Exponential Distribution and the Cental Limit Theory

Overview

For this part of the course project for Coursera Statistical Inference, I am investigating the exponential distribution in R and comparing it with the Central Limit Theorem. I will be answering three questions comparing sample means and variance to their theortetical counterparts. I will be simulating the exponential distribution in R with rexp(n, lambda) where lambda is the rate parameter.

Lambda will set to 0.2 for all of the simulations. I will be taking the distribution of averages of 40 exponentials, and performing 1,000 simulations.

First load reuqired libraries for plotting.

```
library(ggplot2)
```

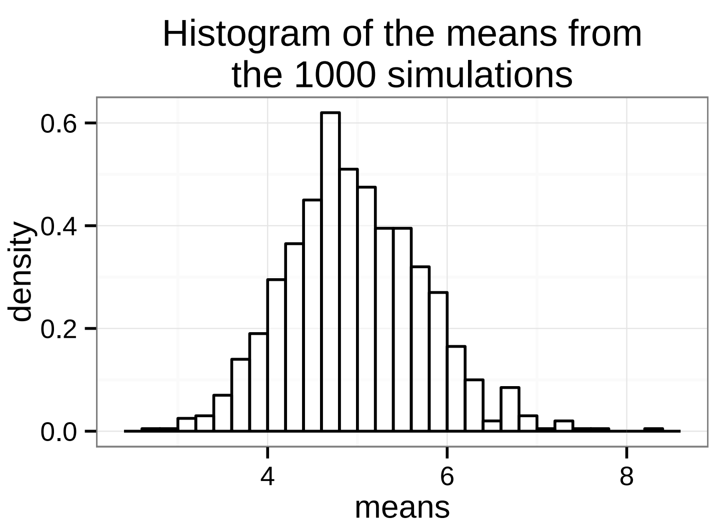
Simulations

Set parameters for simulations using guidelines stated above. Because we are running simulations, we will also set the seed to ensure the data is reproducible.

```
lambda <- 0.2
n <- 40
number_simulations <- 1000
set.seed(12345)
```

Mean of each random exponential is calculated for the thousand simulations and stored in a dataframe called "means". This is done using the values defined above in the calculations.

```
### run simulation getting the mean for 1000 simulations
means <- NULL
for (i in 1:number_simulations) means = c(means, mean(rexp(n, lambda)))
means <- as.data.frame(means)
```



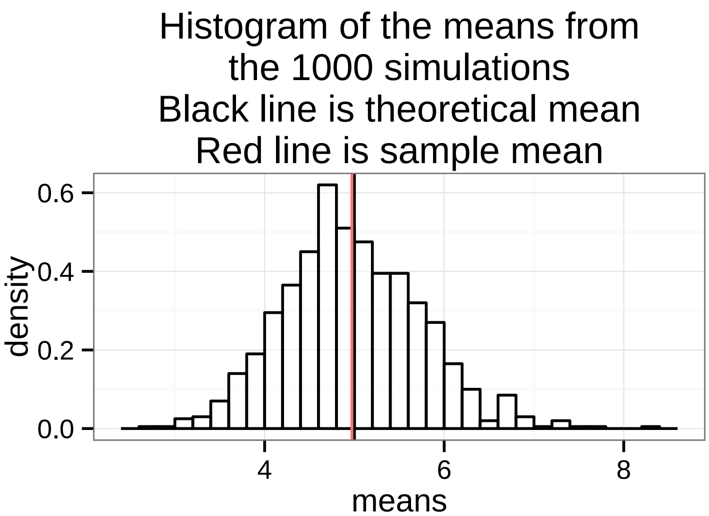
Sample Mean versus Theoretical Mean

Theoretical mean of an exponetial distribution is 1/lambda. This is compared to the mean of the simulation results.

```
### theoretical mean
theor_mean <- 1/lambda
### simulation mean
sample_mean <- mean(means$means)
print(c(theor_mean, sample_mean))
```

```
## [1] 5.000000 4.971972
```

My sample mean (4.972) is very close the the theoretical mean (5).



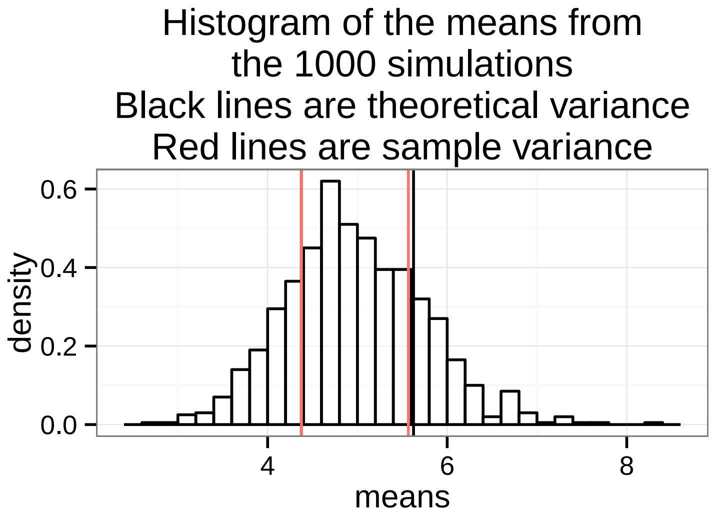
Sample Variance versus Theoretical Variance

Theoretical sampling variance is 1/(n*lambda^2), where n in this case is equal to 40. This is compared to the sample variance for the simulated results.

```
### theoretical variance of an exponetial distribution is 1/(n*lambda^2)
theor_var <-  1/(n * lambda^2)
### simulation variance
sample_var <- var(means$means)
print(c(theor_var, sample_var))
```

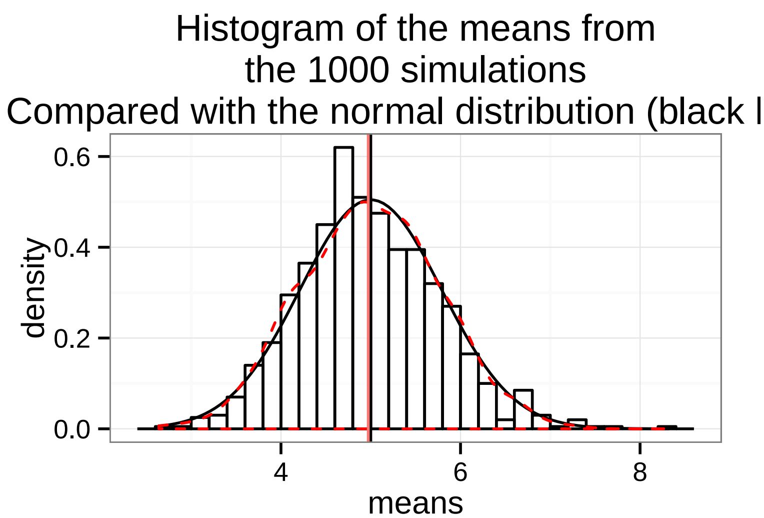
```
## [1] 0.6250000 0.5954369
```

My theoretical variance (0.595) is close to the theoretical variance (0.625).



Distribution

Create a plot to compare my distribution of simulated means to the normal distribution.



The distribution of my simulation means (dotted red line) is very close to the normal distribution (solid black line). Taking a greater number of simulations would results in the simulation means distribution falling even closer to that of the normal distribution.

Conclusions

The distribution of the averages of the exponential distribution is nearly normal. A exponential random distributoin have a minimum value of zero and a long skew tail. This distribution looks entirely different to the almost normal distribution obtained after taking the average of a thousand simulations. Thus I have demonstated that with larger sample size, the distribution of averages become closer to normal distribution even using the exponential distribution, which is the Central Limit Theorem.