Course 6 - Project - Part 1

Bruno Assis

19/03/2017

Exponential distribution compared to Central Limit Theorem

Overview

In this project you 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 of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

We'll load `ggplot2` library and use a fixed seed so the results can be reproduced:

```r
library(ggplot2)
set.seed(42)
```

Simulations

Let's set some constants for these simulations:

```r
lambda <- 0.2
n.distributions <- 40
n.simulations <- 1000
```

Sample Mean vs. Theoretical Mean

```r
calculate.mean <- function(x) {
  mean(rexp(n.distributions, lambda))
}

means <- sapply(X = 1:n.simulations, FUN = calculate.mean)
```

Let's take a look at some of the means and the average of the means. We can also plot a histogram with the means:

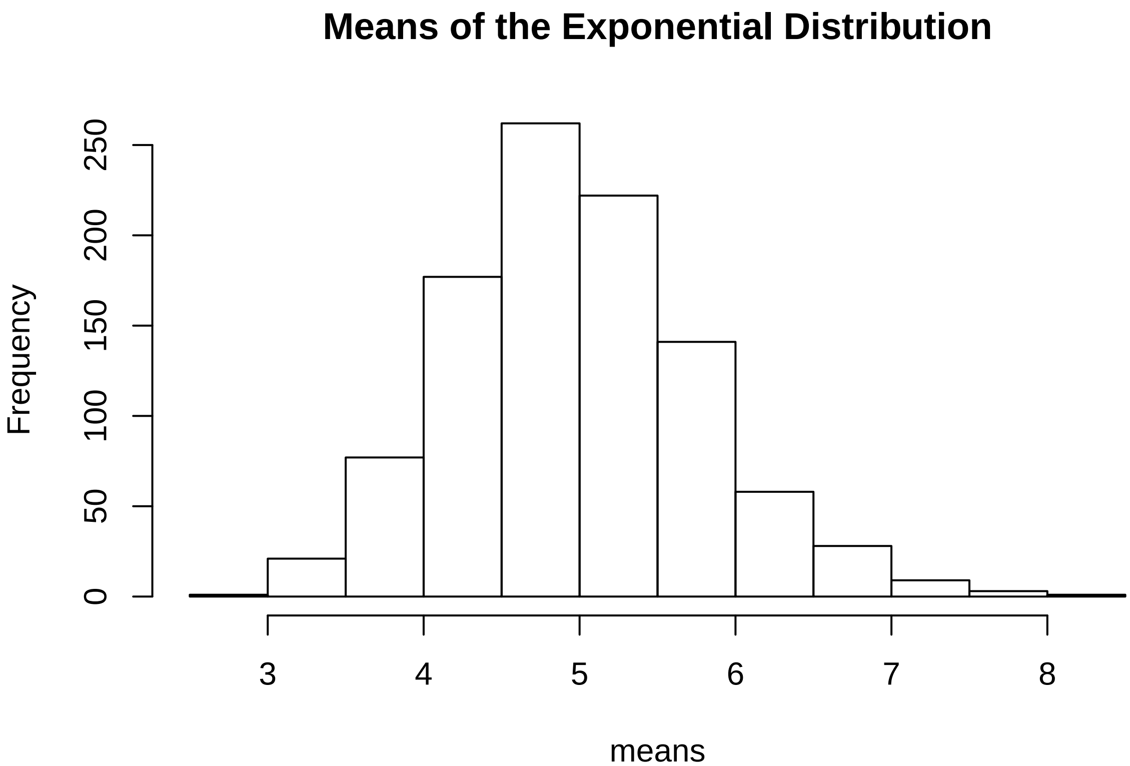
```r
head(means)
```

```
## [1] 4.915756 6.941835 4.775331 5.310784 7.002644 5.320620
```

```r
mean(means)
```

```
## [1] 4.986508
```

```
hist(means, main="Means of the Exponential Distribution")
```



We can now compare the average of the means to `1 / lambda` to see if they converge:

```
theoretical.mean <- 1/lambda
sample.mean <- mean(means)
```

It is, so, centered in the theoretical mean.

Sample Variance vs. Theoretical Variance

We can also compare the theoretical variance (`((1/lambda)^2)/n.distributions`) with the variance of the means to see if they converge:

```
theoretical.variance <- ((1/lambda)^2)/n.distributions
sample.variance <- var(means)
```

Which they do.

Distribution

We can see the distribution of the means looks like a normal distribution. We'll now plot the histogram with both the theoretical and sample functions to compare them.

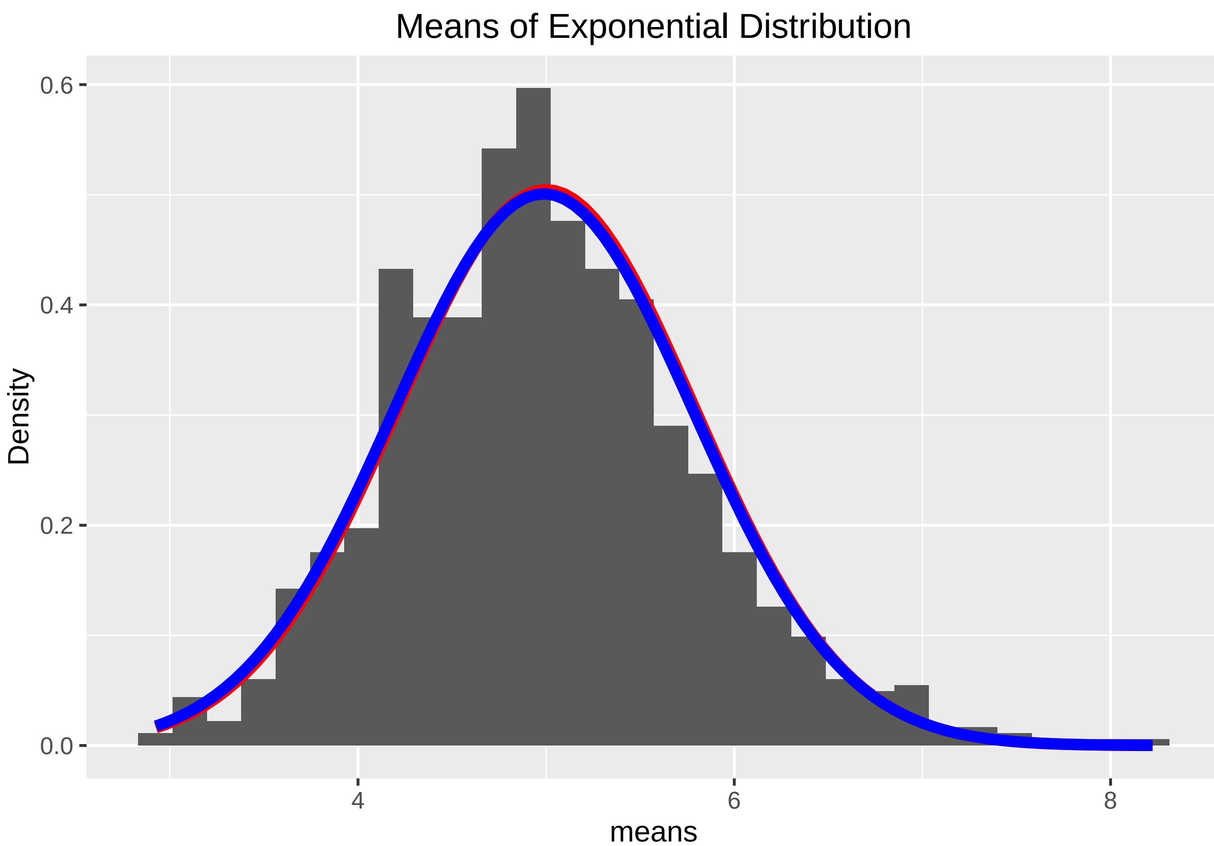
```
theoretical.normal.function <- stat_function(fun = dnorm,
                                             args = list(mean = theoretical.mean,
                                                         sd = sqrt(theoretical.variance)),
                                             color = "red",
                                             size = 2.0)
```

```r
sample.normal.function <- stat_function(fun = dnorm,
                                        args = list(mean = sample.mean,
                                                    sd = sqrt(sample.variance)),
                                        color = "blue",
                                        size = 2.0)

comparison.plot <- ggplot(data.frame(means), aes(x = means)) +
                   geom_histogram(aes(y = ..density..)) +
                   labs(title = "Means of Exponential Distribution", y = "Density") +
                   theoretical.normal.function +
                   sample.normal.function

comparison.plot
```

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



Comparing the sample (blue line) with the theoretical (red line) distribution we can see they look alike.

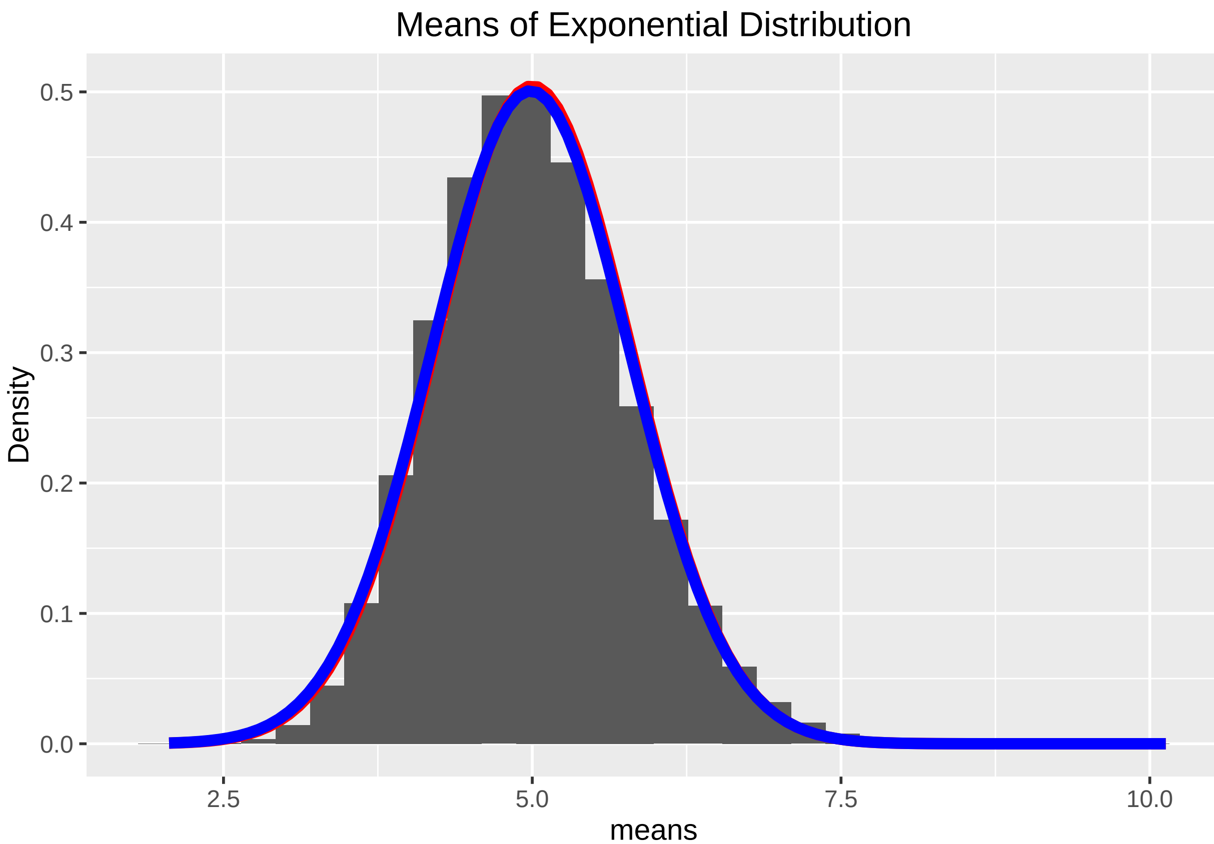
We can run the same experiment with a higher number of simulations, and see it converges even more:

```r
n.simulations <- 1000000
```

Rerunning this same experiment, with a higher number of simulations, we get the following plot:

```r
comparison.plot
```

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


Means of Exponential Distribution
Density
0.5
0.4
0.3
0.2
0.1
0.0
2.5
5.0
7.5
10.0
means