

Introduction

This project is about simulation and inference and stuff

```
summary(cars)
```

```
##           speed           dist
##  Min.      : 4.0    Min.      : 2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean    : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.    :120.00
```

This summation expression $\sum_{i=1}^n X_i$ appears inline.

$$\sum_{i=1}^n X_i$$

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

require(gridExtra)

## Loading required package: gridExtra
```

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

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

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

sample mean

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

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

This distribution has a population mean of λ and a standard deviation of $\sigma = \lambda$

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

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

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.

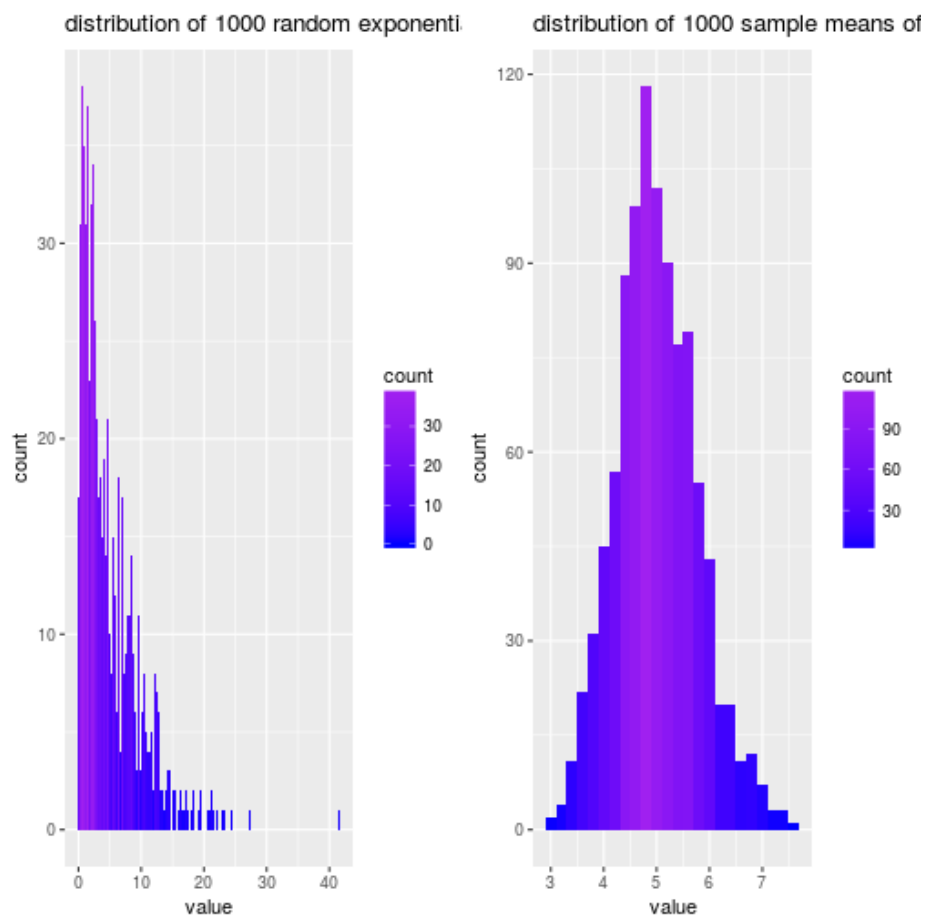


Figure 1: plot of chunk plot