Statistical Inference-Course Project Part1

Praveen Balasubramanian

Tuesday, March 17, 2015

PART 1

This is the course project for the statistical inference class. In it, you 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.

You will create a report to answer the questions. Please use knitr to create the reports and convert to a pdf. (I will post a very simple introduction to knitr). Each pdf report should be no more than 2 pages with 3 pages of supporting appendix material if needed (code, figures, etcetera). 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. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

```
#setwd("C:\\Users\\Praveen\\Desktop")
library(lattice)
library(gplots)
```

```
## Warning: package 'gplots' was built under R version 3.1.3
```

```
## KernSmooth 2.23 loaded
## Copyright M. P. Wand 1997-2009
##
## Attaching package: 'gplots'
##
## The following object is masked from 'package:stats':
##
##     lowess
```

```
lambda <- 0.2
n <- 40
simulations <- 1:2000
set.seed(52)
means <- data.frame(x = sapply(simulations, function(x) {mean(rexp(n, lambda))}))
```

```
mean(means$x) # sampling mean = 5.000542
```

```
## [1] 5.017351
```

```
(1/lambda) # theoretical mean = 5
```

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

The distribution is centered at 5.000542 (sampling mean) which is similar to the theoretical mean which is 5

2. Show how variable it is and compare it to the theoretical variance of the distribution.

```
var(means$x) # sampling var = 0.6564438
```

```
## [1] 0.6365326
```

```
sd(means$x) # sampling sd = 0.8102122
```

```
## [1] 0.7978299
```

```
((1/lambda)/sqrt(n))^2 # theoretical var = 0.625
```

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

```
(1/lambda)/sqrt(n) # theoretical sd = 0.7905694
```

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

Sampling variance and standard deviation are 0.6564438 and 0.8102122 respectively.

Theoretical variance and standard deviation are 0.625 and 0.7905694 respectively.

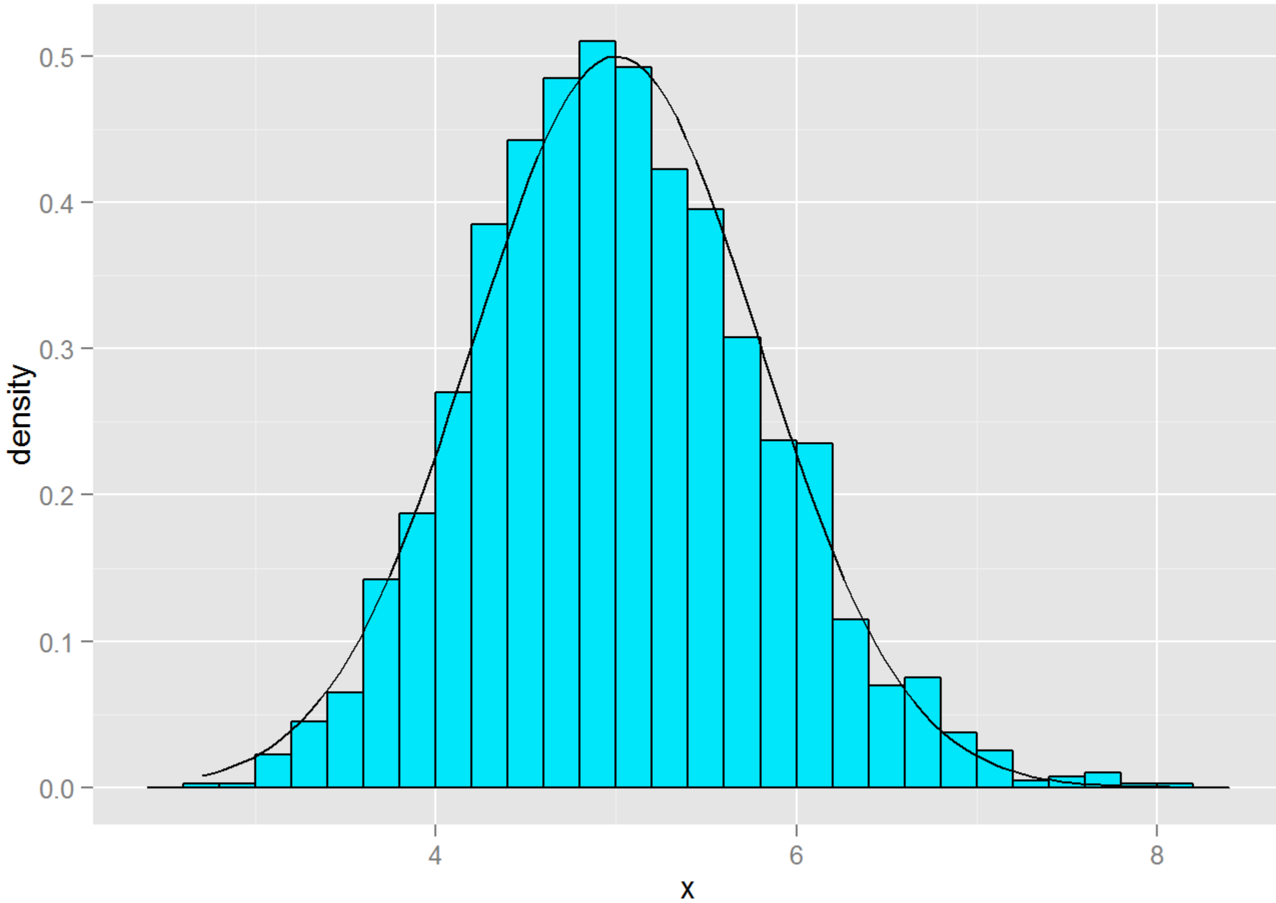
3. Show that the distribution is approximately normal.

Below is a histogram plot of the means of the 2000 simulations of rexp(n, lambda). It is overlaid with a normal distribution with mean 5 and standard deviation 0.7909. Yes, the distribution of our simulations appears normal.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.2
```

```
ggplot(data = means, aes(x = x)) +
    geom_histogram(aes(y=..density..), fill = I('#00e6fa'),
                   binwidth = 0.20, color = I('black')) +
    stat_function(fun = dnorm, arg = list(mean = 5, sd = sd(means$x)))
```



4. Evaluate the coverage of the confidence interval for 1/lambda: $X \pm 1.96 S_n$

```
# 95% confidence interval (Z=1.96)
mean(means$x) + c(-1,1)*1.96*sd(means$x)/sqrt(nrow(means)) # 4.965033 5.036051
```

```
## [1] 4.982385 5.052318
```

Note both sampling and theoretical mean are contained within confidence interval.