Statistical Inference Course Project 1

York Huang

Sunday, March 13, 2016

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.

Prepare Data

Create matrix with 1000 simulations, each has 40 exponential samples.

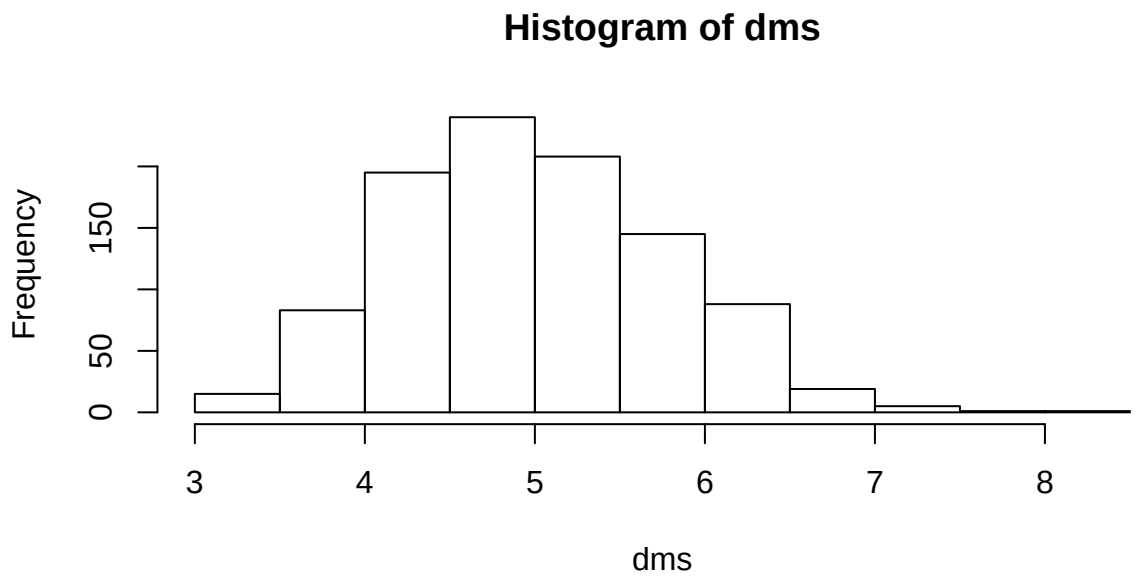
```
library(knitr)
library(ggplot2)

set.seed(1)
lambda <- 0.2
exps <- matrix( rexp(40*1000, rate=lambda), 1000)
```

Exploration of Data

Let's calculate the means of each simulations and then draw the histogram.

```
dms <- apply(exps,1, mean)
hist(dms)
```



The summary of means is below.

```
summary(dms)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.085   4.417   4.924   4.990   5.530   8.059
```

Sample Mean versus Theoretical Mean

The theoretical mean of the exponential distribution is 1/lambda as below

```
theory_mean = 1 / lambda
print(paste("Theoretical mean = ", 1/lambda))
```

```
## [1] "Theoretical mean =  5"
```

The sample mean is as below.

```
sample_mean = mean(dms)
print(paste("Sample mean = ", sample_mean))
```

```
## [1] "Sample mean =  4.99002520077716"
```

The sample mean is very close to theoretical mean.

Sample Variance versus Theoretical Variance

The theoretical standard deviation is 1/lambda. So the theoretical variance of simulation with n=40 samples is 1/(lambda^2 * n).

```
theoretical_variance = 1 /(lambda^2 * 40)
theoretical_sd = 1/lambda/sqrt(40)
print(paste("theoretical_variance = ", theoretical_variance ))
```

```
## [1] "theoretical_variance =  0.625"
```

```
print(paste("theoretical_deviation = ", theoretical_sd ))
```

```
## [1] "theoretical_deviation =  0.790569415042095"
```

The sample variance is as below.

```
sample_variance = var(dms)
sample_sd = sd(dms)
print(paste("sample_variance = ", sample_variance))
```

```
## [1] "sample_variance =  0.617707174842697"
```

```
print(paste("sample_deviation = ", sample_sd))
```

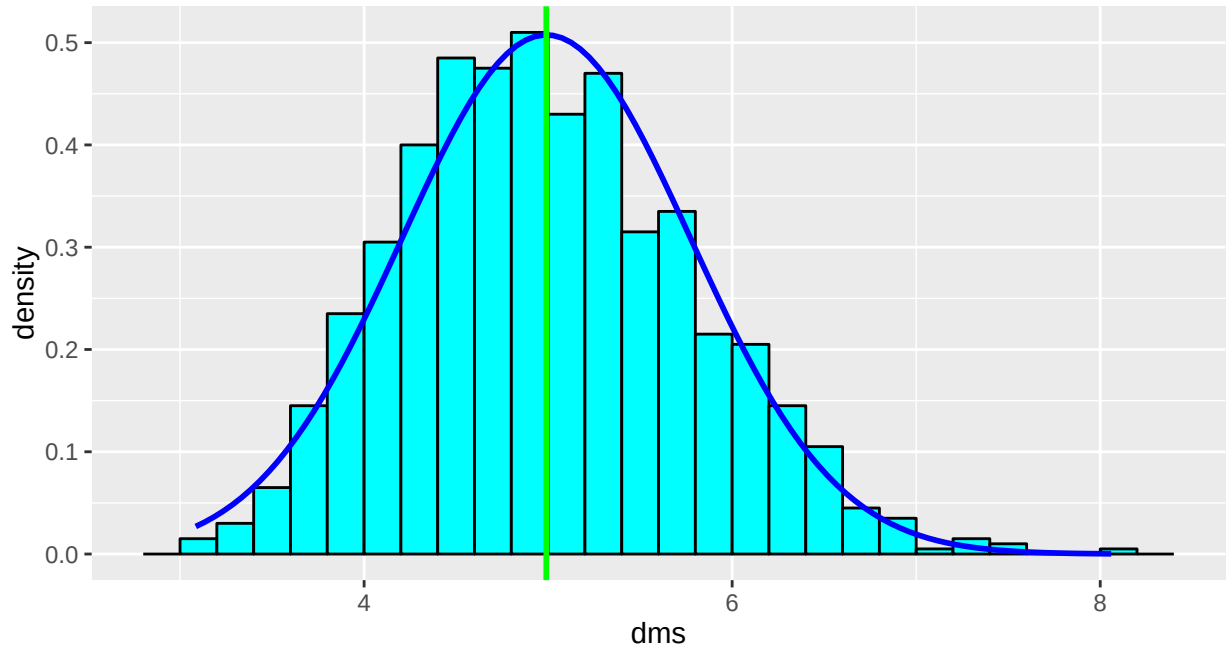
```
## [1] "sample_deviation =  0.785943493415841"
```

The sample variance is close to theoretical variance.

Distribution

In the picture below, the actual exponential distribution and the normal distribution with same mean and deviation are drawn together.

```
dms_frame <- data.frame(dms)
ggplot(dms_frame, aes(x=dms)) +
  geom_histogram(binwidth=lambda, fill="cyan",color="black",aes(y = ..density..))  +
  stat_function(fun=dnorm, args=list(mean=sample_mean, sd=sample_sd), color = "blue", size = 1.0) +
  geom_vline(xintercept=sample_mean,size=1.0,color="green")
```



From the picture below, we can see how close is the exponential and normal distributions.