Statistical Inference Course Project, Exponential Distribution

Morozov Gleb

Saturday, May 23, 2015

In this project we 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. Set lambda = 0.2 for all the simulations. We will investigate the distribution of averages of 40 exponentials. We will do a thousand simulations.

```
library(dplyr)
library(ggplot2)
```

Simulations.

```
nosim <- 1000
n <- 40
lambda <- 0.2
means.dist <- rexp(n*nosim, lambda) %>% matrix(., nosim) %>% apply(., 1, mean)
```

Sample and Theoretical Mean.

Theoretical mean:

```
theory.mean <- 1/lambda
theory.mean
```

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

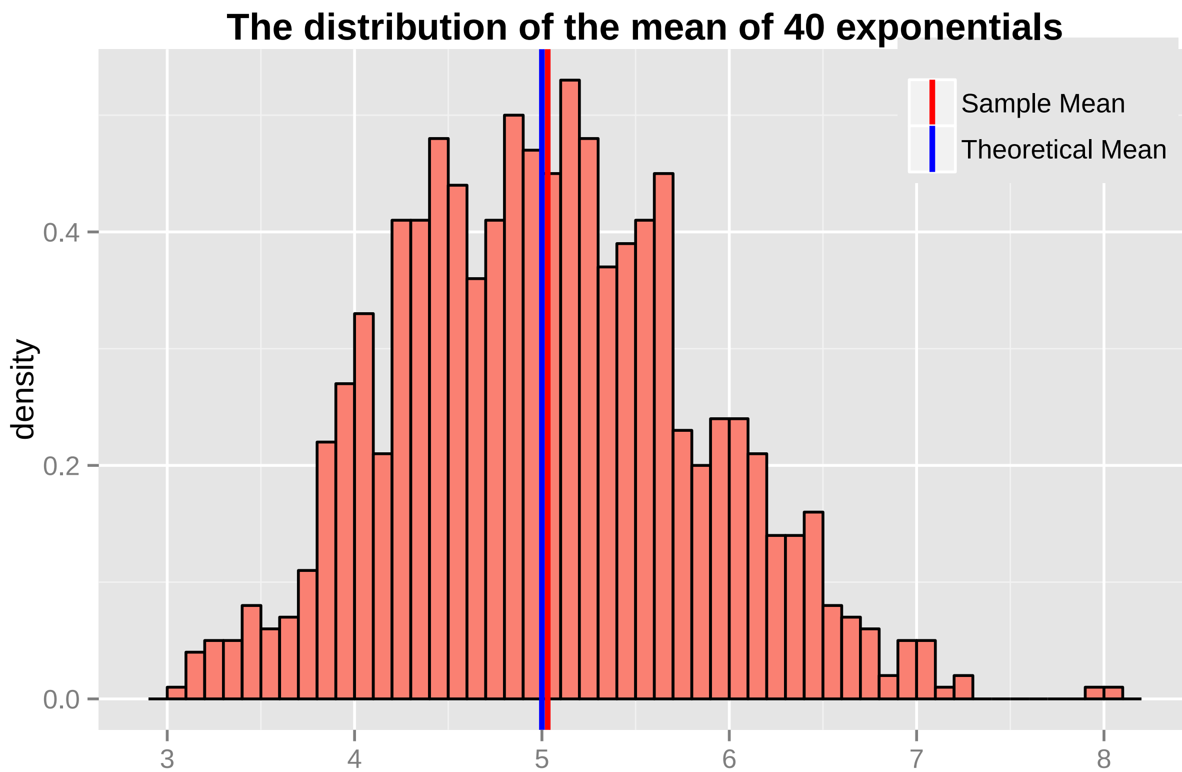
Sample mean:

```
sample.mean <- mean(means.dist)
sample.mean
```

```
## [1] 5.030827
```

```
g <- ggplot(data.frame(x = means.dist), aes(x = x)) +
        geom_histogram(binwidth = 0.1, colour = "black", fill ="salmon",
                        aes(y = ..density..))
g + geom_vline(aes(xintercept = c(theory.mean, sample.mean), colour =
                                c("Theoretical Mean", "Sample Mean")),
                size = 1, show_guide = T) +
        scale_color_manual(name = "", values =
                                c("Theoretical Mean" = "blue", "Sample Mean" =
```

```
                                              "red" )) +
        theme(legend.position = c(0.86, 0.91), legend.background =
                     element_rect(fill = "grey90")) +
        labs(title = "The distribution of the mean of 40 exponentials", x = "") +
        theme(plot.title = element_text(face="bold"))
```



We can see that the sample mean and the theoretical mean is very close.

Sample and Theoretical Variance.

Theoretical Variance.

```
theoretical.var <- ((1/lambda)^2)/40
theoretical.var
```

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

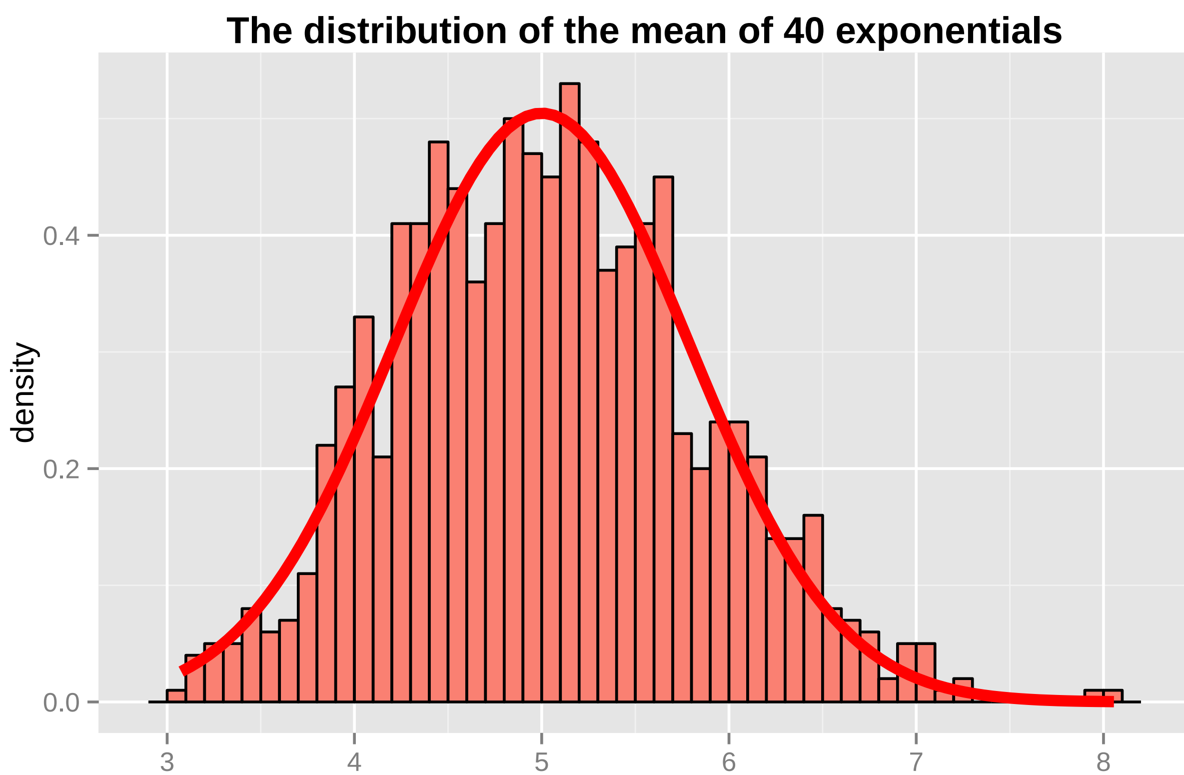
Sample Variance.

```
sample.var <- var(means.dist)
sample.var
```

```
## [1] 0.6538313
```

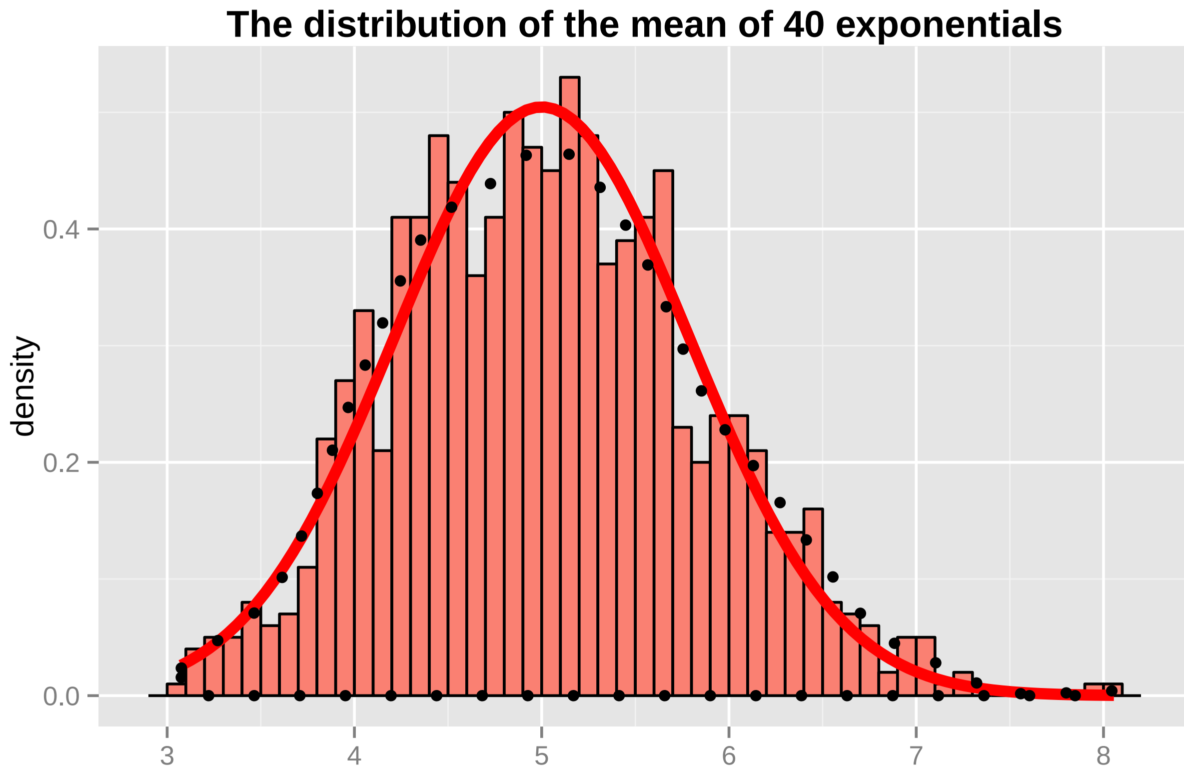
We can see that the sample variance and the theoretical variance are very close. Variance is a measure of the spread of the data from the mean. We observe from the figure that the spread of the sample data and theoretical values are also very close.

```
g + stat_function(fun = dnorm, colour = "red", size = 2, arg =
                          list(mean=1/lambda, sd=(1/lambda/sqrt(40)))) +
        labs(title = "The distribution of the mean of 40 exponentials", x = "") +
        theme(plot.title = element_text(face="bold"))
```



Disbutrion.

```
g + stat_function(fun = dnorm, colour = "red", size = 2, arg =
                          list(mean=1/lambda, sd=(1/lambda/sqrt(40)))) +
        geom_density(size = 2, colour = "black", linetype = "dotted") +
        labs(title = "The distribution of the mean of 40 exponentials", x = "") +
        theme(plot.title = element_text(face="bold"))
```



From Central Limit Theorem, we know that the distribution of a sample consisting n independently identically distributed random variables closely resembles normal distribution. In this simulation exercise, we compare the simulated data with the theorical values. Their first and second central moment are very close to each other, theoretical and sample density graphics are close to each other which verifies the conclusion of Central Limit Theorem.