P1_Simulartion_Exercise

Jerad Acosta

December 21, 2014

Load Appropriate Libraries and set seed for reproducibility

```
library(ggplot2); library(dplyr)
set.seed(741)
```

Creating 1000 Simulations:

To simulate 1000 exponential distributions with size 40 and **lambda** 0.2 we create a variable, **nosim**, to represent the number of simulartions,**n** for the size of each distribution and set **lambda = 0.2**

```
nosim <- 1000
n <- 40
lambda = 0.2
```

Next we create the simulated exponential distributions by filling a matrix with dimensions 1000 by 40 where each of the 1000 rows represents a set of 40 variables (columns)

```
simulation <- matrix(rexp(nosim * n, lambda), nosim)
print(paste(nrow(simulation),"rows","by", ncol(simulation),"columns"))
```

```
## [1] "1000 rows by 40 columns"
```

The Matrix was filled in with 40,000 variables of an exponential distribution using the function **rexp(nosim * n, lambda)** ***NOTE:*** That since each variable is iid the order in which we chose our groups of 40 variables does not matter as long as it is not related to their values This is the of taking 1000 samples of size 40 from the exponential distribution*

Q1 with regard to the measured mean and the expected or theoretical mean

First we take the mean of each distribution as well as their cumulative mean and variability

```
colMeans <- apply(simulation, 1, mean)
meanAll <- mean(colMeans)
sdAll <- sd(colMeans)
```

```
## The observed mean of samples is 5.00668039159687
```

Where the theoretical mean is

$$\frac{1}{\lambda} = \frac{1}{0.2} = 5.$$

Q2 compare variance of sample means with theoretical

```
## the observed standard deviation of means is 0.777705716296829
```

While the theoretical standard deviation is

$$\frac{\sigma}{\sqrt{n}} = \frac{1/\lambda}{\sqrt{40}} = 0.7906.$$

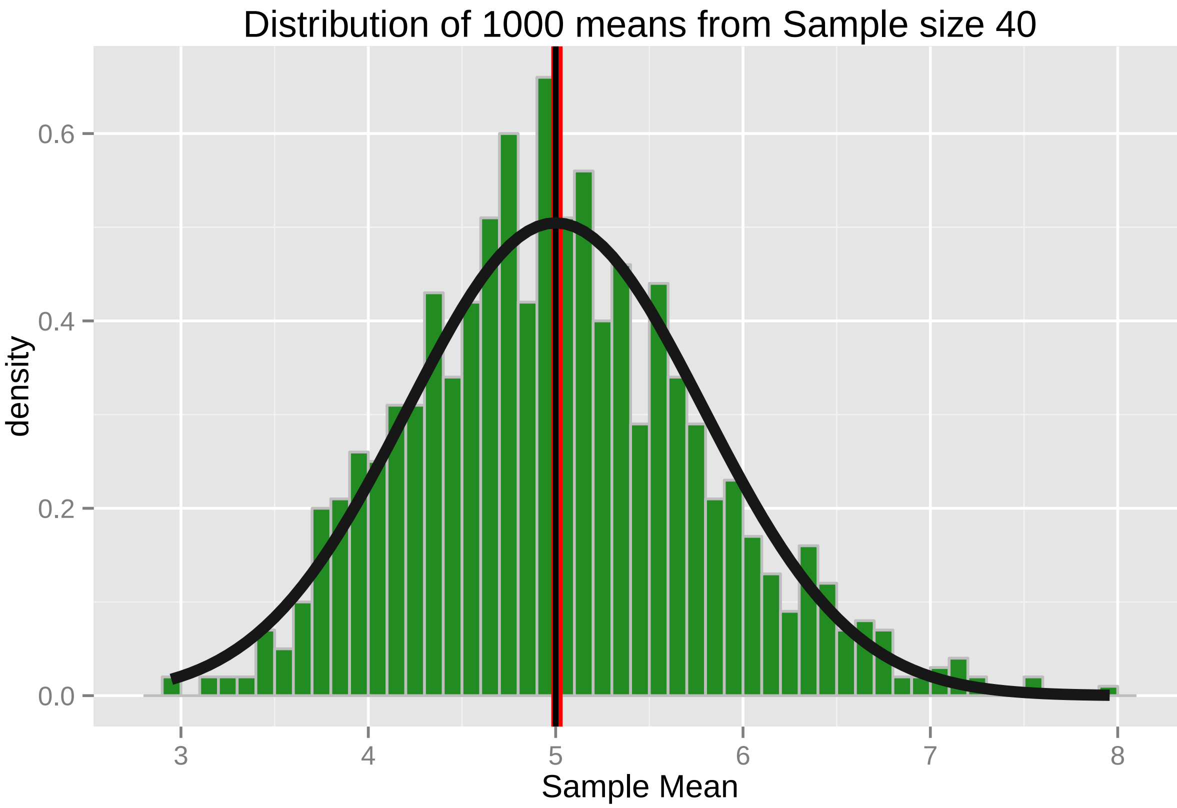
The Variance of our sample means is 0.6048 With a theoretical mean of

$$\frac{\sigma^2}{n} = \frac{1/\lambda^2}{40} = 0.625.$$

Q3 Show the distribution is approximately normal

Now we Use ggplot2 to plot the density of the sample means with the normal distribution $N\left(1/\lambda, \frac{1/\lambda}{\sqrt{40}}\right)$ superimposed.

```r
# First we plot the relative density of each sample mean
ggplot(data = data.frame(colMeans), aes(x = colMeans)) +
    geom_histogram(binwidth = 0.1, aes(y= ..density..), fill = "forest green",
    color = "gray") + xlab("Sample Mean") +
# Place a Yellow vertical line to show where the distribution is centered
    geom_vline(xintercept = meanAll, size = 2, color = "red") +
# Use a Black Line to represent the theoretical mean
    geom_vline(xintercept = 5, size = 1, color = "black") +
# Superimpose the Normal Distibution
    stat_function(fun = dnorm, color = "gray9", size = 2,
    arg = list(mean = 1/lambda, sd = 1/lambda*(1/sqrt(n)))) +
    ggtitle("Distribution of 1000 means from Sample size 40")
```



It is quite apparent that both the **sample mean** or *center of the distribution*, in **red** as well as the **variation** are well modeled by a normal distribution which is outlined in **black** along with a black **theoretical** or *true* **mean**

But just as extra fun:

We can use the confidence interval for $1/\lambda$:

$$\overline{x} \pm 1.96\frac{s}{\sqrt{n}}$$

to see just how many of our samples captured the theoretical mean with their appropriate confidence intervals with a T value of 1.96

```
simulation <- data.frame(simulation)
sds <- apply(simulation, 1, sd)
simulation <- data.frame(simulation,colMeans, sds)
simulation <- simulation %>% mutate(CI =
                (colMeans - 1.96 * sds / sqrt(n) < 1/lambda) &
                (colMeans + 1.96 * sds / sqrt(n) > 1/lambda))
percent <- mean(simulation$CI)
```

According to these calculations 93.6% of the samples means contained the true, theoretical mean