Testing the Central Limit Theorem with the R Exp distribution (A Simulation Exercise)

Juan Jose

Monday, March 16, 2015

Overview

This analysis seeks to compare the exponential distribution in R, rexp, with the Central Limit Theorem. To do so, we will compare the sample means of 40 exponentials a total of 1000 times.

First load all the necessary libraries

```
library(ggplot2)
```

Simulations

For the purpose of this study, we will define the rate parameter of the R rexp function as 0.2

```
lambda = 0.2
#this is a place holder for the means we will collect during the
simulations.
all.the.means = NULL
#num.sims is the number of simulations we will run, in this case 1000
num.sims = 1000

#set a seed to make it reproducible
set.seed(500)

for(i in 1:num.sims) all.the.means = c(all.the.means,
mean(rexp(40,lambda)))

#store the simulation results in a data frame
df = data.frame(x=1:num.sims, y=all.the.means)
```

Analyzing The Mean (Answer to Question 1)

We know the population mean of the exponential distribution is 1/lambda. In our case, lambda = 0.2 so the theoretical mean is 5.

Theoretical Mean

```
pop.mean = 1/lambda
pop.mean
```

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

Mean of all the Simulations

```
sample.mean = mean(all.the.means)
sample.mean
```

```
## [1] 5.010562
```

Our sample mean is close to the theoretical mean.

Analyzing The Variance (Answer to Question 2)

Since our lambda is 0.2, we know our theoretical sample variance is

Sample Variance

```
theoretical.var = ((1/lambda)/sqrt(40))^2
theoretical.var
```

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

Variance of the Simulations

```
sample.var = var(all.the.means)
sample.var
```

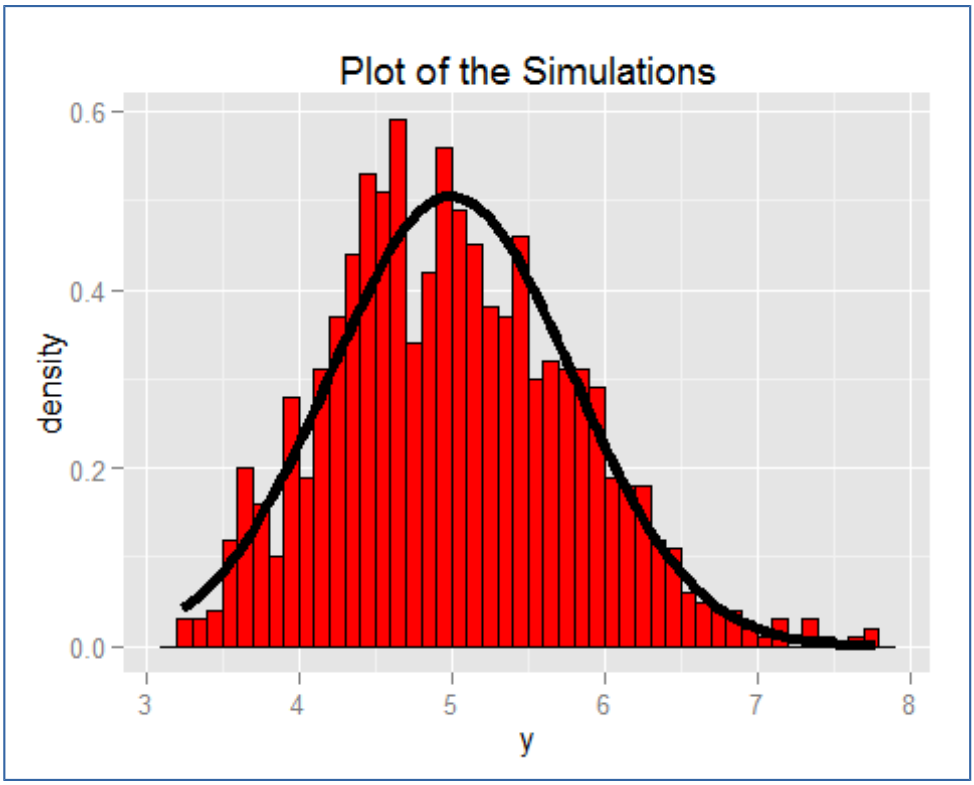
```
## [1] 0.6201215
```

Our sample variance to the theoretical variance.

Show that the distribution is approximately normal.

From the histogram, we can see that the shape resembles that of a normal distribution with a mean of (1/lambda) and standard deviation of (1/lambda)/sqrt(40)), where lambda = 0.2.

```
#histgoram of the means
g <- ggplot(df, aes(x=y)) + geom_histogram(aes(y=..density..),
binwidth=0.1, fill="red", color="black") + labs(title="Plot of the
Simulations")
g = g + stat_function(fun=dnorm, arg=list(mean=(1/lambda),
sd=(1/lambda)/sqrt(40)), size=2)
print(g)
```


Plot of the Simulations
density
0.6
0.4
0.2
0.0
3
4
5
6
7
8
y