Statistical Inference (statinference-011) Course Project - Part 1

Esmit Pérez

February 16, 2015

Part 1

Overview

This is the report for 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. A simulation exercise.
2. Basic inferential data analysis.

The format and formulas here included are based off the outline of the project.

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 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.

```
# load libraries, setup variables
library(ggplot2)
lambda <- 0.2
mu <- 1/lambda
stdDev <- 1/lambda
numExponentials <- 40
numSimulations <- 1:1000

# we're dealing with random data, so always set seed to make it reproducible.
set.seed(909)

# obtains the mean of running rexp with 40 exponentials and given lambda
cfunc <- function(v) {mean(rexp(numExponentials, lambda))}

#  for each entry in array of size 1000, run the function
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(cfunc()))

dat <- data.frame(x = mns)
```

Sample Mean versus Theoretical Mean

Theoretical mean is 1/lambda:

```
mu
```

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

The sample mean is:

```
mean(dat$x)
```

```
## [1] 4.960913
```

Sample Variance versus Theoretical Variance

Theoretical variance is $\mu/\sqrt(n)$:

```
mu/sqrt(numExponentials)
```

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

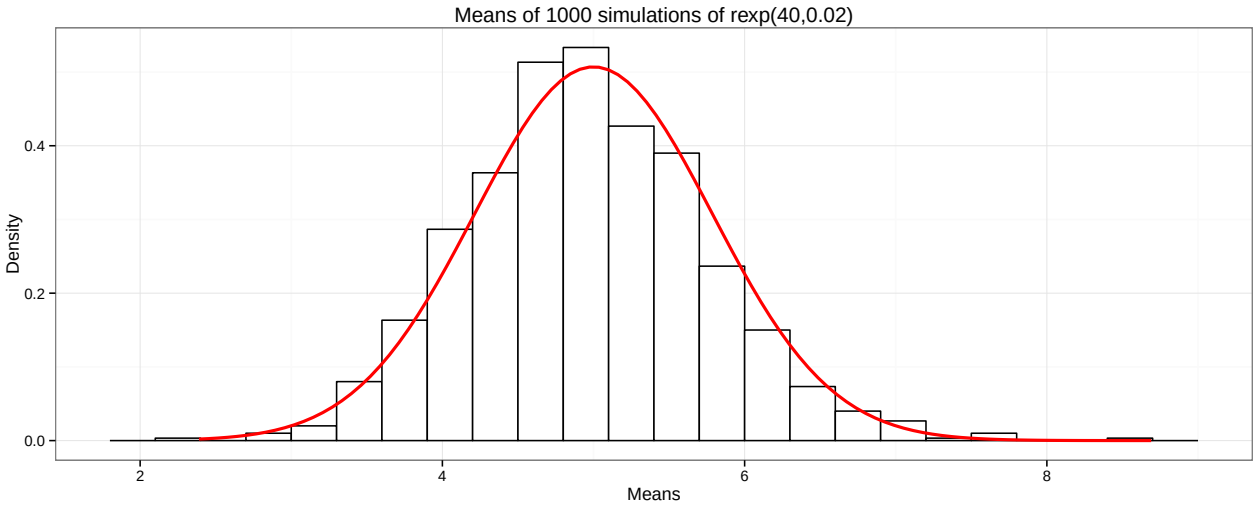
The sample variance S is:

```
var(dat$x)
```

```
## [1] 0.619083
```

Distribution

The following graph shows how the mean values of 1000 simulations approximate the normal distribution (curve in red). For complete code and output see appendix, fig 1.1



Appendix

Part 1 - Supporting figures.

Figure 1.1

```
# plot and customize graph
g <- ggplot(dat, aes(x = x))
g <- g + geom_histogram(binwidth=.3, colour = "black", fill="white",
        aes(y = ..density..))
g <- g + ggtitle("Means of 1000 simulations of rexp(40,0.02)")
g <- g + xlab("Means") + ylab("Density")

g <- g + stat_function(fun = dnorm, arg=list( mean= mu, sd=sd(dat$x) ),
        color="red", size=1)
g <- g + theme_bw()

g
```