Statistical Inference - Final Project: Part 1 Simulation Exercise - Exponential Distribution with R

Anastasia Thöns

December 26, 2016

Introduction

In the following the exponential distribution is investigated in R and compared with the Central Limit Theorem. The **exponential distribution** is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate, with a probability density function (pdf):

$$f(x) = \lambda e^{-\lambda x}$$

where λ is the rate parameter and $x > 0$. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$.

The **central limit theorem** states that the sampling distribution of the mean and the variance of any independent, random variable will be normal or nearly normal, if the sample size is large enough. How large is "large enough"? The answer depends on two factors.

- Requirements for accuracy. The more closely the sampling distribution needs to resemble a normal distribution, the more sample points will be required.
- The shape of the underlying population. The more closely the original population resembles a normal distribution, the fewer sample points will be required.

In practice a sample size of 30 is large enough when the population distribution is roughly bell-shaped. Others recommend a sample size of at least 40. But if the original population is distinctly not normal (e.g., is badly skewed, has multiple peaks, and/or has outliers), researchers like the sample size to be even larger.

The project aims to illustrate the above based on the exponential distribution.

Implementation in R

The exponential distribution is simulated in R with the function `rexp(n, lambda)` where `lambda` is the rate parameter.

The value of lambda for all of the simulations is set to $\lambda = 0.2$.

```
lambda = 0.2
```

The distribution of averages of 40 exponentials is investigated. In total 1000 simulations are perfomed. The sampling distribution of the mean is obtained by

```
mns.exp = NULL
for (i in 1 : 1000) mns.exp = c(mns.exp, mean(rexp(40, lambda)))
```

and its distribution is presented in the following histogram where the red line indicate the theoretical mean of exponential distribution $\mu = 1/\lambda = 5$. We can see that the mean sample mean is

```
mean(mns.exp)
```

```
## [1] 4.996388
```

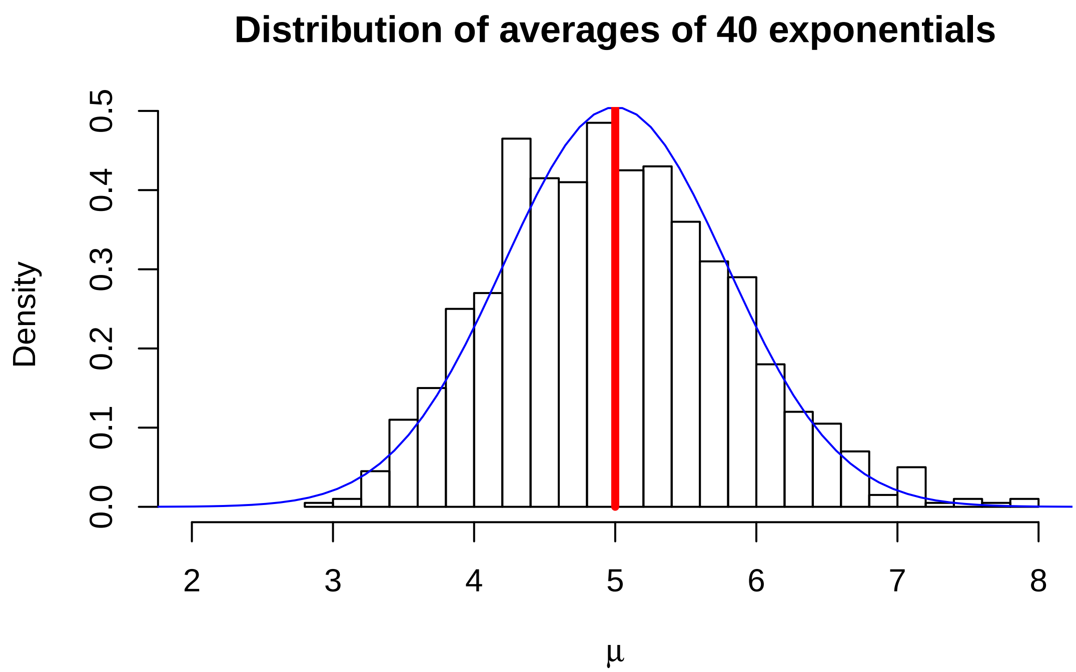
and its standard deviation is

```
sd(mns.exp)
```

```
## [1] 0.8302554
```

The blue line on the graphic represents normal distribution `hx` with $\mu = 1/\lambda = 5$ and $\sigma = 1/\lambda/\sqrt{N} = 0.79$.

```
x <- seq(0, 10, length=100)
hx <- dnorm(x,  mean = 1/lambda, sd = 1/lambda/sqrt(40))
```



The sampling distribution of the standard deviation is obtained by

```
mns.var = NULL
for (i in 1 : 1000) mns.var = c(mns.var, sd(rexp(40, lambda)))
```

and its distribution is presented in the histogram below. Here again the red line indicate the theoretical standard deviation of exponential distribution $\sigma = 1/\lambda = 5$ and mean of sample variance is

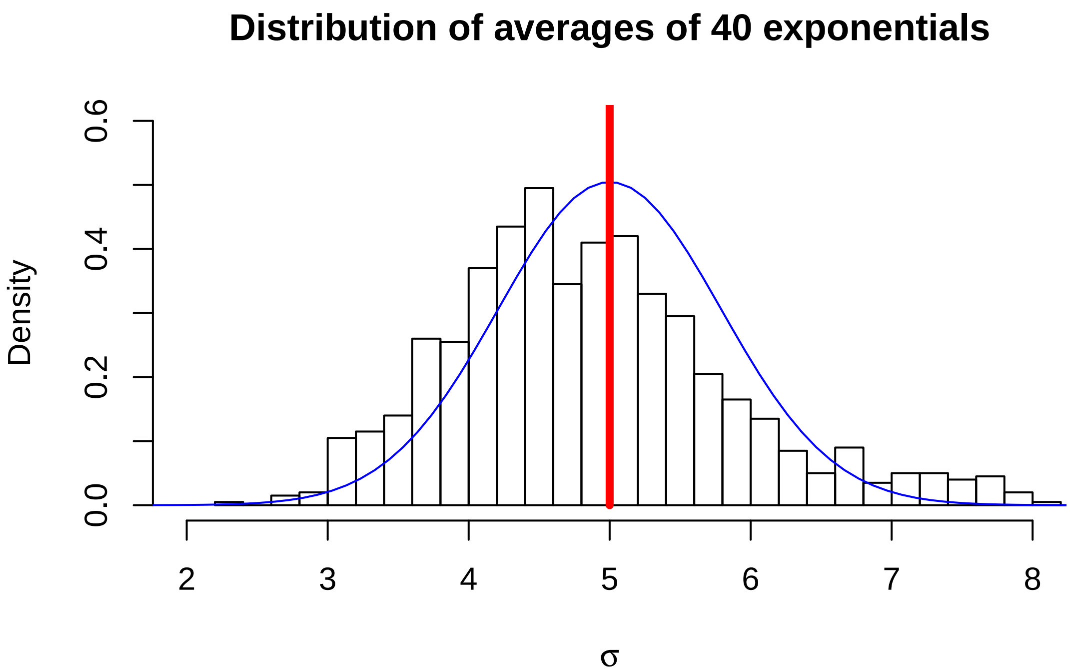
```
mean(mns.var)
```

```
## [1] 4.867591
```

and its standard deviation is

```
sd(mns.var)
```

```
## [1] 1.012526
```



The blue line on the graphic represents normal distribution `hx` as above.

Conclusions

From the graphics and values on the samples we can see that the distribution of sample mean as well as sample variance are approximately normal. The obtained values are:

	Mean	Standart deviation
Theoretical values	5.00	0.79
Sample mean	5.00	0.83
Sample variance	4.87	1.01

Used sources

1. Stat Trek: http://stattrek.com
2. WikipediA: https://en.wikipedia.org/wiki/Central_limit_theorem
3. Wolfram MathWorld: http://mathworld.wolfram.com/ExponentialDistribution.html