Exponential Distribution vs Central Limit Theorem

Ricardo Fernandez

Overview

We will investigate the exponential distribution in R and compare it with the Central Limit Theorem. For accomplish this purpose we will Illustrated the properties of the distribution by simulation and associated explanatory text and graphics.

Simulation

In order to create the dataset we will run a serie of 1000 simulations with 40 sampling size (observations) and set the exponential distribution to `rexp(40, 0.2)`. The first step will be to define the parameters to generate our sample: `simulations`, `lambda` and `observations`. For be sure that always we generate the same random values we will set the seed.

```
simulations <- 1000; lambda <- 0.2; observations <- 40
set.seed(108)
# generate the sample, size 'simulations' and fill it with the mean of the rexp
sample = rep(0,simulations)
for (i in 1:simulations) sample[i] = mean(rexp(observations,rate=lambda))
```

For an easier overview of the generated data we can check the appendix Figure 1

Sample mean and the theoretical mean of the distribution

As is shown on the following code the value of the sample mean is `5.058108` and the theoretical value is `5`. The mean of exponential distribution and the theoretical mean of the distribution are quite close despite the mean of the sample It is slightly larger.

```
# mean of the sample data
mean_sample <- mean(sample)
> mean_sample
[1] 5.058108

# theoretical mean
theoretical_mean <- 1/lambda
> theoretical_mean
[1] 5
```

Variability, sample variance compared to the theoretical variance of the distribution

As is shown below the value of the sample variance is `0.6435606` and the theoretical is `0.625`, this results implies that the mean from the sample is very close to the theoretical mean of normal data. In this case the difference between both is produced for the behaviour of the random data generated for the exponential distribution.

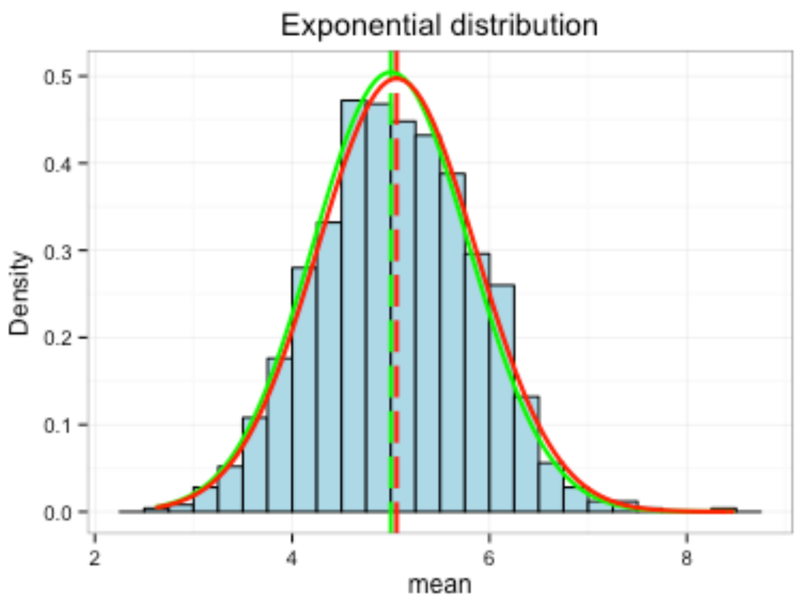
```
# variance for the sample data
sample_variance <- var(sample)
> sample_variance
[1] 0.6435606

# theorical_variance
theoretical_variance <- (1/lambda)^2/observations
> theoretical_variance
[1] 0.625
```

For a graphical view of the sambpled and theoretical mean and the variance we can check the Figure 2 from the Apendix.

Distribution approximately normal

In order to prove that the distribution is approximately normal, first we will create a histogram with a normal distribution and see how fits with the exponential distribution, our sample. Is shown in `red` the theoretical curve and in `green` the generated from the sampled data.



Based on the results shown in the graph and considering both curves we can conclude that the distribution of the sample mean can be considered as a normal distributed.

Confidence interval

We will compare now the 95% confidence interval for both theoretical and sampled data and prove that they are quite close. This result is expected if the exponential distribution from the random sample is equivalent to the theoretical.

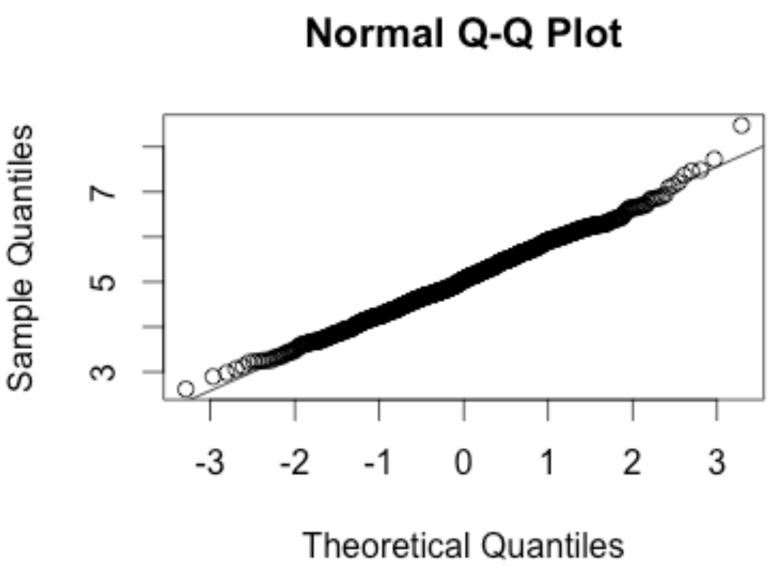
```
sample_ci <- round (mean(sample) + c(-1,1)*1.96*sd(sample)/sqrt(observations),3)
> sample_ci
[1] 4.809 5.307

theoretical_ci <- theoretical_mean + c(-1,1)*1.96*sqrt(theoretical_variance)/sqrt(
observations)
> theoretical_ci
[1] 4.755 5.245
```

Based on the results we can conclude that the confidence interval for the sample mean `[4.809, 5.307]` and the theoretical confidence interval `[4.755, 5.245]` are quite close.

Normal Q-Q Plot Quantiles

Finally the Normal Q-Q plot for quantiles that shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values.



Concusions

Considering the results and all the previously exposed data we can conclude that the exponential distribution is approximately normally distributed and its mean and variance are close enough for consider it equivalent.

Apendix

Figures

Figure 1.

The following figure shows the data generated for the exponential distribution sample.

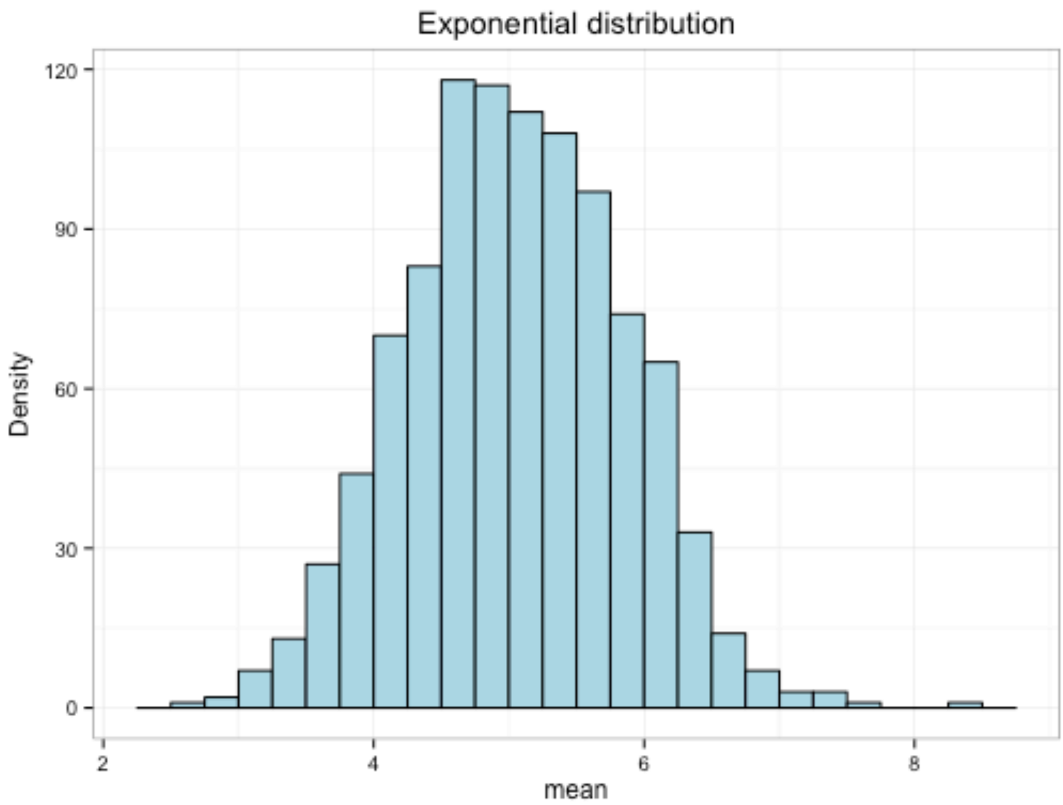
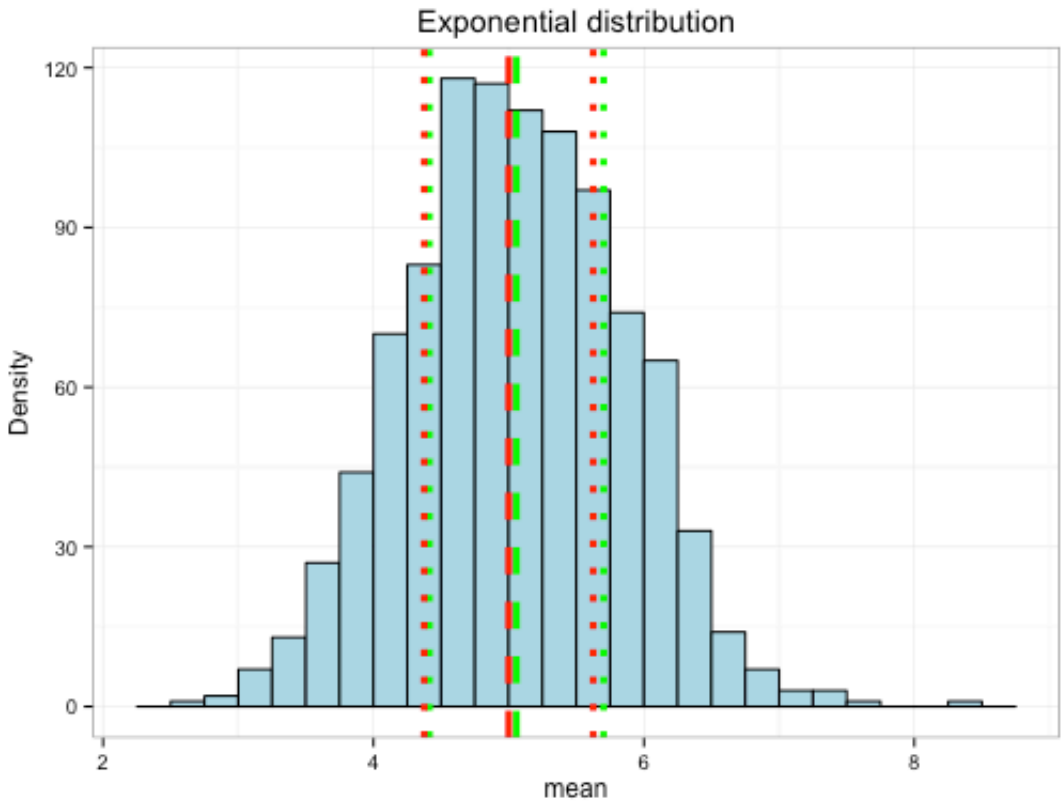


Figure 2.

The following figure shows the mean and variance for both theoretical and random exponential distribution sample.

- In **dashed lines** is shown the **mean**:
    - Red: theoretical
    - Green: sample
- In **doted lines** is shown the **variance**:
    - Red: theoretical
    - Green: sample



Code

```
# ggplot plotting library
library(ggplot2)
# defined values for the simulation
simulations <- 1000; lambda <- 0.2; observations <- 40;
# set the seed so alway are generated the same randome values
set.seed(108)
# generate the sample, size 'simulations' and fill it with the mean of the rexp
sample = rep(0,simulations)
for (i in 1:simulations) sample[i] = mean(rexp(observations,rate=lambda))
# sample plot with histogram chart, bandwidth set between 0.1 - 0.5 show the
# distribution correctly.
ggp <- ggplot() + aes(sample) +
  geom_histogram(binwidth=0.25, colour="black", fill="lightblue") + theme_bw() +
  labs(x = "mean", y = expression("Density")) +
  labs(title=expression("Exponential distribution"))
print(ggp)
## Sample mean and the theoretical mean of the distribution
# mean for the sample data
sample_mean <- mean(sample)
# theoretical mean
theoretical_mean <- 1/lambda
## Variability, sample variance compared to the theoretical variance of the distri
bution
```

```
# variance for the sample data
sample_variance <- var(sample)
# theorical_variance
theoretical_variance <- (1/lambda)^2/observations
ggp <- ggplot() + aes(sample) +
  geom_histogram(binwidth=0.25, colour="black", fill="lightblue") + theme_bw() +
  # vertical lines for Sample and Theoretical Mean and Variance
  geom_vline(xintercept = sample_mean, colour="green", linetype = "dashed", size =
 1.5) +
  geom_vline(xintercept = theoretical_mean, colour="red", linetype = "dashed", siz
e = 1.5) +
  geom_vline(xintercept = sample_mean+sample_variance, colour="green", linetype =
"dotted", size = 1.5) +
  geom_vline(xintercept = sample_mean-sample_variance, colour="green", linetype =
"dotted", size = 1.5) +
  geom_vline(xintercept = theoretical_mean+theoretical_variance, colour="red", lin
etype = "dotted", size = 1.5) +
  geom_vline(xintercept = theoretical_mean-theoretical_variance, colour="red", lin
etype = "dotted", size = 1.5) +
  labs(title=expression("Exponential distribution"), x = "mean", y = expression("D
ensity"))
print(ggp)
## Distribution approximately normalplotdata
# plot the normal distributed and the sample data
plotdata <- data.frame(sample)
ggp <- ggplot(plotdata, aes(x =sample)) + theme_bw() +
  geom_histogram(aes(y=..density..), binwidth=0.25, colour="black", fill = "lightb
lue") +
  geom_vline(xintercept=theoretical_mean,size=1.0, color="red",linetype ="longdash
") +
  geom_vline(xintercept=sample_mean,size=1.0, color="green",linetype = "longdash")
 +
  stat_function(fun=dnorm,args=list( mean=theoretical_mean, sd=sqrt(theoretical_va
riance)),color = "red", size = 1.0) +
  stat_function(fun=dnorm,args=list( mean=sample_mean, sd=sqrt(sample_variance)),c
olor = "green", size = 1.0) +
  labs(title=expression("Exponential distribution"), x = "mean", y = expression("D
ensity"))
print(ggp)
# confidence interval
sample_ci <- round (mean(sample) + c(-1,1)*1.96*sd(sample)/sqrt(observations),3)
theoretical_ci <- theoretical_mean + c(-1,1)*1.96*sqrt(theoretical_variance)/sqrt(
observations)
# Normal Q-Q Plot
qqnorm(sample)
qqline(sample)
```