Introduction

This project is about simulation and inference and stuff

For random variables drawn from a (continuous) exponential distribution of rate λ, the propability dirstribution function is given by

$$P(x) = \lambda e^{-\lambda x} \quad (x \geq 0).$$

This distribution has a population mean of $\mu = 1/\lambda$ a population variance of $1/\lambda^2$, and a standard deviation of $\sigma = 1/\lambda$.

CLT The central limit theorem states that the mean of a sample of iid random variables should itself behave like a random **normal** variable, with mean equal to the mean of the original population, and variance equal to **something about the standard error**

Simulation

```
#opts_chunk$set(fig.width=8, fig.height=8,dpi=144)
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
require(dplyr)
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
require(ggmap)
```

```
## Loading required package: ggmap
```

```r
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```
## Warning: package 'gridExtra' was built under R version 3.1.3
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

generate two datasets, the first is a sample of 1000 iid random variables drawn from an exponential distribution with rate $\lambda = 0.2$. The second dataset consists of 1000 sample means, where each sample consists of 40 iid random variables drawn from an exponential distribution with rate $\lambda = 0.2$

```r
set.seed(5074491)
lambda<-0.2
samples1k<-rexp(1000,rate=lambda)
means1kx40<-replicate(1000,mean(rexp(40,rate=lambda)))
# data1k<- data.frame(samples1k)
# names(data1k)<-'expSamples'
# data1kx40<-data.frame(means1kx40)
# names(data1kx40)<-'meanExp40Samples'

#names(data)<-c('expSamples','meanExp40Samples')
```

sample mean

The population mean for an exponential distribution is equal to $1/\lambda$, which in this case is $1/0.2 = 5$. The sample mean for the 1000 random variables is

```r
mean(samples1k)
```

```
## [1] 5.174467
```

the mean of the 1000 sample means is

```r
mean(means1kx40)
```

```
## [1] 5.000418
```

moose

```
# plot distributions
hist_exp<-ggplot(aes(x=samples1k,fill=..count..)) + geom_histogram(binwidth=0.2) +
labs(title='distribution of 1000 random exponential variables',x='value',y='count') +
scale_fill_gradientn(colours=c('blue','purple'))
```

```
## Error: ggplot2 doesn't know how to deal with data of class uneval
```

```
hist_expMean<-ggplot(aes(x=means1kx40,fill=..count..)) + geom_histogram(binwidth=0.2) +
labs(title='distribution of 1000 sample means of 40 random exponential variables',x='value',
```

```
## Error: ggplot2 doesn't know how to deal with data of class uneval
```

```
grid.arrange(hist_exp,hist_expMean,nrow=1)
```

```
## Error in arrangeGrob(...): object 'hist_exp' not found
```

```
# add curves to these, the population distribution and also the exponential and gausssian f
```

Part 1: Simulation Exercise Instructions In this project you 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. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the **properties of the distribution of the mean of 40 exponentials** . You should

Show the sample mean and compare it to the theoretical mean of the distribution. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. Show that the distribution is approximately normal.