Statistical Inference - Course Project - Part 1

Miguel Corral

23th September 2015

Overview

This report shows the comparison made betweeen the exponential distribution and the Central Limit Theoreme.

This analysis will show:

1. The sample mean and compare it to the theoretical mean of the distribution.
2. How variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. That the distribution is approximately normal.

Parameters for the analysis

- The exponential distribution is simulated in R with 'exp(n, lambda)' where lambda is the rate parameter.
- The mean of the exponential distribution is 1/lambda.
- The standard deviation is also 1/lambda.
- Lambda will be set to 0.2 for all the simulations.
- The analysis will be done upon the averages of 40 exponentials.
- There will be done 10 thousand simulations.

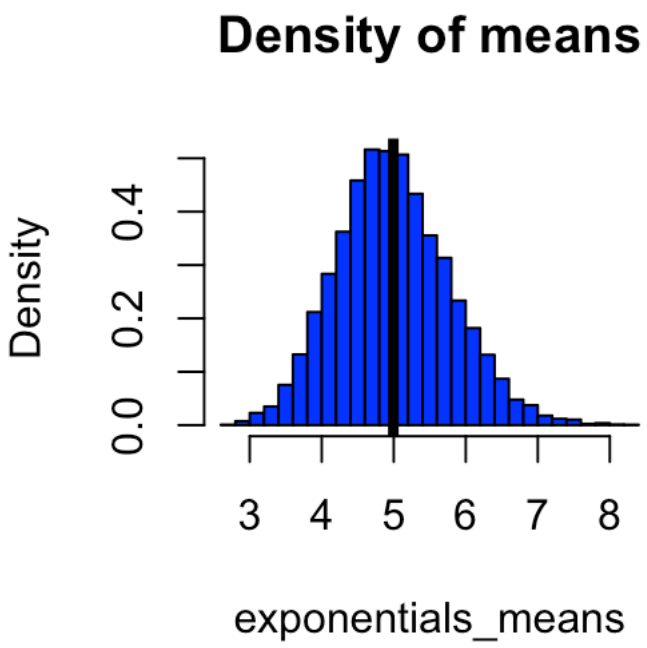
Simulations

```
# set seed for reproducability
set.seed(43)
# set lambda to 0.2
lambda <- 0.2
# 40 samples
n <- 40
# 10000 simulations
no_simulations <- 10000
# generate matrix of exponentials
exponentials <- replicate(no_simulations, rexp(n, lambda))
# calculate mean of exponentials
exponentials_means <- apply(exponentials, 2, mean)
```

1. Sample Mean versus Theoretical Mean

```
## [1] 4.994552
```

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



As seen above, the sample distribution mean (4.9945519) is very close to the theoretical mean (5)

2. Sample Variance versus Theoretical Variance

Let's first calculate the sample variance:

```
# standard deviation of distribution
distribution_standard_deviation <- sd(exponentials_means)
distribution_standard_deviation
```

```
## [1] 0.7882976
```

```
# variance of distribution
sample_variance <- distribution_standard_deviation^2
sample_variance
```

```
## [1] 0.6214131
```

Now let's calculate the theoretical variance:

```
# standard deviation from exponential distribution
theoretical_standard_deviation <- (1/lambda)/sqrt(n)
theoretical_standard_deviation
```

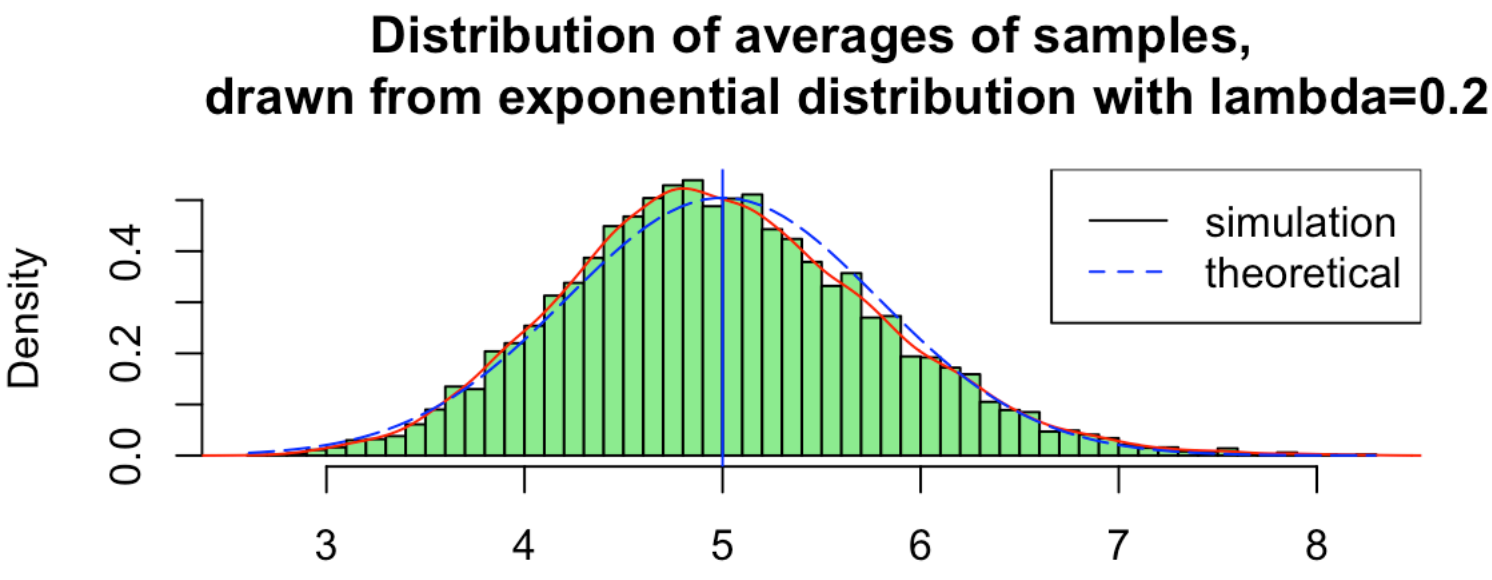
```
## [1] 0.7905694
```

```
# theoretical variance
variance_theory <- ((1/lambda)*(1/sqrt(n)))^2
variance_theory
```

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

The standard deviation of the distribution is 0.7882976 and the theoretical standard distribution is 0.7905694. The theoretical variance is 0.625and the sample variance of 40 exponentials is 0.6214131 that is very close.

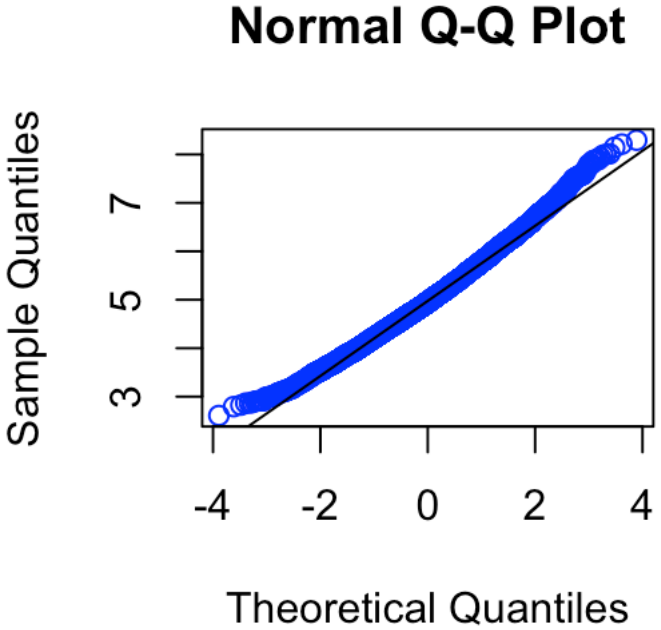
Graphical representation



As shown in the graph, the calculated distribution of means of samples of exponential distributions, overlaps quite nice with the normal distribution.

3. Distribution

The central limit theorem exposes that the averages of samples follow normal distribution. The figure above shows the histogram of the density and the normal deensity with theoretical mean and variance values.



The q-q plot suggests the distribution of averages of 40 exponentials is very close to a normal distribution.