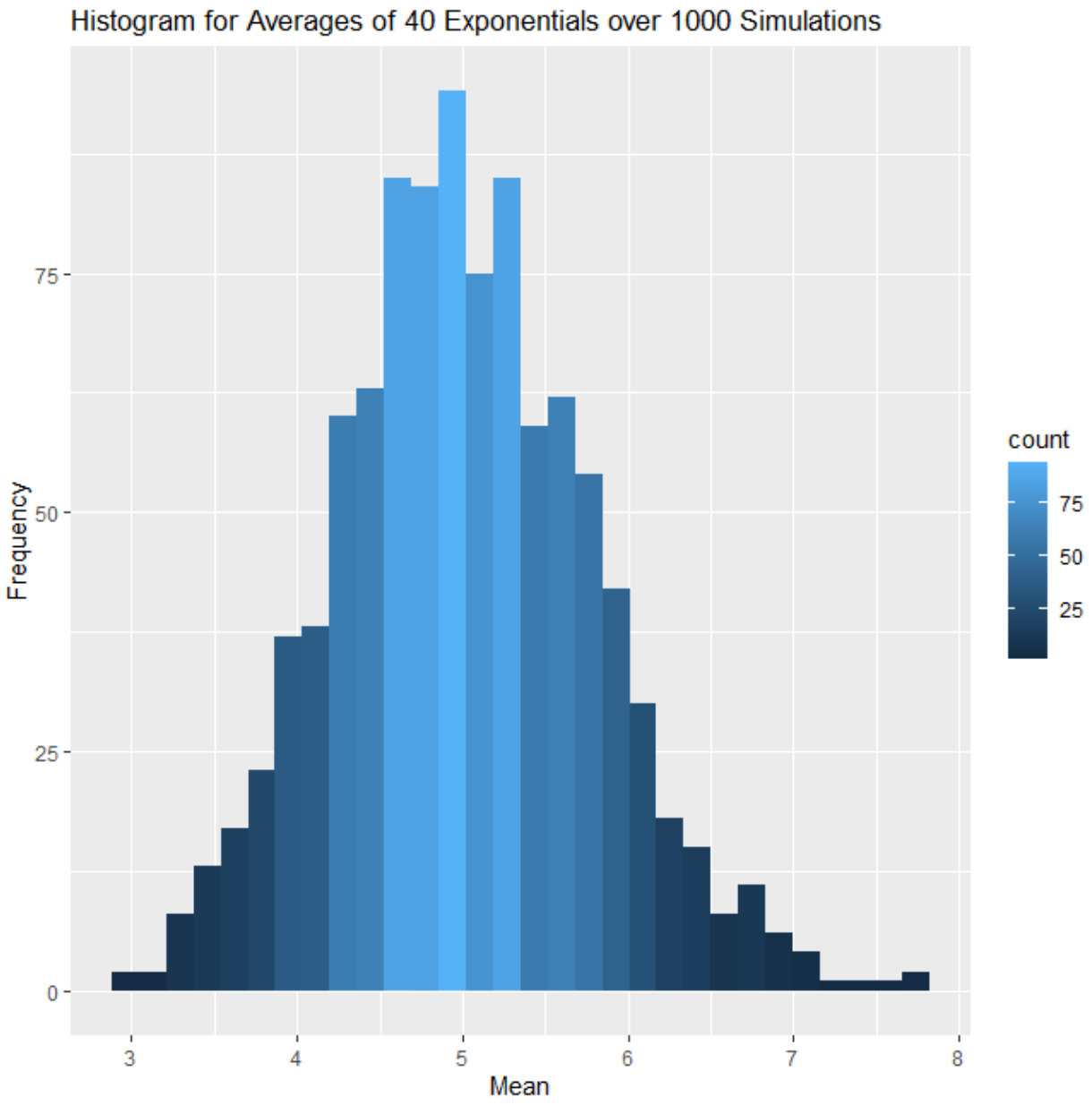
Investigate thethe exponential distribution in R and compare it with the Central Limit Theorem

author: "P.Chat."
date: "July 24 2020"

Overview
The purpose of this data analysis is to investigate the exponential distribution and compare it to the Central Limit Theorem. For this analysis, the lambda will be set to 0.2 for all of the simulations. This investigation will compare the distribution of averages of 40 exponentials over 1000 simulations.
Simulations
Set the simulation variables lambda, exponentials, and seed.
```{r}
ECHO=TRUE
set.seed(1337)
lambda = 0.2
exponentials = 40
```
Run Simulations with variables
```{r}
simMeans = NULL
for (i in 1 : 1000) simMeans = c(simMeans, mean(rexp(exponentials, lambda)))
```
Sample Mean versus Theoretical Mean
Sample Mean
Calculating the mean from the simulations with give the sample mean.
```{r}
mean(simMeans)
```
Theoretical Mean
The theoretical mean of an exponential distribution is lambda^-1.
```{r}
lambda^-1
```

```
#### Comparison
There is only a slight difference between the simulations sample mean and the exponential distribution theoretical mean.
```{r}
abs(mean(simMeans)-lambda^-1)
```
Sample Variance versus Theoretical Variance
Sample Variance
Calculating the variance from the simulation means with give the sample variance.
```{r}
var(simMeans)
```
Theoretical Variance
The theoretical variance of an exponential distribution is
(lambda * sqrt(n))^-2.
```{r}
(lambda * sqrt(exponentials))^-2
```
Comparison
There is only a slight difference between the simulations sample variance and the exponential distribution theoretical variance.
```{r}
abs(var(simMeans)-(lambda * sqrt(exponentials))^-2)
```
Distribution
This is a density histogram of the 1000 simulations. There is an overlay with a normal distribution that has a mean of lambda^-1 and standard deviation of (lambda*sqrt(n))^-1, the theoretical normal distribution for the simulations.
```{r}
library(ggplot2)
ggplot(data.frame(y=simMeans), aes(x=y)) +
  geom_histogram(aes(y=..density..), binwidth=0.2, fill="#0072B2",
                 color="black") +
  stat_function(fun=dnorm, arg=list(mean=lambda^-1,
                                    sd=(lambda*sqrt(exponentials))^-1),
                size=2) +
  labs(title="Plot of the Simulations", x="Simulation Mean")
```
```


Histogram for Averages of 40 Exponentials over 1000 Simulations
Frequency
0
25
50
75
3
4
5
6
7
8
Mean
count
75
50
25