# Validation of Central Limit Theorem

#### Coursera: Statistical Inference Project 1

#### Eric Hullander

## Exponential Distribution

* .

The mean of 40 samples are taken from an exponential distribution described above. A simulation is run and the mean is calculated 10000 times. The mean and standard deviation of the means are then calculated from the distribution of sample means. As can be seen, the histogram of sample means approximates the normal curve with and sd, .

## Loading required package: ggplot2

## 

Mean of Sample Means

## [1] 4.972821

Theoretical Mean

## [1] 5

Standard Deviation of Sample Means (aka the standard error)

## [1] 0.803485

Theoretical Standard Deviation of Sample Means

## [1] 0.7905694

The sample mean is a good predictor of the population mean and with a standard error of about .8. This value agrees with the theoretical mean and calculated standard error.

N=10000  
n=40  
m=5  
ypop=rexp(N, 1/m)  
  
y=data.frame()  
for(i in 1:1000)  
 {  
 ysamp=sample(ypop,n, replace=FALSE)  
 y[i,1]=mean(ysamp)  
 y[i,2]=var(ysamp)  
 ##print(x)  
 }  
  
  
ggplot(y, aes(V1)) + stat\_function(fun = dnorm, args = list(mean = 5, sd = 5/sqrt(n))) + geom\_histogram(aes(y=..density..), col="red",   
 fill="green",   
 alpha = .2) +  
 labs(title="Distribution of Means") +  
 labs(x="Mean", y="Count")