Investigation of exponential distribution with R

Roman Preobrazhensky

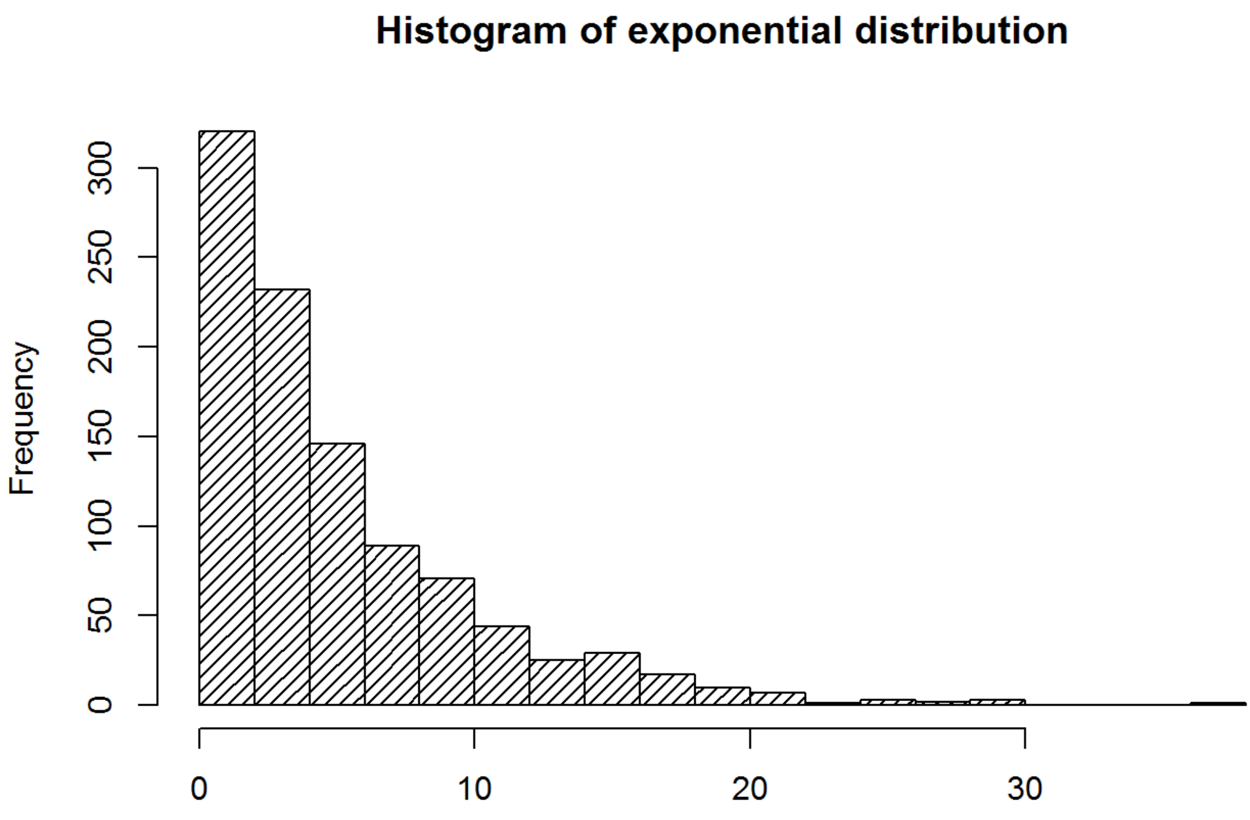
Overview

In this document we will investigate the distribution of the mean of 40 exponentials using Central Limit Theorem (CLT) and Law of Large Numbers (LLN). We will calculate its mean and variance and compare them with corresponding theoretical values.

Exponential distribution

For this study we will use $\lambda = 0.2$ as a rate parameter for exponentials. Let's first take a look at exponential distribution itself. Concretely, at a histogram of 1000 exponentials.

```
hist(rexp(1000, rate=0.2), main = "Histogram of exponential distribution", xlab = NULL,
density = 20, breaks = 20)
```



Note that it is far from being bell-shaped. According to definition of exponential distribution, its theoretical mean is $M = 1/\lambda = 5$ and its variance is $Var = 1/\lambda^2 = 25$.

Mean and variance

The Law of Large Numbers tells us that mean of a large number of samples from the distribution should be close to the theoretical mean of this distribution. Let's create a large (1000 observations) sample of means of 40 exponentials and look at its mean value.

```
means <- NULL
for (i in 1:1000) means <- c(means, mean(rexp(40, rate=0.2)))
mean(means)
```

```
## [1] 5.0157
```

As one can see, sample mean is fairly close to the theoretical mean of the exponential distribution. Now let's take a look at variance of distribution of means.

```
var(means)
```

```
## [1] 0.6329684
```

We can see that variance of sample mean is much smaller that variance of population. However, we can estimate sample variance from population variance. According to CLT, sample standard deviation should be equal to standard error of the mean $SD = \sigma/\sqrt{n}$ and theoretical variance of sample means should be σ^2/n, where $\sigma^2 = 1/\lambda^2$ is a variance of population. So, theoretical variance of sample means is $1/\lambda^2/40 = 0.625$ and our estimated variance is quite close to this value.

Let's show that the average of variances of random samples is a good estimator for a population variance. For this purpose we will create a set of variances of 1000 random samples of 40 exponentials and take its mean.

```
vars <- NULL
for (i in 1:1000) vars <- c(vars, var(rexp(40, rate=0.2)))
mean(vars)
```

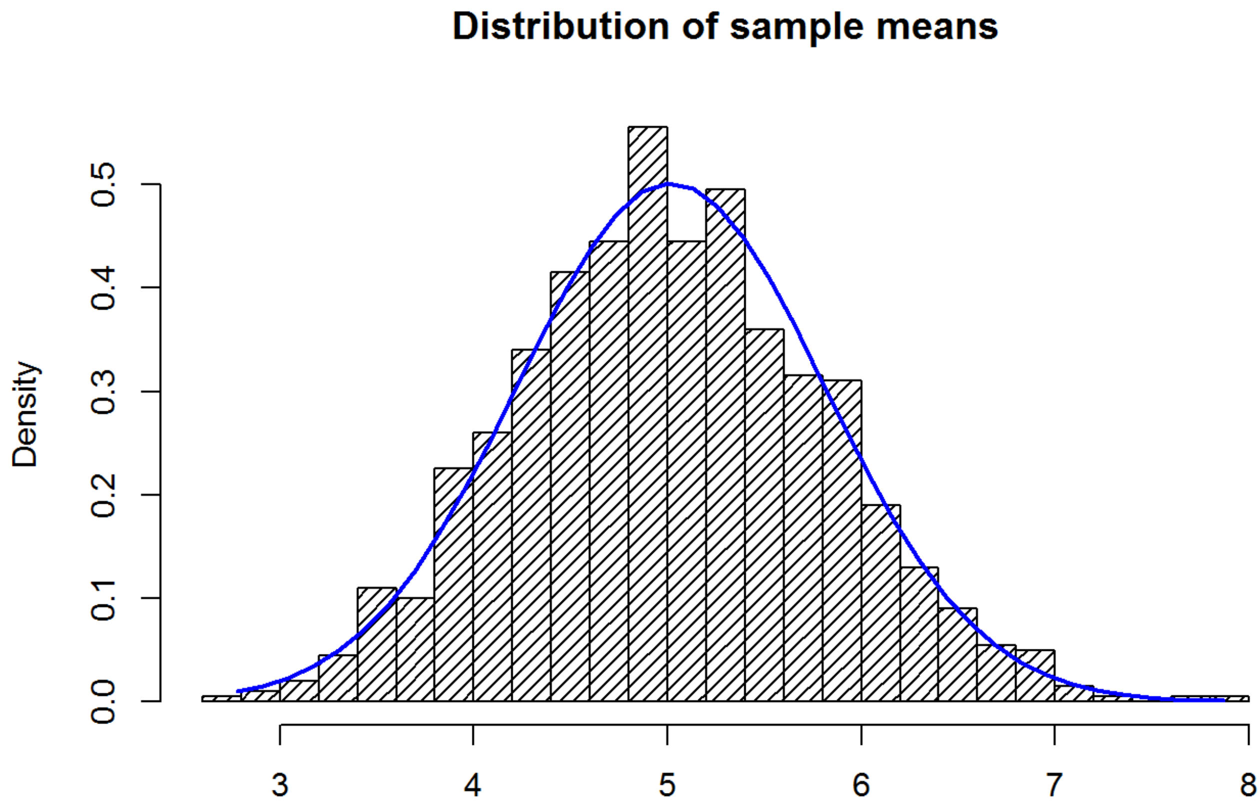
```
## [1] 25.10691
```

As one can see, this is close to theoretical variance of exponential distribution, and hence sample variance if confirmed to be a good estimator of population variance.

Shape of the distribution

Let's now discuss CLT. It states that distribution of sample means should have be close to a normal distribution. Well, since we already have such sample distribution (in the variable `means`), we can plot a histogram of it and overlay this histogram with a normal density plot to make sure that CLT is true. We will calculate our sample distribution mean and standard deviation to create an appropriate density plot.

```
m <- mean(means)
std <- sd(means)
h <- hist(means, density = 20, breaks = 25, prob=T, xlab = NULL, main = "Distribution o
f sample means", freq = F)
xfit <- seq(min(means), max(means), length=40)
yfit <- dnorm(xfit, mean = m, sd = std)
lines(xfit, yfit, col="blue", lwd = 2)
```



As one can see, the distribution is very close to the plotted bell-shaped curve of the normal distribution and it confirms the CLT.

Conclusions

We've shown that in conformity with LLN the mean of a distribution of sample means is very close to the mean of population distribution, and variance of sample means distribution is close to its theoretical estimate too, although it is very different from variance of exponential distribution. We have also confirmed the CLT by showing that the distribution of sample means is very close to a normal distribution.