Peer Assessment for Coursera Statistical Inference Class

Edoardo "Dado" Marcora

2015-04-26 - Part 1

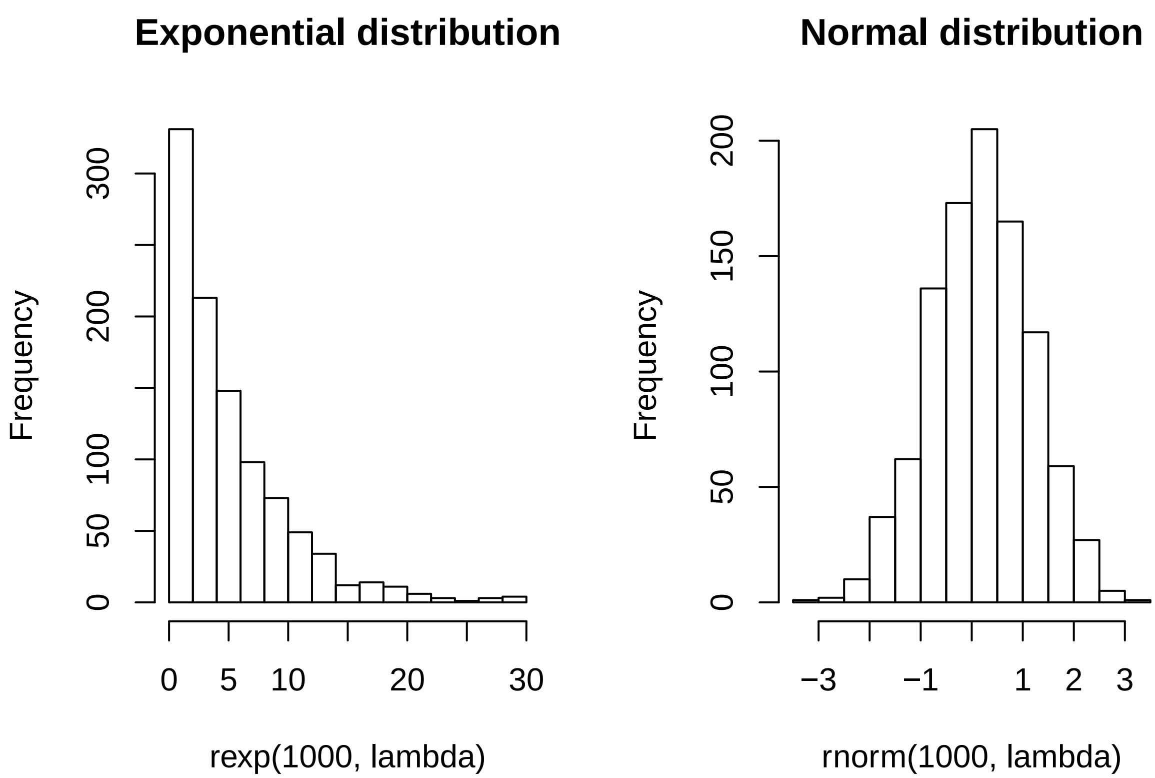
In this project we investigate the exponential distribution in the context of the central limit theorem (CLT) using R. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter and n is the number of random exponentials. The mean of this distribution is 1/lambda and the standard deviation is also 1/lambda.

We will investigate the sampling distribution of the mean of random exponentials by running a large number of simulations in R. In this project, lambda is set to 0.2 and the sample size to 40 for all of the 1,000 simulations that we will run.

```
lambda = 0.2 # rate parameter for the exponential distribution
n = 40 # sample size
b = 1000 # number of simulations
```

First, we look at the exponential distribution using an histogram. It is most definitively not bell-shaped like the normal distribution.

```
par(mfrow=c(1,2))
hist(rexp(1000, lambda), main="Exponential distribution")
hist(rnorm(1000, lambda), main="Normal distribution")
```



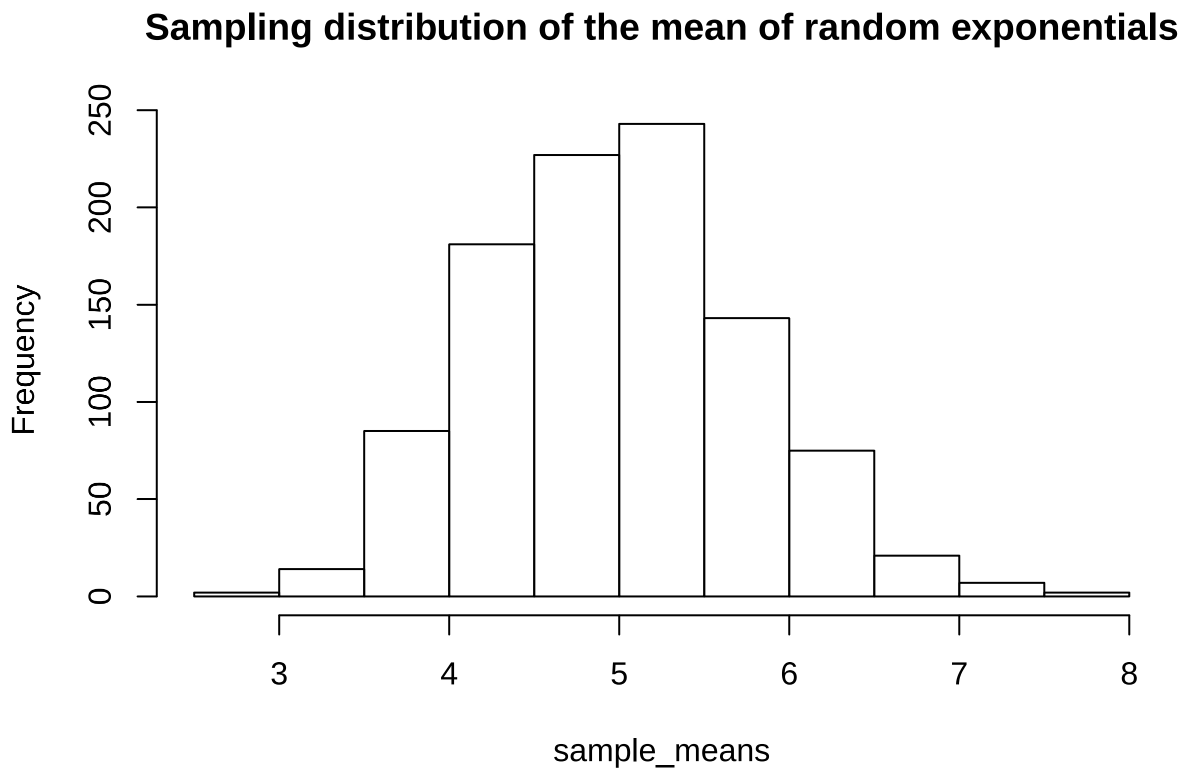
However, according to the CLT, the sampling distribution of the mean of random exponentials should be normal with mean XXX and standard deviation XXX.

In order to verify this, we first collect the means of 1,000 samples of random exponentials, each of size 40.

```
sample_means = NULL
for (i in 1 : b) sample_means = c(sample_means, mean(rexp(n, lambda)))
```

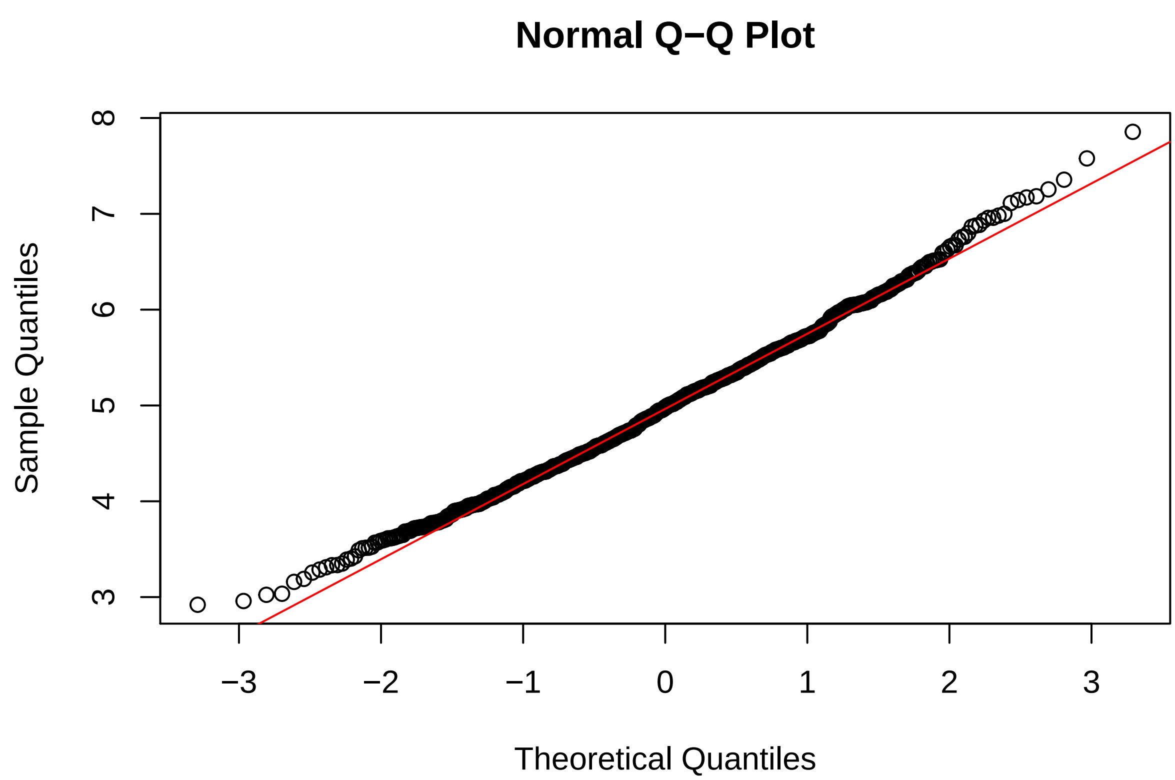
Second, we inspect the distribution of these sample means, which (in accordance to the CLT) is bell-shaped.

```
par(mfrow=c(1,1))
hist(sample_means, main = "Sampling distribution of the mean of random exponentials")
```



We then further verify that this distribution is approximately normal by inspecting a normal Q-Q plot.

```
qqnorm(sample_means)
qqline(sample_means, col=2)
```



As expected, the normal Q-Q plot reveals that the distribution of these sample means is approximately normal with mean:

```
(sample_means_mean = mean(sample_means))
```

```
## [1] 4.984523
```

and standard deviation:

```
(sample_means_sd = sd(sample_means))
```

```
## [1] 0.7744163
```

Therefore, in accordance with the CLT, the empirical mean of this distribution (sample_means_mean) is very similar to the theoretical mean (1/lambda):

```
(theoretical_mean = 1/lambda)
```

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

and the empirical standard deviation of this distribution (sample_means_sd) is very similar to the theoretical standard deviation, i.e., the standard error of the mean (1/(lambda*sqrt(n)))

```
(theoretical_sd = 1/(lambda*sqrt(n)))
```

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