Statistical inference course project part I

Halley Wang

October 25, 2015

This report shows the relationship between the theoritical mean and variance with the sample mean ar

```
## Loading the necessary packages
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## Initialize the sample size, times of simulations and parameters for lambda:
sampleSize = 40
n = 1000
lambda = 0.2

## For controlling the random variables:
set.seed(123)
```

```
ExpDisSample = function(size = sampleSize, rate = lambda) {
        smpl = rexp(n = size, rate = rate)
        return (smpl)
}
```

This process generates a sample with 40 observations.

```
## make n times of stimulation:
ExpDisMeanStimulation = numeric()
for (i in 1:n) {
        ExpDisMeanStimulation = c(ExpDisMeanStimulation, mean(ExpDisSample()))
}

## Convert into a dataframe
ExpDisStimulation = data.frame(mean = ExpDisMeanStimulation)
head(ExpDisStimulation, 10)
```

```
##       mean
## 1  4.811212
## 2  5.360077
```

```
## 3  4.592871
## 4  4.900051
## 5  5.516619
## 6  5.612835
## 7  4.914947
## 8  4.504420
## 9  4.425668
## 10 5.014880
```

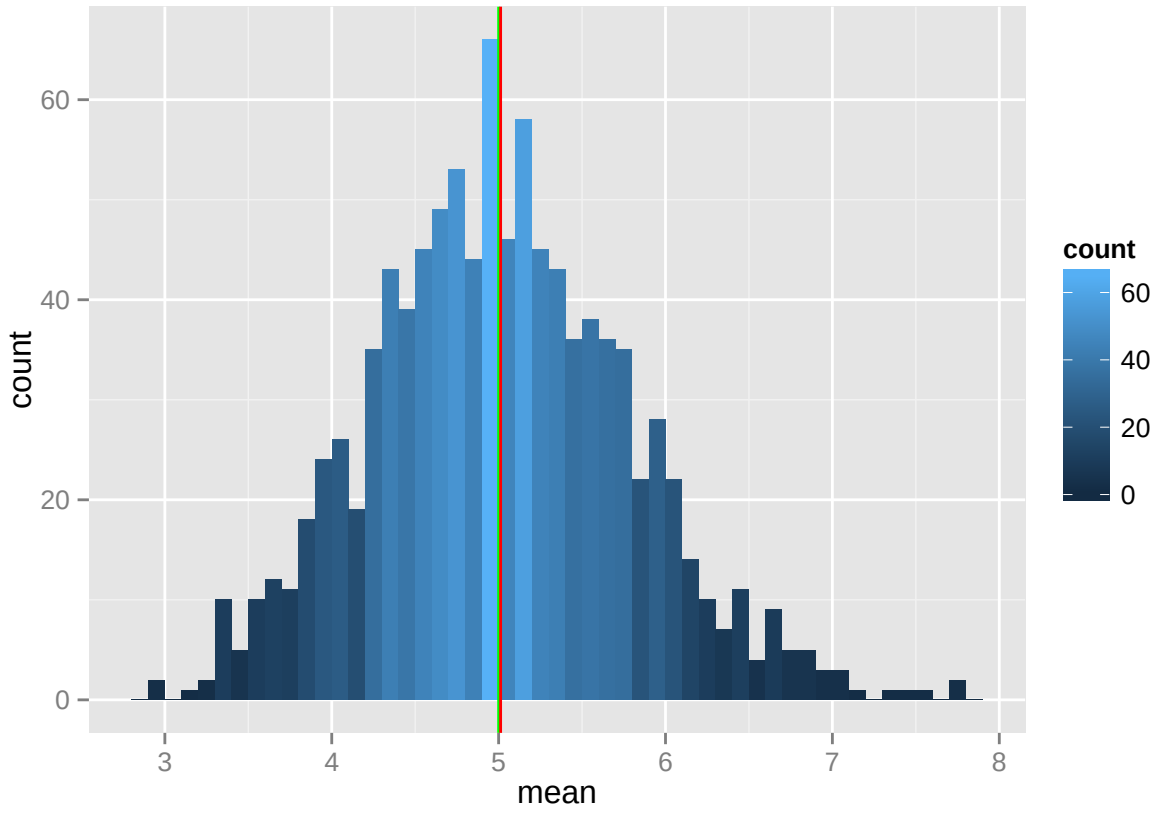
This process simulates 1000 replication of sample with 40 observation by repeating the process above

```
## Calculate the sample mean and sample variance:
sampleMean = mean(ExpDisStimulation$mean)
sampleVariance = var(ExpDisStimulation$mean)
```

We know that the random variable has mean lambda^-1 = 5, while the sample variance is lambda^-2 = 25
In comparison, the sample mean is 5.0119113, the sample variance is 0.6004928.

```
## Create ggplot object:
g1 = ggplot(data = ExpDisStimulation, aes(x = mean))

## Filling colors and
g1 + geom_histogram(binwidth = 0.1, aes(fill = ..count..)) + geom_vline(x = c(lambda^-1, mean(ExpDisStim
```



As we can see from the graph, the sample mean distribution is roughly Gaussian, which agrees on the
The Green line represents the theoritical mean, while the red represents the sample mean. As we can