Coursera Statistical Inference Project 1

Lisa Yuan

January 28, 2016

Overview

In this project, I investigated the exponential distribution in R and compared it with Central Limite Theorem. The exponential distribution is simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of the exponential distribution is 1/lambda and standard deviation is also 1/lambda. The lambda is set to 0.2 for all of the simuations. I investigated the distribution of averages of 40 exponentials and I did 1000 simuations.

Simulations

```
### loading ggplot2 for plotting
library(ggplot2)

nsim <- 1000
n <- 40
lambda <- 0.2

### Create a matrix with 1000 simuations each with 40 exp generated.
simdata <- matrix(rexp(nsim*n, rate=lambda), nsim)

### /Calculate the mean of the 40 exp of 1000 simuations
simdata_means <- apply(simdata, 1, mean)
```

1. Sample mean Versus Theoretical mean of the distribution

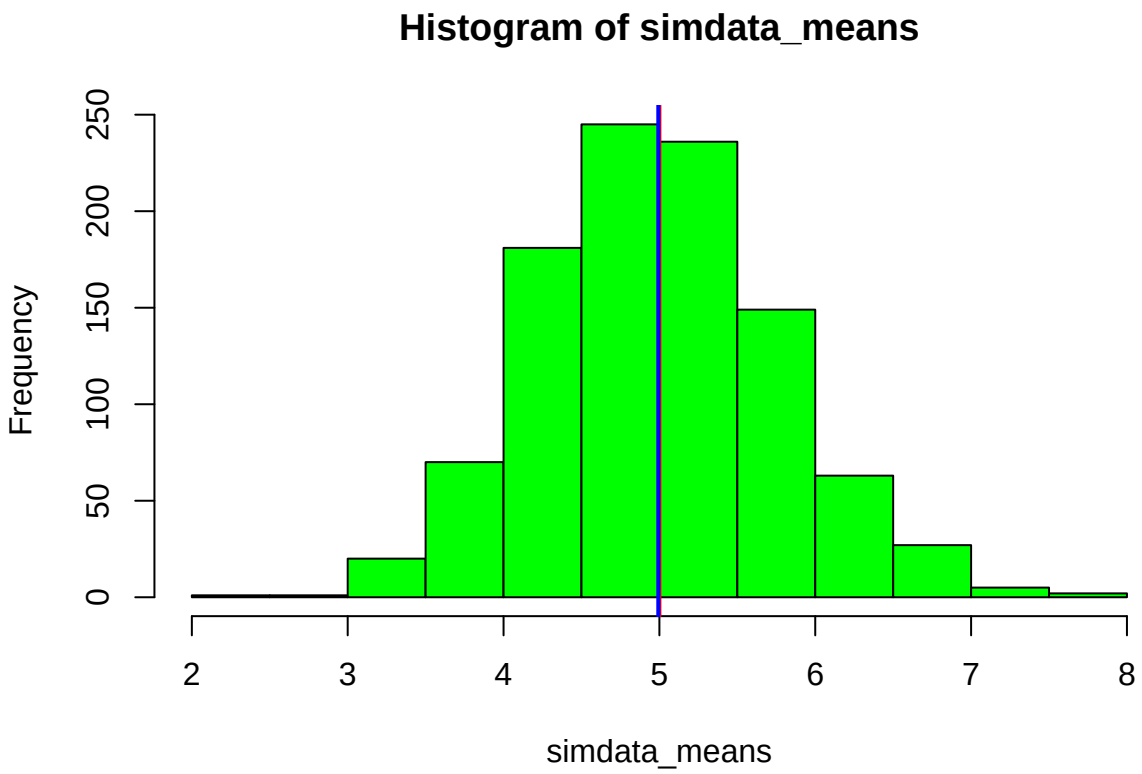
```
### Calculate theoretical mean
t_mean <- 1/lambda
t_sd <- 1/lambda/sqrt(n)
print (paste("Theoretical mean of the distribution = ", t_mean))
```

```
## [1] "Theoretical mean of the distribution =  5"
```

```
### Calculate the sample mean of the distribution based on simulation
s_mean <- round(mean(simdata_means), 3)

print (paste("Sample mean of the distribution based on simulation = ", s_mean))
```

```
## [1] "Sample mean of the distribution based on simulation =  4.993"
```



Both number and plot show that the sample mean and the theoretical mean are very close.

2. Sample variance Versus theoretical variance of the distribution

```
### Calculate the theoretical variance of the distribution
t_var <- (1/lambda)^2/n

print (paste("Theoretical variance = ", t_var))
```

```
## [1] "Theoretical variance =  0.625"
```

```
print (paste("Theoretical standard deviation = ", round(t_sd, 3)))
```

```
## [1] "Theoretical standard deviation =  0.791"
```

```
### Calculate the sample variance
s_var <- round(var(simdata_means), 3)
s_sd <- sqrt(s_var)
print (paste("Sample variance based on simulation = ", s_var))
```

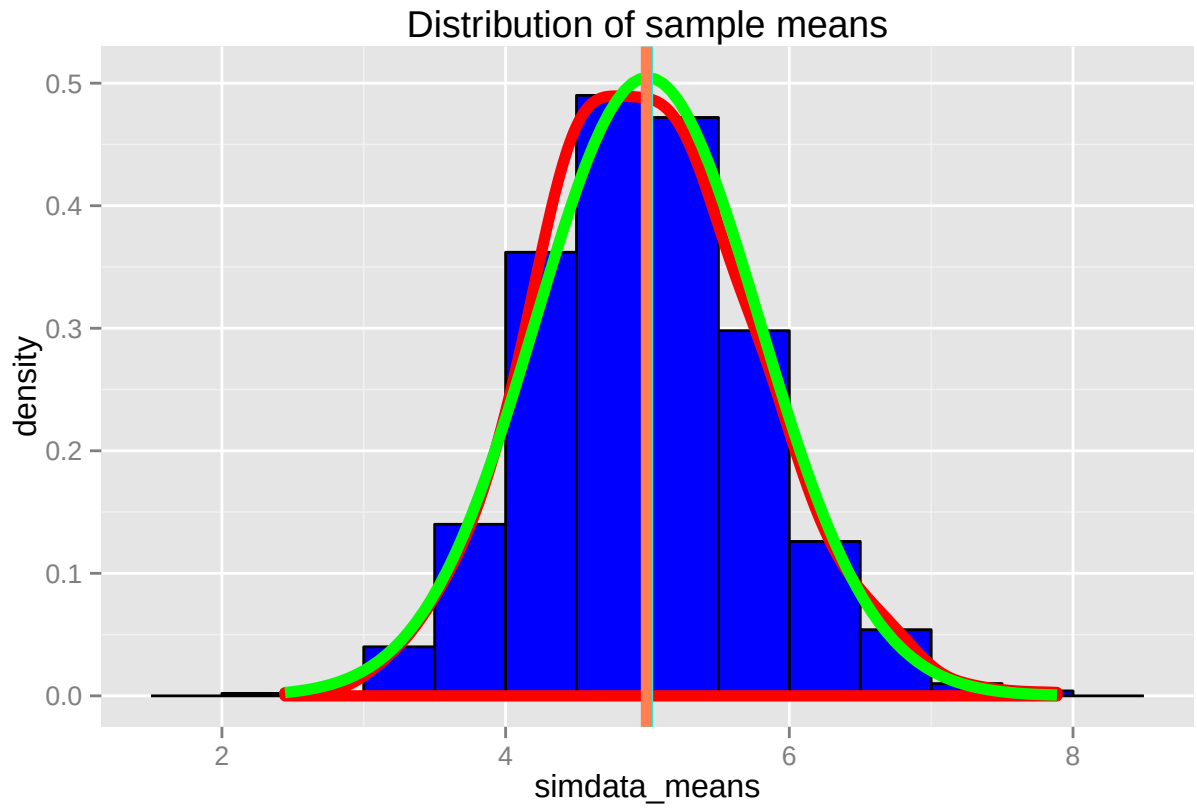
```
## [1] "Sample variance based on simulation =  0.601"
```

```
print (paste("Sample standard deviation based simuation = ", round(s_sd,3)))
```

```
## [1] "Sample standard deviation based simuation =  0.775"
```

The variance of the sample mean and the variance of the theoretical distribution are very close. Also the standard deviation of the sample mean and the standard deviation of the theoretical distribution, again very close.

3. Distribution



We can easily observe from the project results that the sample mean is very close to the theoretical mean and the sample variance is close to the theoretical variance.

From the central limit theorem we know that the mean of samples will follow a normal distribution. The plot produced shows that the calculated distribution of means of random sampled exponantial eistributions, overlaps quite nicely withe the normal distribution with the expected values based on the given lambda.