Statistical Inference - Exponential Distribution Simulation

Dung

August 17, 2017

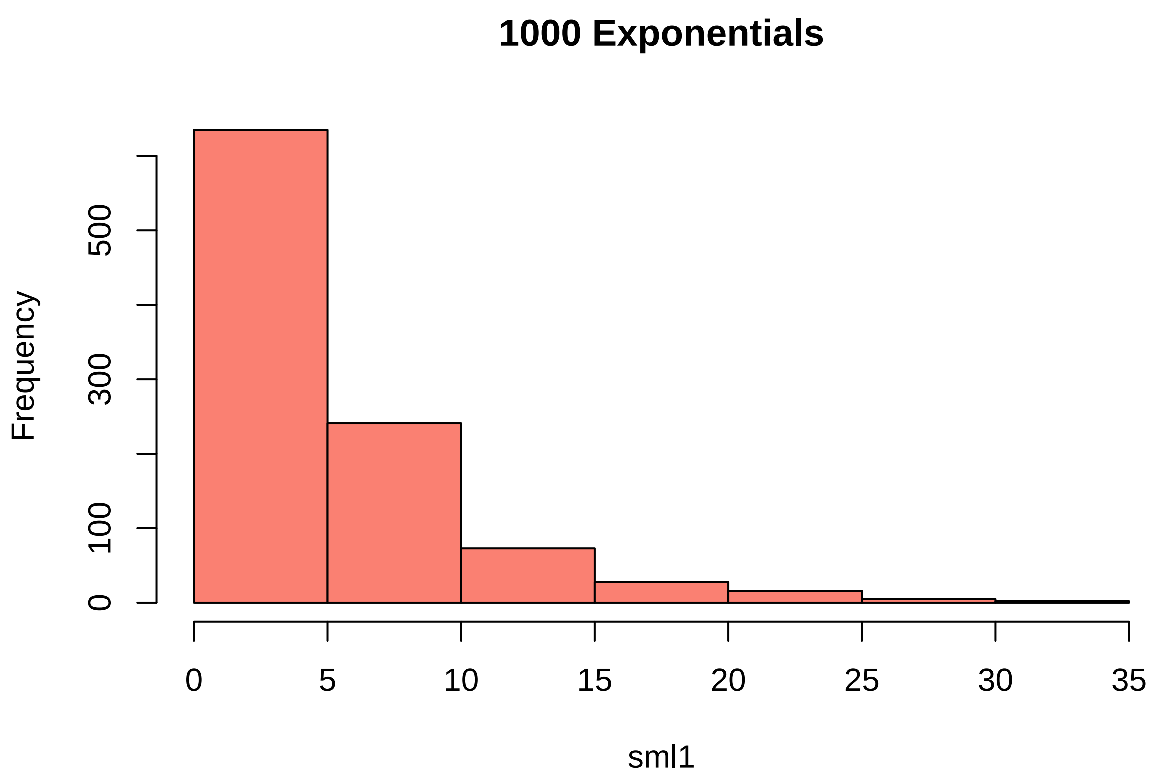
Purpose

In this exercise, I was going to simulate a distribution of 1000 exponentials and a distribution of 1000 means of 40 exponentials with lambda = 0.2. Characteristics from simulations will be examined and discussed.

Simulations

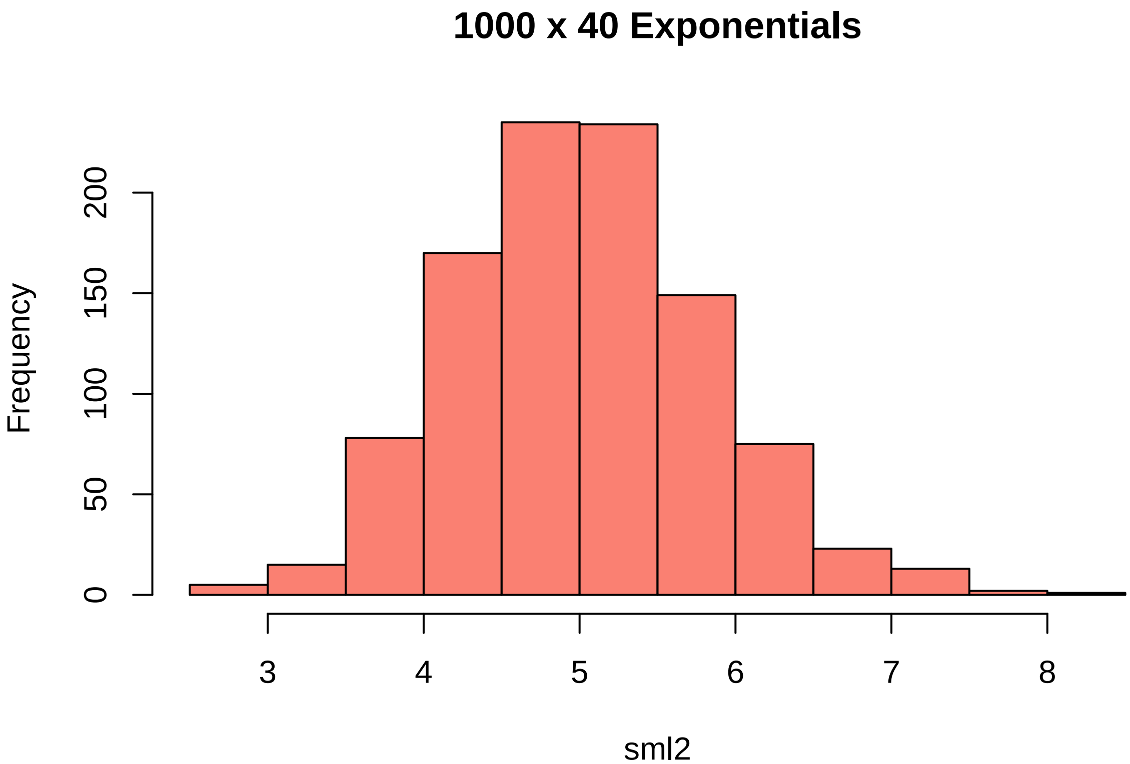
The following figure shows the simulated distribution of 1000 exponentials with lambda = 0.2. Let's call this Simulation 1.

```
sml1 = rexp(1000,0.2)
hist(sml1, main = "1000 Exponentials", col = "salmon")
```



The following figure shows the simulated distribution of 1000 means of 40 exponentials with lambda = 0.2. Let's call this Simulation 2

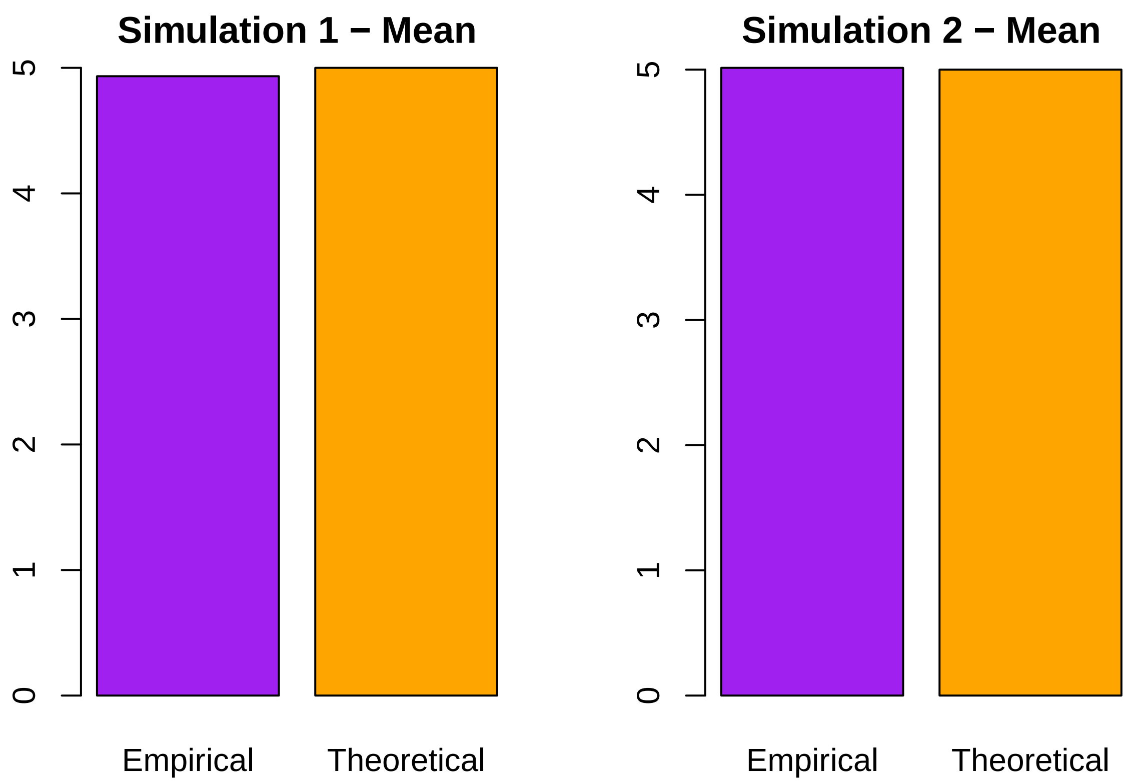
```
sml2 = NULL
for (i in 1 : 1000) sml2 = c(sml2, mean(rexp(40,0.2)))
hist(sml2, main = "1000 x 40 Exponentials", col = "salmon")
```



Discussion

1. Empirical mean versus theoretical mean

```
par(mfrow = c(1,2), mar = c(4,4,2,1))
barplot(height = c(Empirical = mean(sml1), Theoretical = 5),
        col = c("purple", "orange"), main = "Simulation 1 - Mean")
barplot(height = c(Empirical = mean(sml2), Theoretical = 5),
        col = c("purple", "orange"), main = "Simulation 2 - Mean")
```



In simulation 1, I simulated the exponential distribution with n = 1000. In theory, the population mean of the exponential distribution is 1/lambda, or 5 in this case since lambda = 0.2.

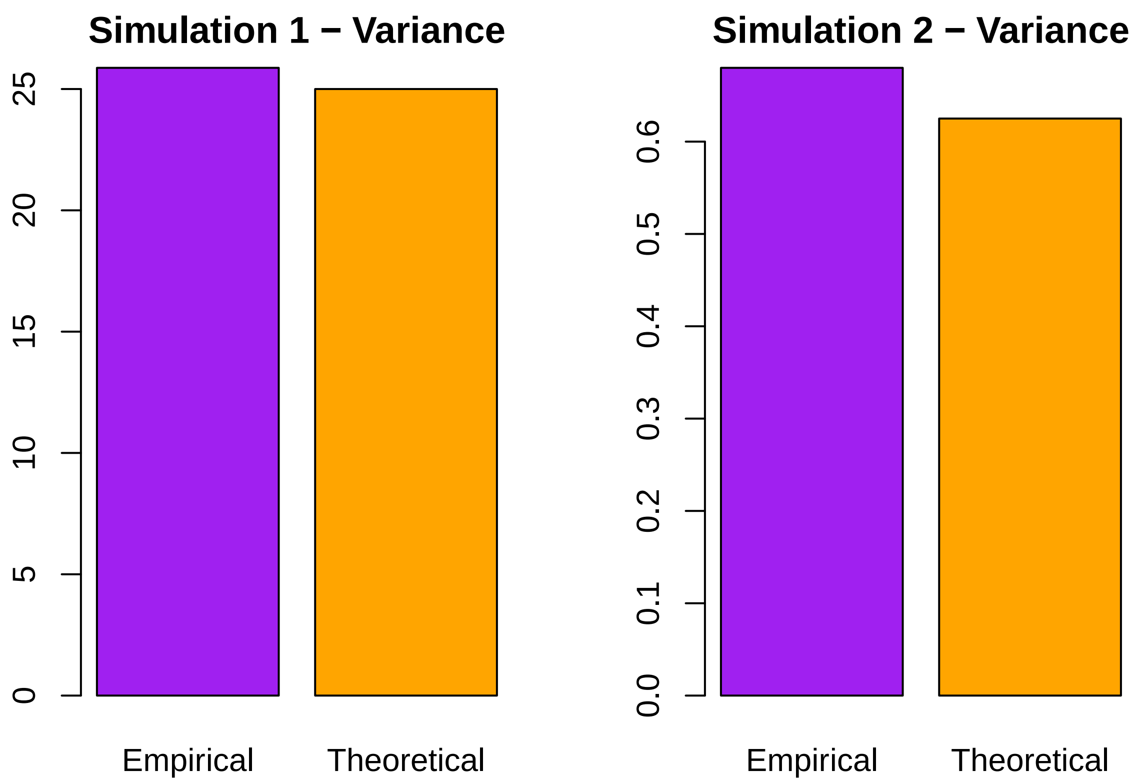
The simulation 2 is the empirical distribution of means of 40 exponentials. The mean of 1000 sample means of 40 exponentials with lambda = 0.2 should be an estimate of the population mean of the exponential distribution with mean equal 5.

```
data.frame(EmpiricalMean = c(mean(sml1), mean(sml2)), TheoreticalMean = c(5,5),
           row.names = c("Simulation 1", "Simulation 2"))
```

```
##              EmpiricalMean TheoreticalMean
## Simulation 1      4.933297               5
## Simulation 2      5.014257               5
```

2. Sample variance versus theoretical variance

```
par(mfrow = c(1,2), mar = c(4,4,2,1))
barplot(height = c(Empirical = var(sml1), Theoretical = 25),
        col = c("purple", "orange"), main = "Simulation 1 - Variance")
barplot(height = c(Empirical = var(sml2), Theoretical = 0.625),
        col = c("purple", "orange"), main = "Simulation 2 - Variance")
```



Since the variance of the exponential distribution is 1/lambda to the power of 2, the true variance of the exponential distribution is 25. The sample of 1000 exponentials should estimate this population variance.

A random sample of exponentials with size 40 was drawn and its mean was calculated. After repeating this process 1000 times, the distribution of the sample means appears more Gaussian and has the variance equal to the variance of the original data that these samples were drawn from divided by the sample size of 40. This yields the variance of 0.625. The following figure and table compares the empirical results from the simulations with their theoretical expectations discussed above.

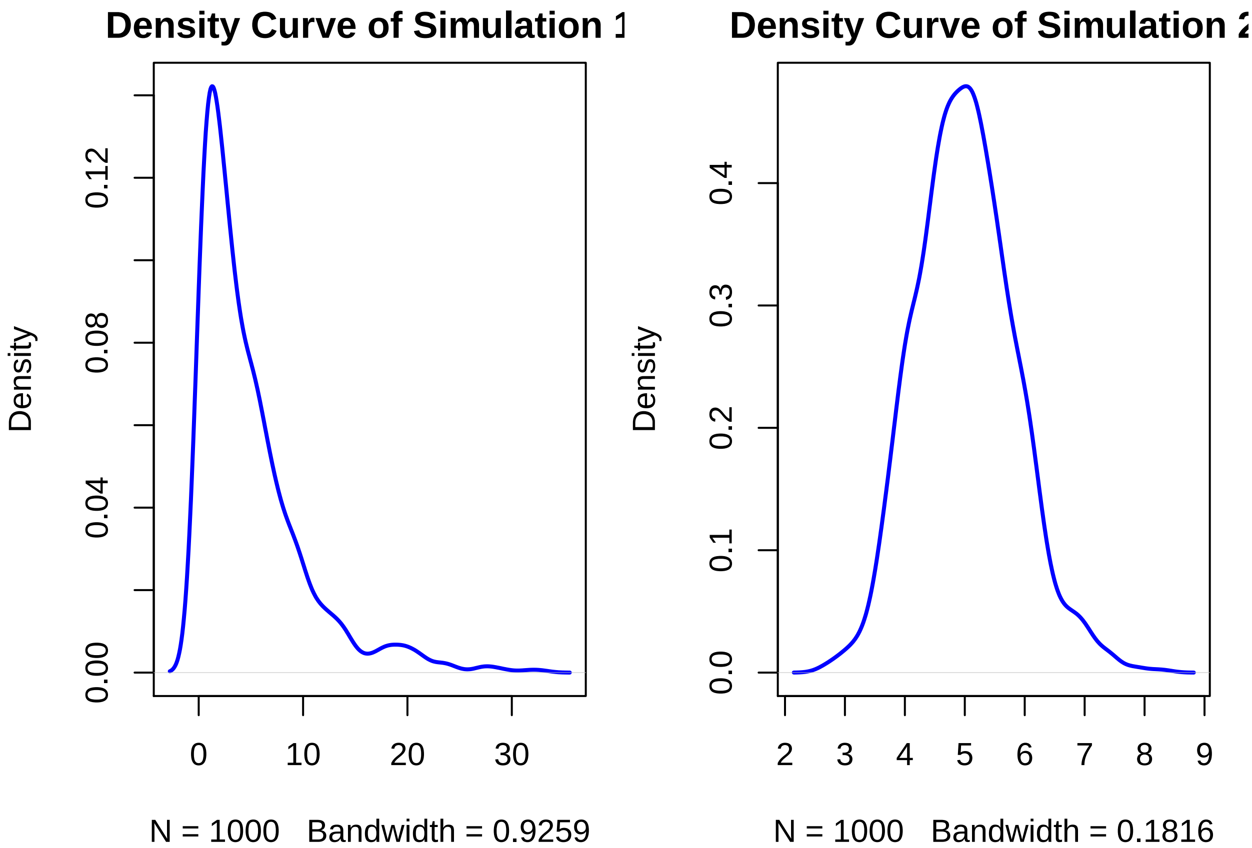
```
data.frame(EmpiricalVariance = c(var(sml1), var(sml2)), TheoreticalVariance = c(25,0.625),
           row.names = c("Simulation 1", "Simulation 2"))
```

```
##              EmpiricalVariance TheoreticalVariance
## Simulation 1        25.8732528              25.000
## Simulation 2         0.6799096               0.625
```

3. Density curves

The distribution of 1000 means of 40 exponentials looks more Gaussian than the distribution of 1000 exponentials, which is predicted by the Central Limit Theorem.

```
par(mfrow = c(1,2), mar = c(4,4,2,1))
plot(density(sml1), col="blue", lwd=2, main = "Density Curve of Simulation 1")
plot(density(sml2), col="blue", lwd=2, main = "Density Curve of Simulation 2")
```


Density Curve of Simulation 1
Density
0.00
0.04
0.08
0.12
0
10
20
30
N = 1000 Bandwidth = 0.9259
Density Curve of Simulation 2
Density
0.0
0.1
0.2
0.3
0.4
2
3
4
5
6
7
8
9
N = 1000 Bandwidth = 0.1816