Coursera: Statistical Inference - Course Project

Konstantin Greger

2015/01/24

Part 1: The Exponential Distribution in R in Comparison with the Central Limit Theorem

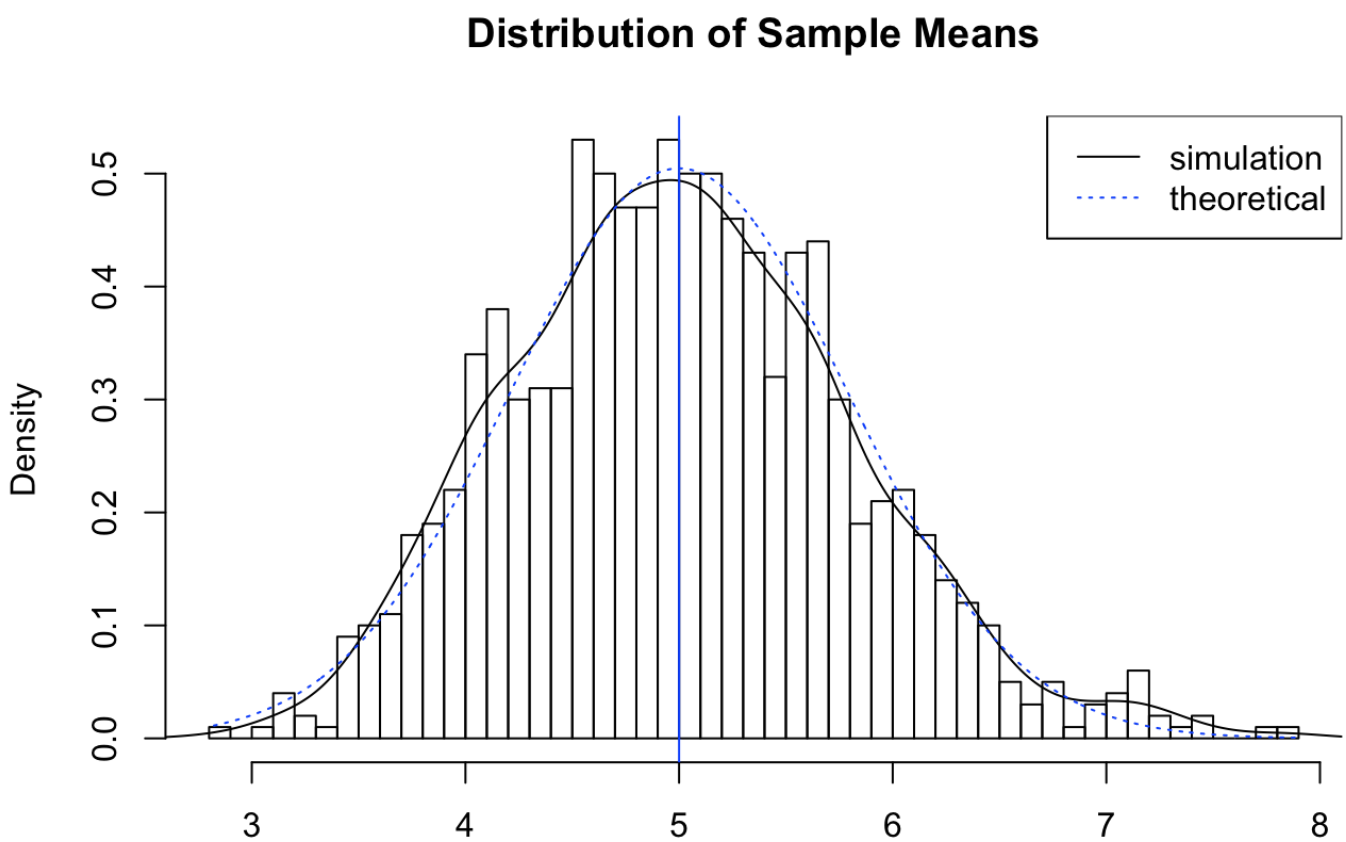
In R the exponential distribution can be simulated using the command `rexp(n, lambda)` with `lambda` \(\lambda\) being the rate parameter. For the purpose of this simulation we set \(\lambda=0.2\).

1,000 averages of 40 exponentials can be simulated as follows:

```
# always set ranodomizer seed for reproducible results
set.seed(281179)
# define simulation parameters
lambda <- .2
numberOfSimulations <- 1000
sampleSize <- 40
# run simulation
data <- matrix(rexp(numberOfSimulations * sampleSize, rate = lambda),
               numberOfSimulations,
               sampleSize)
```

The distribution of the means of each sample of 40 randoms from the exponential distribution is distributed as follows:

```
# calculate the sample means
rowMeans <- rowMeans(data)
# plot the histogram of the sample means
hist(rowMeans,
     breaks = 50,
     prob = TRUE,
     main = "Distribution of Sample Means",
     xlab = "")
# add the density of the sample means
lines(density(rowMeans))
# add the theoretical center of distribution
abline(v = 1/lambda, col = "blue")
# add the theoretical density of the sample means
xfit <- seq(min(rowMeans), max(rowMeans), length = numberOfSimulations)
yfit <- dnorm(xfit, mean = 1 / lambda, sd = (1 / lambda / sqrt(sampleSize)))
lines(xfit, yfit, pch = 22, col = "blue", lty = 3)
# add legend
legend('topright',
       c("simulation", "theoretical"),
       lty = c(1, 3),
       col = c("black", "blue"))
```



The distribution of the simulated sample means is centered at 5.0079, the theoretical mean of the exponential distribution is \(\lambda^{-1}\) = 5. The variance of the simulated sample means is 0.6373 as opposed to the theoretical variance of the exponential distribution of \(\sigma^2 / n = 1/(\lambda^2 n) =

1/(0.04 \times 40)\) = 0.625.

The Central Limit Theorem states that the means of samples follow a normal distribution. For comparison the figure above also shows the density computed using the histogram and the normal density plotted resulting from the theoretical mean and variance values in blue.

Another way to visualize this is the q-q plot below, which in this case suggests the normality of the simulation distribution:

```
# create a qq plot
qqnorm(rowMeans)
qqline(rowMeans)
```