Statistical Inference Project Part 1

Li Xiaowei

Overview

This report is the part 1 of the statistical inference project. We'll explore the exponential distribution in this report.

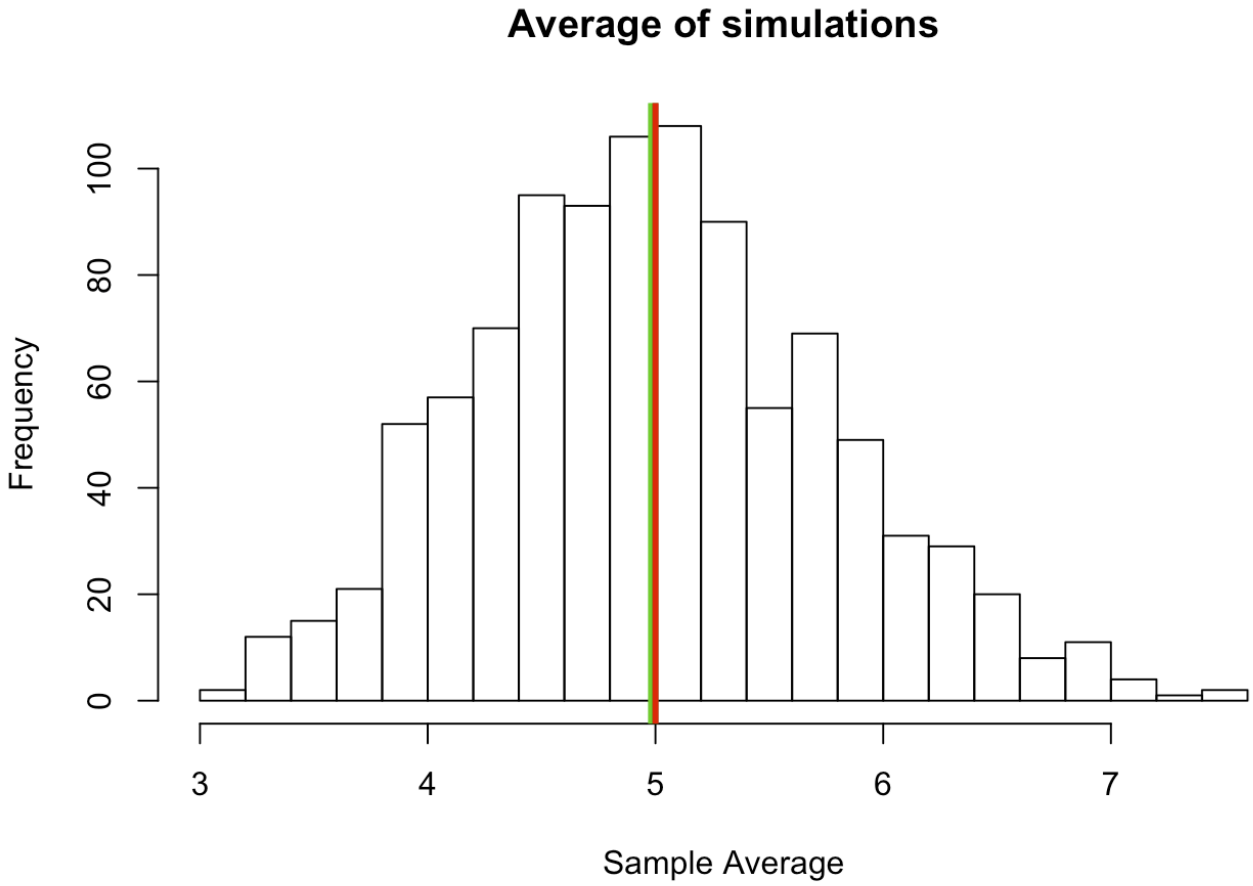
Exponential distribution

The mean and standard deviation of an exponential distribution are both 1/rate parameter. In this report we use lambda = 0.2 to represent the rate parameter. We will invetigate the distribution of averages of 40 exponentials, with 1000 simulations.

Sample mean compare to theoretical mean

The expected value of the mean is 1/0.2, which is **5**, shown as the vertical red line. The distribution is centred at **4.99** as shown by the green vertical line, which is almost equal to the expected value. In below plot I made the green line thicker in order to show both lines.

```
set.seed(1)
lambda = 0.2
nosim <- 1000
n <- 40
exp =  NULL
for (i in 1:1000) exp = rbind(exp, rexp(n, lambda))
expMean <- apply(exp, 1, mean)
hist(expMean, main = "Average of simulations", xlab = "Sample Average", breaks
= 30)
abline(v = mean(expMean), col = 3, lwd = 5)
abline(v = 1/lambda, col = 2, lwd = 3)
```



Variance of sample means compare to theoretical variance by CLT

The standard deviation of the exponential distribution is 1/0.2 = 5. By the central limit theorem, the expected standard deviation of the sample means should be 5/sqrt(40) = **0.7905694**. It's very close to **0.7817394** which can be calculated by below code.

```
sd(expMean)
```

```
## [1] 0.7817394
```

We have made 1000 simulations which is considered as a quite large number of sampling. That's why we have the variance (or standard deviation) of sample means quite close to the expected variance (or standard deviation).

The exponential distribution is approximately normal

The black line in below plot is the sample mean distribution, the grey line is the normal distribution line with mean of 5 and expect standard deviation as depicted in previous part. It's easily seen that they are quite close to each other, meaning the sample mean of the exponential distribution is approxiately normal.

```
dst <- density(expMean)
plot(dst$x, dst$y, type = "n", main = "The distribution of sample mean", xlab =
"Sample Mean", ylab = "Density")
lines(seq(3, 7, length = 100), dnorm(seq(3, 7, length = 100), mean = 5, sd =
(1/lambda)/sqrt(n)), col = grey(.8), lwd = 3)
lines(dst$x, dst$y, lwd = 2)
```