Peer-graded Assignment: Statistical Inference Course Project

Kirill Setdekov

13 may 2019

Overview

In the first part of this assignment, I will be exploring results of the simulation of exponential distribution and properties of the sample mean. Below are results of the simulations and comparisons of means and variance as well as tests for normality.

Part 1. A simulation exercise.

Simulations:

I simulate 1000 means of 40 exponential distribution and save them in a vector **mns**.

```
#load libraries silently
require(knitr)
require(ggplot2)
require(ggpubr)
require(gridExtra)
require(dplyr)
```

```
set.seed(42)
n <- 40
lambda <- 0.2
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(n, lambda)))
```

Sample Mean versus Theoretical Mean:

Comparison of Sample mean an theoretical mean in the table below:

```
theory_mean <- 1/lambda
sample_mean <- mean(mns)
kable(cbind(theory_mean,sample_mean), digits = 3)
```

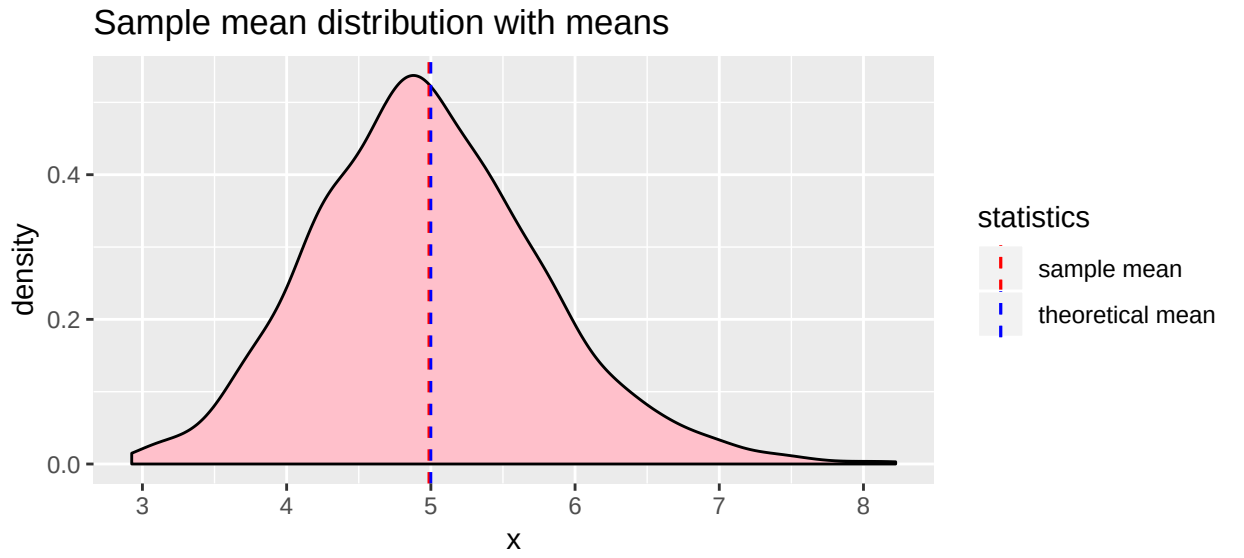
theory_mean	sample_mean
5	4.987

They are very close.

For more detail, I plot the distribution of simulated means together with lines for sample and theoretical mesns below:

```
ggplot(data.frame(x = mns), aes(x = x)) +
geom_density(fill = "pink") +
geom_vline(aes(xintercept = mean(mns),color = "sample mean"),
           linetype="dashed") +
```

```
geom_vline(aes(xintercept = 1 / lambda,color = "theoretical mean"),
           linetype="dashed")+
scale_color_manual(name = "statistics",
                   values = c('sample mean' = "red", 'theoretical mean' = "blue")) +
labs(title = "Sample mean distribution with means")
```



Sample Variance versus Theoretical Variance:

Below is a calculation and comparison of sample variance and theoretical variance. They are close as the number of simulations (1000) is significant.

```
theory_variance <- (1/lambda)/sqrt(40)
sample_variance <- sd(mns)
kable(cbind(theory_variance,sample_variance), digits = 3)
```

theory_variance	sample_variance
0.791	0.797

Distribution

Based on the **Shapiro-Wilk's test** the distribution of the means of 40 exponential random variables are different from normal as p-value<0.05.

```
shapiro.test(mns)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mns
## W = 0.99084, p-value = 6.987e-06
```

Also, looking at the **qq-plot** below, we can see, that the right tail of the distribution is significantly different from the normal one. We have a large sample and we can not assume normality.

```
ggqqplot(mns)
```