Statistical Inference Course Project - simulation exercise

Zanin Pavel

February 27, 2016

Link to project on GitHUB

Overview

In this project we are going to 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.
We will set lambda = 0.2 for all of the simulations. Also we will investigate the distribution of averages of 40 exponentials. We will do a thousand simulations.

1. Simulations

```r
# loading neccesary libraries
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.3
```

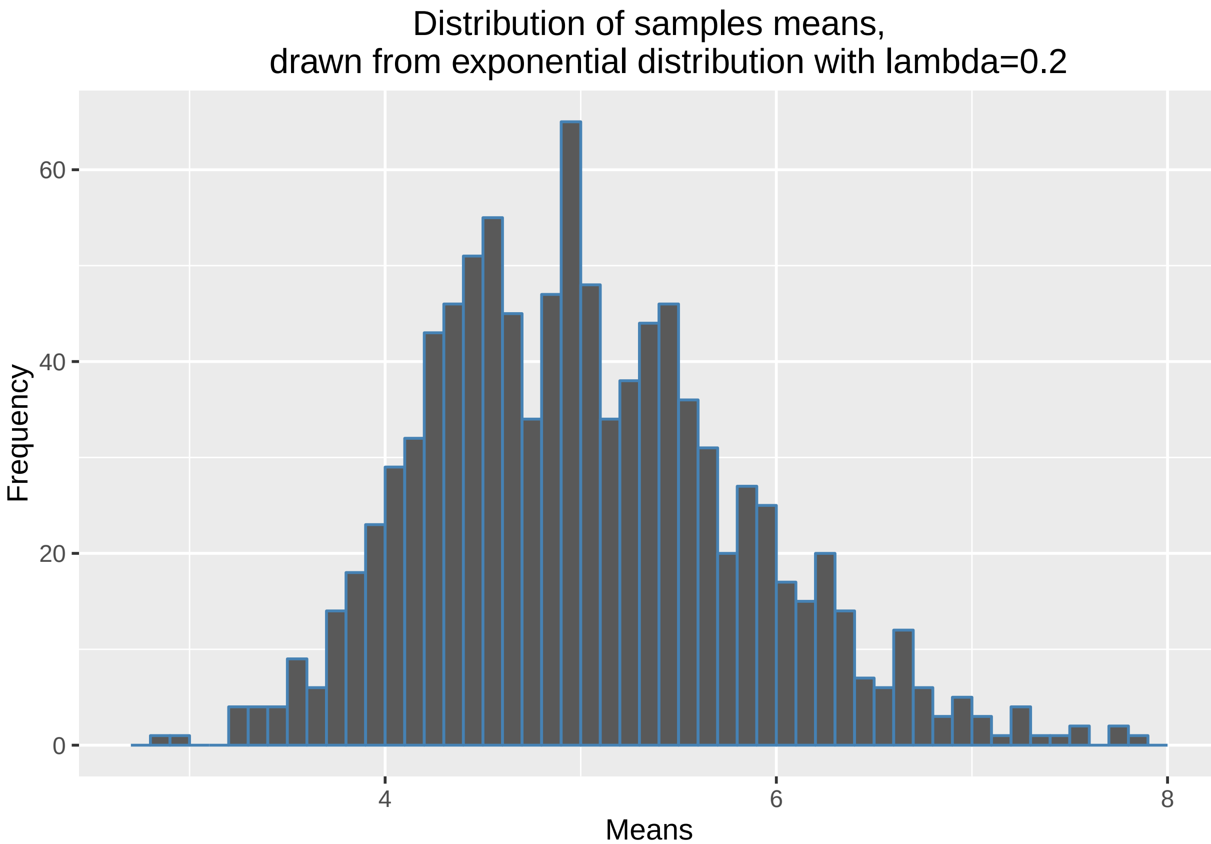
```r
# set constants
lambda <- 0.2 # rate parameter for exponential distribution
n <- 40 # number of exponetials
numberOfSimulations <- 1000 # number of simulations

# set the seed to create reproducability
set.seed(123456)

# creating matrix with observations
exponentialDistributions <- matrix(data=rexp(n * numberOfSimulations,
                                              lambda),
                                   nrow=numberOfSimulations)

# creating data frame with means of rows in the matrix
exponentialDistributionMeans <- data.frame(means=apply(exponentialDistributions, 1, mean))

# plotting the means
ggplot(data = exponentialDistributionMeans, aes(x = means)) +
              geom_histogram(binwidth=0.1, color = "steelblue") +
              labs(title = "Distribution of samples means,
    drawn from exponential distribution with lambda=0.2") +
              labs(x = "Means") +
              labs(y = "Frequency")
```



2. Sample Mean versus Theoretical Mean

The theoretical mean μ of a exponential distribution of rate λ is

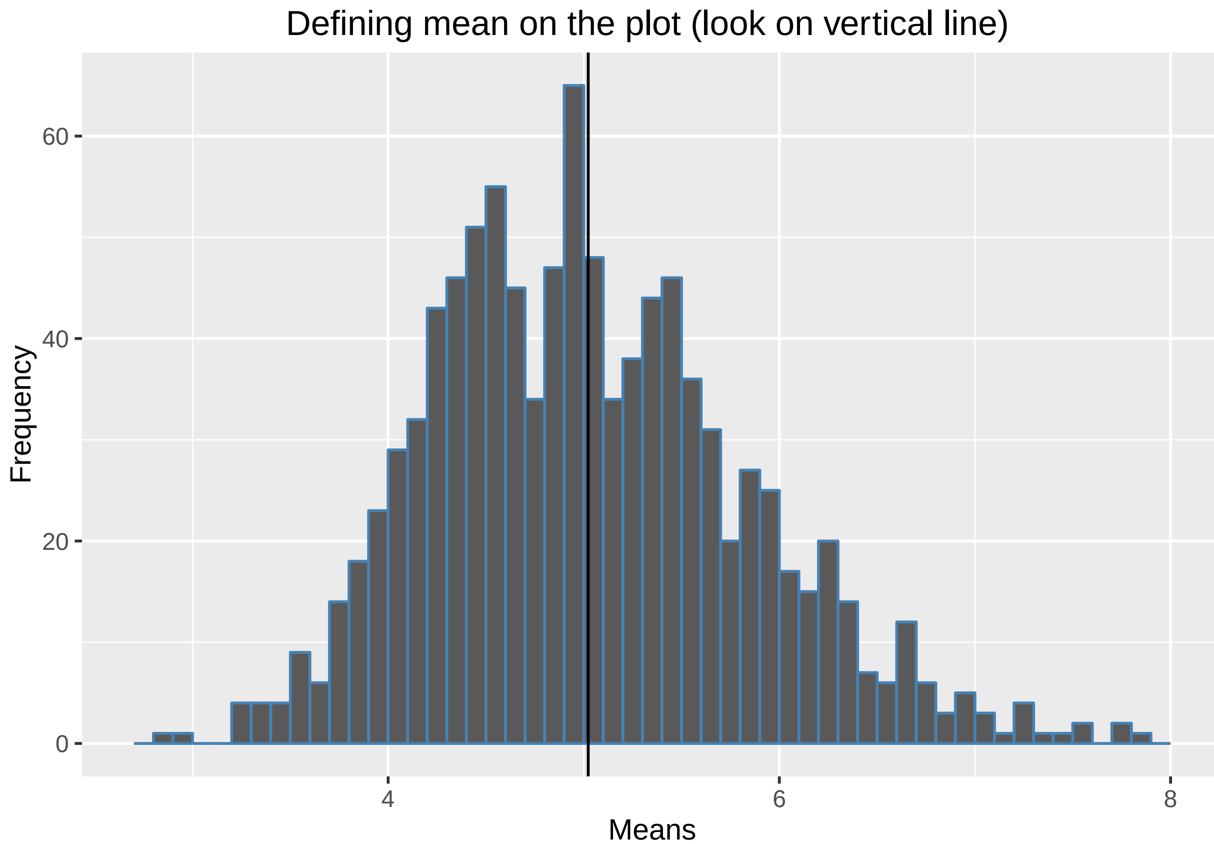
$\mu = \frac{1}{\lambda}$, then μ for λ=.2 equal:

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

Lets define sample mean and compare it with theoretical:

```
# calculating sample mean
SampleMean <- mean(exponentialDistributionMeans$means)
SampleMean
```

```
## [1] 5.022915
```



Infer: So, as you can see the theoretical mean (5) and sample mean (5.0229151) are very close.

3. Sample Variance versus Theoretical Variance

The theoretical standard deviation σ of a exponential distribution of rate λ is: $\sigma = \frac{1/\lambda}{\sqrt{n}}$, then σ for λ=.2 and n=40 equal:

```
## [1] 0.7905694
```

The theoretical variance $Var = \sigma^2$. Then its equal:

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

Lets define the sample variance (Var_x) and the sample standard deviation σ_x.

```
sd_x <- sd(exponentialDistributionMeans$means)
sd_x
```

```
## [1] 0.8097816
```

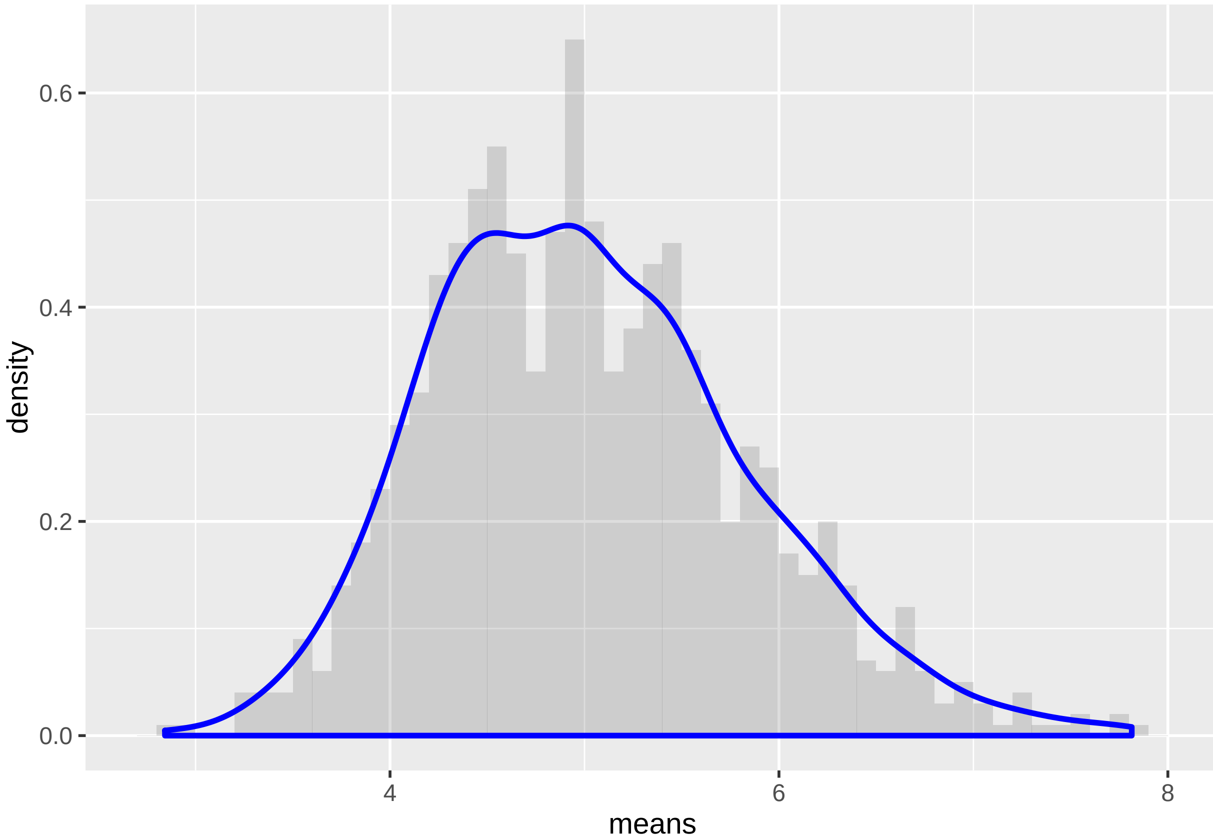
```
Var_x <- var(exponentialDistributionMeans$means)
Var_x
```

```
## [1] 0.6557463
```

Infer: As you can see standard deviation and variance in theory (0.7905694 and 0.625) and in the sample (0.8097816 and 0.6557463) are pretty close.

4. Distribution

Due to the Central Limit Theorem, the averages of samples should follow a normal distribution. Lets show it.



Infer: As shown on the plot, the calculated distribution of means of random sampled exponantial distributions pretty close, in accordance to the Central Limit Theorem, to the normal distribution. Its what we wanted to prove.