Part A-Exponential Distribution vs Central Limit Theorem

A Lopez

20 October 2015

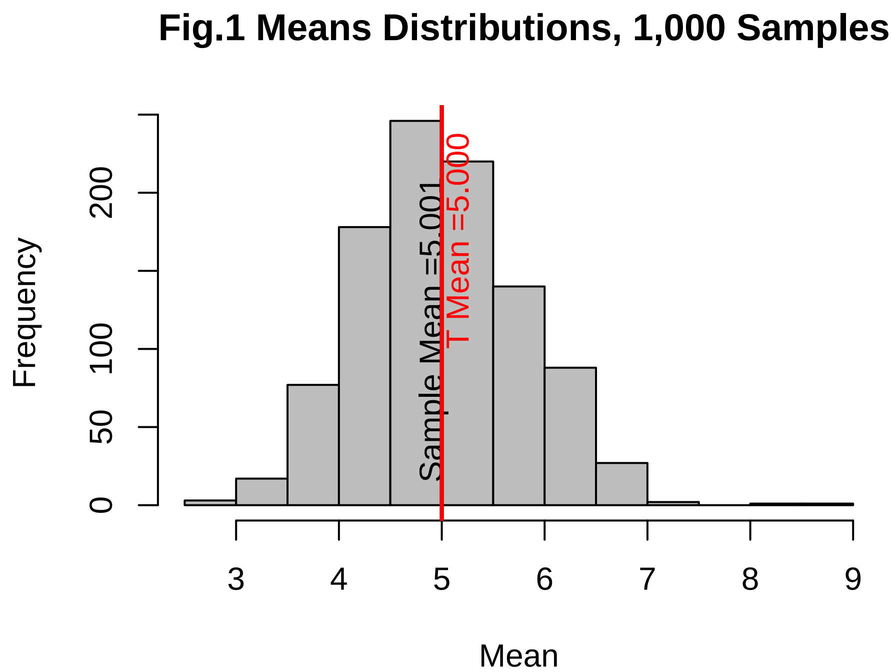
Objective: The aim of this project is to explore the exponential distribution in R and compare it with the Central Limit Theorem. For this project Lambda is set to 0.2 for a thousand simulations. The project considers the distribution of averages of 40 exponentials.

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

Simulations and Plot:

```
# Data: Parameters for the Exponential Distribution
set.seed(900) # Set the seed for random number generation
lambda <- 0.2   # parameter of the distribution, rate parameter
N_sim <- 1:1000 # Number of simulations
N_exp <- 40     # 40 exponentials
#Computations of theoretical and sample mean
means <- data.frame(c = sapply(N_sim, function(c) {mean(rexp(N_exp, lambda))})) #1000 simulations
tMean <- round(1/lambda,4) # mean of the exponential distribution rounded to 4 digits
sample_mean <- round(mean(means$c),4) #sample mean rounded to 4 digits
```

Fig 1 depicts a histogram of the samples and plots both the theoretical and the sample mean.



Analysis: The Sample Mean versus the Theoretical Mean

A) Let μ be the expected theoretical mean, given by: $\mu = \frac{1}{\lambda}$

replacing values we can calculate μ :

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

Let $\bar{X}$ be the average sample mean of 1000 simulations of 40 randomly sampled exponential distributions. Computing the mean of the samples we obtain (see code above) $\bar{X}$:

```
## [1] 5.0001
```

Conclusion: Note that the plot and the calculation shown above verify that the values of the expected and the sample mean are very similar with a negligible difference, proving that the simulations are very accurate.

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

Expected Values:

Let the std deviation of a exponential distribution of rate λ be σ. Therefore: $\sigma = \frac{1/\lambda}{\sqrt{n}}$

σ :

```
## [1] 0.7906
```

And let the variance Var of standard deviation σ be $Var = \sigma^2$

Var :

```
## [1] 0.6250484
```

Sample Calculations:

Let Var_x be the variance of the average sample mean of 1000 simulations of 40 randomly sampled exponential distribution, and sd_x the corresponding standard deviation. # Variance of the simulations:

```
sd_x<-round(sd(means$c),4)# standard deviation
sd_x;
```

```
## [1] 0.7921
```

```
var_x<- round(var(means$c),4)
var_x;
```

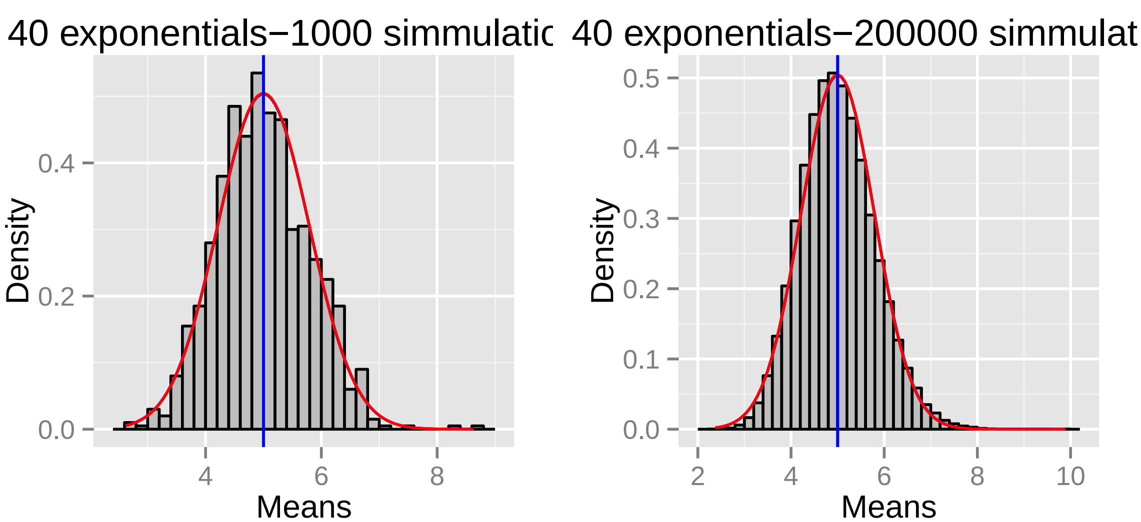
```
## [1] 0.6274
```

Analysis: The standard deviation values are very close, and since the variance is the square of the standard deviations, a negligible difference is expected, demonstrating again the accuracy of the simulations.

3- Show that the distributionis approximately normal

Focus on the difference between the distributions of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

```
## Loading required package: ggplot2
```



Analysis: For lambda=0.2 ,the sample distribution is approximately normal with Mean and Variance values matching the Theoretical mean and variance with a very accurate approximation. To further elaborate, Should we increase the sample size from 1000 to x, eg 200000, the distribution of means shape a bell curve of a normal distribution, therefore concluding that it is approximate normal distribution.

CONCLUSION

Summary Results for Q1, Q2, and Q3

```
##          Expected Value Simulation
## Mean          5.0000000     5.0025
## Std Dev       0.7906000     0.7921
## Variance      0.6250484     0.6274
```

1- We see above the distribution is centered, with the theoretical mean vs sample mean presenting a negligible difference. The simulation method is accurate and yields only a difference of .0001 for a 1000 simulations.
2- The theoretical variance is 0.637 and the sample variance 0.6250, also proving a high degree of accuracy using the sample std deviation and variance.
3- The graph above shows the distribution approximately normal for 1000 samples and continues the same shape for higher numberof samples,demonstrating again a consistent tendency and approximation to normal distribution.