SI Course Project I

DM Rochman

August 20, 2015

=================================

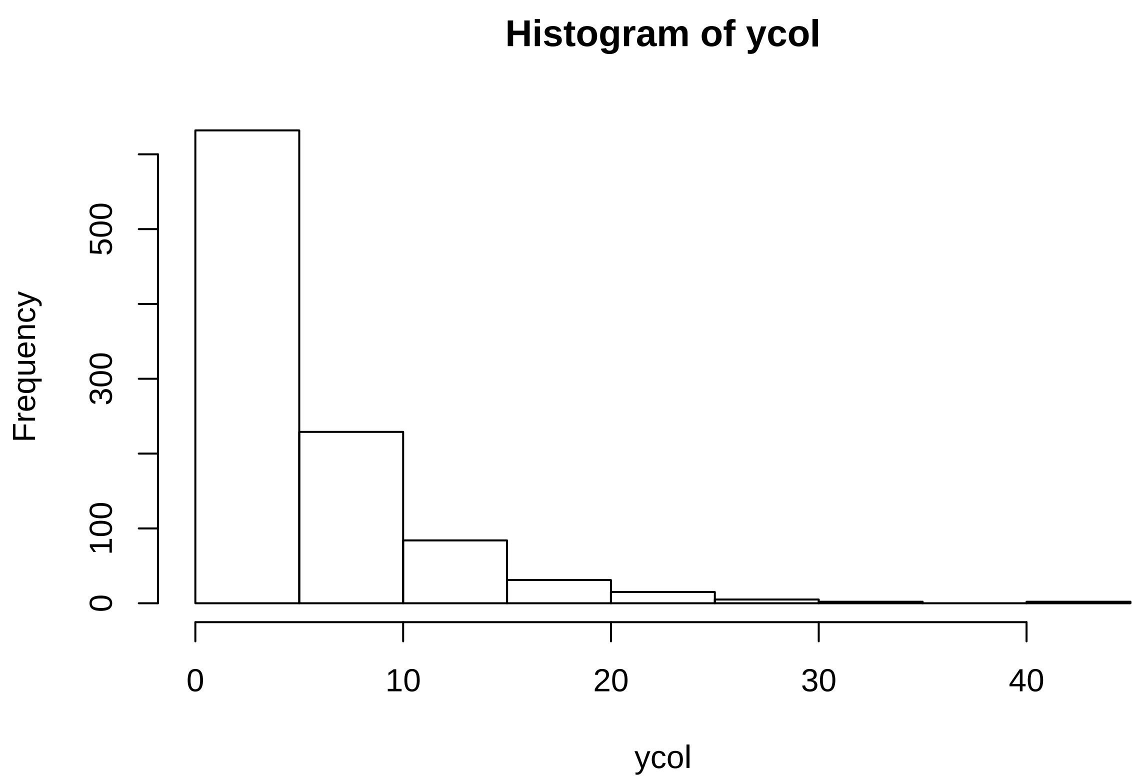
OVERVIEW

This report discusses a study of a random exponential distribution (EX) and how it does/does not compare with the Central Limit Theorem (CLT). Roughly, the central limit theorem states that the distribution of the sum (or average) of a large number of independent, identically distributed variables will be approximately normal, regardless of the underlying distribution.We shall investigate how the sample mean of EX compares to the theoretical distribution of the CLT.

SIMULATION

In order to begin the comparisons, we must first simulate the theoretical model of a population, in this case, an exponential distribution mass. A simulation was run of a sample of random exponential functions, a thousand times, lambda = 1/.20.A descriptive analysis followed.

```
ycol <- replicate(1000,rexp(1, .20))
hist(ycol); mean(ycol); var(ycol); sd(ycol)
```



```
## [1] 5.079691
```

```
## [1] 27.51807
```

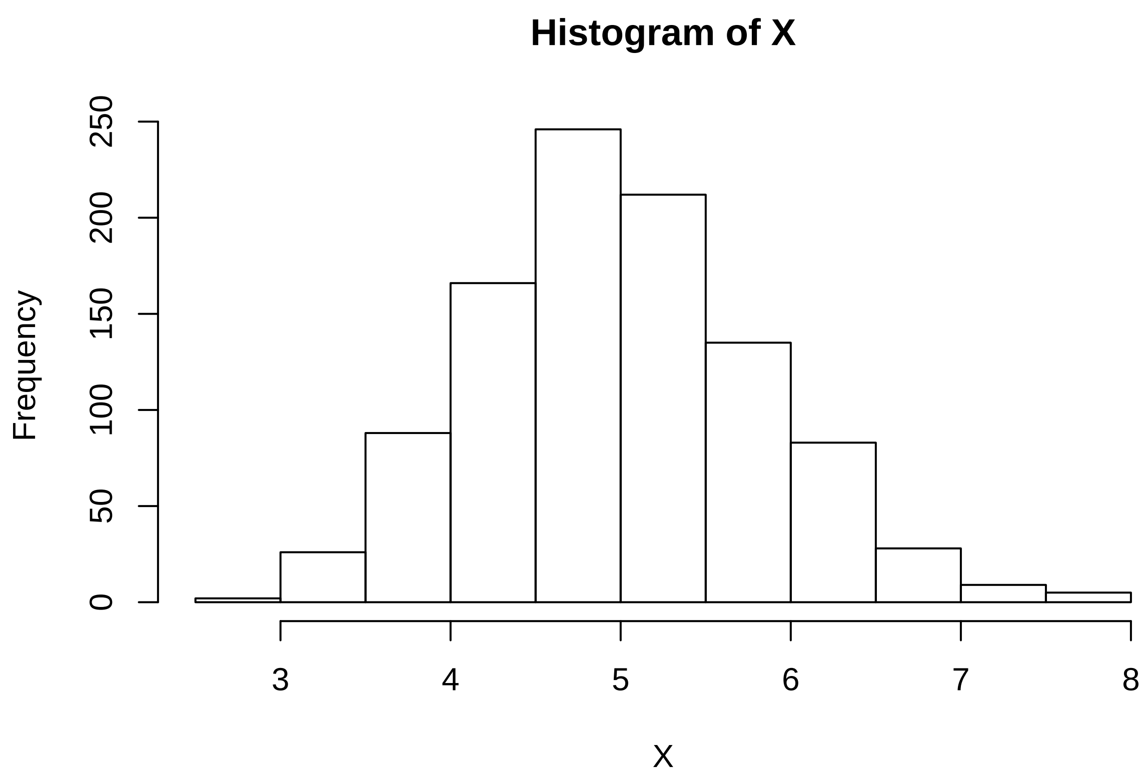
```
## [1] 5.245767
```

Next, a simulation was ran of the same exponential function 1000 times, but this time samples of n = 40 were drawn. Then the averages of each sample were recorded, and then a mean of the sample distribution was taken. The R package, matrixStats was used to column variances. Again, descriptives follow.

```
xcol <- replicate(1000, rexp(40, rate = .20))
dim(xcol)
```

```
## [1]   40 1000
```

```
X <- .colMeans(xcol, 40, 1000, na.rm = FALSE)
hist(X); mean(X); var(X); sd(X)
```




```
## [1] 4.988945
```

```
## [1] 0.6984646
```

```
## [1] 0.835742
```

SAMPLE vs.THEORETICAL MEANS

As we can see from the results of both simulations, the mean of the theoretical population, with a gamma or lambda rate of 1/.20, is approximately 5 (~ 4.83), and the mean of the sample distribution is almost the same (~ 4.99). The exponential distribution is a continuous analogue of a geometric distribution, and it has the key property of being memory less. The exponential distribution is frequently used in the Poisson process, while the sampling distribution of its mean gives a fine example of the CLT at work.

SAMPLE vs. THEORETICAL VARIANCES

Holding true to the CLT, and also the parameters associated with sampling distributions, there is a stark contrast between the variances of the two simulations. The theoretical model's variance (~ 24.67) accounts for the wide distribution of random numbers passed through the exponential function, where as the sampling model has a much tighter variance of (~ 0.62) around the mean. It must be noted that this is not quite an "apple to apple" comparison as the variance of the sampling model is taken from a vector of the average of forty samples of 1000 simulations.

DISTRIBUTION

There is also a clear distinction in the distribution and deviation from the mean between the two simulations. The theoretical shows the "ski ramp" shape of the exponential function while the sample distribution is clearly Gaussian. The standard deviation of the theoretical simulation, 1/Lambda is approximately the same as its mean (~ 4.85) while the sample distribution simulation shows a much tighter number (~ 0.748).