Statistical Inference Simulation Exercise

Harland Hendricks

June 27, 2018

The instructions for this assignment are located here.

Built with R version 3.5.0 with the following system:

```
##          sysname           release           version          nodename
##        "Windows"          "10 x64"     "build 17134" "DESKTOP-TPCQ5AJ"
##          machine             login              user    effective_user
##         "x86-64"           "harla"           "harla"           "harla"
```

```
## [1] 0
```

Load the required libraries

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

Comparing R's Exponential Distribution and the Central Limit Theorem

Simulation

For our experiment we will use

$$\lambda = 0.2$$

and run 1000 simulations for the Exponential Distribution and store the results in an object named ed, this represents the Law of Large Numbers distribution that can demonstrate the Central Limit Theorem.

```
set.seed(1234)
n <- 1000
l <- 0.2
ed <- rexp(n, l)
```

We will also run 1000 simulations to sample the mean of 40 observation from the Exponential Distribution and store this result in the object named mns.

```
set.seed(1234)
mns = NULL
for (i in 1 : 1000) mns <- c(mns, mean(rexp(40, l)))
```

We will also run 1000 simulations to sample the variance of 40 observation from the Exponential Distribution and store this result in the object named vars.

```r
set.seed(1234)
vars = NULL
for (i in 1 : 1000) vars <- c(vars, sd(rexp(40, l))^2)
```

Sample Mean vs Theoretical Mean

We know that the theoretical mean of the Exponential Distribution is

$$1/\lambda$$

We can calcualte the theoretical mean and compare it to the mean of the 1000 observations in the Exponential Distribution stored in object ed:

```r
##Theoretical Mean
tm <- 1/l
print(tm)
```

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

```r
##Expenetial Distribution Mean from 1000 observations
meaned <- mean(ed)
print(meaned)
```

```
## [1] 5.003067
```

```r
##Difference in means
tm-meaned
```

```
## [1] -0.003066873
```

These means are close when looking at 1000 observations, but what if we compare the mean from a sample of only 40 observations from the Exponential Distribution that was simulated 1000 times and stored in the object mns to the theoretical mean and compare the difference of means to the previous calculation?

```r
##Mean from sample of 40 observations from Exponential Distribution simulated 1000 times
meanmns <- mean(mns)
print(meanmns)
```

```
## [1] 4.974239
```

```r
##Difference from theoretical
tm - meanmns
```

```
## [1] 0.02576123
```

The result is a little larger, but still pretty close. Figure one displays the Eponential Distribution and means.

```r
hist(ed, breaks = 200, main = "Mean Comparisons", xlab = "Exponential Distribution Observation
     Zoomed in Around Mean", xlim = 4.5:5.5)
abline(v = tm, col = "black")
abline(v = meaned, col = "blue")
abline(v = meanmns, col = "red")
legend("topright", legend = c("Theoretical", "1000 Obs", "40 Obs"), fill = c("black", "blue", "red"))
```

Sample Variance vs Theoretical Variance

We know that the theoretical standard deviation of the Exponential Distribution is

$$1/\lambda$$

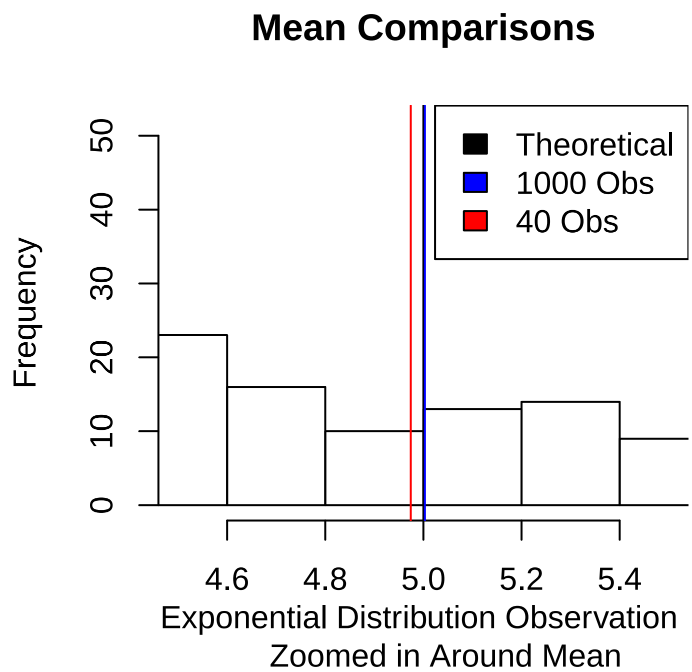

Figure 1: Figure One - Theoretical, 1000 observations, and 40 observations Means for Exponential Distribution

and that variance is

$$\sigma^2$$

We can calcualte the theoretical variance and compare it to the variance of the 1000 observations in the Exponential Distribution stored in object ed:

```
##Theoretical variance
tv <- (1/l)^2
print(tv)
```

```
## [1] 25
```

```
##Expenetial Distribution variance from 1000 observations
vared <- sd(ed)^2
print(vared)
```

```
## [1] 25.5704
```

```
##Difference in means
tv-vared
```

```
## [1] -0.5703997
```

These variances are close when looking at 1000 observations, but what if we compare the variance from a sample of only 40 and observations from the Exponential Distribution that was simulated 1000 times and stored in the object mns to the theoretical variance and compare the difference of variances to the previous calculation?

```
##Mean of variances from sample of 40 observations from Exponential Distribution simulated 1000 times
varmns <- mean(vars)
```

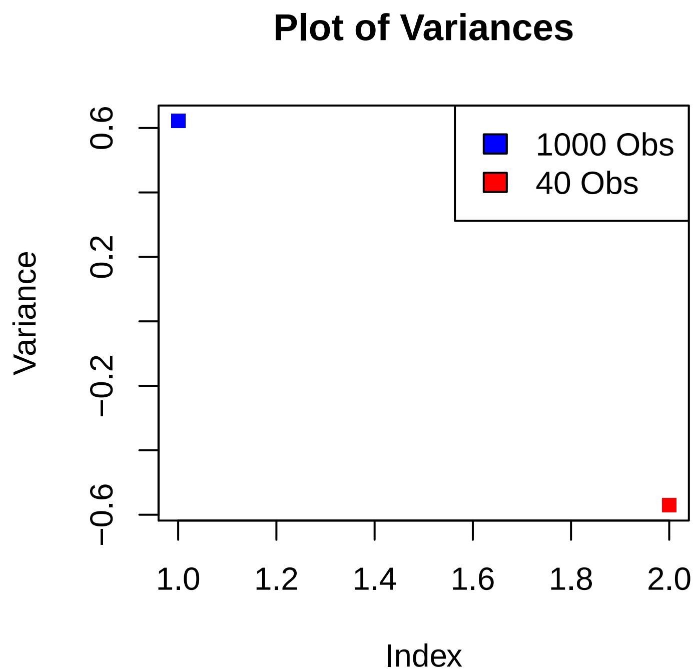

Figure 2: Figure Two - Densities for 1000 observations and 40 observations for Exponential Distribution

```
print(varmns)
```

```
## [1] 24.37801
```

```
##Difference from theoretical
tvv <- tv - varmns
tv - varmns
```

```
## [1] 0.6219853
```

```
##Difference from theoretical for 1000 observations
tvthous <- tv - vared
```

The result is a little further from theoretical, but still pretty close. Figure two compares the difference for variance between the Theoretical Variance, 1000, and 40 observation variances.

```
set.seed(1234)
fed = NULL
for (i in 1 : 1000) fed <- c(fed, rexp(40, 1))
```

```
plot(c(tvv, tvthous), col = c("blue", "red"), pch = 15, main = "Plot of Variances", ylab = "Variance")
legend("topright", legend = c("1000 Obs", "40 Obs"), fill = c("blue", "red"))
```

Distribution

The Central Limit Theorem (CLT) states that the distribution of averages of independent and identically distributed (iid) variables becomes that of a standard normal as sample size increases.

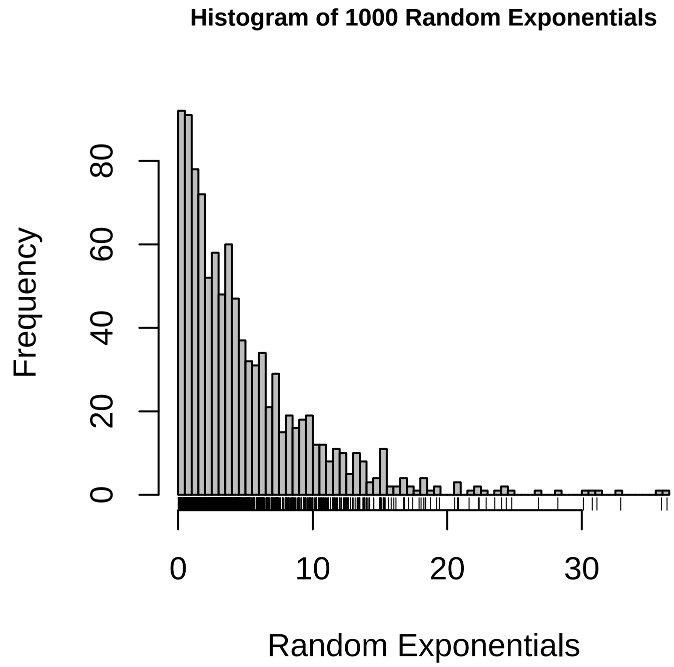

Figure 3: Figure Three - histogram of 1000 Random Exponentials

The large distribution of random exponentials is depicted in Figure Three

```
hist(ed, col = "grey", main = "Histogram of 1000 Random Exponentials", ylab = "Frequency",
     xlab = "Random Exponentials", breaks = 100, cex.main = 0.75)
rug(ed)
```

Now look at the distribution of averages from 40 observed random expenetials simulated 1000 times in Figure Four.

```
hist(mns, col = "grey", main = "Histogram of means of 40 Random Exponentials Simulated 1000 Times", ylal
     xlab = "Observed Means", breaks = 100, cex.main = 0.5)
rug(mns)
```

Now look at the same graph, except the observations were increased from 40 to 400, or by a factor of 10. Figure Five depicts the distribution of averages from 400 observed random expenetials simulated 1000 times.

```
set.seed(1234)
mnsfh = NULL
for (i in 1 : 1000) mnsfh <- c(mnsfh, mean(rexp(400, 1)))
```

```
hist(mnsfh, col = "grey", main = "Histogram of means of 400 Random Exponentials Simulated 1000 Times", y
     xlab = "Observed Means", breaks = 100, cex.main = 0.5)
rug(mnsfh)
```

Figure Five shows that as sample size increases, the distribution becomes more normal.

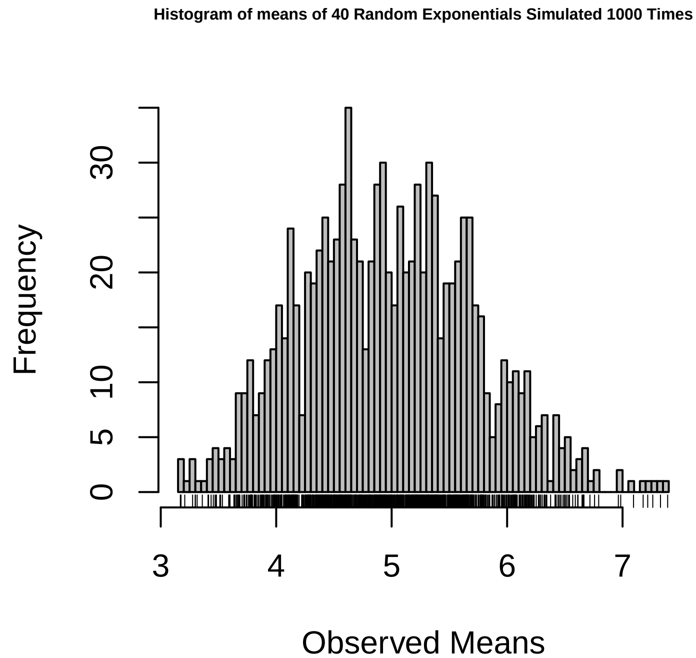

Figure 4: Figure Four - histogram of means of 40 observations simulated 1000 times

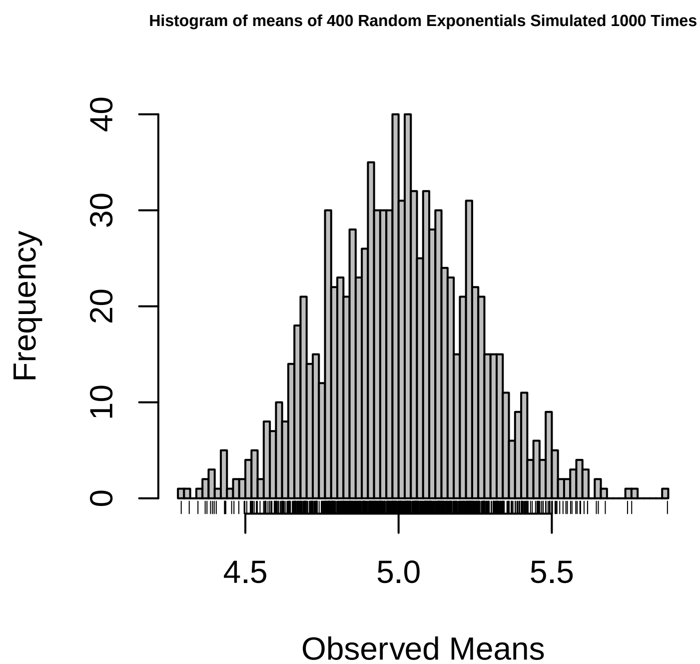

Figure 5: Figure Five - histogram of means of 400 observations simulated 1000 times