

Investigation of Exponential Distribution and Central Limit Theorem in R

Ron He

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 $1/\lambda$. This project will use `lambda = 0.2` for all of the simulations, and then investigate the distribution of averages of 40 exponentials.

Construct the Data frame

The `rexp` function can generate a thousand times of simulation, with each simulation collecting 40 samples. The observation matrix is constructed by such matrix with the row as the simulation number. Then apply means to all of those simulations we can have a single mean column consists of the means of the 40 samples.

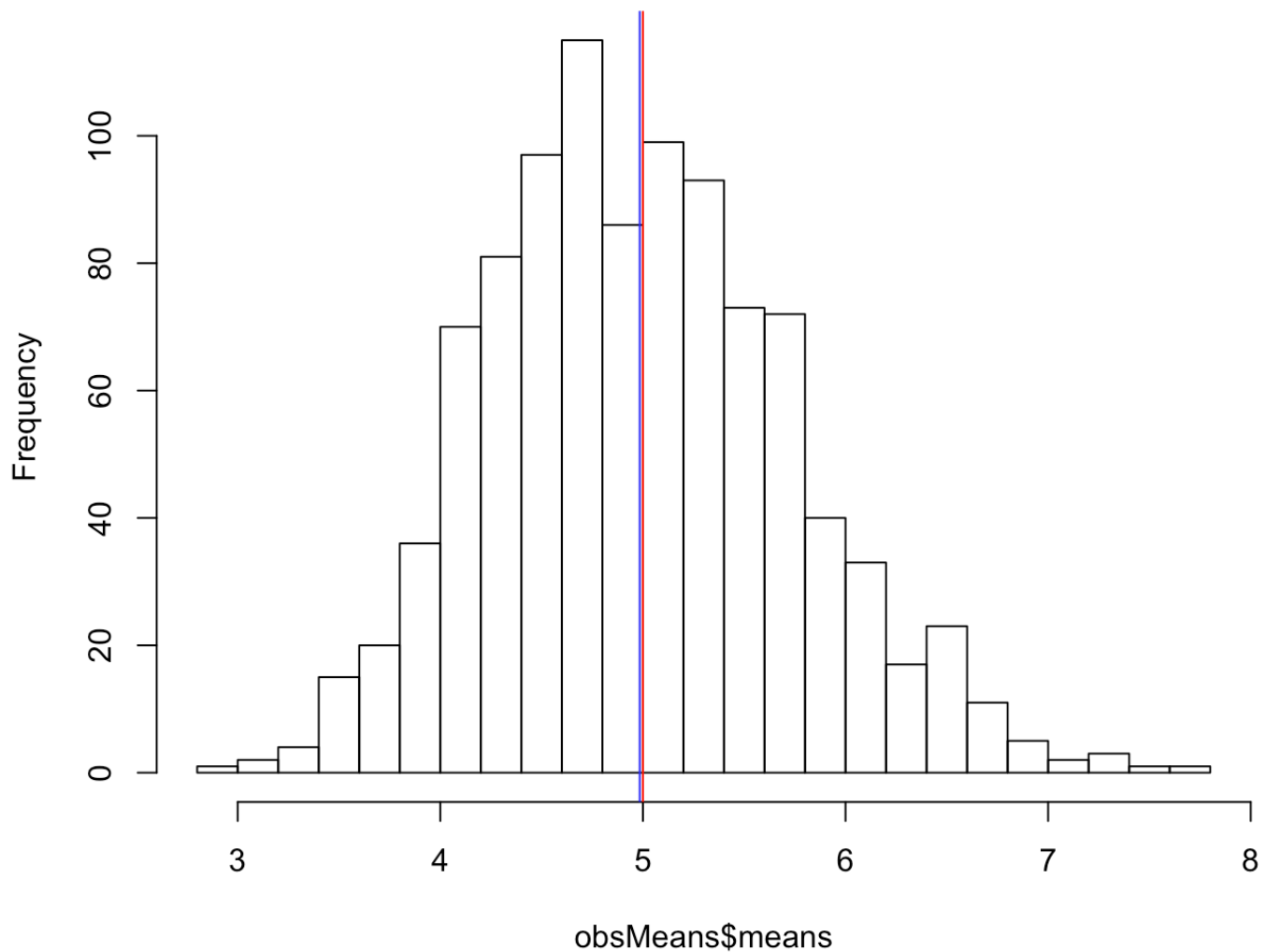
```
lambda<-0.2
n<-40
simNum<-1000
obs <- matrix(data=rexp(n = n * simNum, rate = lambda), nrow=simNum)
obsMeans <- data.frame(means = apply(X = obs, MARGIN = 1, FUN = mean))
```

The mean of the sample means and theoretical mean of the sample means comparison

A histogram can be generated by the means column of the `obsMeans`

```
hist(obsMeans$means, breaks=20)
abline(v= 1/lambda, col = "red", lwd =1 )
mean40 <- mean(obsMeans$means)
abline(v= mean40, col = "blue", lwd =1 )
```

Histogram of obsMeans\$means



The mean of the the samples is

```
## [1] 4.984133
```

The theoretical means is

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

The variance of the sample means and the theoretical variance of the distribution comparison.

The variance of the means is:

```
var(obsMeans$means)
```

```
## [1] 0.5698822
```

The theoretical variance of the distribution is square of the standard deviation divided by sample amount:

```
(1/lambda)^2/40
```

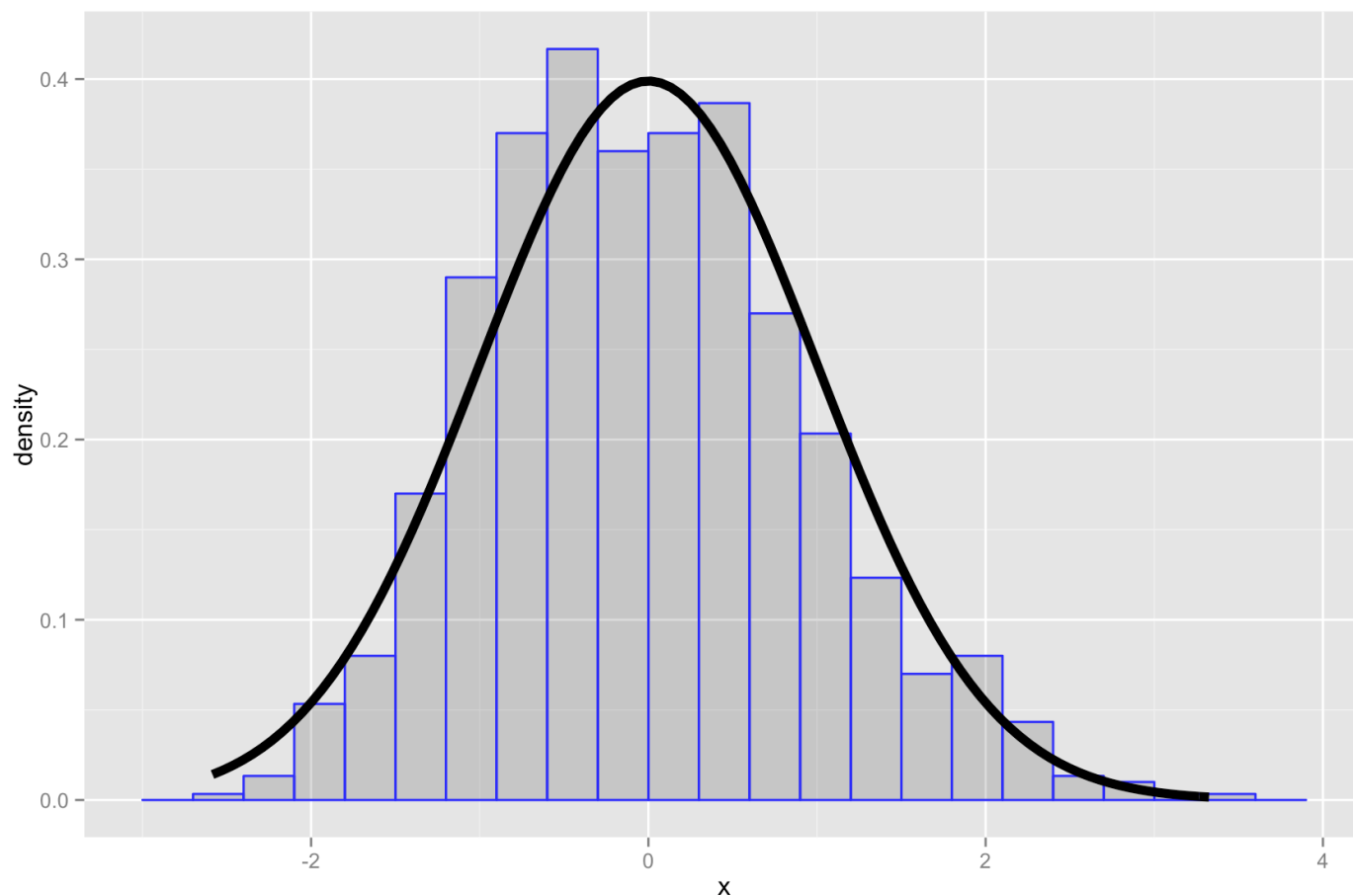
```
## [1] 0.625
```

Normal Distribution of the Means.

The mean of the sample means is $1 / \lambda$, the mean of a distribution of means taken from an exponential distribution should also be $1 / \lambda$. The standard deviation of the sample means is $1 / \lambda$, the standard deviation of a distribution of means should be $1 / (\sqrt{n} * \lambda)$.

To find out whether the means are normally distributed, we will do some conversion to the means and compare with the normal distribution curve. If a population is normally distributed, then after it is subtracted the distribution mean, then divided by the standard deviation, it should be a curve with mean of 0 and standard deviation of 1. The following curve shows that the distribution is the case:

```
library(ggplot2)
cfunc <- function(x, n) lambda * sqrt(n) * (mean(x) - (1/lambda))
dat <- data.frame(
  x = apply(obs, 1, cfunc, 40)
)
g <- ggplot(dat, aes(x = x)) + geom_histogram(alpha = .20, binwidth=.3, colour =
"blue", aes(y = ..density..))
g <- g + stat_function(fun = dnorm, size = 2)
g
```



Summary

In this demonstration, we show that that averages of iid samples converge to the population means that they are estimating, which is the LLN rule. We also show that the averages are approximately normal, with distributions centered at the population mean and with standard deviation equal to the standard error of the mean, which is the The CLT.