

Project1_simulation

Jinhai_Yu

11/13/2020

Overview

The task is to practice the simulation of exponential distribution in R and compare it with the Central Limit Theorem. The report will show the sample mean and variance, and the comparison with the theoretical values, also to see if the distribution is close to normal or not.

Load libraries

```
library(ggplot2)
```

Simulation

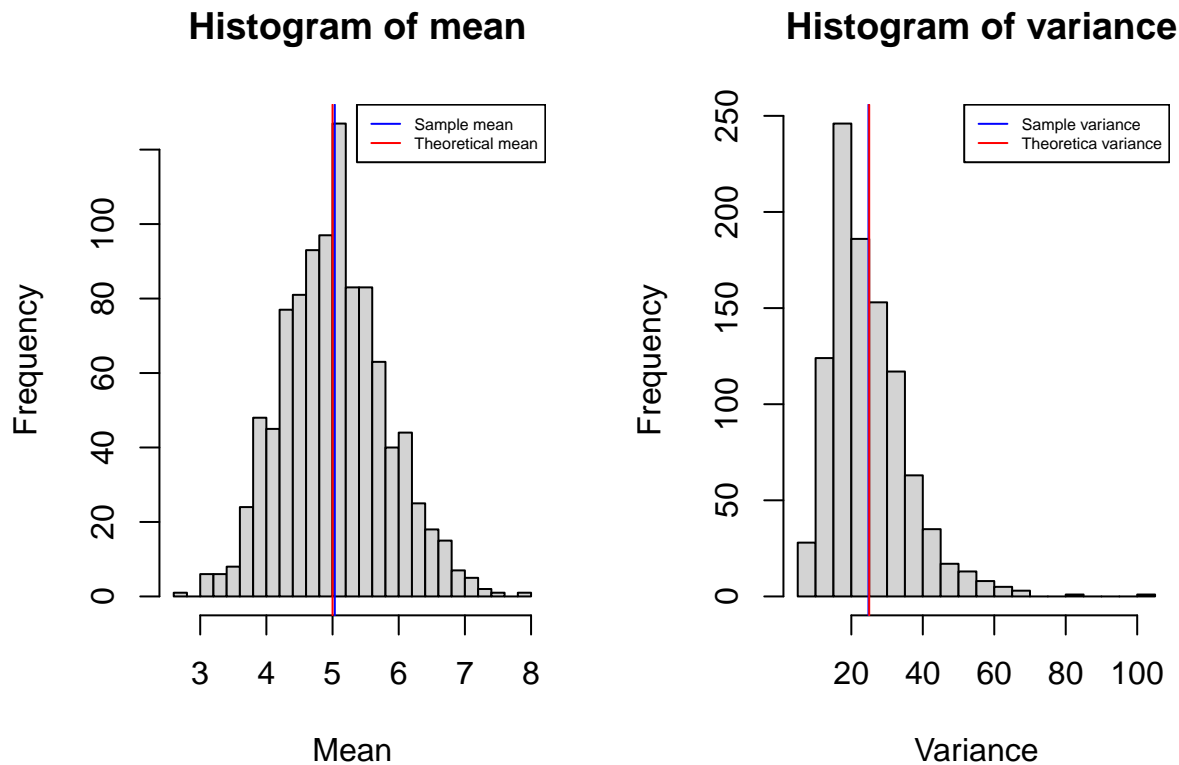
The exponential distribution is $\text{rexp}(n, \text{lambda})$, with $n = 40$, $\text{lambda} = 0.2$, $\text{mean} = 1/0.2$, $\text{sd} = 1/0.2$; and the simulation times are 1000.

```
set.seed(2020)
lambda <- 0.2
n <- 40
rep <- 1000
sim1 <- NULL
sim2 <- NULL
for(i in 1:rep) sim1 <- c(sim1, mean(rexp(n, lambda)))
for(i in 1:rep) sim2 <- c(sim2, var(rexp(n, lambda)))
mean_sim <- mean(sim1)
var_sim <- mean(sim2)
mean_t <- 1/lambda
var_t <- (1/lambda)^2
```

Plot mean and variance of simulated data, and compared with theoretical values, respectively.

```
par(mfrow = c(1, 2))
hist(sim1, breaks = 30, xlab = "Mean", main = "Histogram of mean")
abline(v = mean_sim, col = "blue")
abline(v = mean_t, col = "red")
legend("topright", col = c("blue", "red"), legend = c("Sample mean", "Theoretical mean"), lty = 1, cex = 1)
```

```
hist(sim2, breaks = 30, xlab = "Variance", main = "Histogram of variance")
abline(v = var_sim, col = "blue")
abline(v = var_t, col = "red")
legend("topright", col = c("blue", "red"), legend = c("Sample variance", "Theoretical variance"), lty = 1)
```



```
cat(paste(c("The mean of sample mean is ", mean_sim, " that is pretty close to the theoretical mean ",
            mean_t), sep = "", collapse = ""))
```

```
## The mean of sample mean is 5.03394824561008 that is pretty close to the theoretical mean 5
```

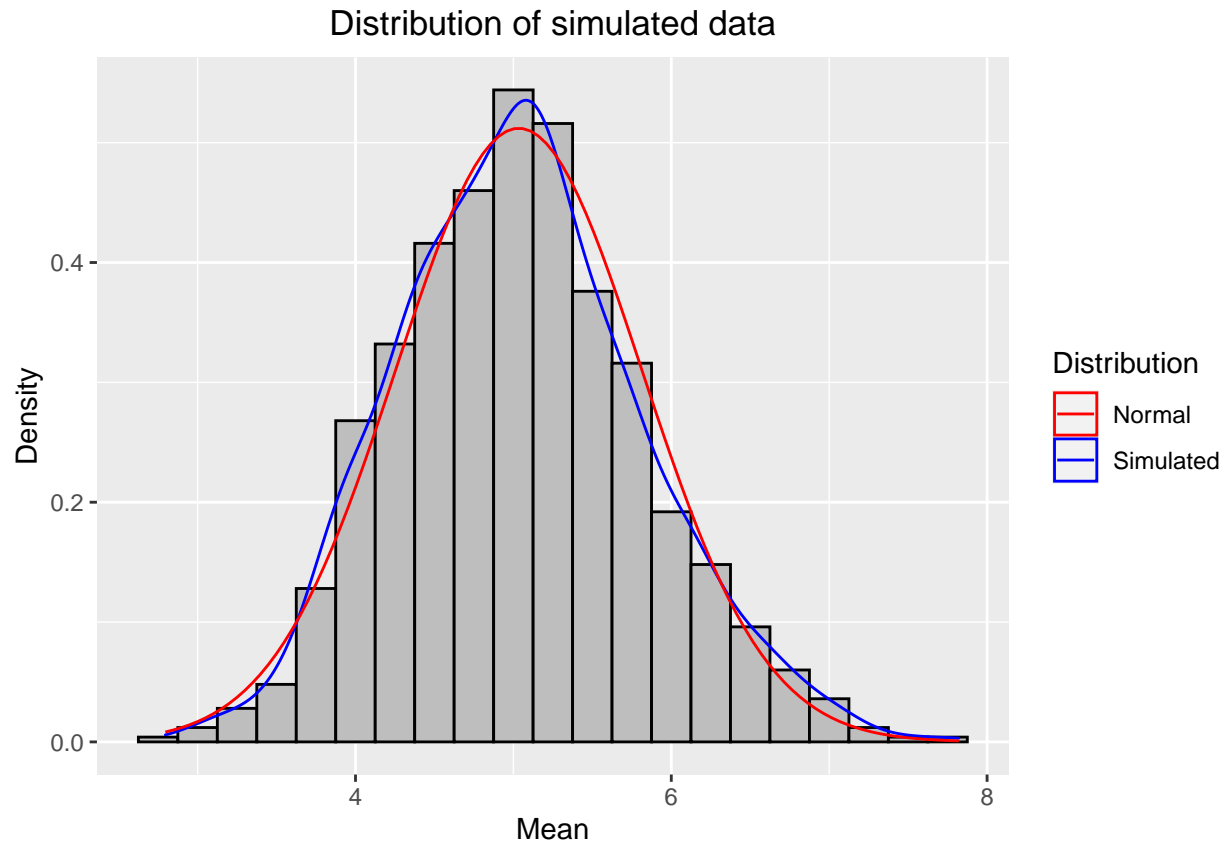
```
cat(paste(c("The mean of sample var is ", var_sim, " that is also close to the theoretical variance ",
            var_t), sep = "", collapse = ""))
```

```
## The mean of sample var is 24.7984586357113 that is also close to the theoretical variance 25
```

Plot mean of simulated data with density and compared with normal distribution
(mean = 5, sd = 5)

```
sim3 <- as.data.frame(sim1)
ggplot(sim3, aes(x = sim1)) +
  geom_histogram(aes(y = ..density..), binwidth = 0.25, col = "black", fill = "gray") +
```

```
geom_density(aes(color = "Simulated")) +
geom_function(aes(color = "Normal"), fun = dnorm, args = list(mean = mean_sim, sd = sd(sim1))) +
labs(x = "Mean", y = "Density", title = "Distribution of simulated data", color = "Legend") +
theme(plot.title = element_text(hjust = 0.5)) +
scale_color_manual("Distribution", values = c("red", "blue"))
```



```
cat(paste("The density line of simulated data is very similar to the normal distribution."))
```

```
## The density line of simulated data is very similar to the normal distribution.
```