Statistical Inference: Course Project 1

VALAT Didier

13/06/2015

Overview

The aim of this project is 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, λ) where λ is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. The variable λ will be set to **0.2** for all of the simulations. We will investigate the distribution of averages of 40 exponentials. Note that we will produce a thousand simulations.

Simulations

First, we will load the required libraries and define the figures size:

```
# Load libraries
library(knitr)
library(ggplot2)
opts_chunk$set(fig.width=6, fig.height=3.5)
```

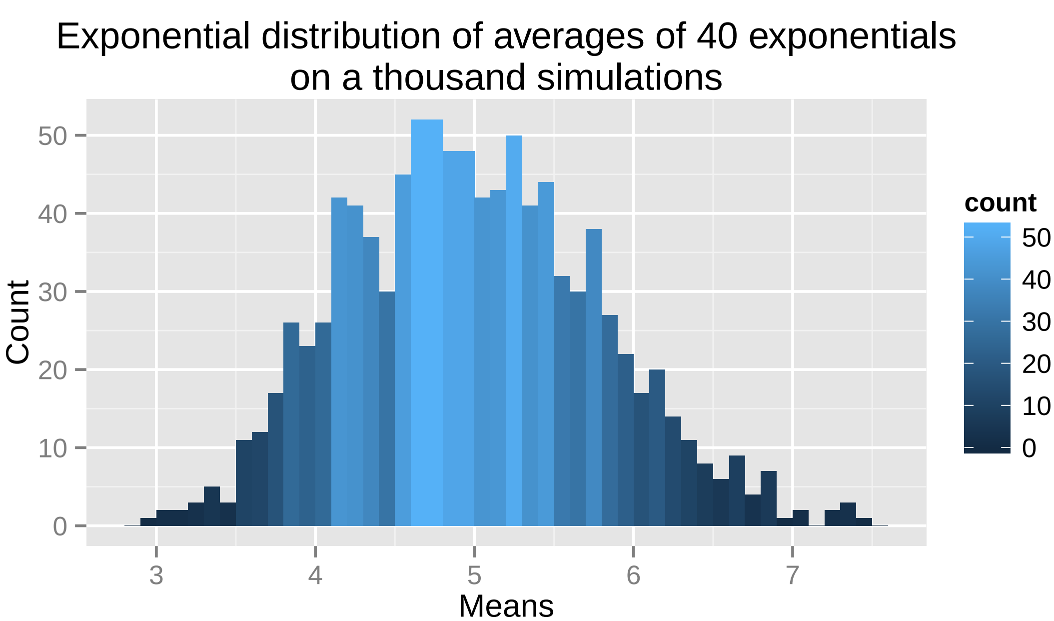
Then, we will define the variables for the study:

```
lambda <- 0.2 # Set lambda for the Exponential Distribution
n <- 40 # Set the number of exponentials
num_of_simul <- 1000 # Set the number of simulations
set.seed(3) # Set the seed for analysis reproducibility
```

After that, we will calculate the Exponential distribution:

```
# Calculate the Exponential distribution using a Matrix of n * number of simulations
exp_distrib <- matrix(data = rexp(n * num_of_simul, lambda), nrow = num_of_simul)
exp_distrib_mean <- data.frame(means = apply(exp_distrib, 1, mean))
```

And then, we will plot it:



Sample Mean versus Theoretical Mean

We will calculate the expected mean μ of an exponential distribution of rate λ using the following formula $\mu = \frac{1}{\lambda}$

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

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

Next, we will get $\bar{X}$ which is the average sample mean of 1000 simulations of 40 randomly sampled exponential distributions:

```
x_bar <- mean(exp_distrib_mean$means)
x_bar
```

```
## [1] 4.98662
```

We can see here that the expected mean and the average sample mean values are very close.

Sample Variance versus Theoretical Variance

We will now calculate the expected standard deviation σ of an Exponential distribution of rate λ with the following formula $\sigma = \frac{1/\lambda}{\sqrt{n}}$

```
std_dev <- 1/lambda/sqrt(n)
std_dev
```

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

Next, we will get the variance Var of the standard deviation σ which is $Var = \sigma^2$

```
var <- std_dev^2
var
```

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

After that, we will calculate the standard deviation σ_x and the variance Var_x of the average sample mean of 1000 simulations of 40 randomly sampled exponential distribution:

```
std_dev_x <- sd(exp_distrib_mean$means)
std_dev_x
```

```
## [1] 0.7910484
```

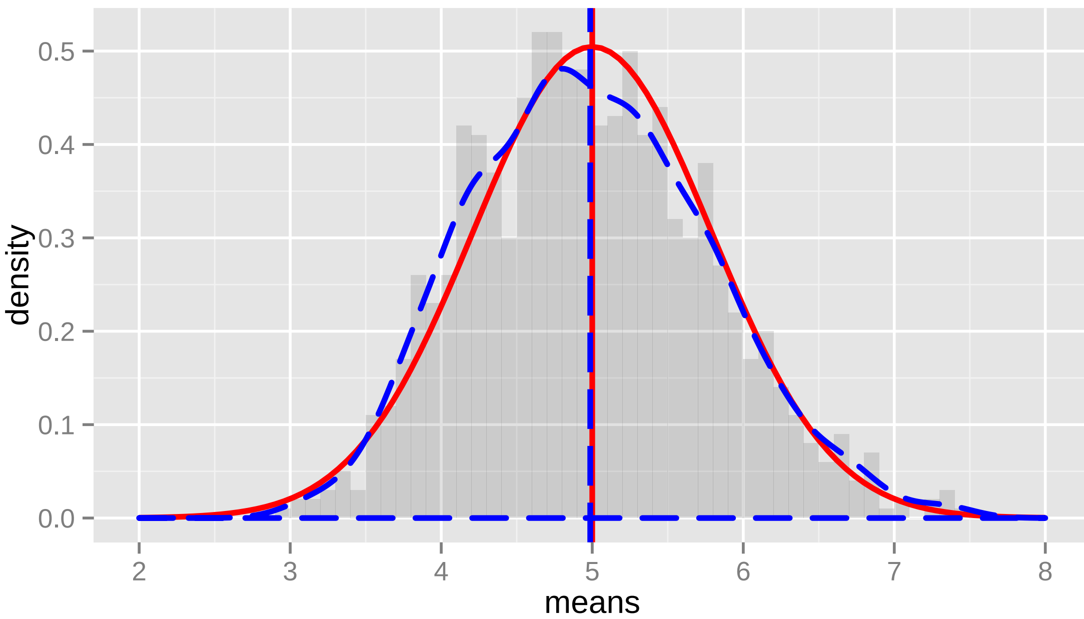
```
var_x <- var(exp_distrib_mean$means)
var_x
```

```
## [1] 0.6257575
```

We can note here that the standard deviations are close too. Indeed, the variance is the square of the standard deviations so the values are quite close.

Distribution

In this last part, we will compare the population means and standard deviation (**blue dashed lines**) with a normal distribution (**red lines**) of the expected values. Lines will be added in order to display the calculated and expected means:



As we can see, the calculated distribution of means of random sampled exponantial distributions quite overlaps with the normal distribution with the expected values based on the given λ.