Statistical Inference Course Project Simulation Exercise

Charles Yoo

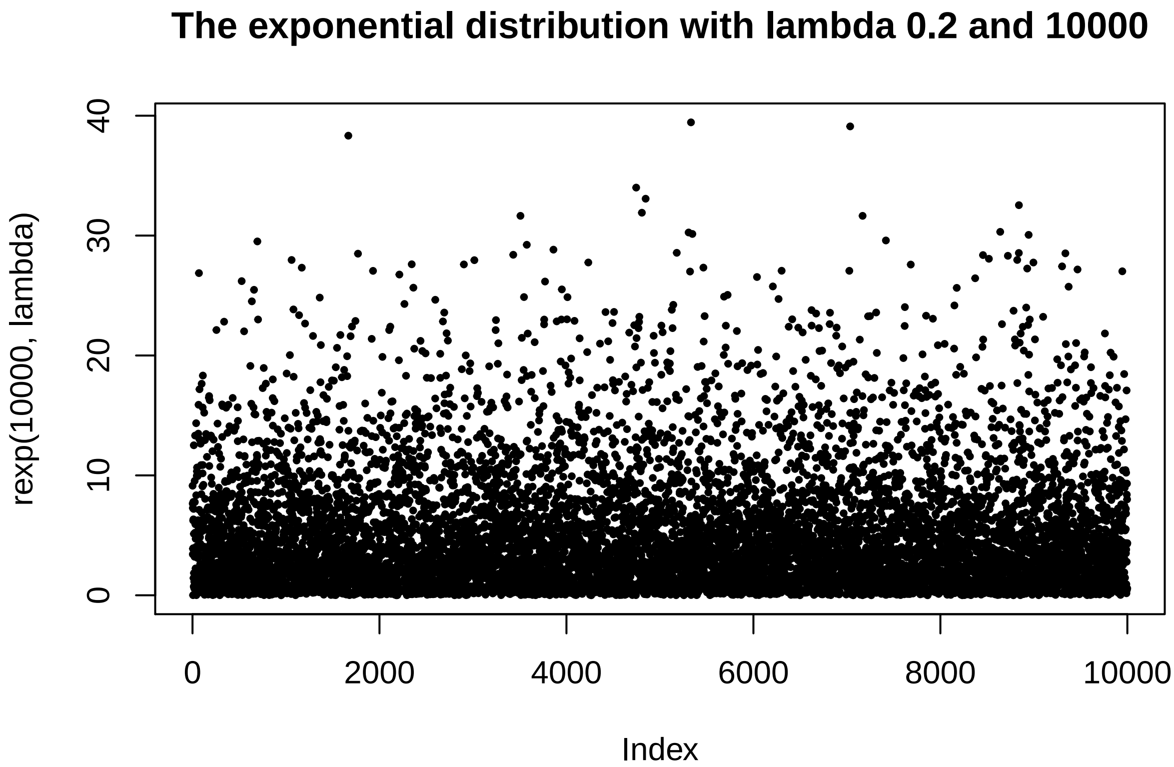
June 19, 2016

Overview

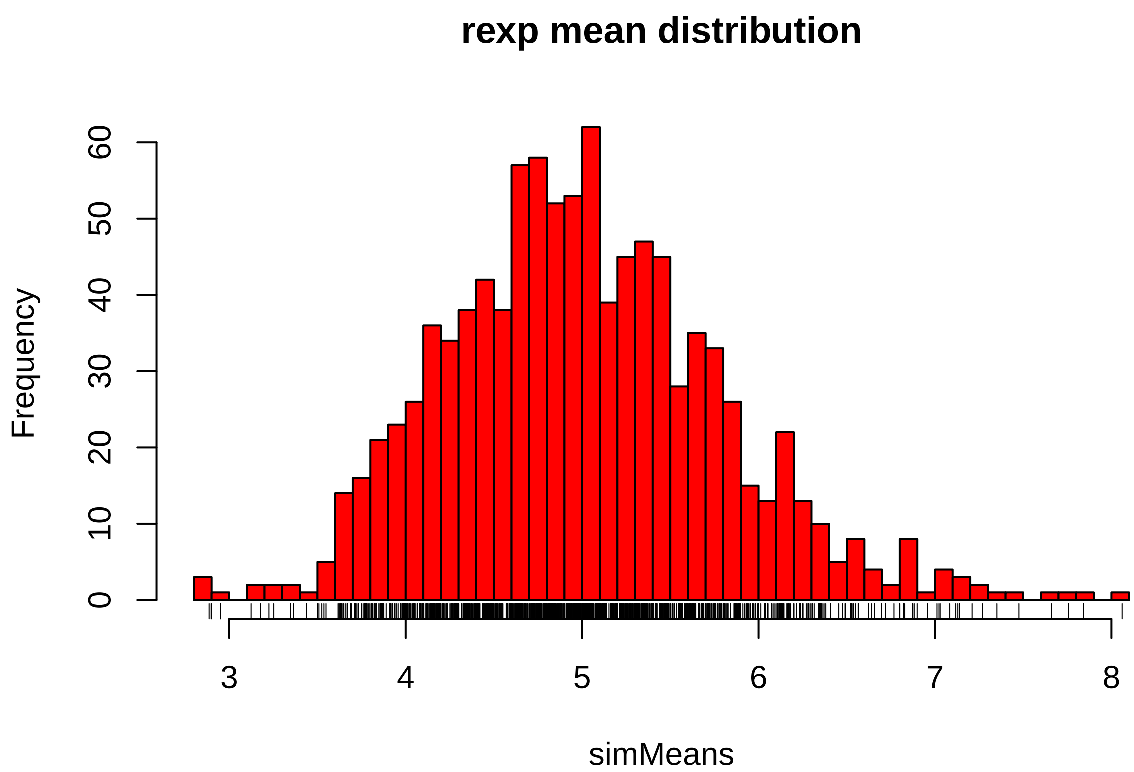
This project will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. We will investigate the distribution of averages of 40 exponentials, requiring a thousand simulations.

Simulations

```
# set seed
set.seed(2016)

# set lambda
lambda <- 0.2

# set number of simulations
sim <- 1000

# set number of exponentials
n <- 40

# the exponential distribution
plot(rexp(10000, lambda), pch = 20, cex = 0.6, main = "The exponential distribution with lambda 0.2 and
```




```
# run simulations with variables
simMeans = NULL
for (i in 1 : sim){
    simMeans = c(simMeans, mean(rexp(n, lambda)))
}

# draw historgram for rexp mean distribution
hist(simMeans, col = "red", main = "rexp mean distribution", breaks = 40)
rug(simMeans)
```



Sample Mean versus Theoretical Mean

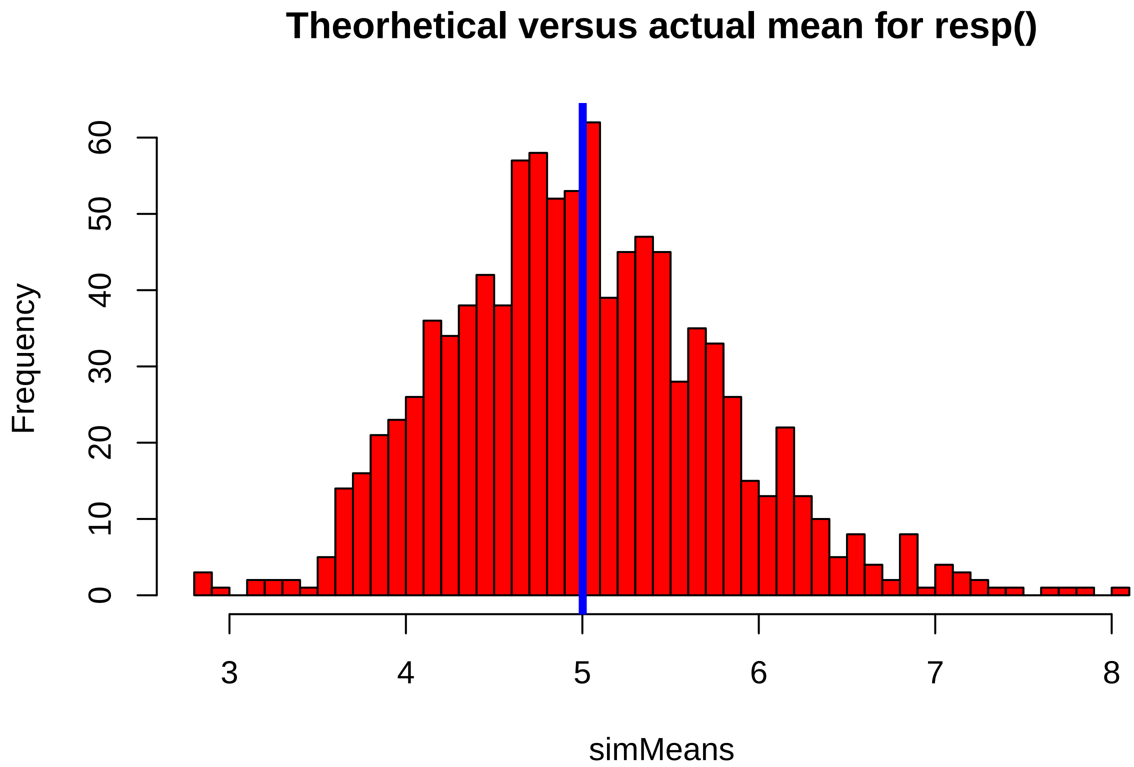
```
# calculate mean from simulations with sample mean
sMean <- mean(simMeans)
paste("Simulated mean", sMean, sep = " ")
```

```
## [1] "Simulated mean 5.00210098581326"
```

```
# calculate the theorhetical mean of an exponential distribution
tMean <- 1/lambda
paste("Theorhetical mean", tMean, sep = " ")
```

```
## [1] "Theorhetical mean 5"
```

```
# draw histogram for theorhetical mean
hist(simMeans, col = "red", main = "Theorhetical versus actual mean for resp()", breaks = 40)
abline(v = mean(simMeans), lwd = "4", col = "blue")
```



Sample Variance versus Theoretical Variance

```
# calculate the simulated standard deviation and variation
sStdDev <- sd(simMeans)
sVar <- sStdDev^2
paste("Simulated standard deviation", sStdDev, sep = " ")
```

```
## [1] "Simulated standard deviation 0.778670624665064"
```

```
paste("Simulated variance", sVar, sep = " ")
```

```
## [1] "Simulated variance 0.606327941716281"
```

```
# calculate the theorhetical standard deviation and variation
tStdDev <- (1/lambda) / sqrt(n)
tVar <- tStdDev^2
paste("Theorhetical standard deviation", tStdDev, sep = " ")
```

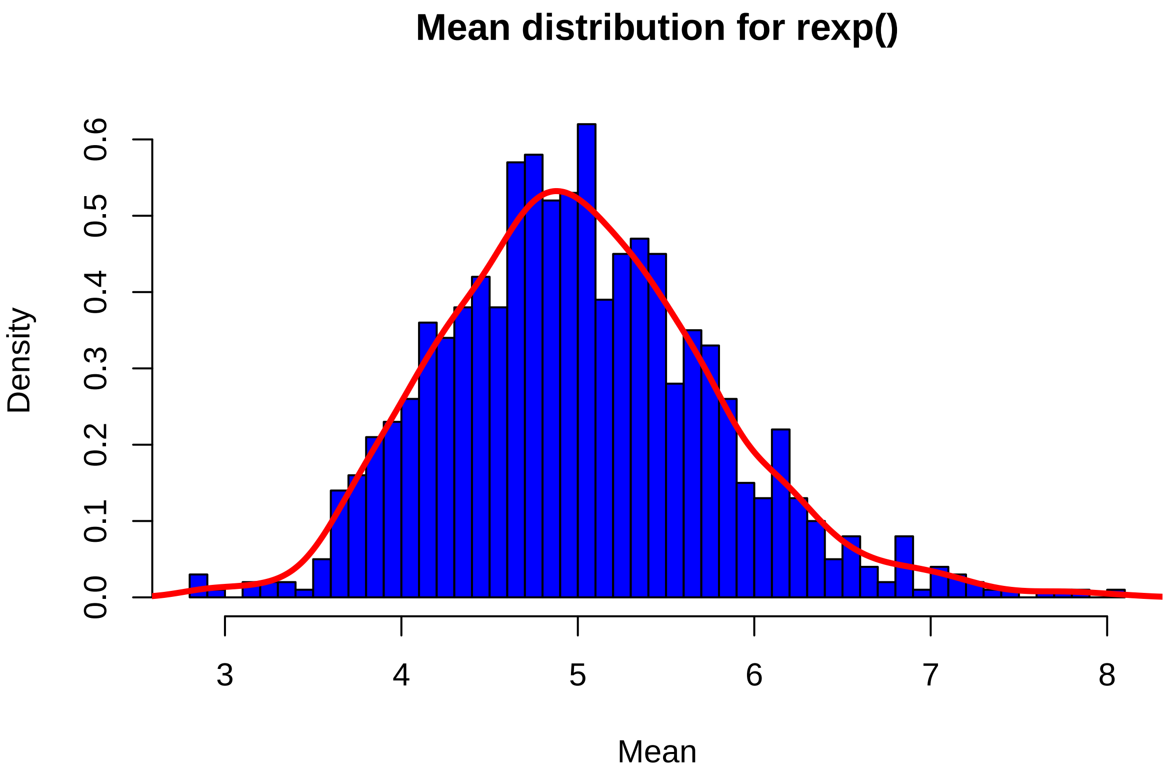
```
## [1] "Theorhetical standard deviation 0.790569415042095"
```

```
paste("Theorhetical variation", tVar, sep = " ")
```

```
## [1] "Theorhetical variation 0.625"
```

Distribution

```r
# draw historgram of the simulation
hist(simMeans, probability = TRUE, col = "blue", xlab = "Mean", main = "Mean distribution for rexp()", b
# draw line to show overlaps with the normal distribution due to the Central Limit Theorem
lines(density(simMeans), lwd = 3, col = "red")
```



The more samples run, the more closer the density distribution will be to the normal distribution bell curve.

The Central Limit Theorem states that the averages of samples should follow a normal distribution

References

https://rpubs.com/calin/111887
https://rpubs.com/schan1031/statinf1
https://github.com/codebender/statistical-inference-course-project/blob/master/exp_distribution_vs_CLT.Rmd