Statistical Inference Project1 - Simulated Data

Diane Clow

August 22, 2015

Introduction

This is the Course Project for Statistical Inference. The assignment asks students to “investigate the exponential distribution in R and compare it with the Central Limit Theorem.”

To show this I will:

- Part 1: Show the sample mean and compare it to the theoretical mean of the distribution
- Part 2: Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution
- Part 3: Show that the distribution is approximately normal

The assingment asks us to run 1000 simulations, with lambda = .2 and the sample size (n) = 40. To make everything consistant I am choosing the seed of 10.

```
set.seed(10)
library(ggplot2)
lambda <- 0.2
num_sim <- 1000
n <- 40
```

Simulations

To gather data for this report 1000 simulations were run. Each simulation took 40 random samples of an exponental distribuation with lambda = .2. Using replicate a 40 by 1000 matrix was returned.

Using the function apply, I was able to get a vector of length 1000, which contained the mean of each simulations.

```
SamExp <- replicate(num_sim, rexp(n, lambda))
MeanExp <- apply(SamExp, 2, mean)
```

Part 1

Part 1 asked to calcuate the sample mean and compare it to the theoretical mean. The sample mean is calculated below using the mean function. The Theoretical mean for an exponental distribution is 1/lambda.

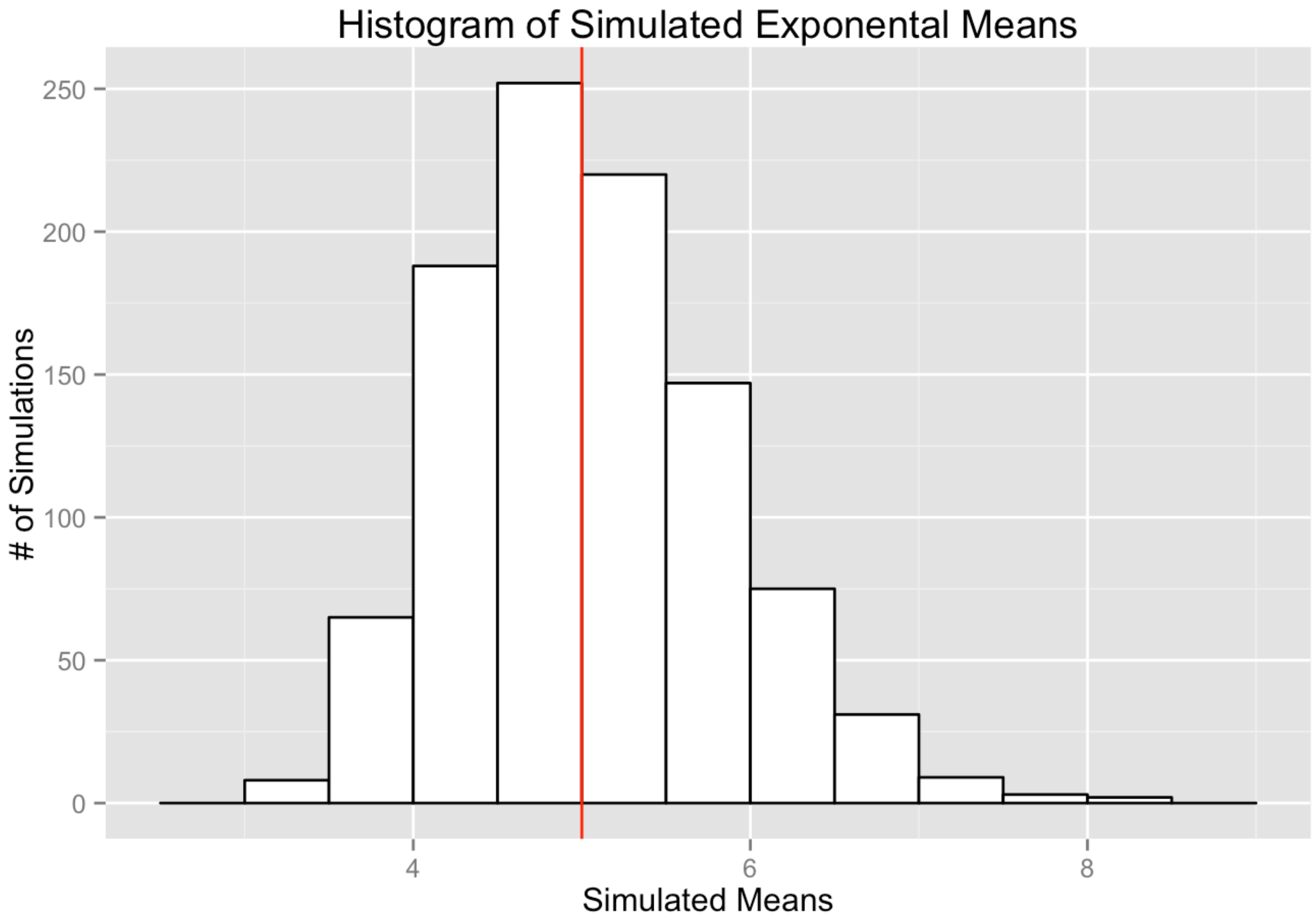
```
##Sample Mean
mean_sample <- mean(MeanExp)
mean_sample
```

```
## [1] 5.04506
```

```
## Theoretical Mean = 1/lambda
mean_theory <- 1/lambda
mean_theory
```

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

```
## histogram showing the two
## add title and lables
ggplot() + aes(MeanExp)+ geom_histogram(binwidth=.5, colour="black", fill="white") + geom_vline(xintercept = 5, colour = "red") + labs(list(title = "Histogram of Simulated Exponental Means", x = "Simulated Means", y = "# of Simulations"))
```



The histogram above represents the sample means for each simulation run. The red line represents the calculated Theoretical mean.

Part 2

Part 2 asks to calculate the sample standard deviation and variance and compare them to the theoretical standard deviation and variance. The theoretical standard deciation is calculated by dividing 1/lambda by the square root of the sample size (n).

```
## Sample Standard Deviation
stdv_sample <- sd(MeanExp)
stdv_sample
```

```
## [1] 0.7982821
## Theoretical Standard Deviation
stdv_theory <- (1/lambda)/sqrt(n)
stdv_theory
```

```
## [1] 0.7905694
```

```
## Sample Variance
var_sample <- stdv_sample^2
var_sample
```

```
## [1] 0.6372544
```

```
## Theoretical Variance
var_theory <- stdv_theory^2
var_theory
```

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

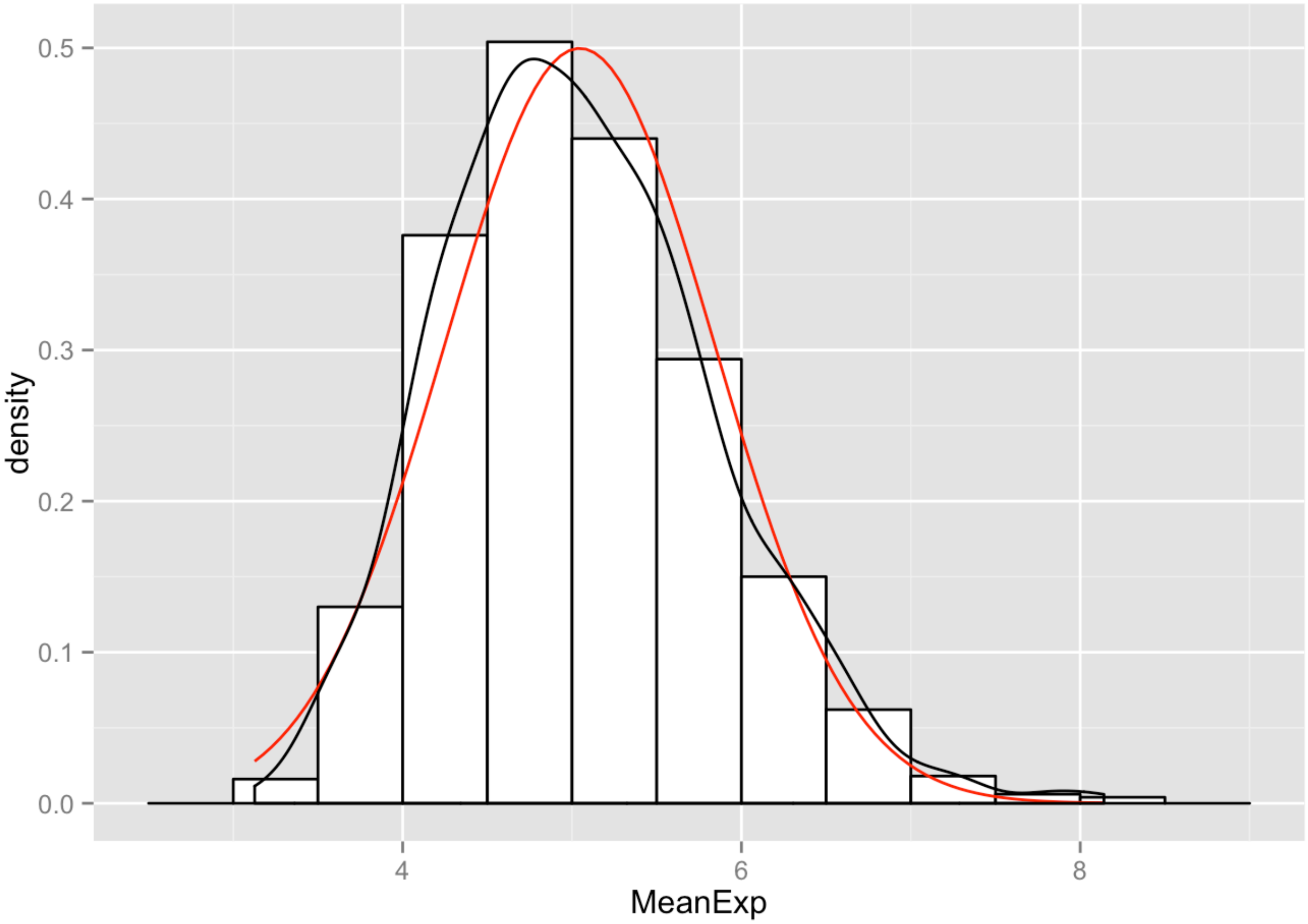
As you can see above the sample and theoretical values are very similar.

Part 3

Part 3 asks to show that the disribution is approximately normal.

Below is the same histogram that was shown above. The red curve is a normal curve with the calculated mean and standard deviation. The black line is the calculated density curve for the sample data.

```
ggplot() + aes(MeanExp)+ geom_histogram(aes(y = ..density..), binwidth=.5, colour="black",
fill="white") + stat_function(fun = dnorm, colour = "red", arg=list(mean=mean(MeanExp), sd=
sd(MeanExp))) + geom_density()
```



As you can see in the graph, the two curves are very similar, showing that the sample distribution is approximatly normal.