Statistical Inference Final Project Part One

Dax Gerts

Saturday, June 20, 2015

Overview

Instructions: "In this project you 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. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations."

This report examines the properties of the exponential distribution, both through numerical and graphical methods. This is accomplished through statistical analysis across a large number of sizeable distributions.

Once the distrubtion has been examined, it will be shown that the distribution is approximately normal.

To begin, the dataset of interest was created using the "rexp" function with values of "n = 40" and "lambda = 0.2"

```
n <- 40
lambda <- 0.2
distribution <- data.frame(x = sapply(1:1000, function(x) {mean(rexp(n, 
lambda))}))
head(distribution)

##          x
## 1 5.494635
## 2 5.840656
## 3 5.169310
## 4 5.073253
## 5 5.170417
## 6 5.923910
```

1. Sample and Theoretical Mean

The sample mean was found using the following code.

```
mean(distribution$x)

## [1] 5.012198
```

sample mean = 5.012198

The theoretical mean, as mentioned in the instructions, is found by taking 1/lambda, as seen below.

```
1/lambda

## [1] 5
```

theoretical mean = 5

```
(5-5.012198)/5

## [1] -0.0024396
```

The difference was marginal, with only a **0.24% error** when comparing the sample and theoretical means.

2. Sample and Theoretical Variance

The sample variance was found using the following code.

```
var(distribution$x)

## [1] 0.621418
```

sample variance = 0.621418

The theoretical variance, as mentioned in the instructions, is found by taking 1/lambda and dividing by the square root of the size of the set, before taking the square of the result.

```
((1/lambda)/sqrt(n))^2

## [1] 0.625
```

theoretical variance = 0.625

```
(0.625-0.621418)/0.625

## [1] 0.0057312
```

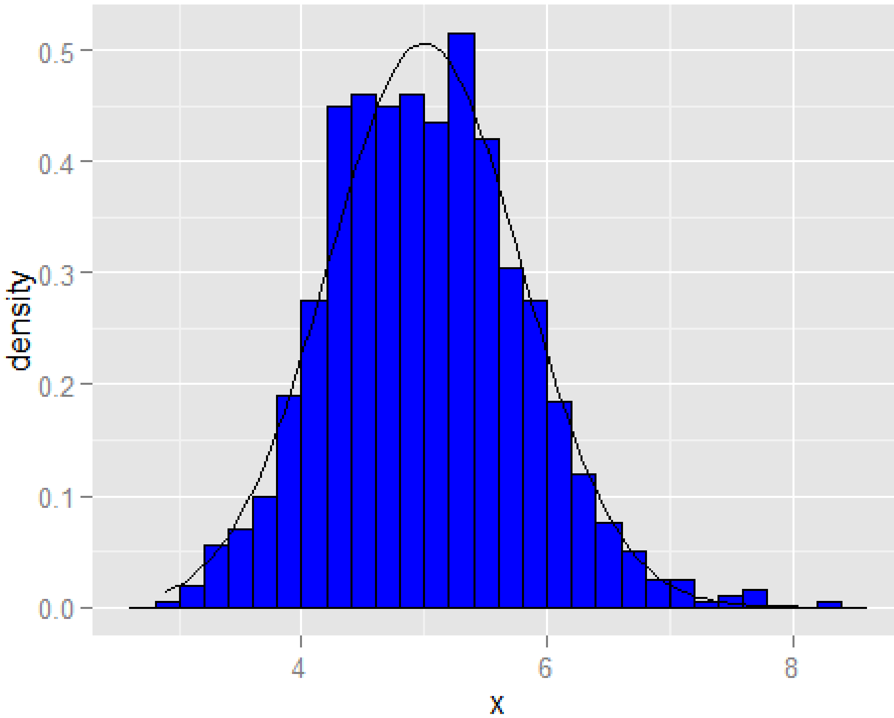
The difference was marginal, with only a **0.57% error** when comparing the sample and theoretical means.

3. Show Aprroximately Normal Distribution

The visual representation of the distrubtion of averages of exponential distributions is seen below.

```
library(ggplot2)
ggplot(data = distribution,aes(x=x)) + 

geom_histogram(aes(y=..density..),fill=I('blue'),binwidth=0.2,color=I('black'
))+
  stat_function(fun = dnorm,arg = list(mean = 5,sd = sd(distribution$x)))
```



This result can be attributed to the properities of the **Central Limit Theorem** which states that as the value of "n" gets larger, the overall distribution begins to approximate a true normal distribution. With n of 40 in this case, the distribution is suitably large enough to approximate normality.