Statistical Inference

Artur Krupa

Saturday, April 25, 2015

Introduction

This is the project for the statistical inference class. In it, I will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

1. Simulation exercises.
2. Basic inferential data analysis.

I will create a report to answer each of the questions. I use knitr to create the reports and convert to a pdf. Each pdf report will be no more than 3 pages with 3 pages of supporting appendix material if needed (code, figures, etcetera).

Part 1: Running simulations

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 also `1/lambda`.

I'll set `lambda = 0.2` for all of the simulations. In this simulation, I will investigate the distribution of averages of 40 `exponential(0.2)`s. Note that I will need to do a thousand or so simulated averages of 40 exponentials, and will generate those as follows:

```
lambda = 0.2
n = 40
nsims = 1:1000
set.seed(820)
means <- data.frame(x = sapply(nsims, function(x) {mean(rexp(n, lambda))}))
head(means)
```

```
##          x
## 1 5.750000
## 2 3.808205
## 3 4.058154
## 4 3.999241
## 5 4.312532
## 6 4.418246
```

Part 2: Question 1 and Question 2

This part will show us the sample mean and compare it to theoretical mean of the distribution

```
mean(means$x)
```

1.) Calculate center of distribution (expected: 5.0):

```
## [1] 4.998812
```

Answer => very close to expected mean!

Proof:

```
1/0.2
```

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

```


sd(means$x)
```

2.) Calculate standard deviation (expected: 0.7909)

```
## [1] 0.7909422
```

Answer => very close to expected deviation!

Proof: Expected deviation

```
(1/lambda)/sqrt(40)
```

```
## [1] 0.7905694
```

```


var(means$x)
```

3.) Using Central Limit Theorem we can calculate the sample variance (expected: 0.625):

```
## [1] 0.6255895
```

Answer => very close to expected variance!

Proof: Expected variance

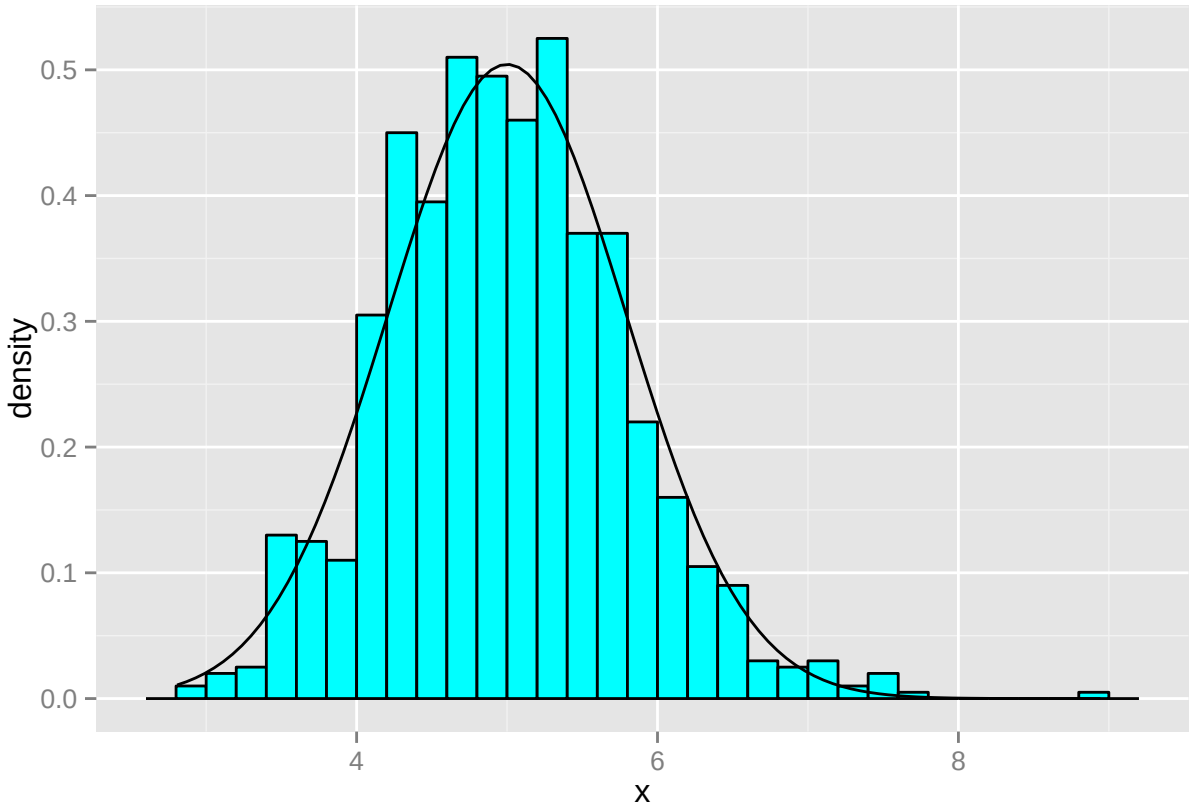
```
((1/lambda)/sqrt(40))^2
```

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

Part 3: Question 3

This part will show us that the distribution is approximately normal

```
library(ggplot2)
ggplot(data = means, aes(x = x)) +
    geom_histogram(aes(y=..density..),
                   fill = I('cyan'),
                   binwidth = 0.20,
                   color = I('black')) +
    stat_function(fun = dnorm, arg = list(mean = 5, sd = sd(means$x)))
```



As we can see the distribution is overlaid with a normal distribution with mean 5 and standard deviation 0.7909.