The Exponential Distribution in R *versus* the Central Limit Theorem (CLT) — Part 1

calaca

March 9th 2016

Overview

This project consists in using simulation to explore inference and do some simple inferential data analysis. The project is divided into two parts:
1. A simulation exercise;
2. Basic inferential data analysis.

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.

Simulations

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

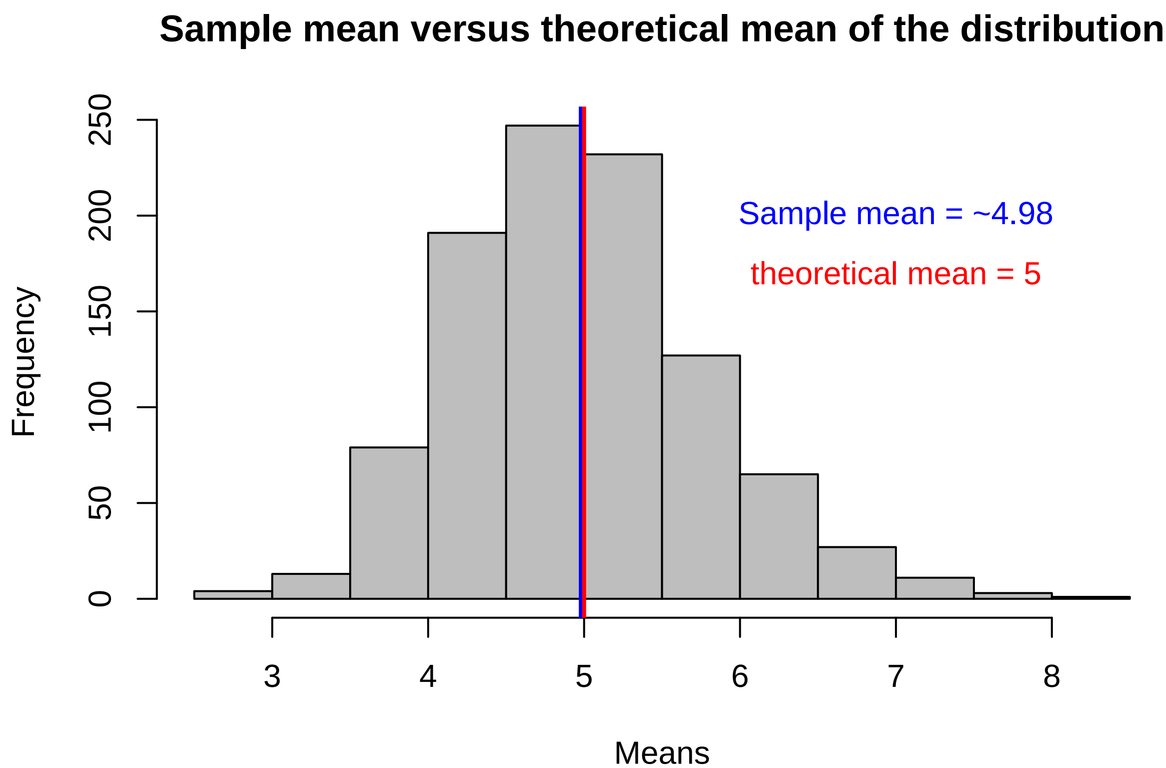
```
set.seed(2016) # Seed
lambda <- .2 # Lambda
n <- 40 # Sample
simulations <- 1000 # Simulations

# Replicating and simulating exponentials
simulated <- replicate(simulations, rexp(n, lambda))

# Mean of the simulated data
simulated_means <- apply(simulated, 2, mean)

# Mean of the simulated means
sample_mean <- mean(simulated_means)
sample_mean
```

```
## [1] 4.979186
```

```
# Mean of the analytical expression
theoretical_mean <- 1/lambda
theoretical_mean
```

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

```r
# Plot
hist(simulated_means, xlab = "Means", main = "Sample mean versus theoretical mean of the distribution",
abline(v = sample_mean, col = "blue", lwd = "2")
abline(v = theoretical_mean, col = "red", lwd = "2")
text(x = 7, y = 200, labels = "Sample mean = ~4.98", col = "blue")
text(x = 7, y = 170, labels = "theoretical mean = 5", col = "red")
```



Observing the code and the plot we can see that, with a thousand simulations, the sample mean is ~4.98 and the theoretical mean of the distribution is 5.

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

```r
# Standard Deviation of the sample means
sd_sample <- sd(simulated_means)
sd_sample
```

```
## [1] 0.7990522
```

```r
# theoretical Standard Deviation
sd_theoretical <- theoretical_mean/sqrt(n)
sd_theoretical
```

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

```r
# Variance of the sample means
var_sample <- sd_sample ^ 2
var_sample
```

```
## [1] 0.6384844
```

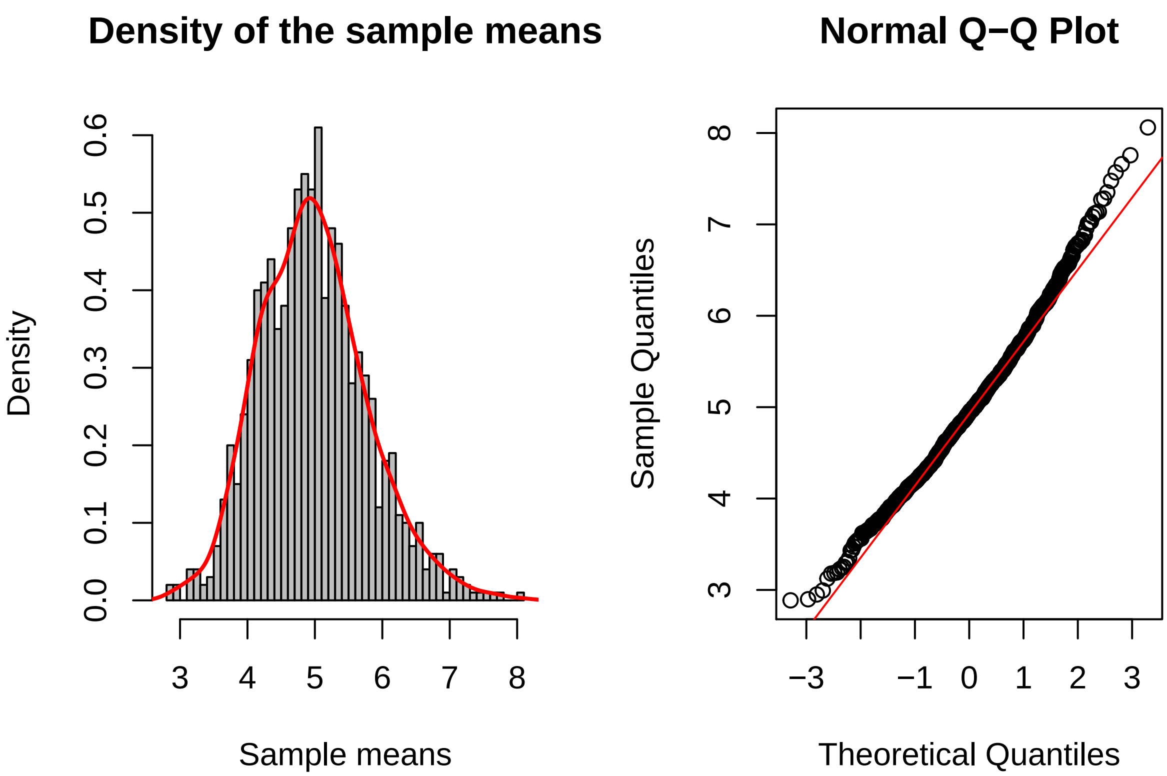
```r
# theoretical variance
var_theoretical <- sd_theoretical ^ 2
var_theoretical
```

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

Therefore, it is possible to note that the sample variance (0.6384844) is very close to the theoretical variance (0.625).

3. Show that the distribution is approximately normal

```r
par(mfrow=c(1, 2))
hist(simulated_means, prob = TRUE, col = "grey", main = "Density of the sample means", ylab = "Density"
lines(density(simulated_means), lwd = 2, col = "red")
qqnorm(simulated_means)
qqline(simulated_means, col = "red")
```



Observing both the density and the normal Q-Q plot, it is possible to see that the distribution of averages of 40 exponentials is close to the normal distribution. If we get more simulations, the density will be even closer to a bell-shaped form.
