Statistical Inference Project 1

Francisco Martín

October 8, 2018

Overview

This is the part 1 of the Statistical Inference final project. It consist on a simulation of the exponential distribution using the R function 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. It will be investigated the distribution of averages of 40 exponentials.

Analysis

For reproducibility, seed is going to be set arbitrarily at 1994 (because I born on this year). As statement says, lambda = .2, the number of exponentials will be n_exp=40 and the number of samples will be n_sample=1000.

```
set.seed(1994)
n_exp <- 40
lambda <- .2
n_sample <- 1000
sample <- replicate(n_sample, rexp(n_exp,lambda))
mean_exp <- apply(sample, 2, mean)
```

Sample mean vs theoretical mean

Once we have our sample ready, we can compare mean agains the theoretical one (1/lambda):

```
theo_mean <- 1/lambda
sample_mean <- mean(mean_exp)
```

Theoretical mean is 5 and sample mean is 5.0118366. As we can see difference is about 0.2%.

Sample vs Theoretical: Standard Deviation and Variance

Once we have seen how close are sample and theoretical means, next step is to calculate standard deviation and variance for the sample and compare with theoretical ones:

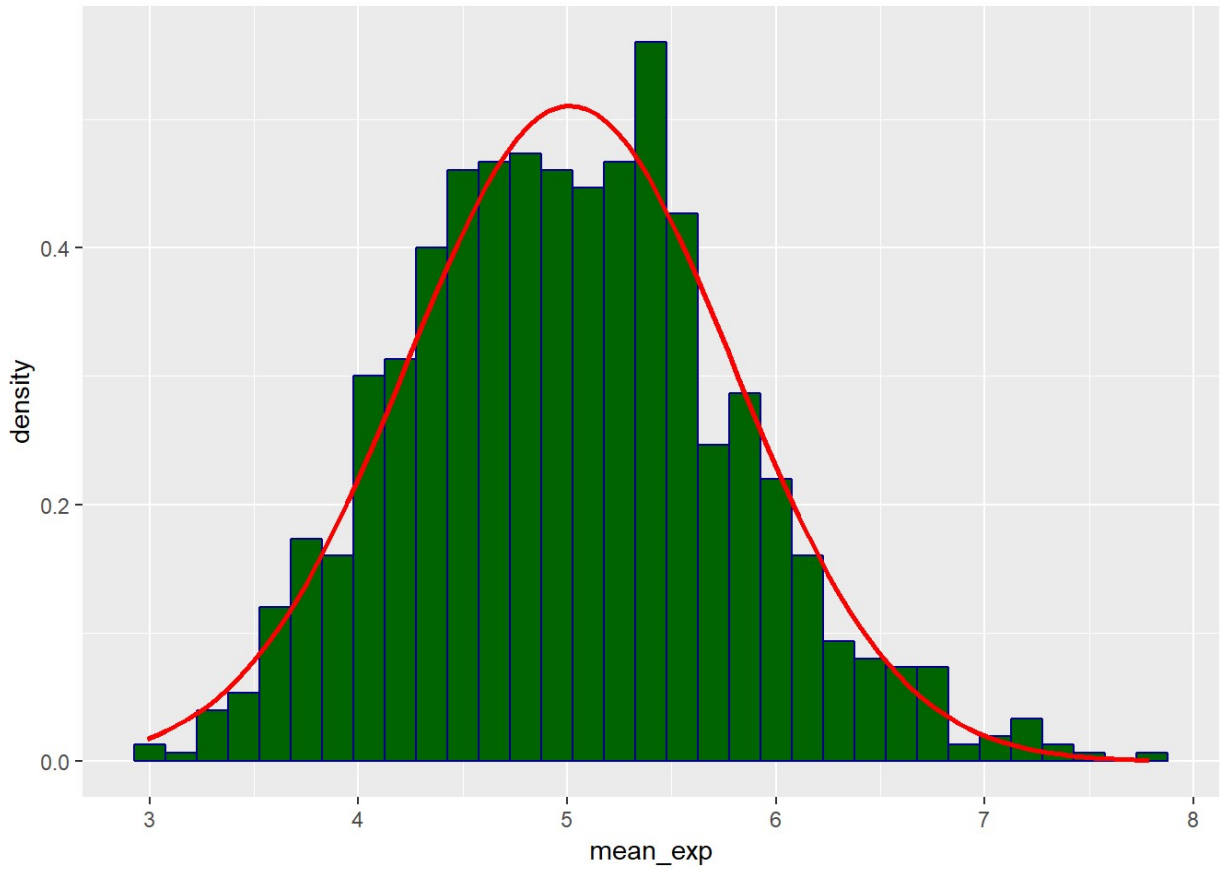
```
theo_sd <- 1/(lambda*sqrt(n_exp))
theo_var <- ((1/lambda)^2)/n_exp
sample_sd <- sd(mean_exp)
sample_var <- var(mean_exp)
```

Just compare: theoretical standard deviation is exactly 0.7905694 and sample's one is 'r sample_sd'. In case of variances, theoretical's one is 0.625 and sample's one is 0.6100856. In case of standard deviation, difference is about 1% and in case of variance it is about 2%, showing their values are pretty close.

Is the distribution approximately normal?

Finally, this is the question we want to answer. Based on Central Limit Theorem, answer must be yes. Let's show it.

```
library (ggplot2)
mean_plot <- data.frame(mean_exp)
g <- ggplot(mean_plot, aes(x = mean_exp))
g + geom_histogram(aes (y = ..density..),colour = "dark blue", fill = "dark gree
n", binwidth = .15) +
        stat_function(fun = dnorm, args = list(mean = sample_mean, sd = sample_s
d), color = "red", size = 1.0)
```



Red line corresponds to a normal distribution with sample_mean and sample_sd. As we can see, this distribution is approximately normal, in agreement with CLT.