Statistical_Inference Course Project Part 1: Simulation Exercise

Author: Michael Nichols

Part 1

In this project, we will conduct an investigation of the exponential distribution in R and compare it with the Central Limit Theorem. This will be done using a thousand simulations of the exponential distribution.

Exponential Distribution in R - rexp(n, lambda) where lambda is the rate parameter - Mean of exponential distribution ~ 1/lambda - Standard deviation ~ 1/lambda.

Simulation Settings - Lambda will be set to 0.2 for all of the simulations. - Investigate the distribution of averages of 40 exponentials - Conduct 1,0000 simulations.

We will illustrate, via simulation and summaries, the properties of the distribution of the mean of 40 exponentials. We will do so through the following: - Show the sample mean and compare it to the theoretical mean of the distribution. - Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. - Show that the distribution is approximately normal, focusing on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

Simulations

```r
#load libraries
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.3
```

```r
#set all variables based on the prompt

# lamdba = 0.2, with 40 exponentials, over 1000 samples
lambda <- .2
n <- 40
simulations <- 1000
```

Sample Mean versus Theoretical Mean

```r
#set seed for reproducible simulations
set.seed(16)

#show distributions of mean over 1,000 simulations
means <- NULL
for (i in 1:simulations) means <- c(means, mean(rexp(40, lambda)))

#mean of sample from the 1,000 simulations
simMean <- mean(means)
print(simMean)
```

```
## [1] 4.994828
```

```r
#theoretical mean: 1/lambda
theoreticalMean <- 1/lambda


#create table of comparisons of the means
MeanTable <- matrix(c(theoreticalMean, simMean),ncol=1,byrow=TRUE)
colnames(MeanTable) <- "Mean"
rownames(MeanTable) <- c("Theoretical","Simulation")
MeanTable<- as.table(MeanTable)

#print table of the means comparison
print(MeanTable)
```

```
##                 Mean
## Theoretical 5.000000
## Simulation  4.994828
```

Sample Variance versus Theoretical Variance

```r
#standard deviation from the simulations
simStandardDev <- sd(means)

#theoretical standard deviation
theoreticalStandardDev <- (1/lambda)/sqrt(n)

#variancestandard deviation from the simulations
simVariance <- simStandardDev^2

#theoretical variance
theoreticalVariance <- theoreticalStandardDev^2

#create table of comparisons of the variances
VarianceTable <- matrix(c(theoreticalStandardDev, theoreticalVariance, simStandardDev, simVariance),ncol
colnames(VarianceTable) <- c("Variance", "Standard Deviation")
rownames(VarianceTable) <- c("Theoretical","Simulation")
VarianceTable<- as.table(VarianceTable)

#print table of variance & standard deviation comparison
print(VarianceTable)
```

```
##              Variance Standard Deviation
## Theoretical 0.7905694          0.6250000
## Simulation  0.7831256          0.6132857
```

Distribution of the Means

According to the Central Limit Theory, given certain conditions, the arithmetic mean of a sufficiently large number of IID random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution.
The following plot displays that with 1,000 simulations of data with exponential distribution, the spread of the means almost identically fall under a normal distribution.

```
SimMeans <- as.data.frame(means)
ggplot(data = SimMeans, aes(x = means)) +
        geom_histogram(binwidth=0.1, aes(y=..density..), fill = "grey")  +
        stat_function(fun=dnorm,args=list(mean=theoreticalMean, sd= theoreticalStandardDev), color = "bl
        stat_function(fun=dnorm,args=list(mean= simMean, sd= simStandardDev), color = "red", size = 1.0)
        geom_vline(aes(xintercept = mean(means)),col='red',size=1) +
        geom_vline(aes(xintercept = 5), col='blue', size = 1) +
        labs(title="Distribution of Means: Exponential Distributions", subtitle="Sample Distribution & M
```