Simulation exercise using the exponential function

Rohil Narula

20 July 2018

Simulation exercise

Overview

This simulation exercise will draw 40 variables from a random exponential distributtion of data with $\lambda = 0.2$. A mean distribution of 1000 simulations will be sampled and the mean and variance of this distribution, calculated. Finally, a simple histogram plot will be used to visualize the distribution, which should be approximately normal from the central limit theorem.

Simulations

In this section, the initial variables are set, then each sample of the simulations will be stored as a row in a matrix. As we have 1000 simulations, the matrix will by default contain 1000 rows (as well as 40 columns).

Variable initialization

Loading initial packages.

```
packages <- c('ggplot2', 'dplyr')
sapply(packages, require, character.only = TRUE, quietly = TRUE)
```

```
## ggplot2   dplyr
##    TRUE    TRUE
```

The initial variables are set to the following values.

```
set.seed(19)
n <- 40
lambda <- 0.2
draws <- 1000
theoreticalMean <- 1 / lambda
theoreticalSD <- 1 / lambda
```

Simulated matrix generation

The matrix of the random draws is calculated using the standard `R` function `rexp()`.

```
simulations <- matrix(rexp(n * draws, lambda), nrow = draws)
str(simulations)
```

```
##  num [1:1000, 1:40] 12.07 10.87 5.77 7.94 3.37 ...
```

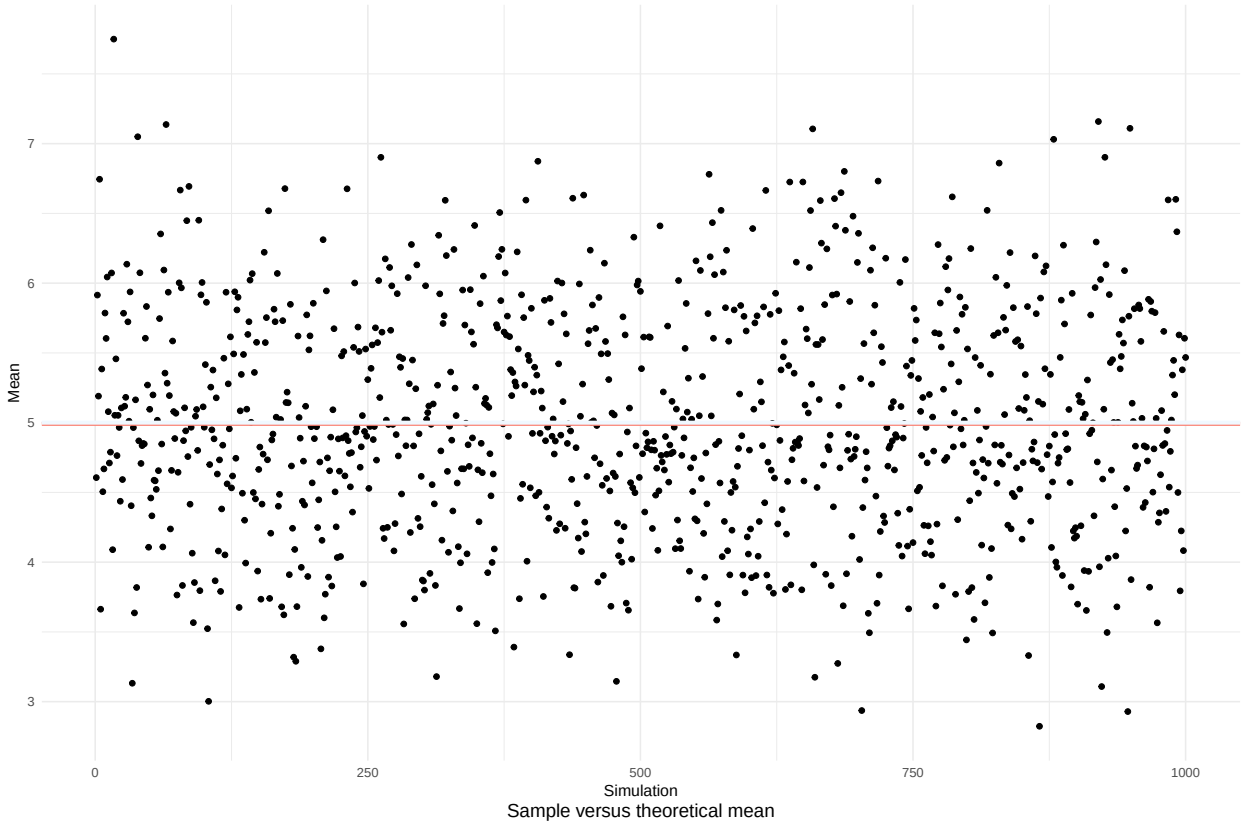
Sample versus theoretical mean

Calculating the means of the simulations.

```
xBar <- rowMeans(simulations)
c(mean(xBar), theoreticalMean)
```

```
## [1] 4.991311 5.000000
```

Plotting the means.



Here, the salmon line represents the actual mean of the distribution sample, and the light blue line represents the theoretical mean. As we can see in the data plot, the theoretical mean of the data and the estimated mean from the sample distribution almost overlap considering 1000 simulations.

Sample versus theoretical variance

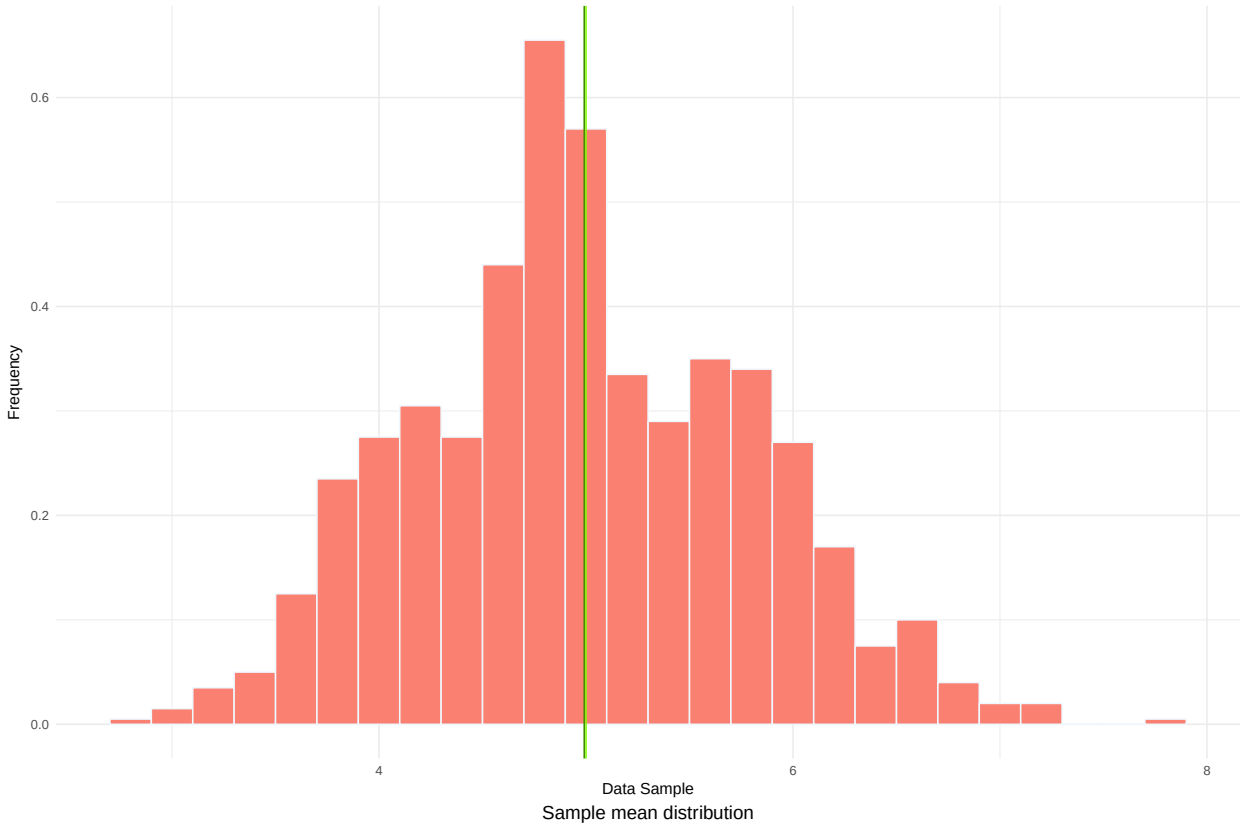
Calculating the variance of the mean distribution.

```
sigma <- sd(xBar) * sqrt(n - 1)
variance <- sigma^2
c(variance, theoreticalSD^2)
```

```
## [1] 25.09842 25.00000
```

Distribution

The following is the plot of the data from the mean distribution of the samples.



Here, the theoretical mean is shown in light green, while the sample mean is shown in deep green. The overlap suggests that the sample mean is a good estimate of the theoretical mean.

Looking at the bell curve, the histogram looks approximately normal, since the number of simulations are sufficiently high for the data to have normal properties. This is an example of the working of the central limit theorem.