Part 1: Exponential Distribution Simulation

Nikolai Alexander

Overview

For *Part 1* of the Course Project, we are investigating exponential distribution and comparing it with the Central Limit Theorem. We are doing this in 3 ways: Comparing the **Sample Mean vs Theoretical Mean**, Comparing the **Sample Variance vs Theoretical Variance**, and lastly, demonstrating that the **Distribution** is approximately normal.

Simulations

Before we run any simulations, we want to make sure this study is reproducible. We can do this by setting the seed.

```
set.seed(19930508)
```

The number the seed is set to does not necessarily matter, as long as you use the same seed for each experiment. I chose my birthday as an easy number to remember.

The next this we need to do is set all necessary variables for the simulation. In this case, we only have 3: **lambda** (l), the **number of exponentials** (n_exp), and the **number of simulations** (n_sim).

```
l = 0.2
n_exp = 40
n_sim = 10^3
```

Now to run the simulations. We can use a *for loop* to get through all 1000 simulations

```
sim <- replicate(n_sim, rexp(n_exp,l))
dim(sim)
```

```
## [1]   40 1000
```

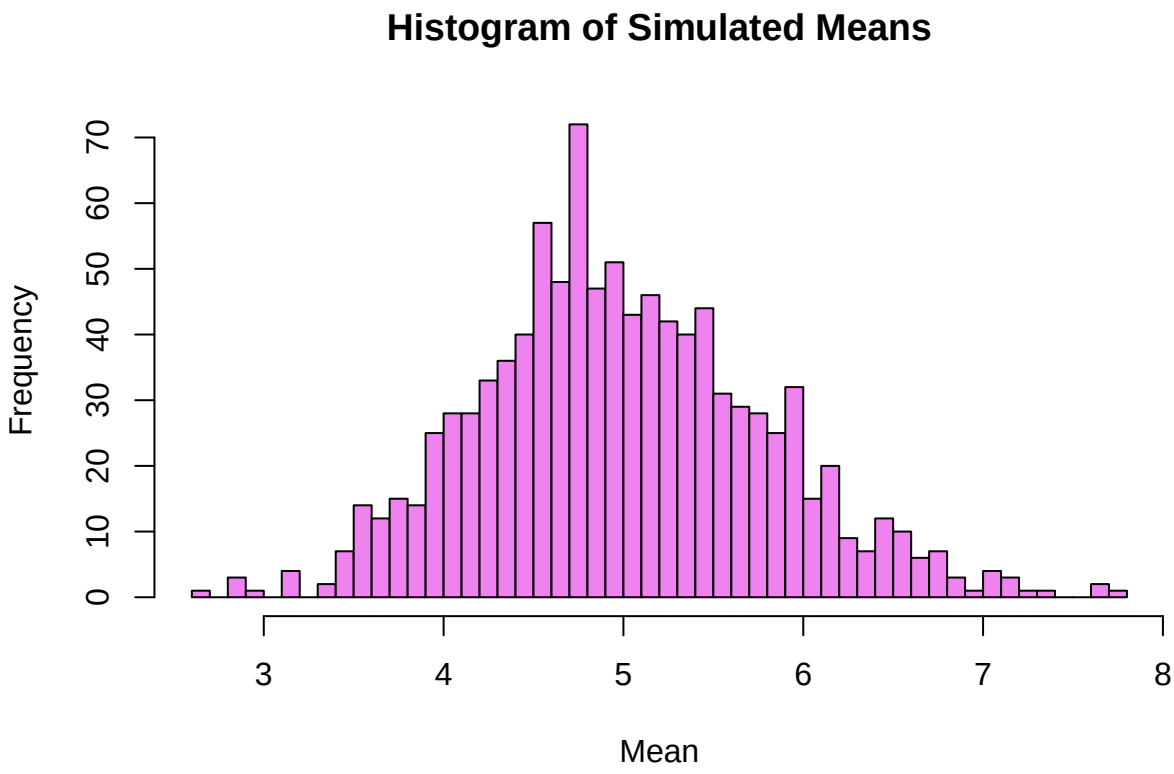
Now that we ran the simulation, we got a 40x1000 matrix of all the possible simulations. From there, need find the mean of all 40 rows for each column.

```
sim_mean <- apply(sim, 2, mean)
head(sim_mean)
```

```
## [1] 6.780708 5.504855 5.307615 4.486720 4.878424 5.222121
```

We can visualize these means in a histogram...

```
hist(
  sim_mean,
  breaks = 50,
  col = c("violet"),
  xlab = "Mean",
  ylab = "Frequency",
  main = "Histogram of Simulated Means"
)
```



Looking at this data, we know the mean is somewhere close to 5. Let's compare it with the theoretical mean.

Sample Mean vs Theoretical Mean

Before we compare the **Sample Mean** to the **Theoretical Mean**, we need to find the **Theoretical Mean**. We know the **mean** is the same as the **expected value**, which for an **exponential distribution** is *1/lambda*.

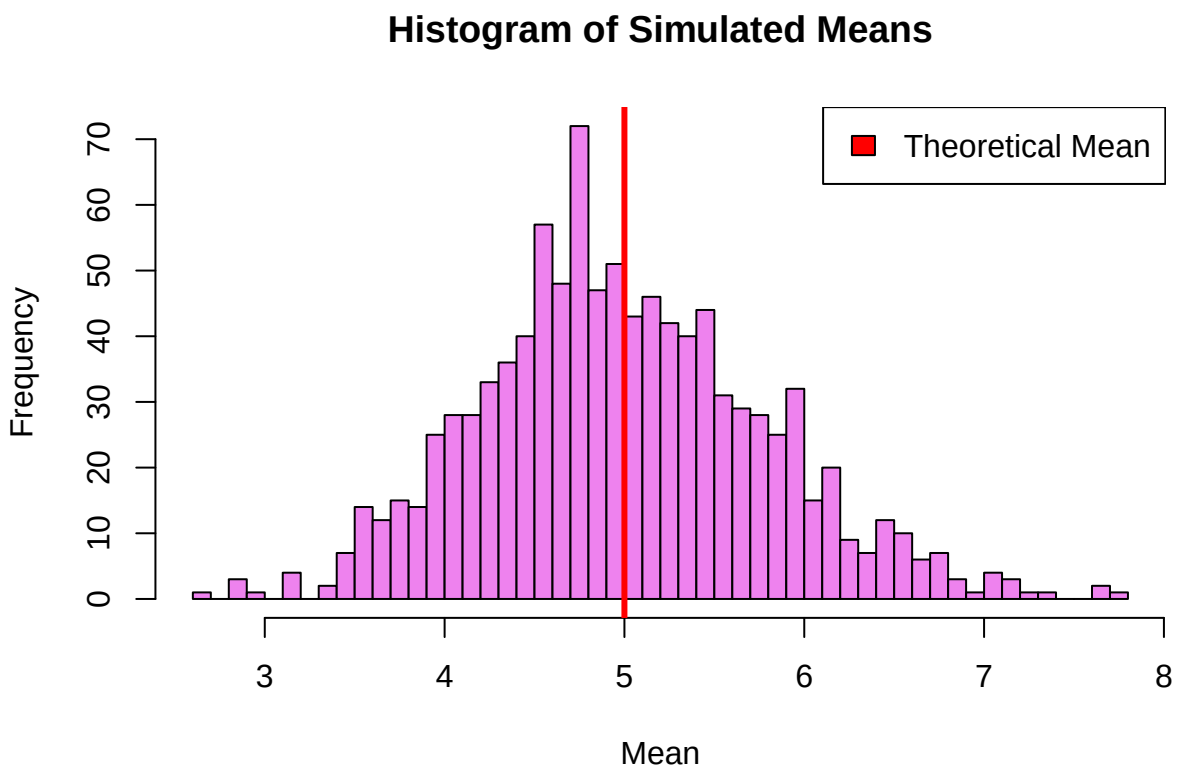
```
theor_mean <- 1/l
theor_mean
```

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

We can visualize this on our previous histogram.

```
hist(
  sim_mean,
  breaks = 50,
  col = c("violet"),
  xlab = "Mean",
  ylab = "Frequency",
  main = "Histogram of Simulated Means"
)
abline(v = theor_mean, col = "red", lwd = 3)
  legend(
  "topright",
  c("Theoretical Mean"),
```

```
  fill = c("red")
)
```

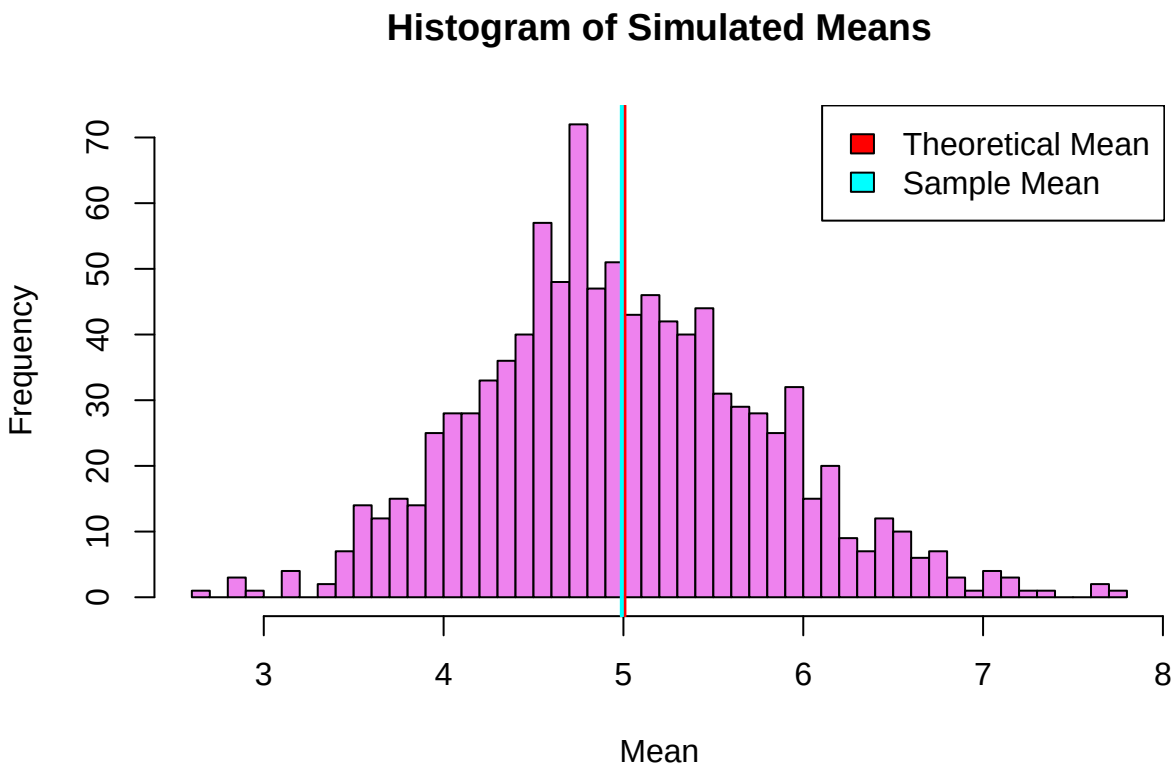


As we can see, the **theoretical mean** is larger than the **sample mean**, but by how much? We need to calculate the **sample mean** to find the difference.

```
samp_mean <- mean(sim_mean)
samp_mean
```

```
## [1] 4.991089
```

```
hist(
  sim_mean,
  breaks = 50,
  col = c("violet"),
  xlab = "Mean",
  ylab = "Frequency",
  main = "Histogram of Simulated Means"
)
abline(v = theor_mean, col = "red", lwd = 3)
abline(v = samp_mean, col = "cyan", lwd = 2)
legend(
  "topright",
  c("Theoretical Mean", "Sample Mean"),
  fill = c("red","cyan")
)
```



This is actually much closer than I anticipated, with a difference of only **0.0089111**. Now how do the **Variances** compare?

Sample Variance vs Theoretical Variance

We know the **variance** is proportional to the **standard deviation** *squared*. For an **exponential distribution**, the **standard deviation** is equal to *(1/lambda)/sqrt(# of experiments)*. Therefore, the **theoretical variance** is...

```
theor_sd = (1/l)/sqrt(n_exp)
theor_var = theor_sd^2
theor_var
```

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

We can find the **sample variance** easily with the *sd()* function

```
samp_sd = sd(sim_mean)
samp_var = samp_sd^2
samp_var
```

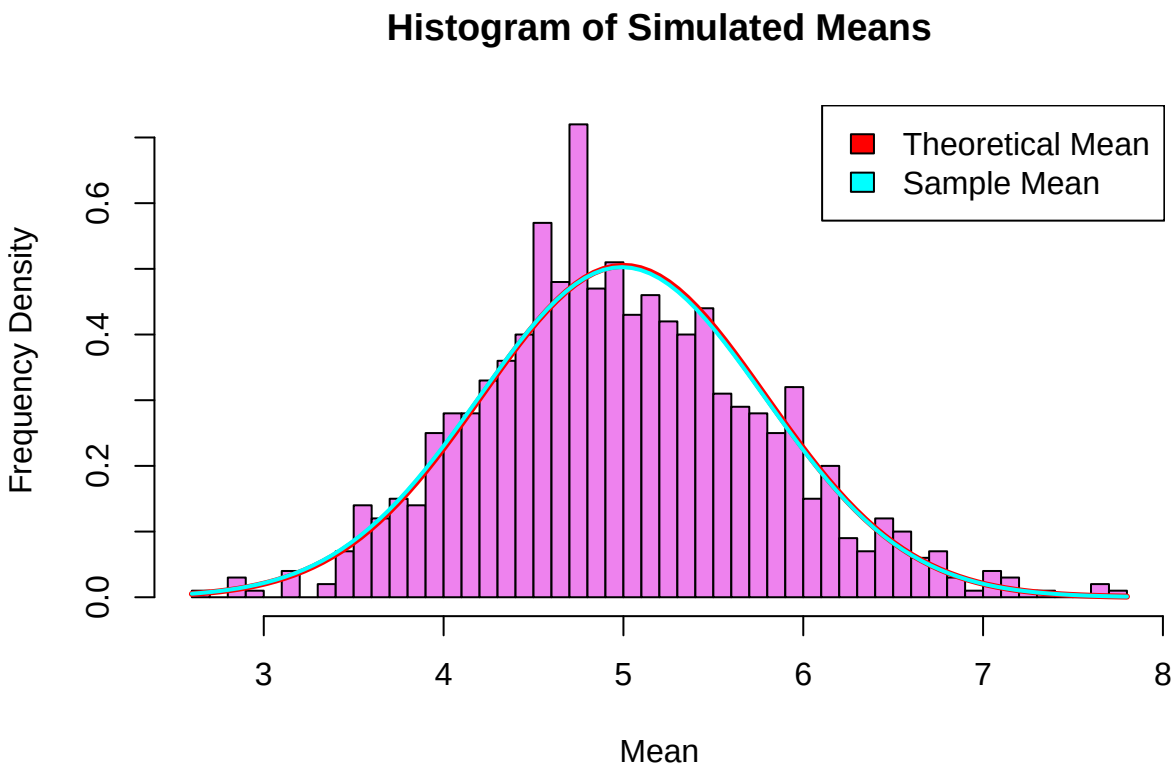
```
## [1] 0.6300698
```

These **variances** have a difference of **0.0050698**.

Distribution

The last thing we need to check, is this a normal distribution? We can figure this out by using the *means* and *standard deviations* from the previous sections.

```r
hist(
  sim_mean,
  probability = TRUE,
  breaks = 50,
  col = c("violet"),
  xlab = "Mean",
  ylab = "Frequency Density",
  main = "Histogram of Simulated Means"
)
curve(
  dnorm(x, mean = theor_mean, sd = theor_sd),
  col = "red",
  lwd = 3,
  add = TRUE
)
curve(
  dnorm(x, mean = samp_mean, sd = samp_sd),
  col = "cyan",
  lwd = 2,
  add = TRUE
)
legend(
  "topright",
  c("Theoretical Mean", "Sample Mean"),
  fill = c("red","cyan")
)
```



As we can see, the normal curve approximately follows the structure of the histogram, minus a few outliers; therefore, we can say the distribution is **normal**