Theoritical Mean is 5, Sample Mean is 4.999702. The difference between the two means is -0.000298, which is -0.00006% of the Theoritical Mean.
2. Theoritical Variance is 0.625, Sample Variance is 0.63353. The difference between the two variances is 0.00853, which is 0.013648% of the Theoritical Variance.
3. Both the distributions follow the same pattern.

###End