

# Analysis on Exponential Distribution and CLT in R

*Thiloshon Nagarajah*

*4/20/2017*

## Overview

In this analysis i will 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  $1/\lambda$ . I have set `lambda = 0.2` for all of the simulations.

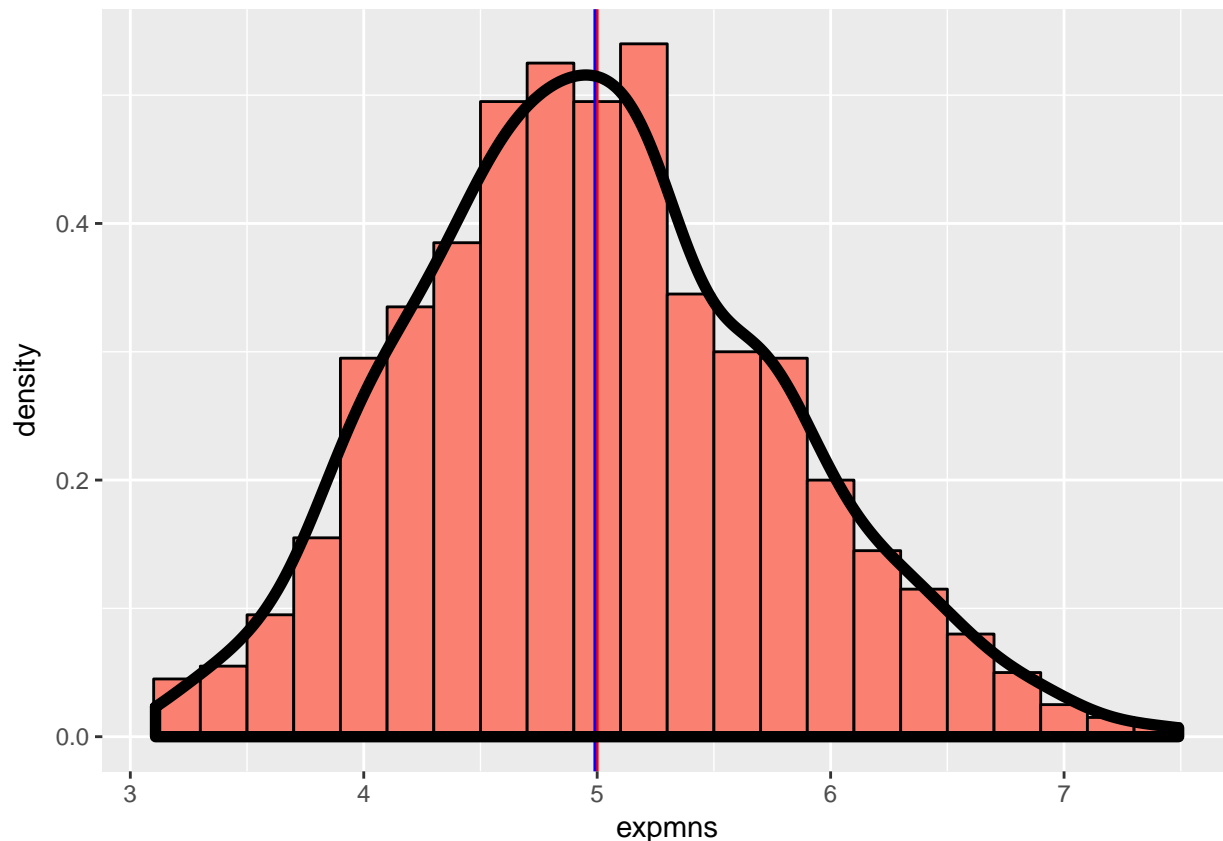
```
set.seed(1)
library(ggplot2)
expmns = NULL
```

Creating a distribution of averages of 40 exponentials of size 1000

```
for (i in 1 : 1000) expmns = c(expmns, mean(rexp(40,0.2)))
```

Plotting the histogram of the distribution

```
g <- ggplot(as.data.frame(expmns), aes(x = expmns))
g <- g + geom_histogram(fill = "salmon", aes(y = ..density..), colour = "black", binwidth = .2)
g <- g + geom_vline(xintercept = 5, colour="red")+ geom_vline(xintercept = mean(expmns), colour = "blue")
g <- g + geom_density(size = 2)
g
```



## 1. Show the sample mean and compare it to the theoretical mean of the distribution.

The theoretical mean is drawn as a vertical line in red and the sample mean is drawn in blue in the above plot. But the lines are overlapping as both values are pretty much same. To be precise let's get the mean of exponential distribution from  $(1/\text{rate parameter})$ .

```
1/0.2
```

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

And also get the mean of the generated exponential.

```
mean(expmns)
```

```
## [1] 4.990025
```

This shows the mean of the sample and theoretical mean are both very much same.

## 2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

The theoretical variance of exponential distribution is given by,

```
((1/0.2)/sqrt(40))^2
```

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

whereas the variance of the generated exponential is given by,

```
var(expmns)
```

```
## [1] 0.6111165
```

We can conclude both variance are very much similar.

## Show that the distribution is approximately normal.

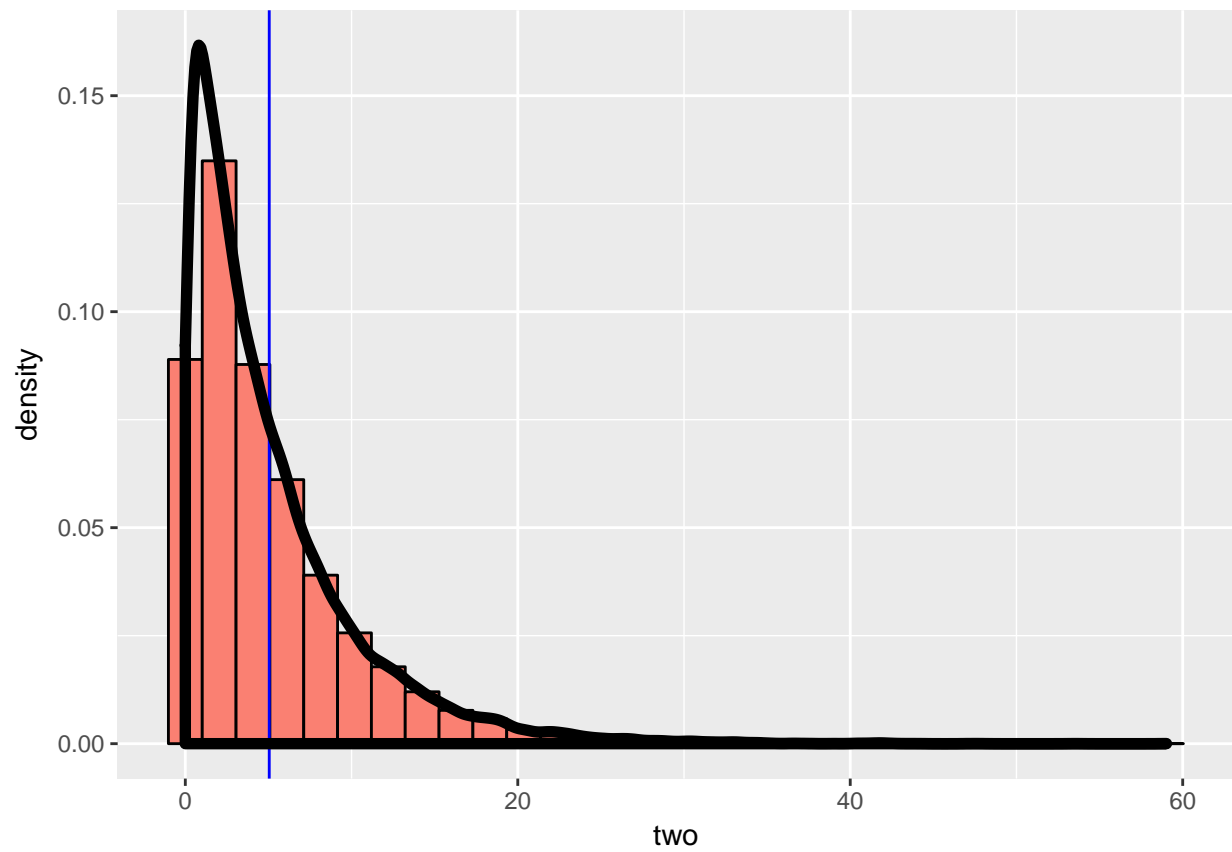
From the initial plot, the density line shows how the plot is slightly left skewed. But it is approximately normal.

To better analyze the CLT, let's create distribution of a large collection of random exponentials and compare with a plot of distribution of a large collection of averages of 40 exponentials.

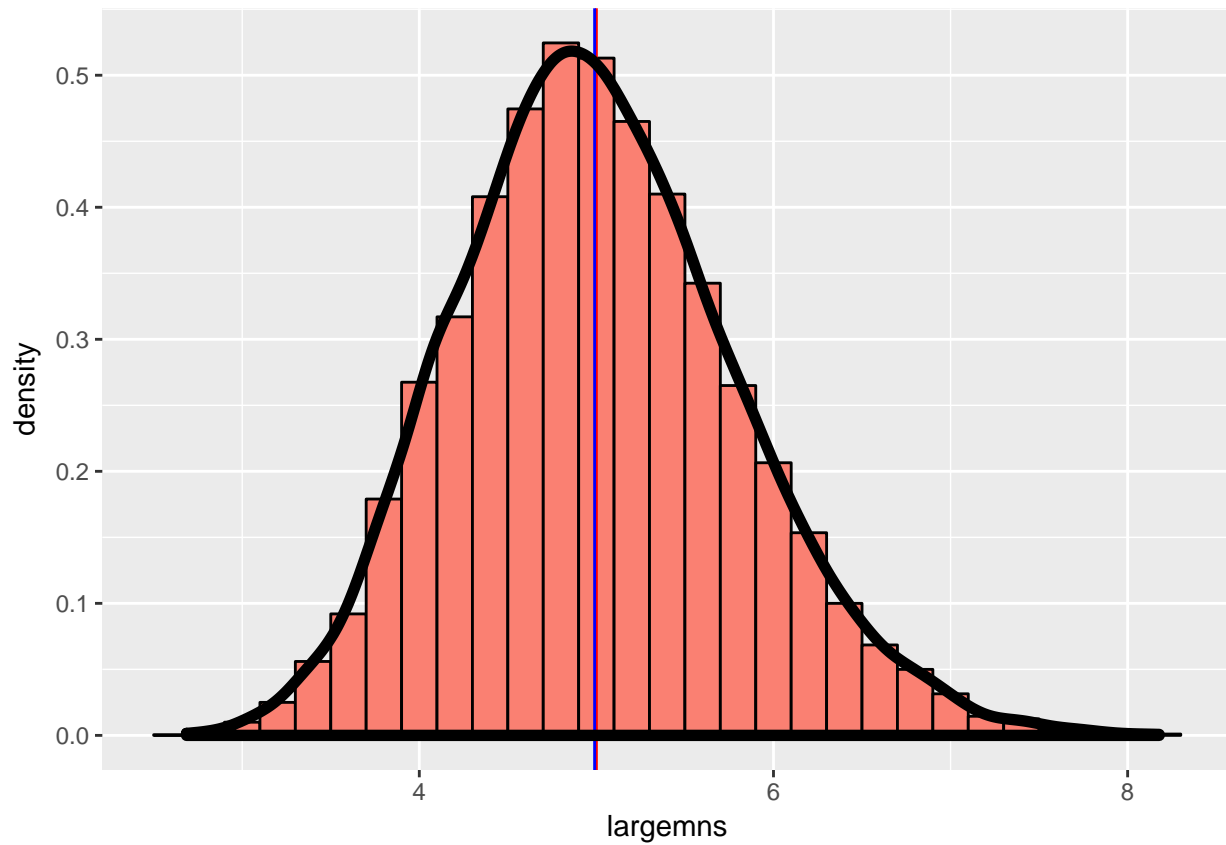
Plot of distribution of a large collection of random exponentials

```
two<-NULL
for (i in 1 : 1000) two = c(two, rexp(40,0.2))
g <- ggplot(as.data.frame(two), aes(x = two))
g <- g + geom_histogram(fill = "salmon", aes(y = ..density..), colour = "black")
g <- g + geom_vline(xintercept = mean(two), colour = "blue")
g <- g + geom_density(size = 2)
g
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Plot of distribution of a large collection of averages of 40 exponentials.



This demonstrates the Central Limit Theorem, distribution of averages of independent and identically distributed (IID) variables becomes that of a standard normal as the sample size increases