Statistical Inference Course Project: Part 1

Dongying

8/20/2020

Simulation exercise

As the project instruction asked, in this project I will 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, λ) where λ is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. Set $\lambda = 0.2$ for all of the simulations. I will investigate the distribution of averages of 40 exponentials. Note that I will do a thousand simulations.

First, let's construct the simulation:

```
set.seed(77481)
lambda <- 0.2
sim <- data.frame(ncol = 2, nrow = 1000)
names(sim) <- c("Index", "ExpMean")
for (i in 1:1000){
        sim[i,] <- c(i, mean(rexp(40, rate = lambda)))
}
```

Sample mean compared to the theoretical mean of the distribution

Here I calculate the sample mean and compared it with the theoretical mean:

```
sample_mean <- mean(sim$ExpMean)
theoretical_mean <- 1/lambda
sample_mean
```

```
## [1] 4.950775
```

```
theoretical_mean
```

```
## [1] 5
```

Sample variance compared to the theoretical variance of the distribution

Here I calculate the sample variance and compared it with the theoretical variance. Noted that, since we are doing 40 exponential distributions, the theoretical standard deviation should be $1/n\lambda$, thus the variance shoule be the square of it:

```
sample_var <- var(sim$ExpMean)
theoretical_var <- (1/lambda)^2/40
sample_var
```

```
## [1] 0.5889956
```

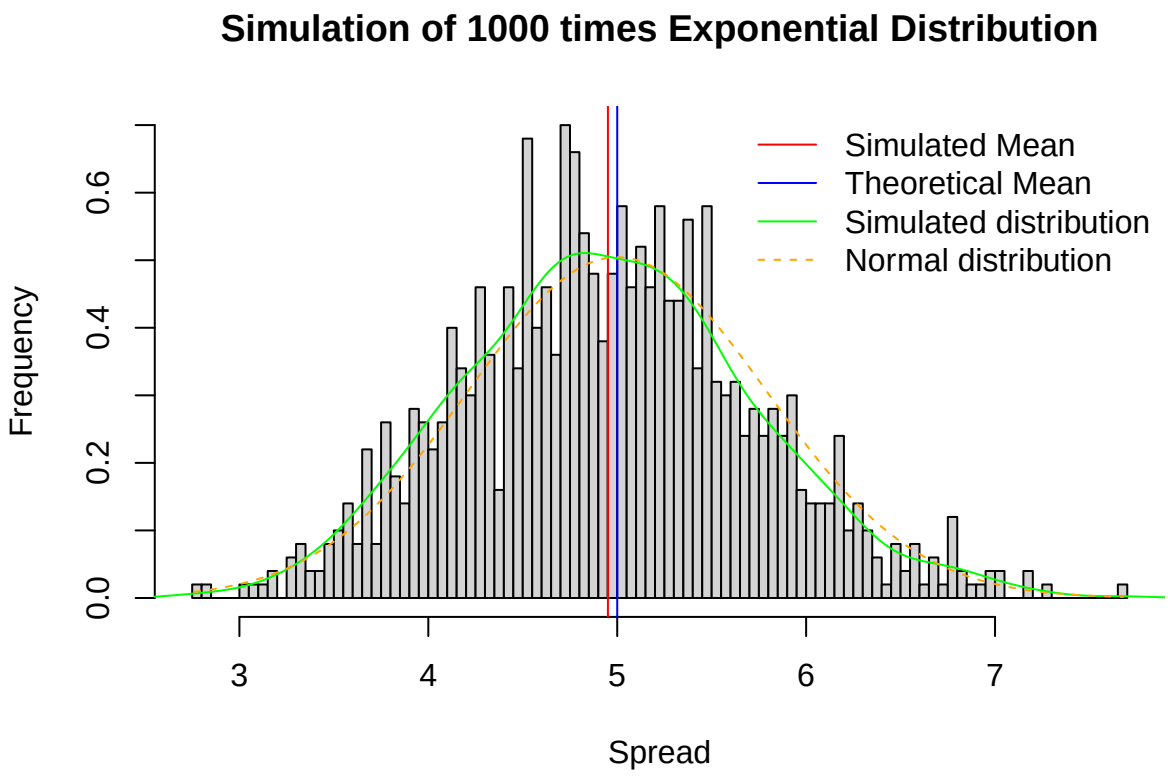
```
theoretical_var
```

```
## [1] 0.625
```

Sample distribution

Here I plot a histogram of the simulated exponential distribution. Due to the central limit theorem, the averages of samples follow normal distribution. The figure below shows the density computed using the histogram and the normal distribution plotted with theoretical mean and variance values. They fall into a similar shape.

```
hist(sim$ExpMean, prob = TRUE, breaks = 100,
        main = "Simulation of 1000 times Exponential Distribution",
        xlab = "Spread", ylab = "Frequency")
abline(v = theoretical_mean, col= "blue", lty = 1)
abline(v = sample_mean, col = "red", lty = 1)
lines(density(sim$ExpMean), col = "green", lty = 1)
xnorm <- seq(min(sim$ExpMean), max(sim$ExpMean), length = 100)
ynorm <- dnorm(xnorm, mean = 1/lambda, sd = (1/lambda/sqrt(40)))
lines(xnorm, ynorm, col = "orange", lty = 2)
legend('topright',
       c("Simulated Mean", "Theoretical Mean",
         "Simulated distribution", "Normal distribution"),
       bty = "n", lty = c(1,1,1,2), col = c("red", "blue", "green", "orange"))
```



Conculsion

Due to the central limit theorem, the averages of 1000 exponential distribution follow normal distribution.