Statistical Inference Part 1

Phil Mai

December 4, 2016

Introduction

```
In this project you will investigate the exponential distribution in R and compare it with the C
entral Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) wh
ere lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the stand
ard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigat
e the distribution of averages of 40 exponentials. Note that you will need to do a thousand simu
lations.

Illustrate via simulation and associated explanatory text the properties of the distribution of
 the mean of 40 exponentials. You should:

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of
 the distribution.
3. Show that the distribution is approximately normal.
```

Prepping the Data

```
setwd("C:/Users/Alex/Desktop/Coursera/Statistical Inference")
library(knitr)
library(ggplot2)
opts_chunk$set(echo=TRUE)
set.seed(1)
```

Setting up the data as noted in the example:

```
numExponentials<-40
numSimulation<-1:1000
lambda<-0.2
mu<-1/lambda
set.seed(999)
```

```
# 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)
```

Questions to answer from Data

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.
2. Show how variable it is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

The theoretical mean was given in the problem as 1 / lambda. The actual mean of the generated data can be calculated by using the apply & mean functions to obtain a mean for each row and then obtaining the mean of those identified from each mean.

Question #1 - Sample Mean vs. Variable Mean

Theoretical mean = 1/lambda

```
mu
```

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

The sample mean is:

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

```
## [1] 5.029028
```

Quesiton #2 - Sample Variance versus Theoretical Variance

Theoretical variance is:

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

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

The sample Variance is:

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

```
## [1] 0.605616
```

Question #3 - Showing the Distribution

```
g <- ggplot(dat, aes(x = x))
g <- g + geom_histogram(binwidth=.3, colour = "black", fill="green",
        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,args=list(mean= mu, sd=sd(dat$x) ),
        color="red", size=1)
g <- g + theme_bw()

g
```