Statistical Inference - Part 1 Simulation Exercise

Andrey Budish

March 31, 2017

Overview

The purpose of this data analysis is to investigate the exponential distribution in R and compare it with the Central Limit Theorem. The distribution will consist of averages of 40 exponentials with thousand simulations. The lambda will be set to 0.2 for all of the simulations.

Simulations

Set a seed, assign variables for sample size and lambda

```
set.seed(999)
n <- 40
lambda <- 0.2
```

Run simulation of exponential distribution of 1000 simulations with these variables and store the means in the **simMns** variable

```
simMns <- NULL
for (i in 1 : 1000) simMns <-  c(simMns, mean(rexp(n, rate = lambda)))
```

Sample mean vs Theoretical mean

Calculate sample mean of 1000 simulations

```
mean(simMns)
```

```
## [1] 5.029028
```

The theoretical mean is:

```
1/lambda
```

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

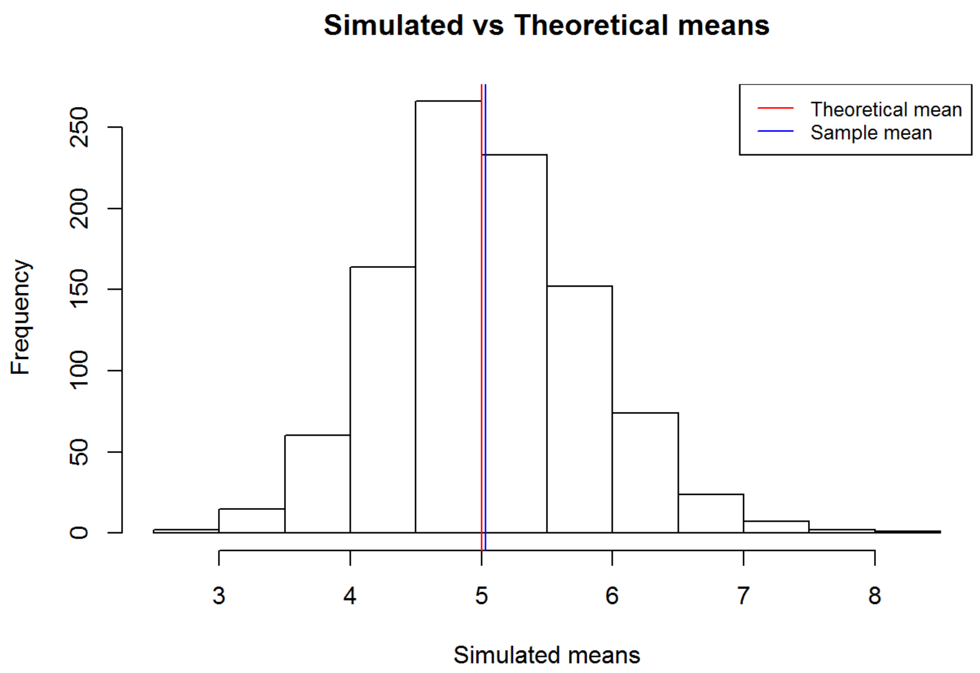
As we can see simulated mean is very close to the theoretical mean.

The figure below shows the distribution of 1000 sample means.
Red vertical line shows the theoretical mean and blue one the simulated mean.

```
hist(simMns, xlab = "Simulated means",
     main = "Simulated vs Theoretical means")
# draw theoretical mean
abline(v = 1/lambda, col = "red", lwd = 1)

# draw simulated mean
abline(v = mean(simMns), col = "blue", lwd = 1)

legend("topright", legend=c("Theoretical mean", "Sample mean"),
       col=c("red", "blue"), lty=1, cex=0.8)
```



Sample variance vs Theoretical variance

Calculate sample variance of 1000 means

```
var(simMns)
```

```
## [1] 0.605616
```

Standard deviation of exponential distribution is 1/lambda. The theoretical variance is equal to standard deviation squared divided by the sample size:

```
(1/lambda)^2/n
```

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

As we can see simulated variance is very close to the theoretical variance.

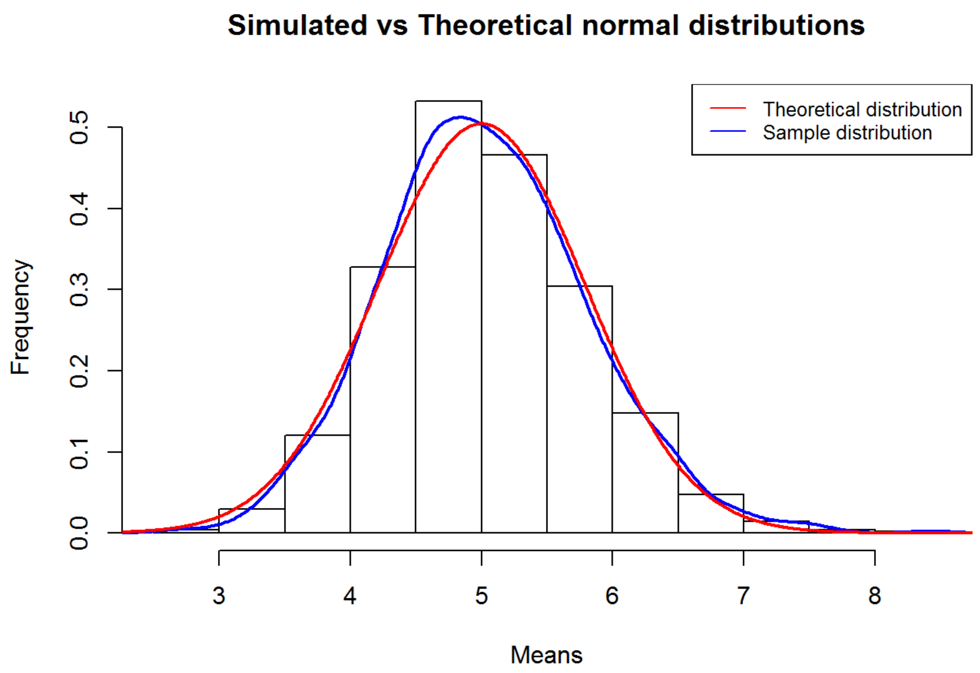
Distribution

Let's draw the histogram of the distribution of simulated means together with related density plot (blue color) and compare it to the density plot with theoretical mean and variance (red color):

```
# histogram of sample means
hist(simMns, xlab = "Means", ylab = "Frequency",
     main = "Simulated vs Theoretical normal distributions",
     prob = TRUE)
# density plot of sample means
lines(density(simMns), lwd = 2, col = "blue")

# density plot with theoretical mean and variance
x <- seq(0,10,length=1000)
y <- dnorm(x, mean = 1/lambda, sd =(1/lambda)/sqrt(n))
lines(x, y, lwd = 2, col = "red")

legend("topright", legend=c("Theoretical distribution", "Sample distribution"),
       col=c("red", "blue"), lty=1, cex=0.8)
```



Due to the big number of simulations our sample mean is close to the theoretical mean and sample variance is close to the theoretical variance. We can see how simulated mean density plot is very similar to the theoretical density plot. They both have normal distribution.

We have just saw the power of the Central Limit Theorem in action: the distribution of means of samples is normal, even though the distribution where they came from is exponential.