Statistical Inference - course project part 1

Sangwon Han

2018-12-18

Overview

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, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

The followings are the instruction for this projecct:

1. lambda = 0.2 for all of the simulations.
2. Investigate the distribution of averages of 40 exponentials
3. Need to do a thousand simulations.

Simulations

Set the conditions

```
lambda <- 0.2
n <- 40
sims <- 1000
```

Set seed for the reproducibility, and simulate exponential distribution

```
set.seed(1)
sim_exp <- replicate(sims, rexp(n, lambda))
dim(sim_exp)
```

```
## [1]   40 1000
```

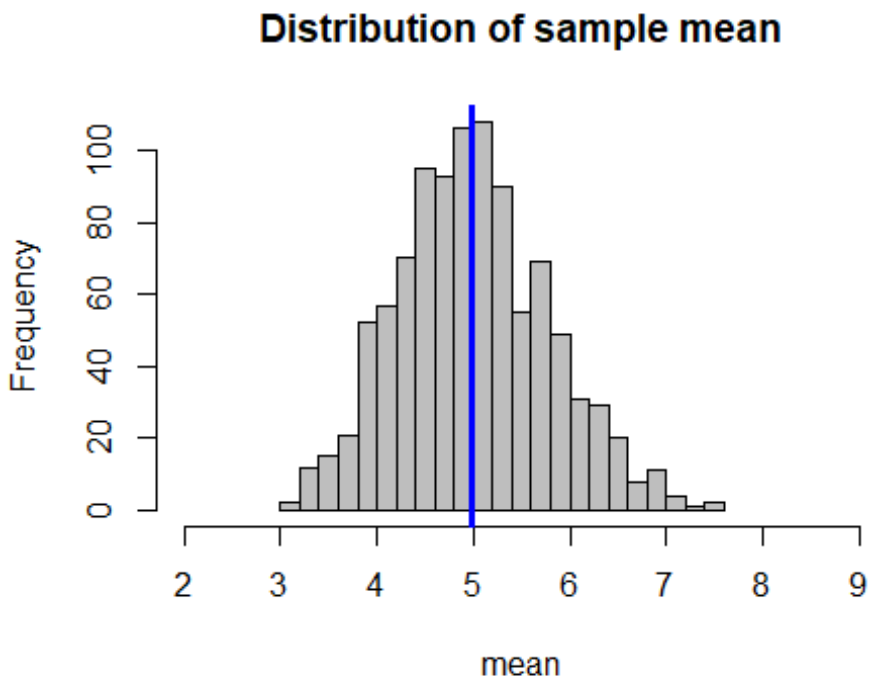
Calculate the mean

```
mean_exp <- apply(sim_exp, 2, mean)
```

1. Sample Mean versus Theoretical Mean

Plot histogram of the means of the exponential distribution.

```r
hist(mean_exp, main = "Distribution of sample mean", xlab = "mean",
    xlim = c(2, 9), breaks = 30, col = "grey")
abline(v = mean(mean_exp), lwd = "3", col = "blue")
```



The sample mean is as follow:

```r
mean(mean_exp)
```

```
## [1] 4.990025
```

The theoretical mean of exponential distribution is 1/lambda.

```r
theo_mean <- 1/lambda
theo_mean
```

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

2. Sample Variance versus Theoretical Variance

The sample variance is as follow:

```r
var(mean_exp)
```

```
## [1] 0.6111165
```

The theoretical variance of exponential distribution is ((1/lambda)/sqrt(n))^2.

```
theo_var <- ((1/lambda)/sqrt(n))^2
theo_var

## [1] 0.625
```

Distribution

The exponential distribution of 1000 simulations is approximately normal. Due to the Central Limit Theorem, the means of the sample simulations should follow a normal distribution.

```
hist(mean_exp, main = "Distribution of sample mean", xlab = "mean",
    xlim = c(2, 9), breaks = 30, col = "grey", probability = TRUE)

# add density plot
lines(density(mean_exp), lwd = 3, col = "red")

# add normal distribution line
x <- seq(min(mean_exp), max(mean_exp), length = 2 * n)
y <- dnorm(x, mean = 1/lambda, sd = sqrt(((1/lambda)/sqrt(n))^2))
lines(x, y, col = "black", lwd = 3, lty = 2)
```