Statistical Inference Course Project - Part 1: Simulation Exercise Instructions

Ely Xavier

March 9, 2018

Overview

In this project, you 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 the exponential distribution is 1/lambda, and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

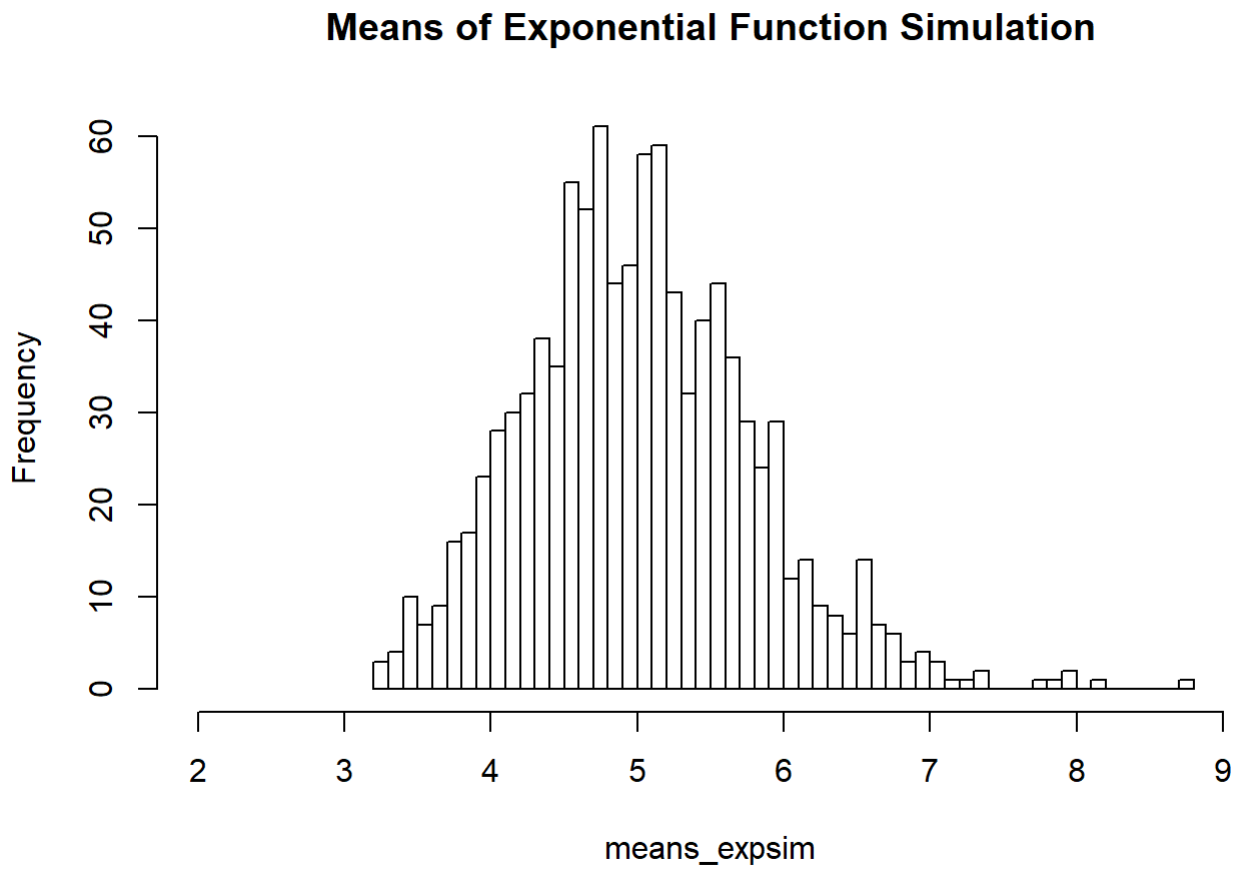
Simulations

The following code includes the rate (lambda), number of exponentials and the number of simulations to run.

Then a numeric vector (‘means’) is generated, which contains the result of the simulations. The following simulations generate the mean.

Finally, a histogram of the simulated mean values is generated.

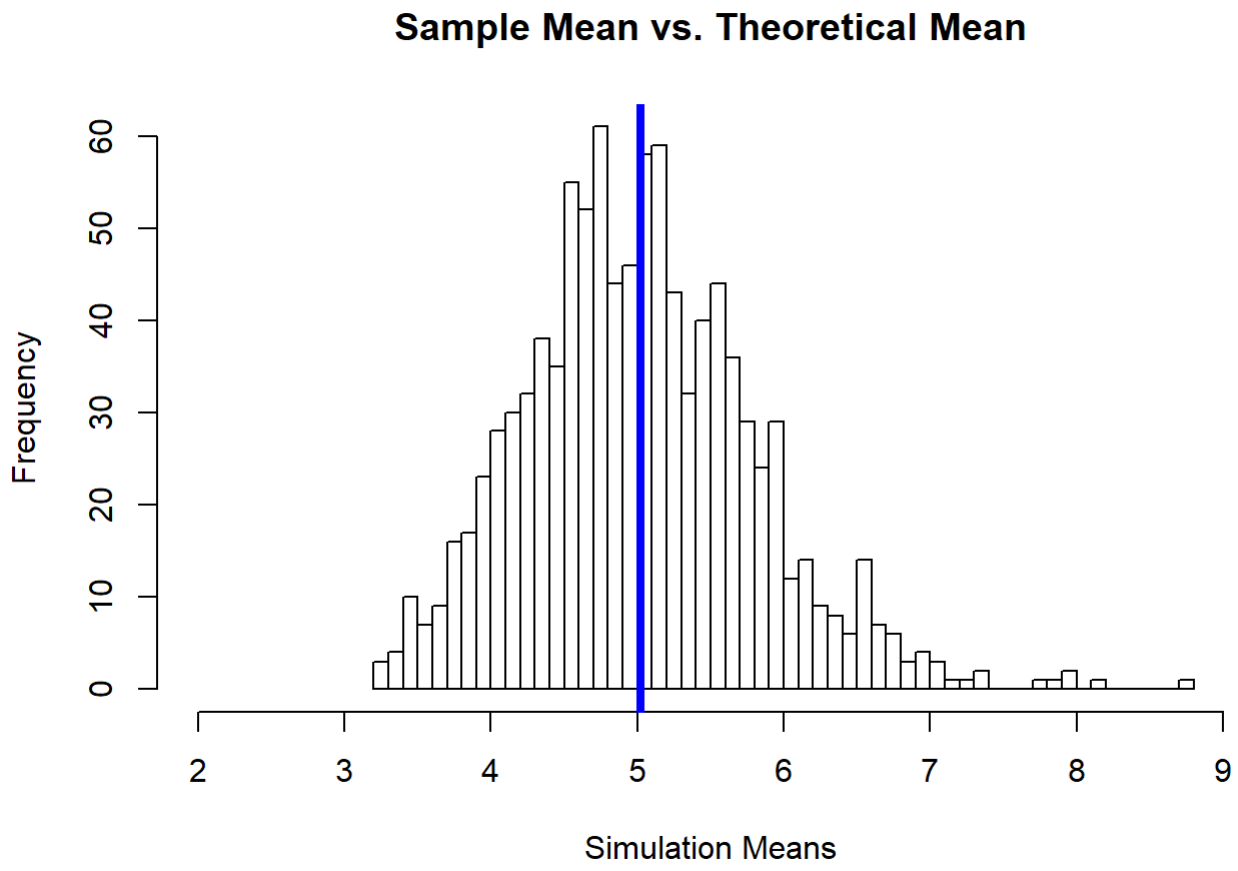
```
# setting seed for reproducibility
set.seed(2018)

# set values as per project instructions
lambda <- 0.2

# number of exponentials
n <- 40

# number of simulations
sims <- 1000

# run simulations
running_sim <- replicate(sims, rexp(n,lambda))

#calculate means of exponential simulations
means_expsim <- apply(running_sim,2, mean)

#Plot means histogram
hist(means_expsim, breaks = 40, xlim = c(2,9), main = "Means of Exponential Function Simulation"
)
```



Question 1: Sample Mean vs. Theoretical Mean

The exponential distribution mean is 1/lambda. For this simulation, lambda is 0.2. The theoretical mean should be 5 (i.e., 1 / 0.2).

```
# plot histogram of the sample means
hist(means_expsim, main = "Sample Mean vs. Theoretical Mean", xlim =  c(2,9), breaks = 40, xlab
 = "Simulation Means")

# plot vertical blue line at mean of samples
abline(v = mean(means_expsim), lwd = "4", col = "blue")
```




```
# calculate the mean of our samples
mean(means_expsim)
```

```
## [1] 5.020107
```

The sample mean is 5.02, very close to our theoretical mean of 5.

Question 2: Sample Variance vs. Theoretical Variance

The standard deviation of the exponential distribution is (1/lambda) / sqrt(n). We will compare this to our simulations.

```
# theoretical variance vs. simulated variance
print(paste ("Theoretical variance is: ", round( (1/lambda)^2/n, 3)))
```

```
## [1] "Theoretical variance is:  0.625"
```

```
print(paste("Actual variance is: ", round( var(means_expsim),3)))
```

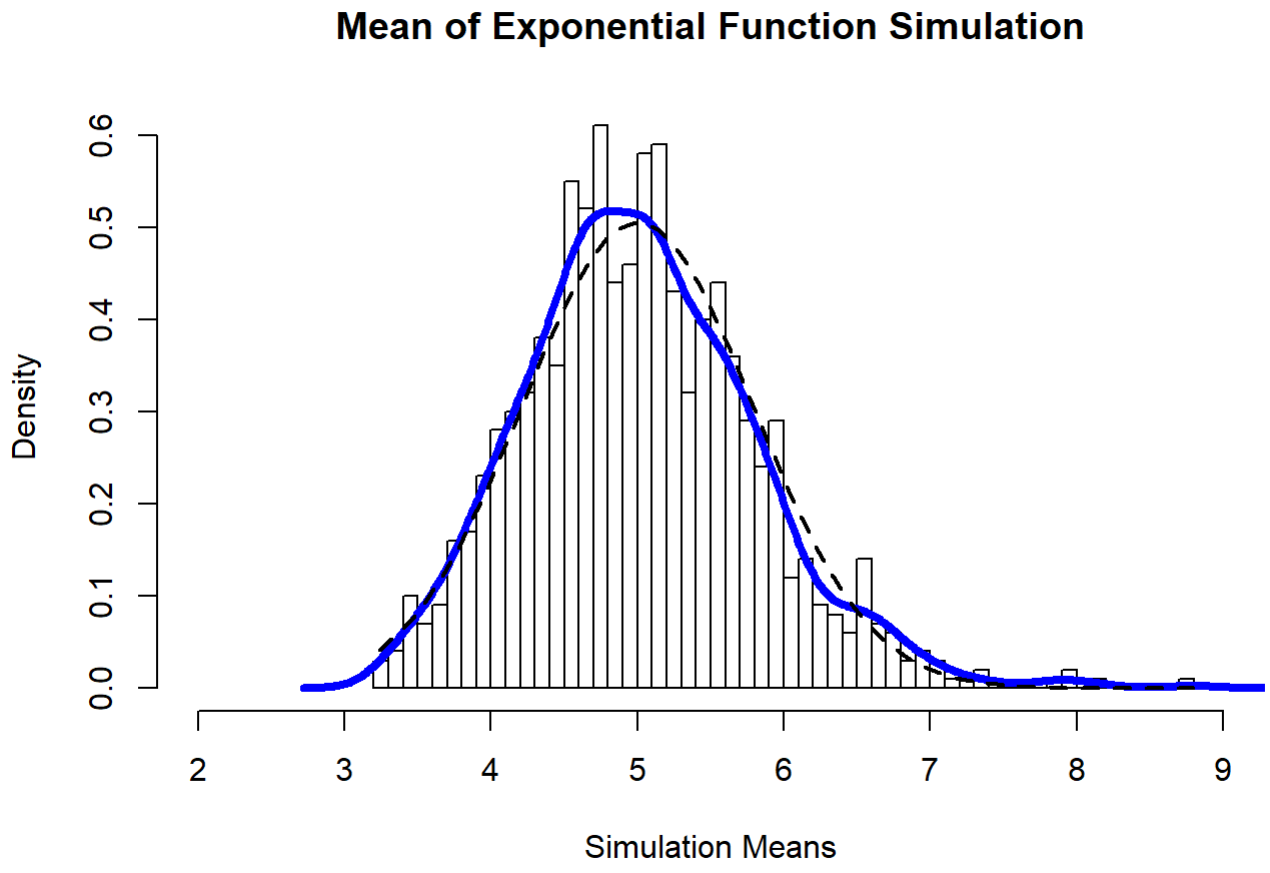
```
## [1] "Actual variance is:  0.626"
```

Based on the previous results, the actual variance is very close to the theoretical variance.

Question 3: Distribution

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

```
# Histogram with distribution curve included
hist(means_expsim, prob=TRUE, main = "Mean of Exponential Function Simulation", breaks = 40, xli
m = c(2,9), xlab = "Simulation Means")
lines(density(means_expsim), lwd=4, col="blue")

# Normal distribution line
x <- seq(min(means_expsim), max(means_expsim), length = 2*n)
y <- dnorm(x, mean = 1/lambda, sd = sqrt(((1/lambda)/sqrt(n))^2))
lines(x,y, pch = 20, lwd = 2, lty = 2)
```



Conclusion

As shown above, the distribution of means of the simulated exponential distributions follows a normal distribution due to the Central Limit Theorem. If the number of samples increase (currently at 1000), the distribution should be even closer to the standard normal distribution (the dotted line, above). The blue line is the simulated curve.