Statistical Inference Course Project

Andrew E. Davidson

February 21, 2015

My R Markdown file borrows heavily from https://github.com/bcaffo/courses/blob/master/06_StatisticalInference/07_Asymp (https://github.com/bcaffo/courses/blob/master/06_StatisticalInference/07_Asymp)

---

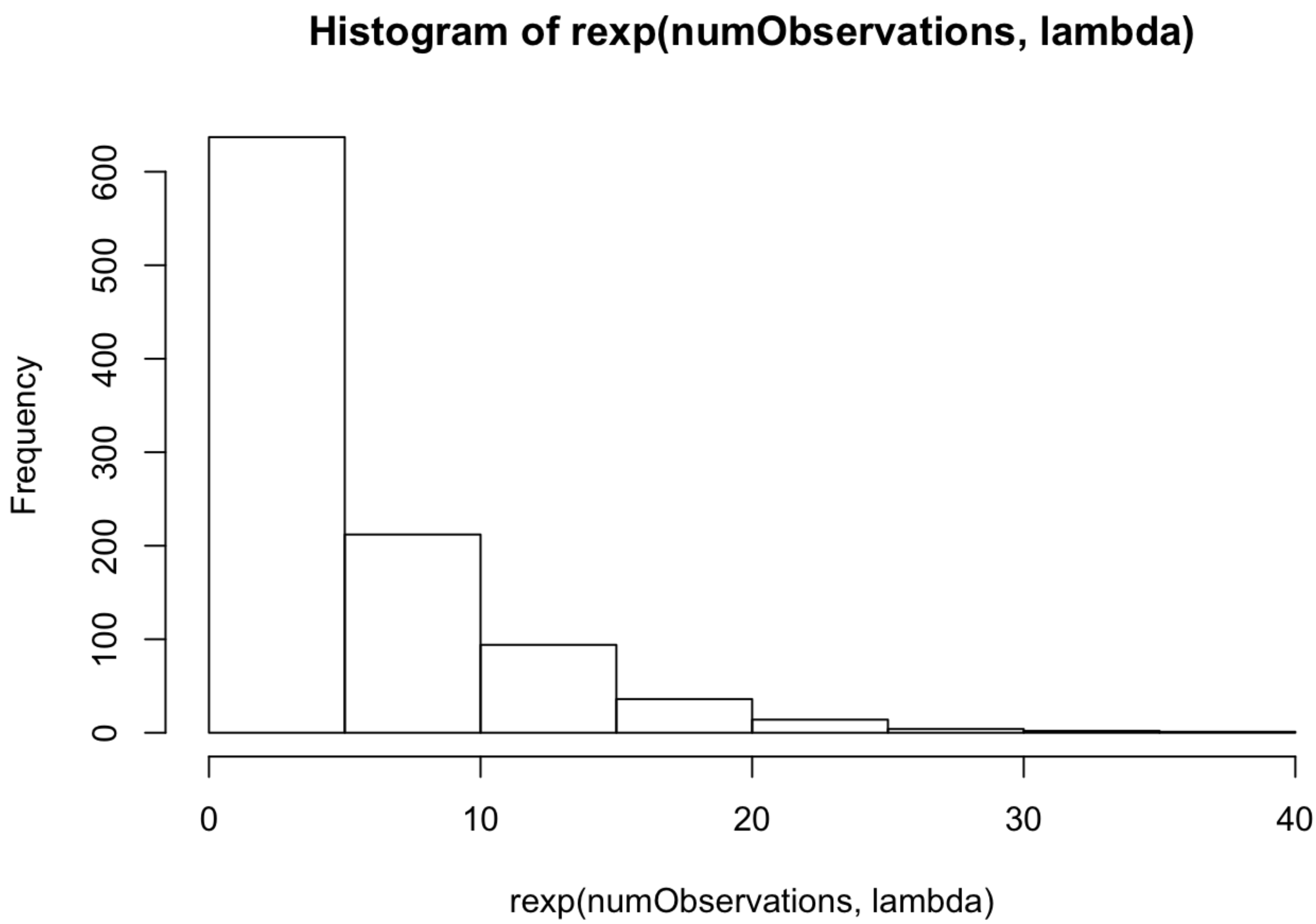
1) Investigate the exponential distribution in R and compare it with the Central Limit Theorem.

“The expotential distrubtion is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate.”"

reference: http://en.wikipedia.org/wiki/Exponential_distribution (http://en.wikipedia.org/wiki/Exponential_distribution)

Clearly expotentially distributed random variables are not “bell shaped.”

```
lambda <- 0.2
numObservations <- 1000
hist(rexp(numObservations,lambda))
```



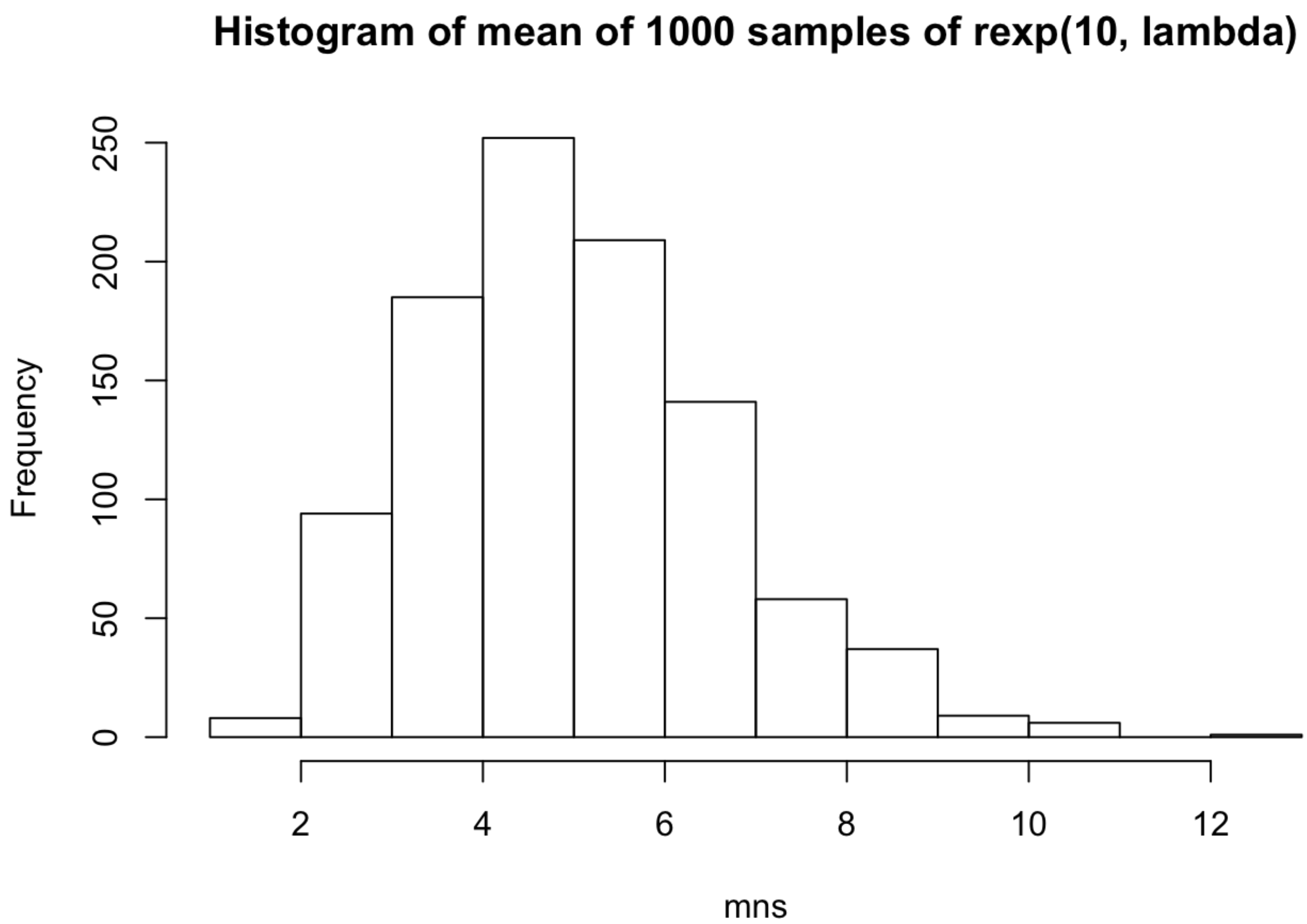
From the Central Limit Theorm we know that as the mean of a sample of expoentially distributed random variable should be normally distributed with the estimated becomming a better approximation of the true population mean as the sample size increase. We will use the following simulation to see if that is true. We'll also see how well the Central Limit Theorm applies to the population variance.

---

1. Show the sample mean and compare it to the theoretical mean of the distribution.

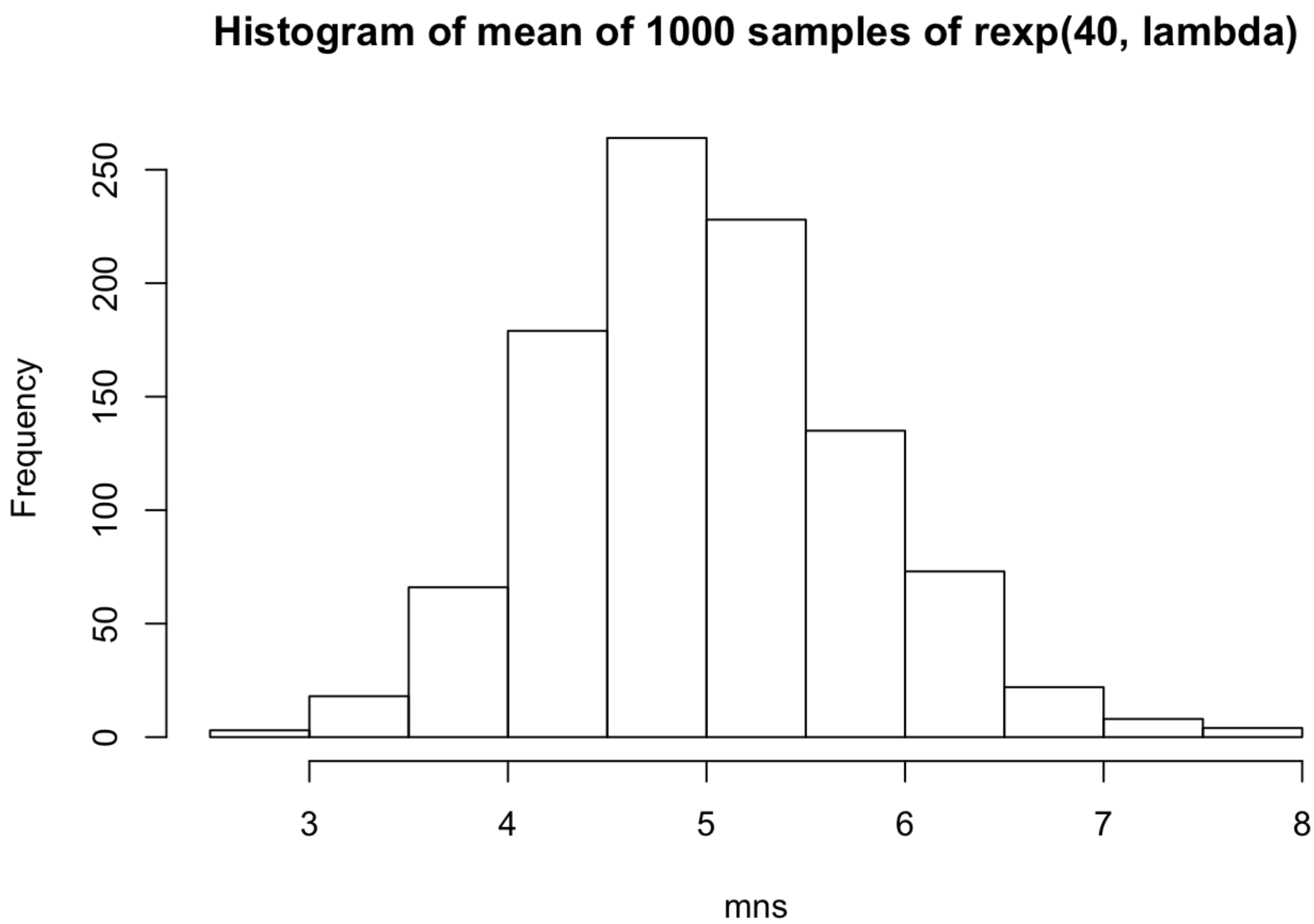
We'll start by drawing 1000 samples each of a relatively small size of 10

```
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(10, lambda)))
hist(mns, main="Histogram of mean of 1000 samples of rexp(10, lambda)")
```



Lets draw another 1000 samples this time lets increase the sample size to 40. Notice how as the sample sized increased the ‘bell shape’ seems more balanced. I.E the center of mass is more evenly distributed around 5.0

```
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(40, lambda)))
hist(mns, main="Histogram of mean of 1000 samples of rexp(40, lambda)")
```



Now, lets compare how the sample mean compares to theoretic mean of the population. We know the population mean = 1/lambda which is 5. Notic the value from our simulation is very close tothe theoretic population mean

```
sampleMean <- mean(mns)
sampleMean
```

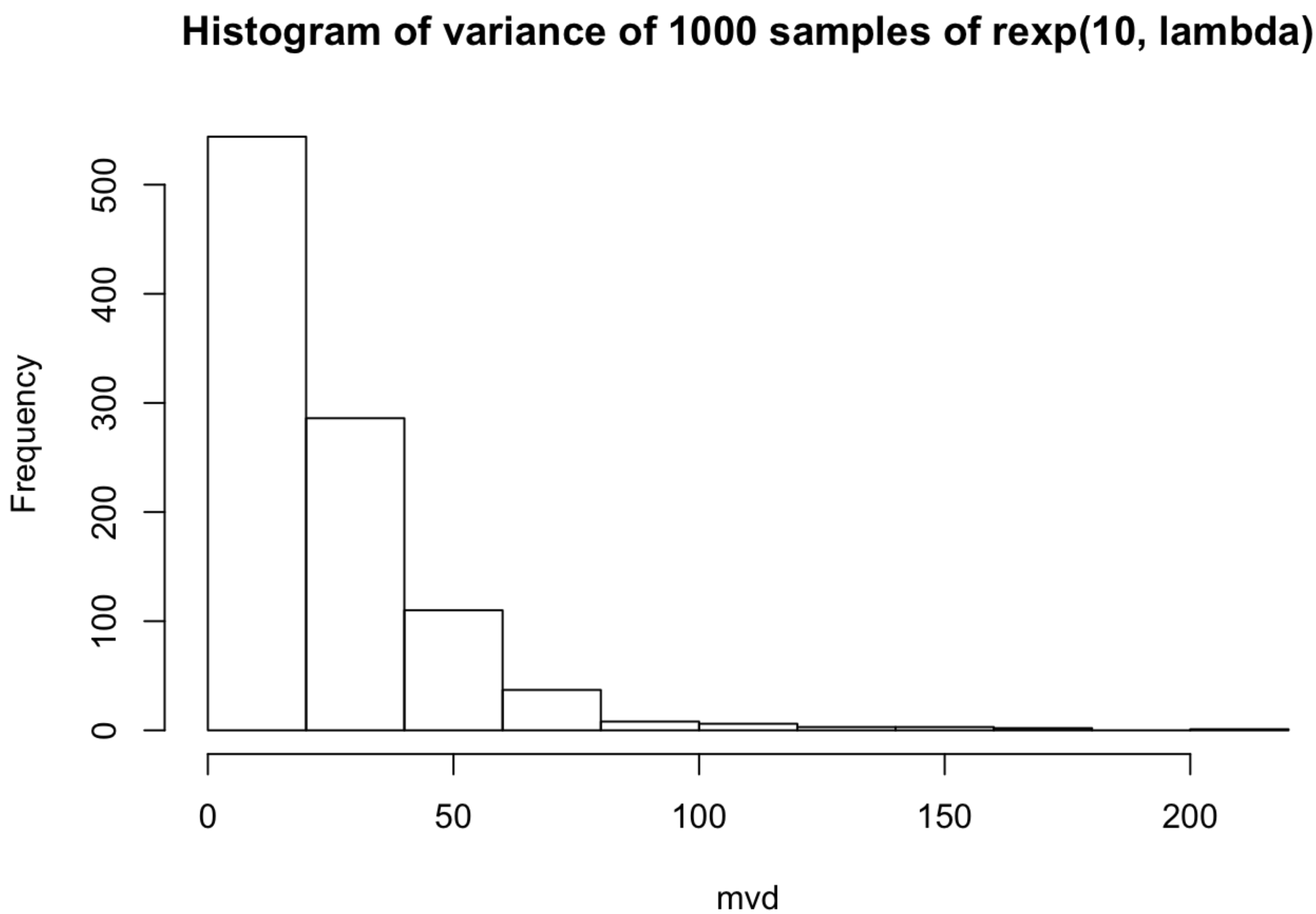
```
## [1] 4.992
```

---

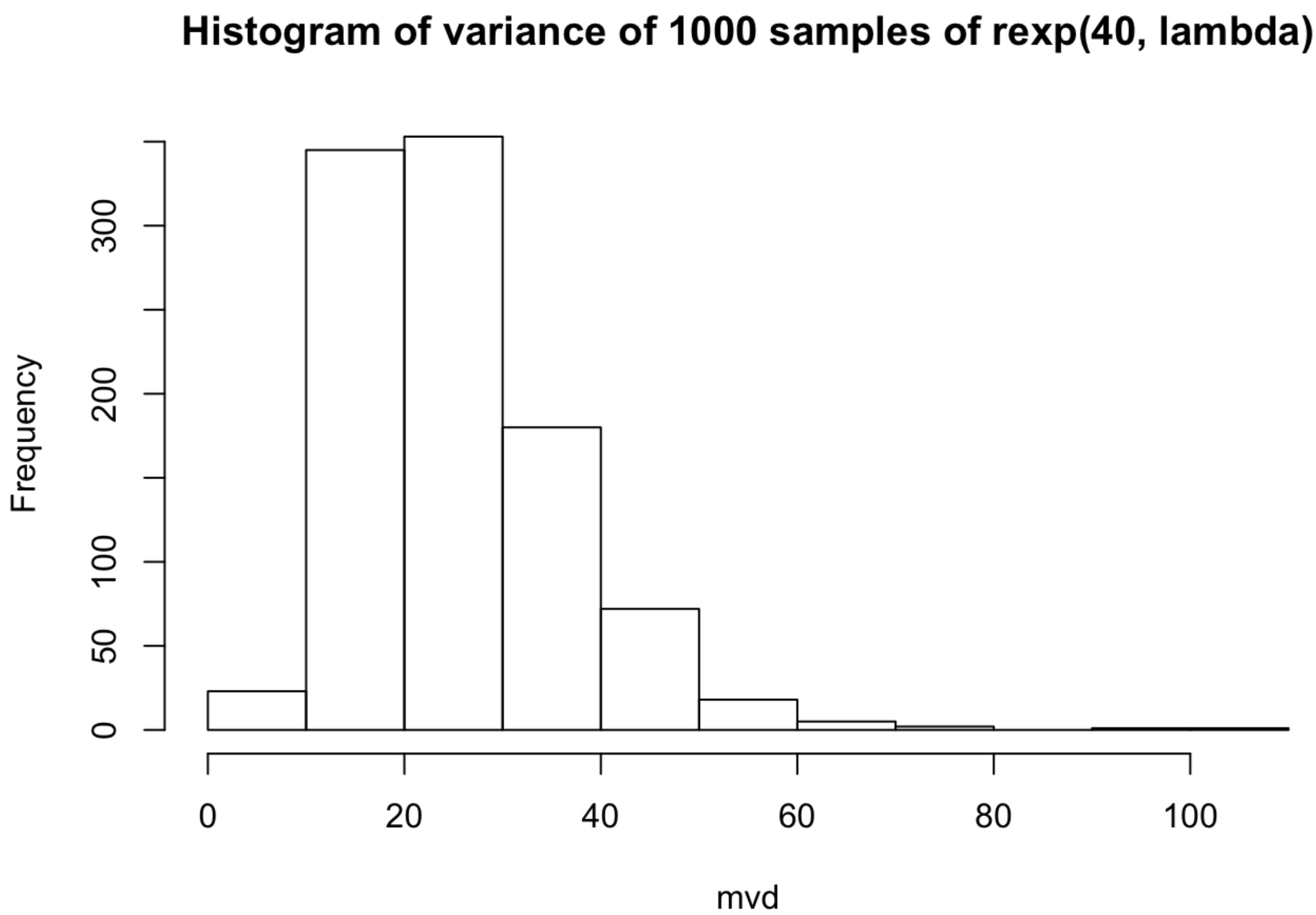
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

Now lets repeat the above simulations this time instead of looking at the sample mean lets look at the variance. Once again the Central Limit Theorm applies. N

```
mvd = NULL
for (i in 1 : 1000) mvd = c(mvd, var(rexp(10, lambda)))
hist(mvd, main="Histogram of variance of 1000 samples of rexp(10, lambda)")
```




```
mvd = NULL
for (i in 1 : 1000) mvd = c(mvd, var(rexp(40, lambda)))
hist(mvd, main="Histogram of variance of 1000 samples of rexp(40, lambda)")
```



clearly as the sample increases the distribution of sample variance become more bell shaped.

Finally lets compare our how or sample variance compares the theoretic population variance of (1/lambda)^2 which is 25. Notice it is very close to the theortical variance of the population

```
sampleMeanVar <- mean(mvd)
sampleMeanVar
```

```
## [1] 25.36
```

---

3. Show that the distribution is approximately normal.

The above graphs clearly show that the mean and variance of a samples of random variable are normally distributed. Our samples where drawn from a population that was expotentially distributed how ever as the sample sizes increased we can clearly see that the distribution of the sample means and variances becomes more bell shaped and become better approximation of the actual population mean and variance