Statistical Inference Course Project

Part 1: Study the mean of 40 exponentially distributed random variables.

To study the distribution of the mean of the exponential distribution, we simulate any samples of size 40.

```
# Lambda
lambda <- 0.2

# Sample Size
N <- 40

# Number of simulations
num_sims <- 10000

# Take all the samples at once, then reshape them into groups of N.
es <- matrix(rexp(N * num_sims, rate=lambda), num_sims, N)

# Now take the means of each group
ms <- apply(es, 1, mean)
```

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

When you sum iid random variables, the mean of the distribution is the sum of the means of the random variables. Thus, the mean of the distribution of the sum of 40 exponentials would be $40 * \frac{1}{\lambda} = 40 * \frac{1}{0.2} = 40 * 5 = 200$. When you divide a random variable by a value, you divide its mean by that value too. To get the mean of 40 exponentials from the sum, we divide by 40: $\frac{200}{40} = 5$. So the theoretical mean of this distribution is just the same as the mean of each exponential sample.

The mean that we get from our experiments is 4.9908033, which is quite close to 5.

2. Show how variable it is and compare it to the theoretical variance of the distribution.

When you sum iid random variables, the variance of the resulting distribution is the sum of the variance of each variable. The standard deviation of a single exponential is $\frac{1}{\lambda}$ so the variance is $(\frac{1}{\lambda})^2 = (\frac{1}{0.2})^2 = (5)^2 = 25$. So the variance of the sum of $40 * 25 = 1000$

When you divide a random variable by a value, you divide its variance by the value squared. To get the mean of 40 exponentials from the sum, we divide by 40, so we divide the variance by 40^2. The result is $\frac{1000}{40^2} = 0.625$.

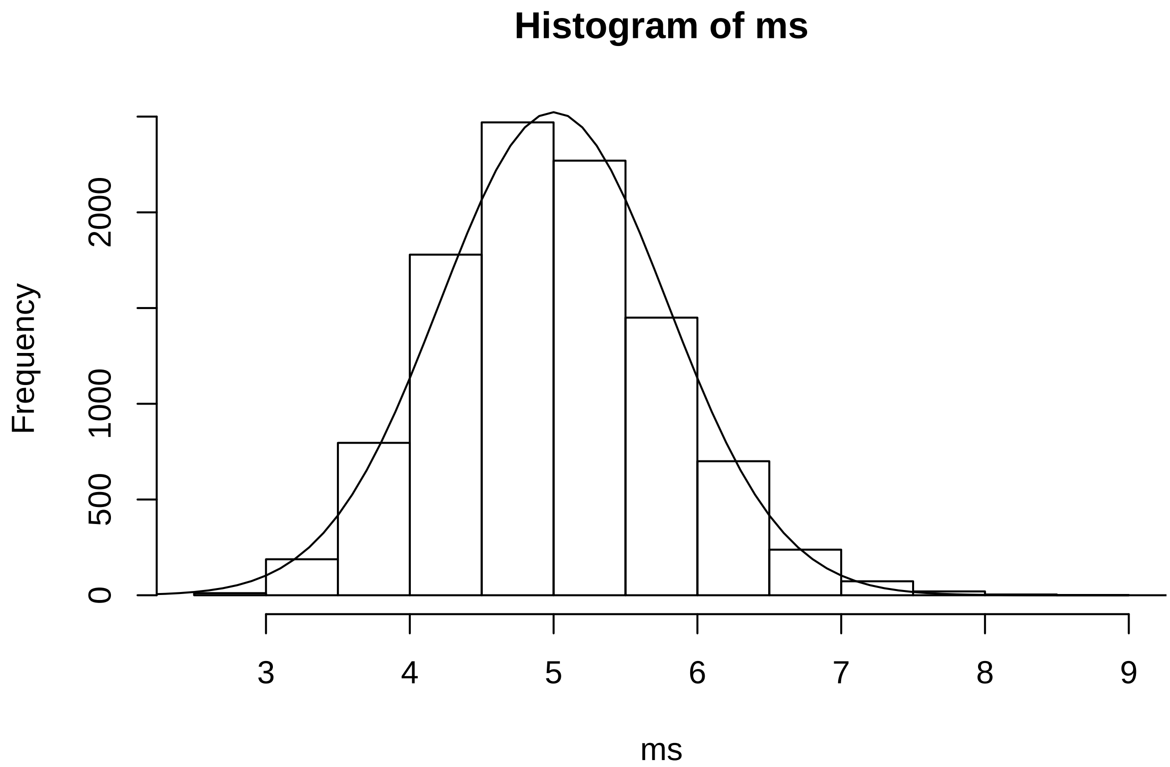
The variance that we get from our experiments is 0.6191159, which is quite close to .625.

3. Show that the distribution is approximately normal. Note that for point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

When you plot the results as a histogram, it produces the characteristic bell shape of a normal distribution.

```
# plot the result
hist(ms)

# overlay a normal
xs <- seq(0, 2/lambda, by=0.1)
lines(xs, num_sims/2 * dnorm(xs, mean=1/lambda, sd=1/lambda/sqrt(N)))
```



This is very different from just plotting the exponentials directly, which doesn't look normal at all:

```r
# plot the result
hist(es)
```