Statistical Inference Course Project 1

Jackson Chou

May 29, 2016

Project Overview

In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations.I will investigate the distribution of averages of 40 exponentials. Note that this will require a thousand simulations.

Simulation

Here we run a thousand simulations of drawing the averages of 40 samples from the exponential distribution.

```
lambda <- 0.2
n <- 40
nsim <- 1000
set.seed(415)

#run 1000 simulations where each observation represents 40 samples drawn from exp distribution

simdata <- matrix(rexp(nsim*n, rate = lambda), nsim)

#calculate the mean for each row
rowmeans <- rowMeans(simdata)
rowmeans <- as.data.frame(rowmeans)
```

Each row in means represents the average of 40 samples.

Sample Mean versus Theoretical Mean

Let's compare our sample mean with the theoretical mean of the distribution. Keep in mind that the expected mean of the exponential distribution of lambda is 1/lambda.

```
theomean <- 1/lambda
theomean
```

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

Here our theoretical mean is 5.

Now let's take a look at our sample mean.

```
round(mean(rowmeans[,1]), 3)
```

```
## [1] 5.022
```

Notice that simulated sample means is 5.022, which is extremely close to our theoretical mean of 5.

Sample Variance versus Theoretical Variance

Let's compare our sample variance with the theoretical variance. The theoretical variance of exponential distribution is 1/lambda^2

```
theovar <- (1/lambda)^2/n
theovar
```

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

Here our theoretical variance is 0.625. Now let's take a look at our sample variance.

```
var(rowmeans[,1])
```

```
## [1] 0.598838
```

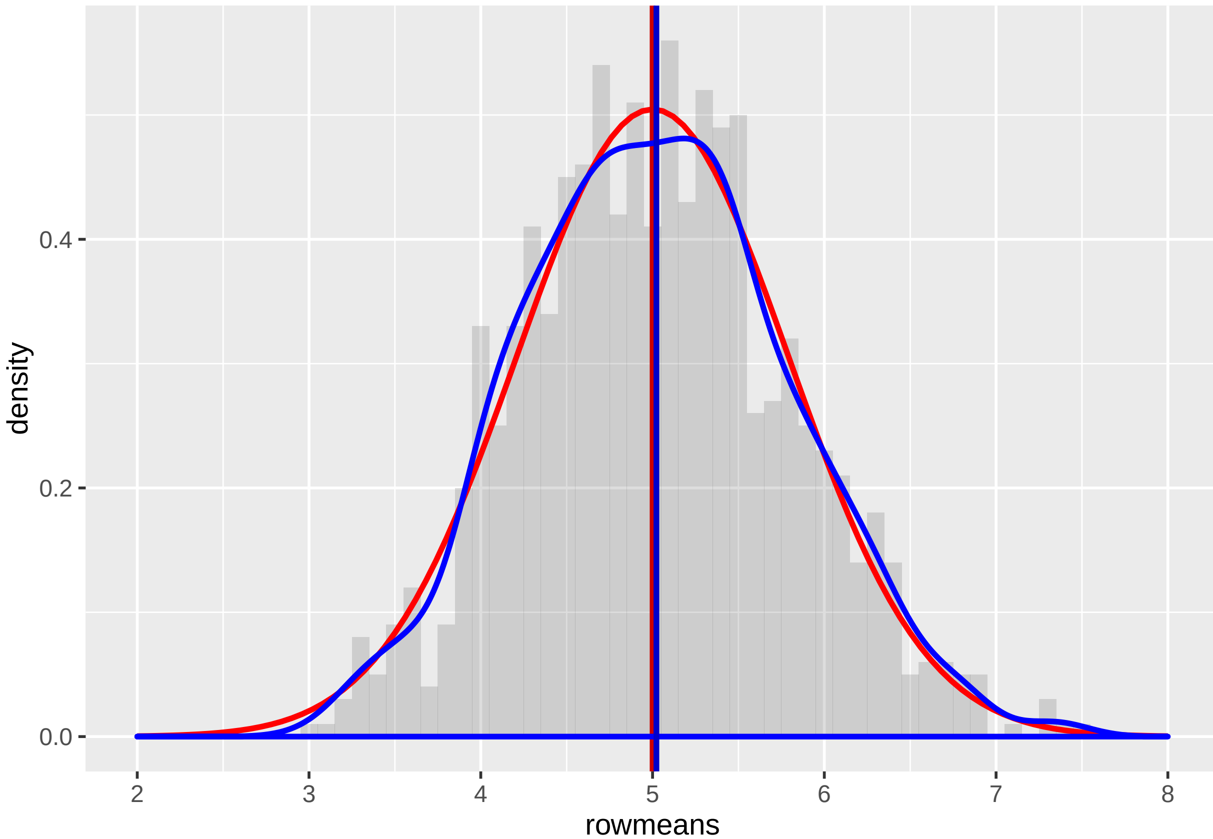
Our sample variance is 0.598, compared to the theoretical variance of 0.625. These two numbers are fairly close.

Distribution is Approximately Normal

Let's compare the distribution of the averages to the Central Limit Theorem. We'll plot the distribution along with the normal distribution.

```
library(ggplot2)

# plot the means
ggplot(data = rowmeans, aes(x = rowmeans )) +
  geom_histogram(binwidth=0.1, aes(y=..density..), alpha=0.2) +
  stat_function(fun = dnorm, args = list(mean = theomean , sd = sqrt(theovar)), colour = "red", size=1)
  geom_vline(xintercept = theomean, size=1, colour="#CC0000") +
  geom_density(colour="blue", size=1) +
  geom_vline(xintercept = mean(rowmeans[,1]), size=1, colour="#0000CC") +
  scale_x_continuous(breaks=seq(theomean-3,theomean+3,1), limits=c(theomean-3,theomean+3))
```



Notice how the blue line (random exponential distribution) overlaps almost entirely with the normal distribution. This shows how the averages of our samples will approximate the normal distribution based on the Central Limit Theorem.