Analysis of exponential distribution and comparing it with Central Limit Theorem

Elena Tikalenko

Overview

Exponential distribution with $\lambda = 0.2$ has the following probability density function: $f_X(x) = 0.2 * e^{-0.2x}$, $x \geq 0$

In further investigation we will illustrate Central Limit Theorem in action: we will investigate the distribution of averages of 40 exponentials and show that it's like that of standard normal distribution.

Note that:

- the mean of exponential distribution $\mu = E[X_i] = 1/\lambda = 5$
- standart deviation $\sigma = 1/\lambda = 5$

Comparing the sample mean and theoretical mean

Theoretical mean for the exponential distribution is $\frac{1}{\lambda} = 5$. Let's calculate the sample mean. For that let's do thousand simulations for averages of 40 exponentials.

```
# For reproducible results
set.seed(1)

lambda <- 0.2

# The number of simulations
nosim <- 1000

# Sample size
n <- 40

# Generate matrix with 1000 rows and 40 columns and then calculate mean of each column
dat <- apply(matrix(data = rexp(nosim * n, rate = lambda), nrow = nosim), 1, mean)
# The sample mean
m <- mean(dat)
```

The sample mean is 4.9900252. In this way we can see that sample mean is quiet close to theoretical mean with great number of simulations (see also the plot "Distribution of sample means" at the end of documents, where the sample and theoretical means marked by blue and red vertical lines).

Comparing the sample variance and theoretical variance

Let's calculate the sample variance.

```
var <- var(dat)
```

The sample variance is 0.6177072.

The theoretical variance is $\frac{\sigma^2}{n} = 0.625$. It's a very close to sample variance.

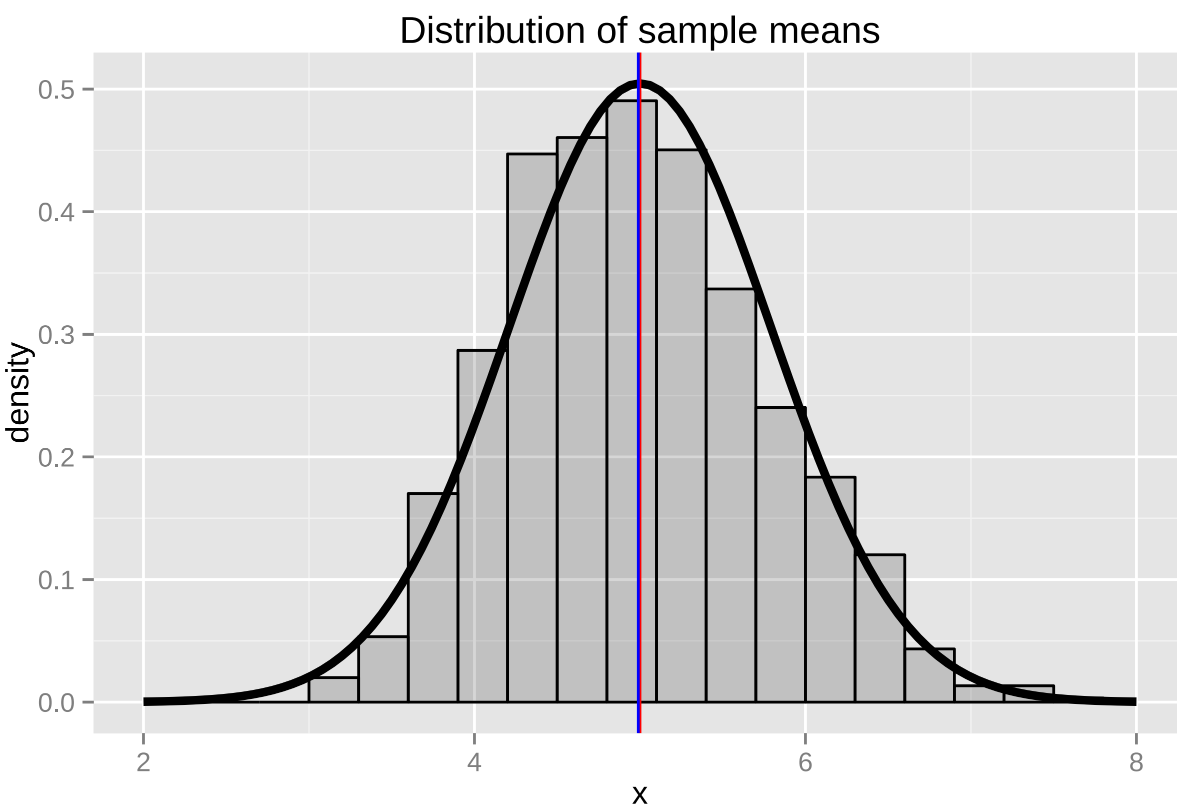
Comparing the distribution of sample mean and normal distribution

Let's create a plot of distribution of sample mean and compare it with normal distribution with $\mu = 5$ and $sd = \frac{\sigma}{\sqrt{n}}$

```
# Convert data for plot to data.frame
dat <- data.frame(x = dat)

library(ggplot2)

# Create plot of distribution of sample means
g <- ggplot(dat, aes(x = x)) +
        geom_histogram(alpha = .20, binwidth=.3, colour = "black", aes(y = ..density..)) +
        xlim(c(2, 8)) + labs(title = "Distribution of sample means")
g <- g + stat_function(fun = dnorm, size = 1.5, args = list(mean = 1/lambda, sd = 1/lambda/sqrt(n)))
g <- g + geom_vline(aes(xintercept = 5, colour = "Mean")) +
        geom_vline(aes(xintercept = m, colour = "Sample Mean")) +
        scale_colour_manual("", breaks=c("Mean", "Sample Mean"),
                            values=c("Mean" = "red","Sample Mean" = "blue"))
g
```



The parent population was an exponential distribution. For n = 40 the distribution of sample mean is quite close to a normal distribution with mean $\mu = \frac{1}{\lambda} = 5$ and standard deviation s $= \frac{\sigma}{\sqrt{n}} = 0.7859435$. The mean of normal distribution marked by red line, the mean of distribution of the sample mean marked by blue line. It's clear that both means are approximately the same.