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:

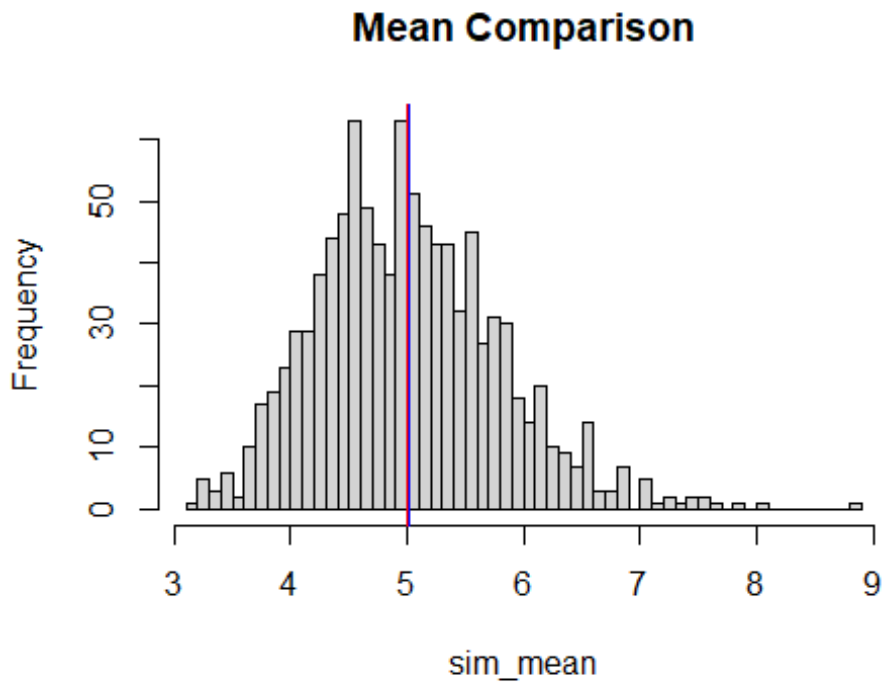
```r
mean_sample <- mean(sim_mean)
mean_sample
```

```
## [1] 5.008682
```

Comparison:

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

```r
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:

```r
var_t <- (lambda*sqrt(n))^-2
var_t
```

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

Sample variance:

```r
var_sample <- var(sim_mean)
var_sample
```

```
## [1] 0.6281728
```

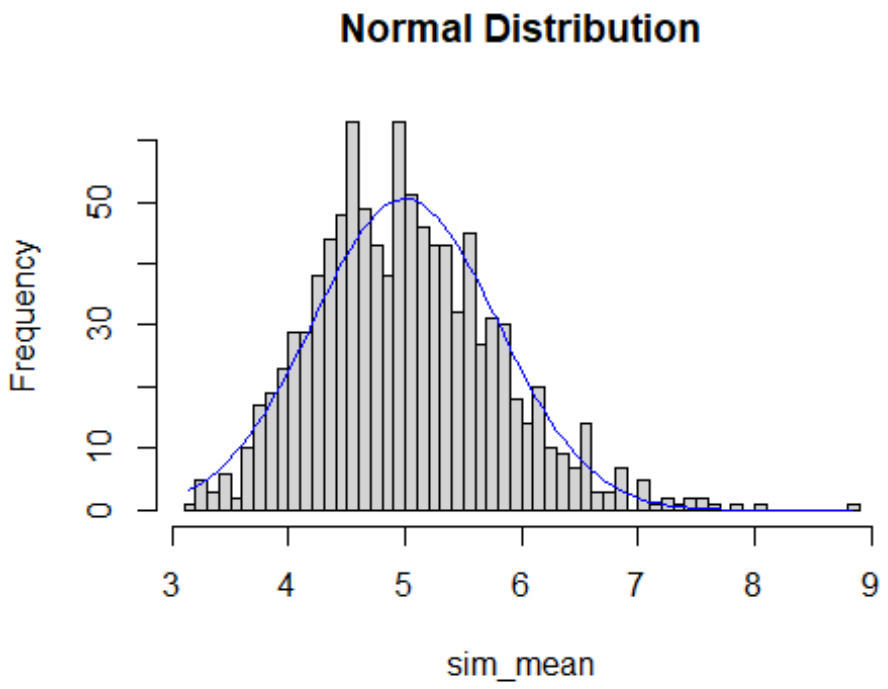
Comparing the two, we can see the theoretical and sample variances are very close

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.