Exponential Distribution Simulation

Dongyuan Wu

2019-05-09

Overview

In this project we will investigate the exponential distribution (set lambda=0.2) in R and compare it with the Central Limit Theorem.

The following things will be done:

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Simulations

First of all, do a simulation 1000 times and each time simulate 40 exponentials (lambda=0.2). Then create a matrix of 1000 rows with the columns corresponding to 40 random simulted exponentials, and calculate the averages of 40 exponentials for each simulation.

```
# set lambda = 0.2
lambda = 0.2
# investigate 40 exponentials
n = 40
# do a thousand simulations
times = 1000
# set a seed
set.seed(1234)
# save the simulation results into a times*n matrix
simulation <- matrix(rexp(n * times, lambda), times, n)
# calculate the averages of 40 exponentials for each simulation
sim.mean <- rowMeans(simulation)
```

Sample Mean versus Theoretical Mean

```
(sample.mean <- mean(sim.mean))
```

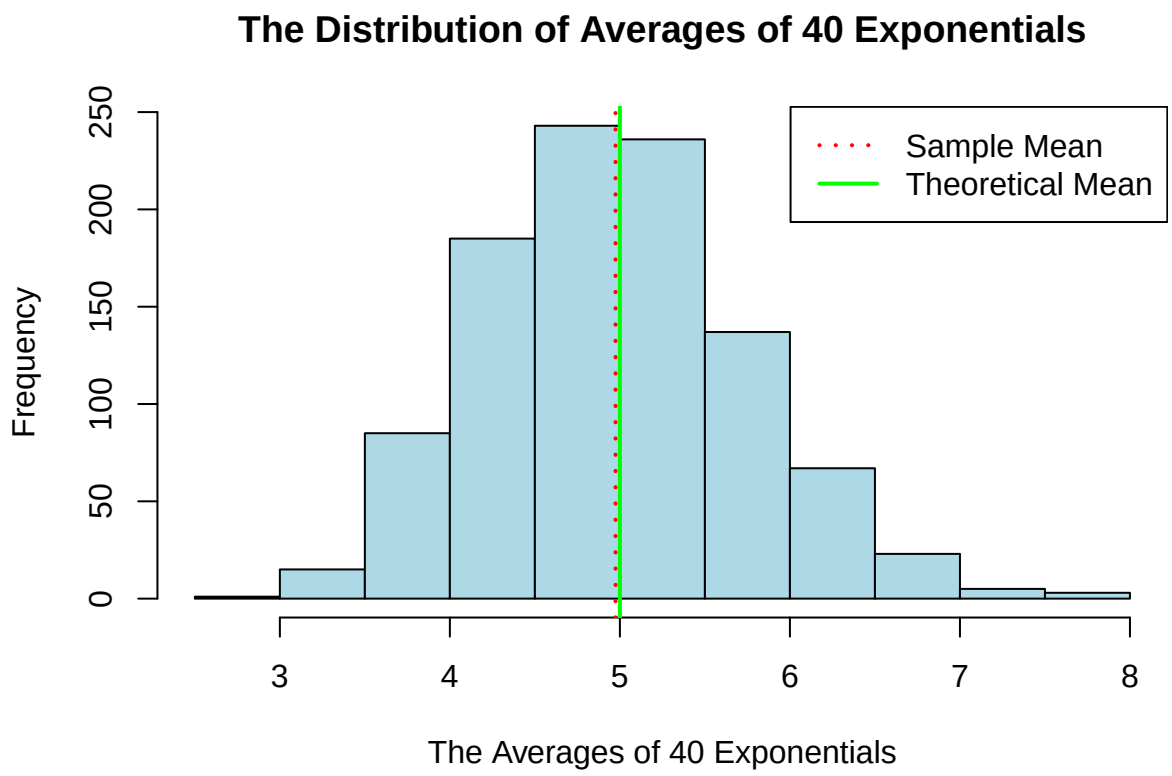
```
## [1] 4.974239
```

```
(theory.mean <- 1/lambda)
```

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

We can find the sample mean is 4.9742, which is very close to the theoretical mean 5. We can draw these two values into the histogram.

```
# the original histogram
hist(sim.mean, col="lightblue", xlab="The Averages of 40 Exponentials",
     main="The Distribution of Averages of 40 Exponentials")
# draw the line for sample mean
abline(v=sample.mean, col="red", lwd=2, lty=3)
# draw the line for theoretical mean
abline(v=theory.mean, col="green", lwd=2)
legend("topright", legend=c("Sample Mean", "Theoretical Mean"),
       col=c("red", "green"), lty=c(3, 1), lwd=2)
```



Sample Variance versus Theoretical Variance

```
(sample.var <- var(sim.mean))
```

```
## [1] 0.5949702
```

```
(theory.var <- 1/(lambda)^2/n)
```

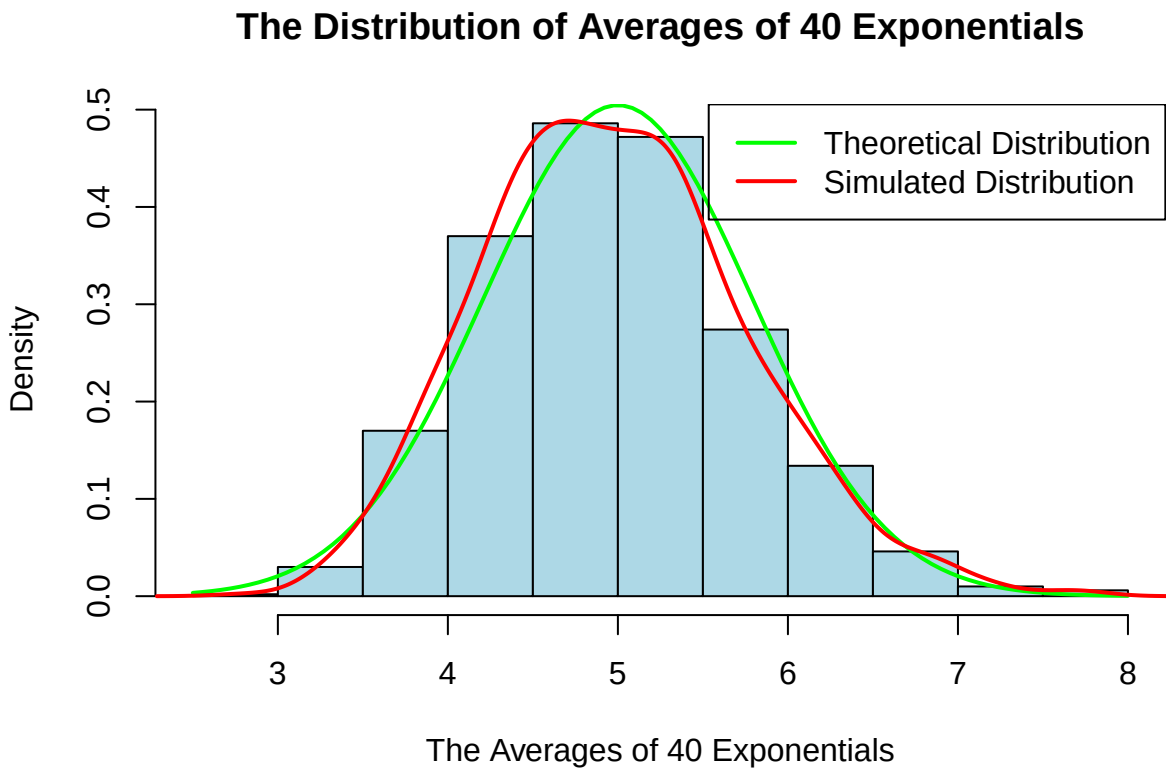
```
## [1] 0.625
```

We can find the sample variance is 0.595, which is very close to the theoretical variance 0.625.

Distribution

We can draw density curves for simulated distribution and theoretical distribution, respectively.

```
x <- seq(0, 8, 0.01)
# the original histogram
hist(sim.mean, freq=FALSE, col="lightblue", xlab="The Averages of 40 Exponentials",
     main="The Distribution of Averages of 40 Exponentials")
curve(dnorm(x, mean=theory.mean, sd=sqrt(theory.var)),
      add=TRUE, col="green", lwd=2)
lines(density(sim.mean), lwd=2, col="red")
legend("topright", legend=c("Theoretical Distribution", "Simulated Distribution"),
       col=c("green", "red"), lwd=2)
```

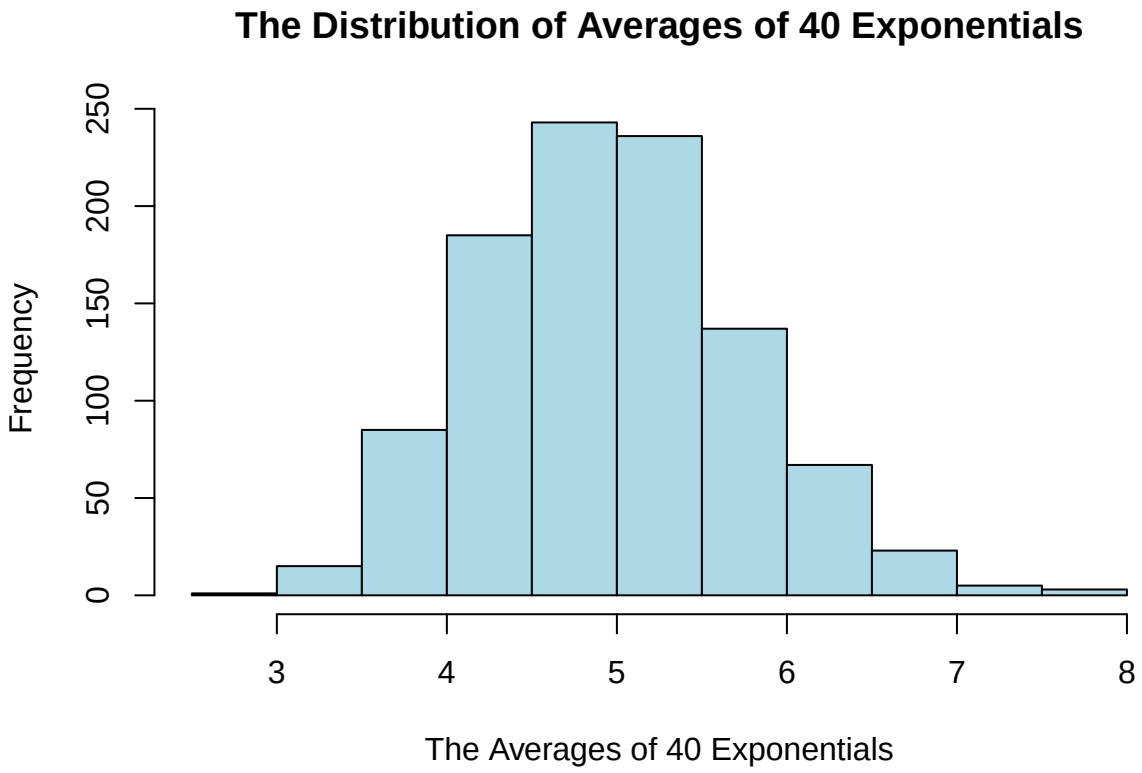


As we can see, the simulated distribution is very similar to the normal distribution. So this distribution is approximately normal.

Appendix

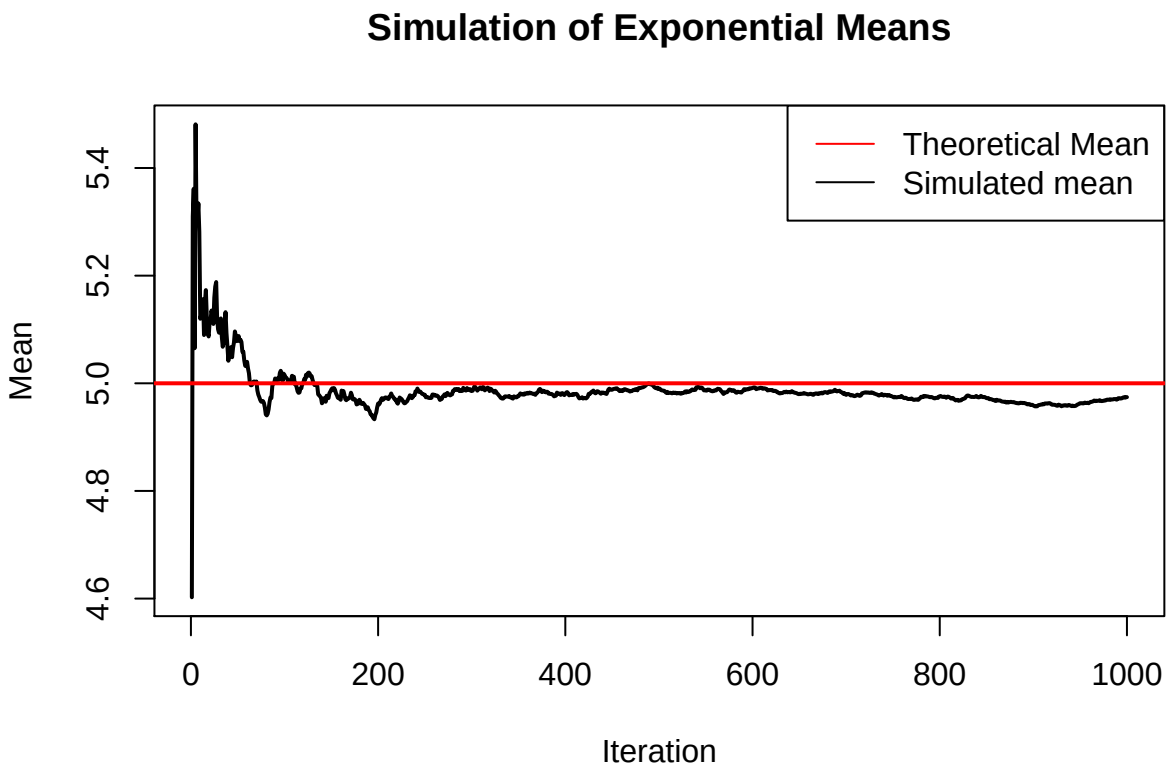
1. Draw a histogram to show the distribution of averages of 40 exponentials.

```
hist(sim.mean, col = "lightblue", xlab="The Averages of 40 Exponentials",
     main="The Distribution of Averages of 40 Exponentials")
```



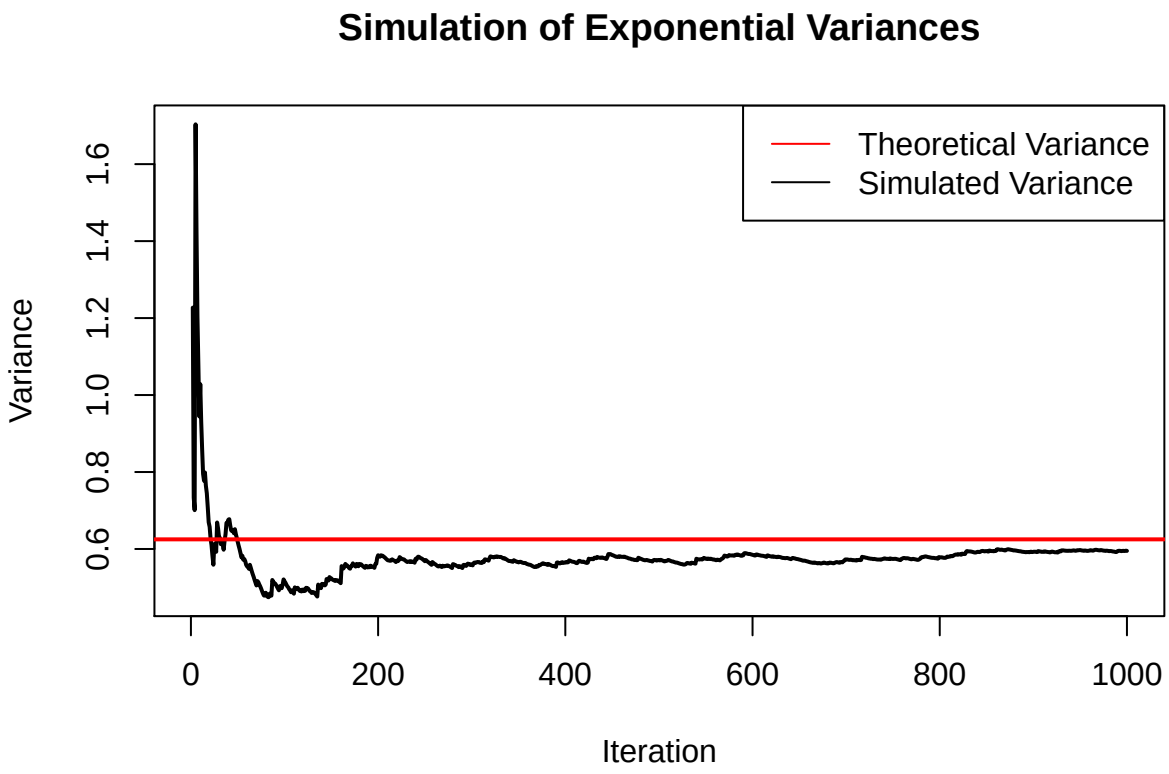
2. Draw a figure shows how the simulated means converge to the theoretical mean according to CLT.

```
cum.mean <- cumsum(sim.mean)/seq_along(sim.mean)

plot(seq_along(sim.mean), cum.mean, type = "l", lwd=2,
     main="Simulation of Exponential Means", xlab="Iteration",
     ylab="Mean")
abline(h=theory.mean, col="red", lwd=2)
legend("topright", lty=c(1,1), col=c("red", "black"),
       legend=c("Theoretical Mean", "Simulated mean"))
```



3. Draw a figure shows how the simulated variances converge to the theoretical variance according to CLT.

```
cum.var <- cumsum((sim.mean-sample.mean)^2)/(seq_along(sim.mean)-1)

plot(seq_along(sim.mean), cum.var,type="l", lty=1, lwd=2,
     main="Simulation of Exponential Variances", xlab="Iteration",
     ylab="Variance")
abline(h=theory.var, col="red",lwd=2)
legend("topright", lty=c(1,1), col=c("red","black"),
       legend=c("Theoretical Variance", "Simulated Variance"))
```


Simulation of Exponential Variances
Theoretical Variance
Simulated Variance
Variance
0.6
0.8
1.0
1.2
1.4
1.6
0
200
400
600
800
1000
Iteration