Statistical Inference: Peer Assessment 1

Exponential distribution

Eddy Delta

01/09/2018

Simulation Exercise Instructions

```
set.seed(9)
lambda <- 0.2
sample_count <- 40
simulation_count <- 1000
```

The goal of this project is to investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with `rexp(n, lambda)` where **lambda** is the rate parameter. The *mean* of exponential distribution is $1/lambda$ and the *standard deviation* is also $1/lambda$. $lambda = 0.2$ is set for all of the simulations. The distribution of averages of 40 exponentials will be analysed within a set of 1000 simulations.

Sample mean and compared to the theoretical mean of the distribution

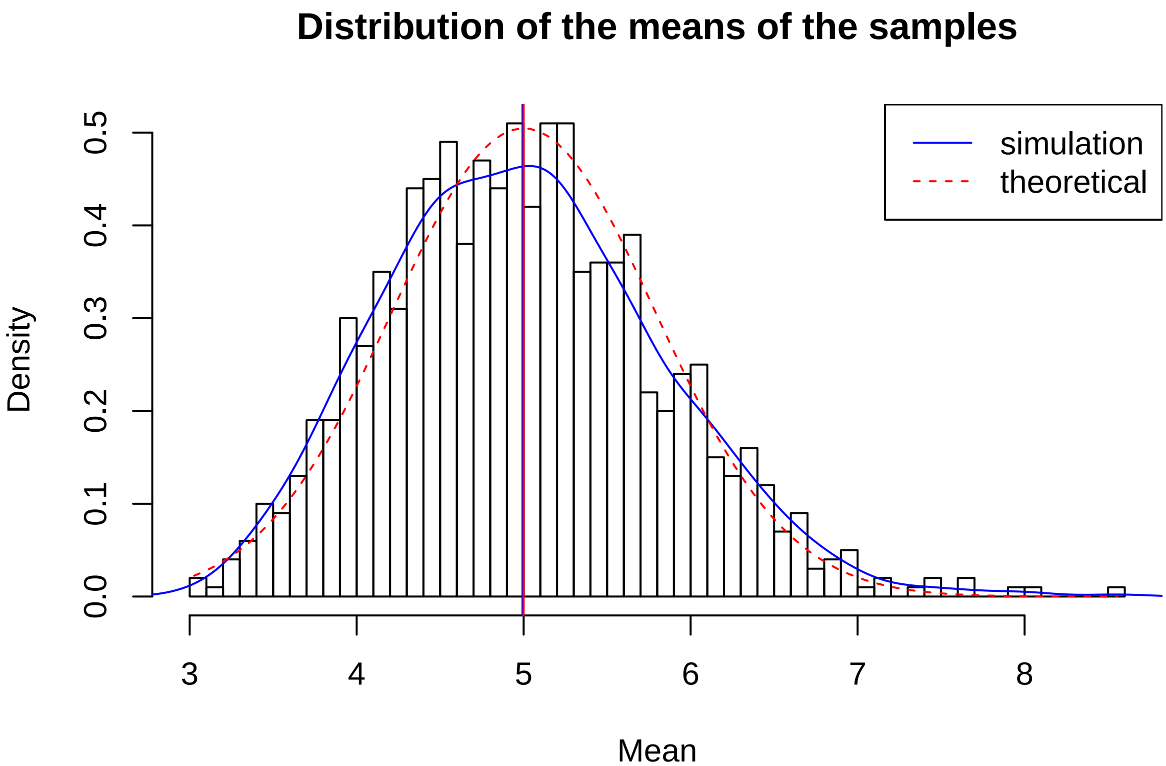
The first step is to create a dataset containing 1000 of 40 numbers generated randomly by the `rexp(n, lambda)` function. This data set is stored in a 1000 x 40 `matrix`. Out of this dataset a vector containing the `mean` of each simulation will be generated.

```
simulations <- matrix(
    rexp(simulation_count * sample_count, rate = lambda),
    simulation_count,
    sample_count)
simulations_means <- apply(simulations, 1, mean)
simulations_mean <- mean(simulations_means)
```

The exponentials are distributed as follow

```
# histogram of the means
hist(simulations_means, breaks = 50, prob = TRUE,
     main = 'Distribution of the means of the samples',
     xlab = 'Mean')
# density of the simulations means
lines(density(simulations_means), col = 'blue')
# center of the simulated distribution
abline(v = simulations_mean, col = 'blue')
# center of theoretical distribution
abline(v = 1 / lambda, col = 'red')
# theoretical density of the means of samples
x <- seq(min(simulations_means), max(simulations_means), length = 100)
y <- dnorm(x, mean = 1 / lambda, sd = (1 / (lambda * sqrt(sample_count))))
```

```
lines(x, y, pch = 22, col = 'red', lty = 2)
legend('topright', c("simulation", "theoretical"), lty = c(1, 2), col = c('blue', 'red'))
```



The distribution of the sample means is centered around 4.993 which is quite close to theoretical center of the mean distribution 5 given by $1/lambda$.

The distribution of the sample means also have a *variance* of 0.683 which isn't far from the theoretical *variance* of $\frac{\sigma^2}{n} = \frac{1/lambda^2}{n} = \frac{1}{lambda^2 \times n} = \frac{1}{0.2^2 \times 40} = 0.625$

In accordance with the Central Limit Theorem the distribution of the sample means follow the normal distribution. We can observe in the histogram above the convergence of the samples means distribution to the normal distribution.

Confidence interval

$H_0 : \mu \neq \bar{X}$ is defined as our null hypothesis and $H_A : \mu = \bar{X}$ our Alternative hypothesis.

The 95% confidence interval for the sample distribution is given by $\mu = \bar{X} \pm Z_{0.95} \times \frac{S}{\sqrt{n}} = 4.993 \pm 1.96 \times \frac{0.826}{31.623}$ which gives a value between 4.942 and 5.044

The average of samples means falls inside our confidence interval 95% of the time, so the H_0 can be rejected.

Conclusion

The means distributions of 40 randomly observed exponentials simulated 1000 times of is quite close to distribution of the means of the theoratical exponential distribution.