Course Project

Zach Dungan

3/1/2022

## Part 1: Simulation Exercises

**Overview:**

Investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution is simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. Investigate the distribution of averages of 40 exponentials. 1000 simulations are needed

**Questions:**

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.

# Set random seed  
set.seed(4569)  
  
# Set variables  
lambda <- 0.2  
n <- 40  
sim <- 1000  
  
# Run simulations  
sim\_run <- replicate(sim, rexp(n, lambda))  
  
# Calculate mean of simulations  
sim\_mean <- apply(sim\_run, 2, mean)

**Question 1: Theoretical Mean vs. Sample Mean**

Theoretical mean:

mean\_t <- lambda^-1  
mean\_t

## [1] 5

Sample mean:

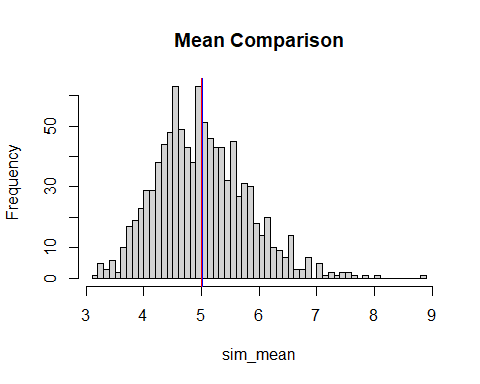
mean\_sample <- mean(sim\_mean)  
mean\_sample

## [1] 5.008682

Comparison:

Below is a histogram comparing the means. The line in red is the theoretical mean, blue is sample mean

hist(sim\_mean, main = "Mean Comparison", breaks=50)  
abline(v = mean\_t, col = "red")  
abline(v = mean\_sample, col = "blue")

 From the graph, we can see the sample mean of 5.0086 is very close to the theoretcial mean

**Question 2: Theoretical Variance vs. Sample Variance**

For theoretical variance, the formula is (lambda\*sqrt(n))^-2:

var\_t <- (lambda\*sqrt(n))^-2  
var\_t

## [1] 0.625

Sample variance:

var\_sample <- var(sim\_mean)  
var\_sample

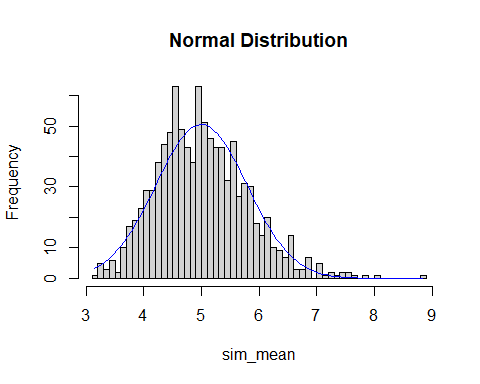
## [1] 0.6281728

Comparing the two, we can see the theoretical and sample variances are very close

**Question 3: Distribution**

This last question will show if the distribution is approximately normal

hist(sim\_mean, main = "Normal Distribution", breaks=50)  
  
x <- seq(min(sim\_mean), max(sim\_mean), length=100)  
y <- dnorm(x, mean=1/lambda, sd=(1/lambda)/sqrt(n))  
lines(x, y\*100, col="blue")



From the graph, we can see that the exponential distribution appears to be normal centered on our mean.