Exponential distribution analysis

Georgy Makarov

April 24, 2020

Overview

This is the part one of the course project of Statistical Inference course. It concerns the questions of exponential distribution analysis. The body of the project describes the calculations and draw conclusions. There are supporting plots in appendix to this project.

Simulations

We run the simulations of exponential distributions with *rexp* function. We setup the parameters according to the instructions given during the course. The lambda is 0.2, sample size is 40, the number of repeats is 1000. We find the mean of replicated distributions and store them as a vector. We set the seed in order to be able to reproduce the data.

```
# set the parameters
lambda <- 0.2
n <- 40
sim <- 1000

# simulate distributions and find means
set.seed(2020)
mean_sim <- replicate(sim, mean(rexp(n, lambda)), simplify = TRUE)
summary(mean_sim)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.796   4.487   5.028   5.034   5.509   7.821
```

Sample mean vs Theoretical mean

Summary of simulations shows that the observed mean is **5.034**. We obtain theoretical mean of exponential distribution from the following equation:

$$\mu = \frac{1}{\lambda}$$

Theoretical mean for our λ is **5.000**. It is obtained from the equation given above:

$$\mu = \frac{1}{0.2} = 5.000$$

The histogram of the simulations and both the observed and theoretical means are given in appendix in *Figure 1.* The observed mean agrees well with the theoretical mean.

Sample variance vs Theoretical variance

Evaluation of variance requires the standard deviation. Theoretical standard deviation is expected to be as follows:

$$sd = \frac{1}{\lambda\sqrt{n}}$$

For our λ and n theoretical standard deviation is **0.791**. We get it from the quation given above:

$$sd = \frac{1}{0.2\sqrt{40}} = 0.791$$

Observed standard deviation is obtained from the code chunk below. Both variances are very close.

```
sd_t <- round((1 / lambda) / sqrt(n), 3) # theoretical sd
sd_o <- round(sd(mean_sim), 3) # observed sd
var_t <- round(((1 / lambda) / sqrt(n))^2, 3) # theoretical variance
var_o <- round(sd(mean_sim)^2, 3) # observed variance
std_dev <- c(sd_t, sd_o)
variance <- c(var_t, var_o)
q2 <- cbind(std_dev, variance)
rownames(q2) <- c("theoretical", "observed")
q2
```

```
##             std_dev variance
## theoretical   0.791    0.625
## observed      0.779    0.607
```

Distributions

We check that sample distribution is close to normal visually. For the purpose of visual comparison we need the parameters of normal distribution. Those are created by the following chunk:

```
x_norm <- seq(min(mean_sim), max(mean_sim), length = 2*n)
y_norm <- dnorm(x_norm, mean = 1 / lambda, sd = (1 / lambda) / sqrt(n))
```

Visual comparison of two distributions in *Figure 2* shows that sample distribution is a little skewed but still very close to normal. According to *CLT* the form of the distribution becomes normal if we have increased number of samples.

Appendix

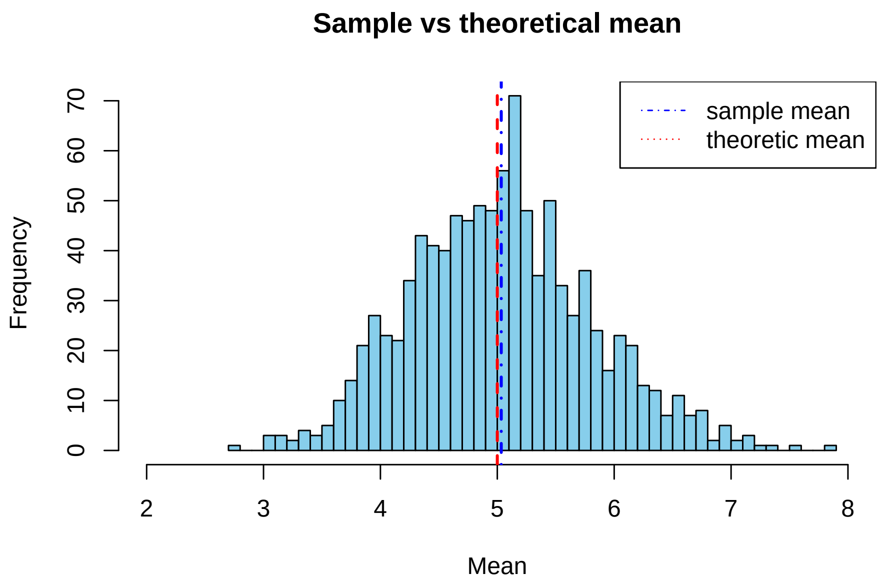

Figure 1. Sample mean vs Theoretical mean

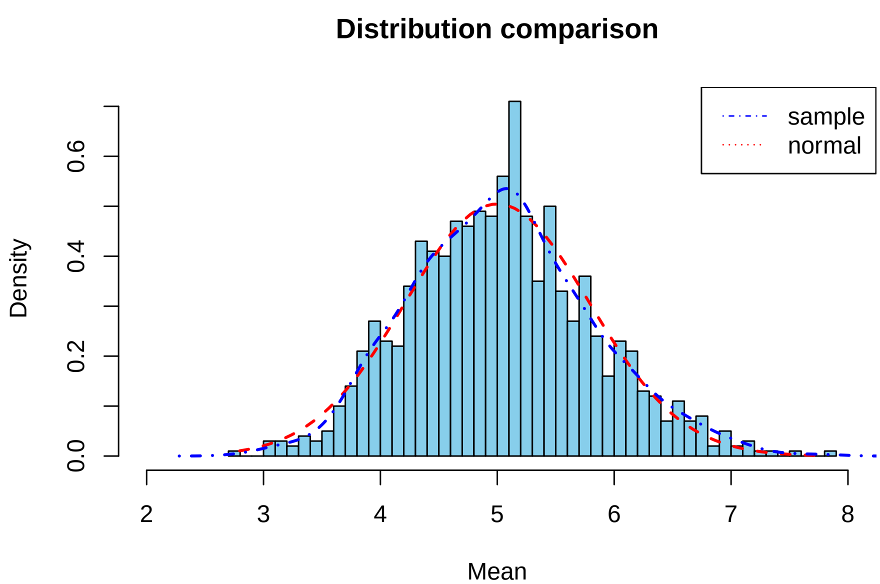

Figure 2. Sample and normal distributions