Statistical Inference Course Project: Part 1

Evan Oman

02/21/2015

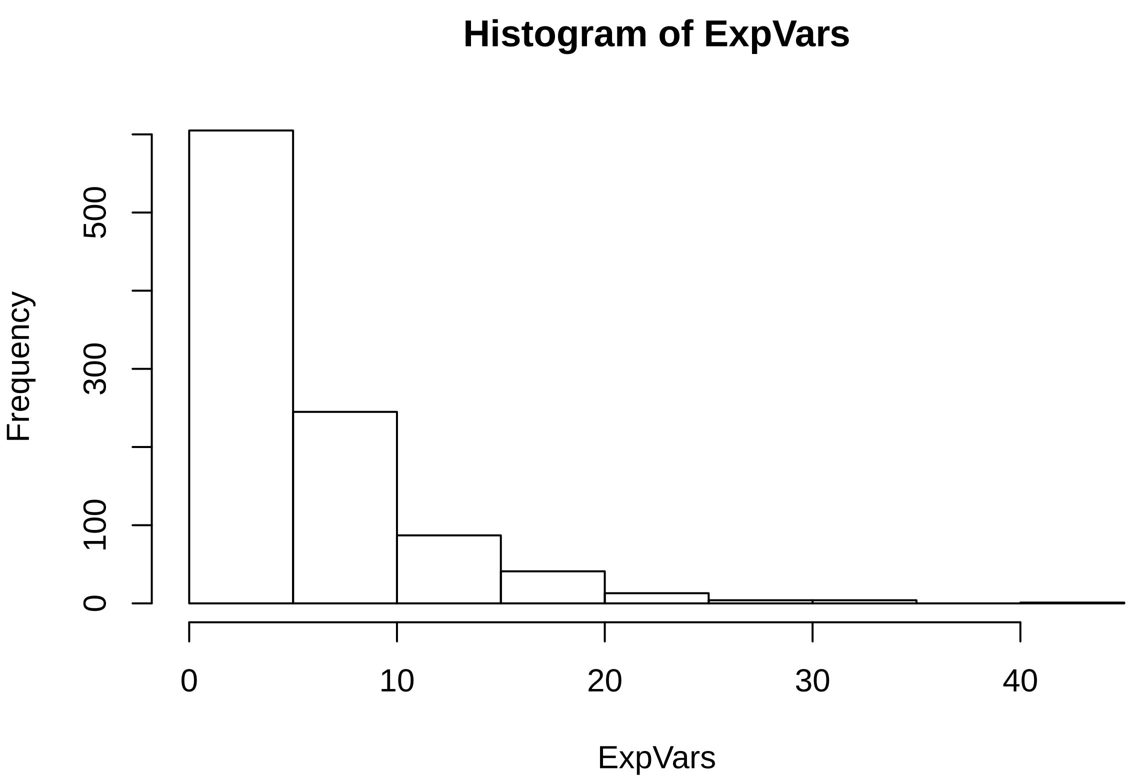
Overview

The Central Limit Theorem states: "that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution"(http://en.wikipedia.org/wiki/Central_limit_theorem). The goal of this project is to experimentally verify this special property using the exponential distribution with parameter $\lambda = 0.2$.

Simulation

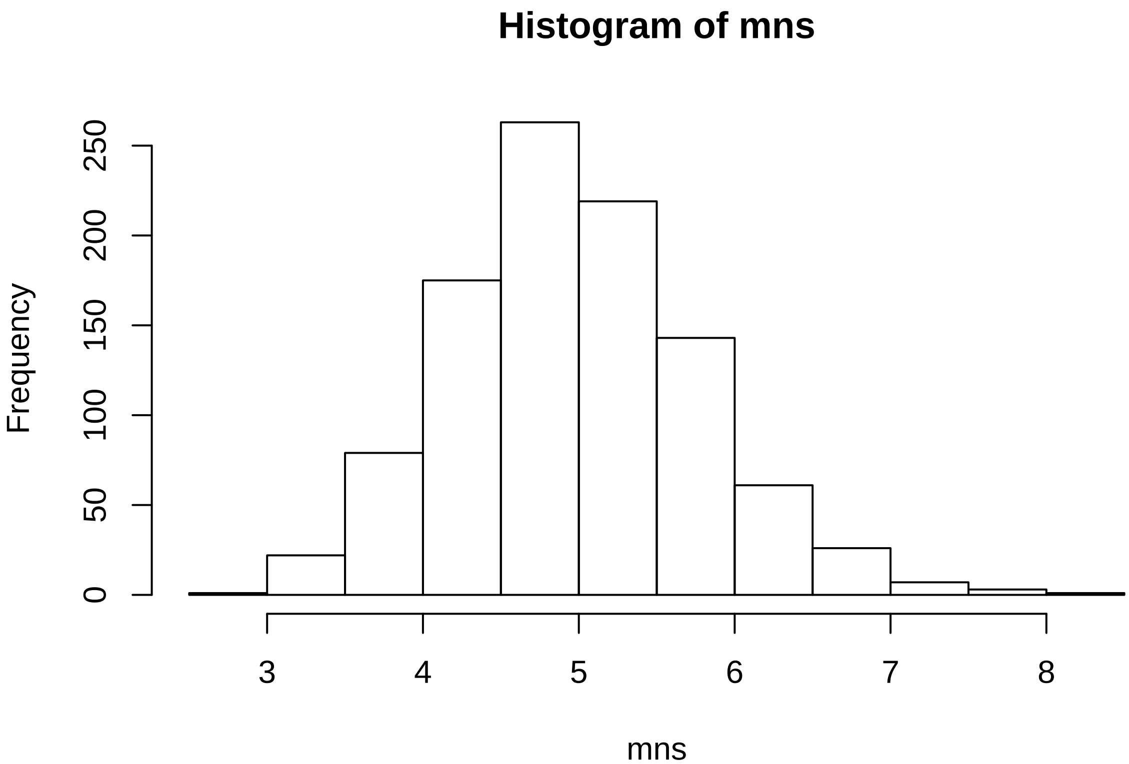
In order to test the Central Limit Theorem, we will need to generate a sufficiently large number of random variables $x \sim E(.2)$(i.e. x is exponentially distributed with $\lambda = .2$), take the mean of these random values, and then consider a large number of these means. The Central Limit Theorem says that these means should start to follow the Normal(or Gaussian) Distribution. First lets check out what a standard distribution of exponential variables looks like:

```
set.seed(42)
ExpVars <- rexp(1000, .2)
hist(ExpVars)
```



To perform the actual simulation we use the following code which creates 1000 variables which are each the mean of 40 random x values such that $x \sim E(.2)$

```
mns <- NULL
for (i in 1 : 1000) mns <- c(mns, mean(rexp(40, .2)))
hist(mns)
```



It is already evident that this distribution of variables is approximately normal based on its general shape.

Sample Mean versus Theoretical Mean

The Theoretical Sample mean is $\mu_M = \mu$, that is the sample mean should be the same as the distribution mean. For the Exponential Distribution, $E(x) = \frac{1}{\lambda} = \frac{1}{.2} = 5$. Using R to find the sample mean, we see that the Sample mean is reasonably close to the Theoretical Mean:

```
mean(mns)
```

```
## [1] 4.9809
```

Looking at the histogram for `mns`, it is clear that the approximate normal curve is centered at about 5.

Sample Variance versus Theoretical Variance

The Theoretical Variance mean is given by $\sigma_M^2 = \frac{\sigma^2}{N-1}$ where N is the sample size. For the Exponential Distribution, $Var(x) = \frac{1}{\lambda^2} = \frac{1}{.2^2} = 25$. Thus we have a sample variance of $Var(x) = \sigma_M^2 = \frac{\sigma^2}{N-1} = \frac{25}{40-1} = .641$. Using R to find the sample variance, we again see that the sample variance is reasonably close to the Theoretical Variance:

```
var(mns)
```

```
## [1] 0.632249
```

Looking at the histogram for `mns`, it is clear that the approximate normal curve has a variance of about .6.

Conclusion

Thus we can see that the sample is following the prescribed formulas and it is evident that the histogram of the sample is essentially a normal distribution. Thus the Central Limit Theorem is experimentally verifiy.