Statistical Inference Project Part 1

Kyle Day

Overview

In this report I will use simulations to demonstrate the following characteristics of the distribution of the mean of 40 exponentials.

```
1. Show sample mean and compare it to the theoretical mean.
2. Show how variable the sample is and compare it to the theoretical variance.
3. Show that the distribution is approximately normal.
```

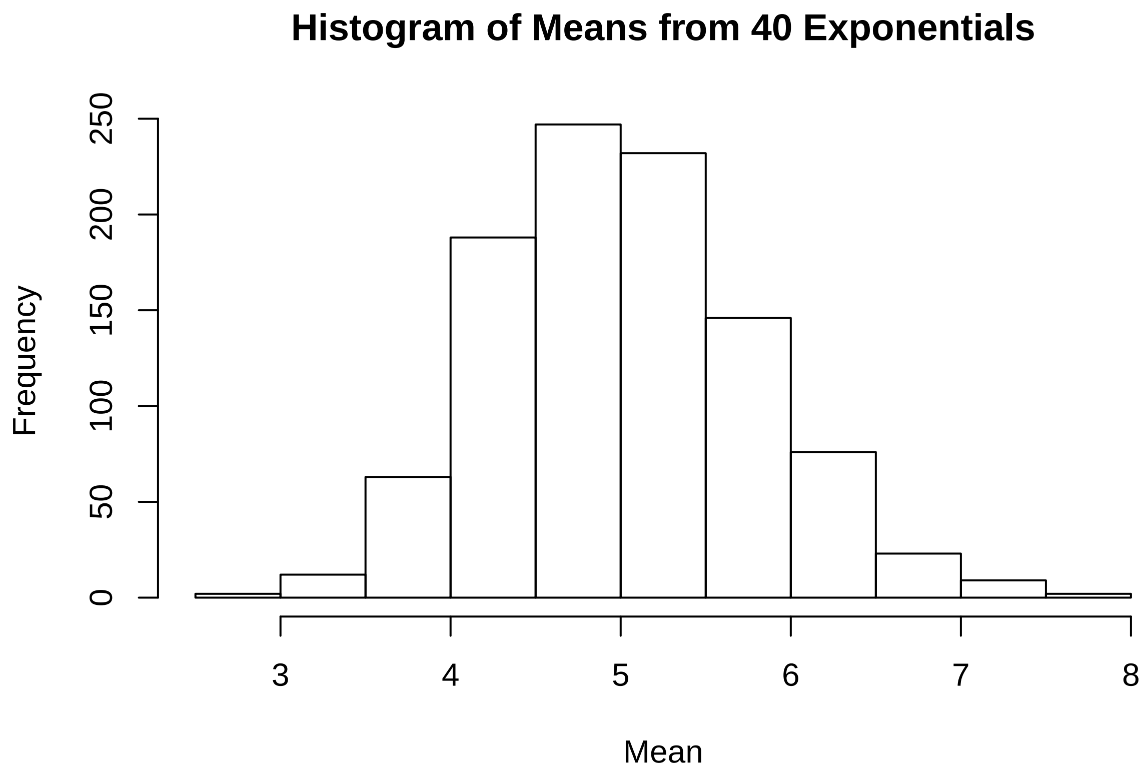
Simulating the data

```
## First we need to create our simulation.
lambda <- 0.2   ##We're told to keep Lambda at 0.2
n <- 40         ##means generated off of sample size 40
sims <- 1000    ## we need 1000 simulations
set.seed(99)   ##needed so that data is reproducible

##create the simulation data
data <- matrix(rexp(n*sims, lambda), sims)
## use n*sims becuase we need 40 unquie samples for each 1000 sims

mean_data <- data.frame() ##a place to put the mean

## Calculate the mean of each row in data and place it in mean_data
for(i in 1:nrow(data)){
    mean_data[i,1] <- mean(data[i,])
}

head(mean_data)
```

```
##         V1
## 1 4.558530
## 2 5.325106
## 3 4.764438
## 4 4.675887
## 5 6.776891
## 6 6.230355
```

Step 1: Show sample mean and compare it to the theoretical mean.

```
hist(mean_data[,1], xlab = "Mean", main = "Histogram of Means from 40 Exponentials")
```




```
## We must point to the column even though there is only one.
## Otherwise hist will try to read the whole data.frame and cause an error
```

The center of this histogram is at 5.

The Mean of our sample should approximate our theoretical mean.

In theory our mean would be 1/lambda.

```
theory_mean <- 1/lambda
actual_mean <- mean(mean_data[,1])
```

Therefor our data seems to support this with an actual mean of `r actual_mean` comparied to the theoretical mean at 5.

Step 2: Show how variable the sample is and compare it to the theoretical variance.**

Theoretically the variance should be (1/lambda)^2/n.

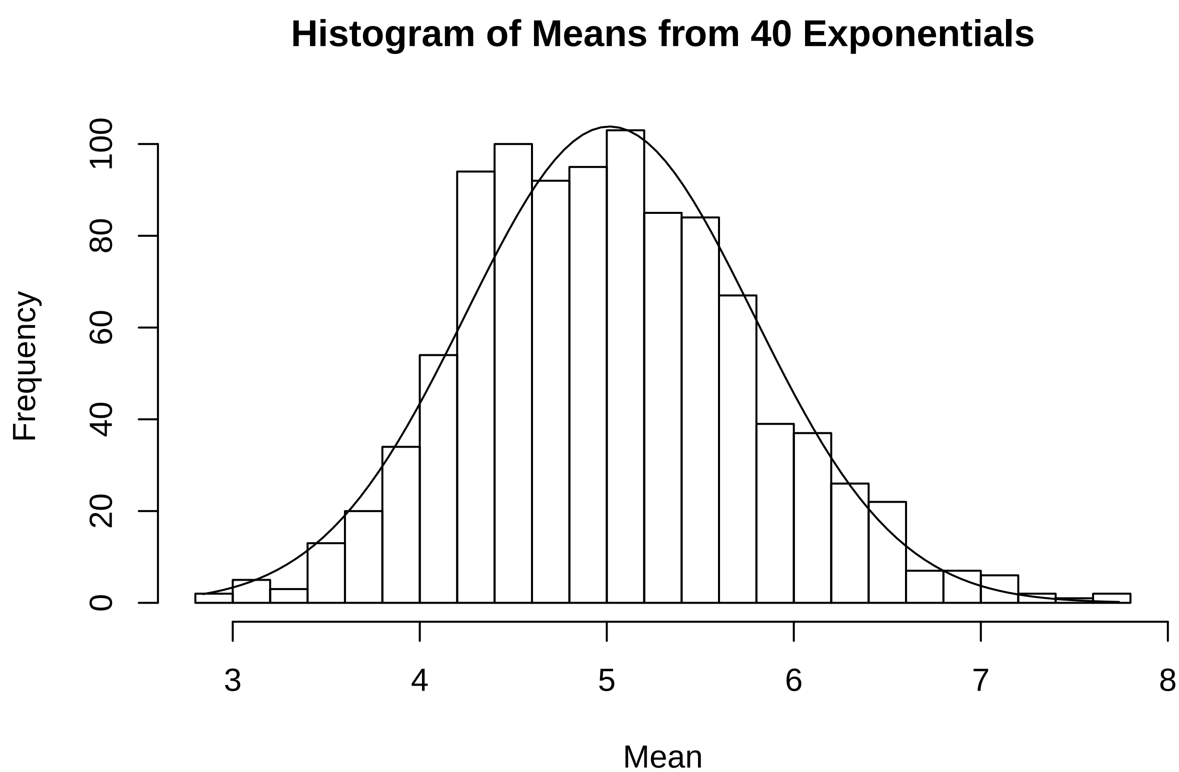
```
theory_var <- (1/lambda)^2/n
actual_var <- var(mean_data[,1])
```

Therefor our data seems to be once again supported with an actual variance of `r actual_var` comparied to the theoretical variance at 0.625.

Step 3: Show that the distribution is approximately normal.

To show this lets see how our data fits a curve that is generated from setting a normal distribution equal to have the same mean and standard deviation of our data. *That is to say what our data would look like if it was normalized with out actually normalizing it.*

```
##first lets show our data, but I want to see more granularity so
##lets set breaks to 25 for mor bins.
main <- "Histogram of Means from 40 Exponentials"
data_plot <- hist(mean_data[,1], xlab = "Mean", main = main, breaks = 25)

## Then lets calculate a few need peices of information.
##we already have our mean, but what of our Standard Deviation.
actual_sd <- sd(mean_data[,1])
##before we create our approximation of our normal we need an x vaule
## to get our x we are you to create a seqence of 100 point from our min to our max.
x <- seq(min(mean_data[,1]), max(mean_data[,1]), length.out= 100)

norm <- dnorm(x, mean = actual_mean, sd = actual_sd)
##before we graph or line we need to know how much to scale by.
multi <- data_plot$counts/data_plot$density

lines(x, norm * multi[1])
```



Our data seems to not fall to far away from our curve meaning that the data is aproximately normally distributed.