Statistical Inference Course Project Part 1

Pradeep K. Pant, ppant@cpan.org (mailto:ppant@cpan.org)

June 12, 2016

Overview

In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem (CLT). 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'll set lambda = 0.2 for all of the simulations. Investigation is done for the distribution of averages of 40 exponentials. I will have to perform a thousand of simulations.

Simulations

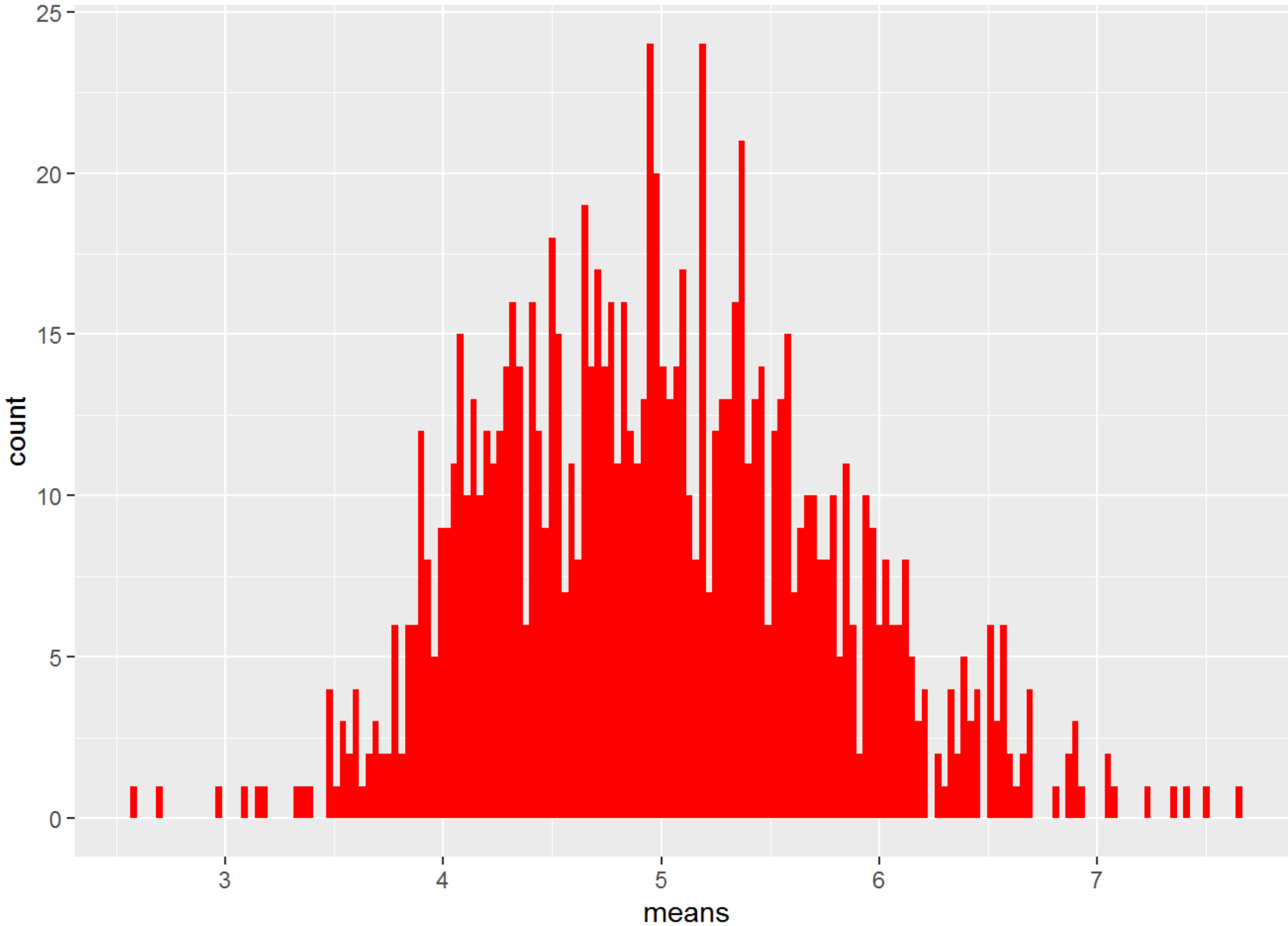
```
# load neccesary libraries
library(ggplot2)

# set constants
set.seed(512324) # set the seed to create reproducability
lambda <- 0.2 # rate parameter
numOfSimulations <- 1000 # no of test
sampleSize <- 40 # number of exponentials

# run the test resulting in n x numOfSimulations matrix

# We'll ise R function rexp which will take n and lambda as parameters
expDis <- matrix(data=rexp(sampleSize * numOfSimulations, lambda ), nrow=numOfSimulations)
expDisMeans <- data.frame(means=apply(expDis, 1, mean))
#expDisMeans <-rowMeans(expDis)
```

Plot the histogram of averages



Sample Mean versus Theoretical Mean

The expected mean mu of a exponential distribution of rate lambda is

```
mu <- 1/lambda
mu
```

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

Let's take average sample mean of 1000 simulations of 40 randomly sampled exponential distributions.

```
avgMean <- mean(expDisMeans$means)
avgMean
```

```
## [1] 4.994854
```

As we can see the expected mean and the avarage sample mean are very close

Sample Variance versus Theoretical Variance

Using the formula for standard deviation and exponential rate

```
sdVal<-1/lambda/sqrt(sampleSize)
sdVal
```

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

Uisng the formula variance is the square of standard deviation

```
Var <- sdVal^2
Var
```

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

Let’s use variance of the average sample mean of 1000 simulations of 40 randomly sampled exponential distribution, and the corresponding standard deviation.

```
# Calculate standard deviation
sdVal1<-sd(expDisMeans$means)
sdVal1
```

```
## [1] 0.7723191
```

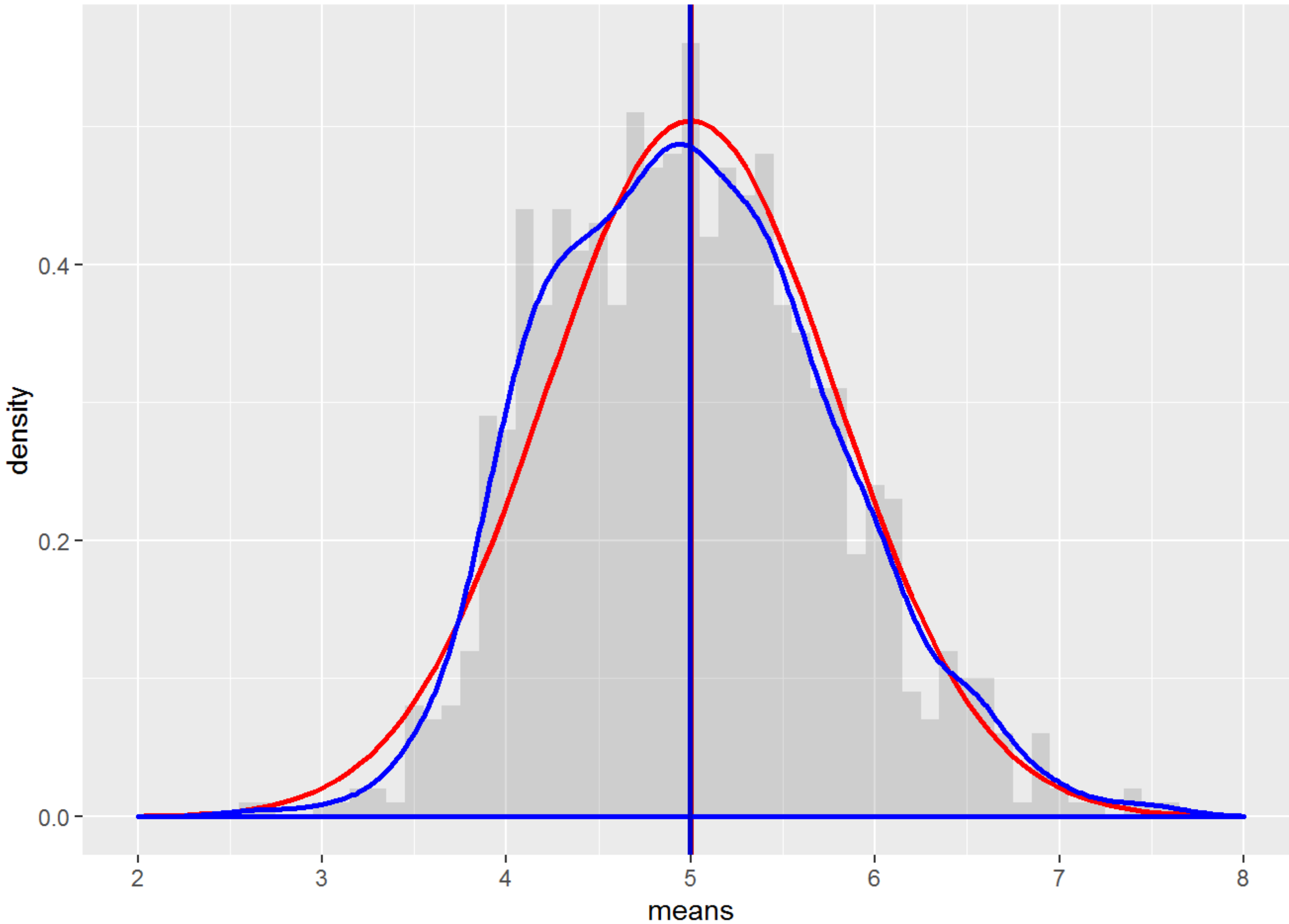
```
# Calculate variance
varVal<-var(expDisMeans$means)
varVal
```

```
## [1] 0.5964767
```

As we can see the standard deviations are close Since variance is the square of the standard deviations.

Distribution

Comparing the population means & standard deviation with a normal distribution of the expected values.



It is quite visible from the graph that the calculated distribution of means of random sampled exponential distributions, overlaps nicely with the normal distribution with the expected values based on the given lambda.