

INFX 573 Lab: Central Limit Theorem

Pierre Augustamar

October 25th, 2016

Collaborators:

Don't forget to list the full names of your collaborators!

Instructions:

Before beginning this assignment, please ensure you have access to R and/or RStudio.

1. Download the `week5a_lab.Rmd` file from Canvas. Open `week5a_lab.Rmd` in RStudio (or your favorite editor) and supply your solutions to the assignment by editing `week5a_lab.Rmd`.
2. Replace the "Insert Your Name Here" text in the `author:` field with your own full name.
3. Be sure to include code chunks, figures and written explanations as necessary. Any collaborators must be listed on the top of your assignment. Any figures should be clearly labeled and appropriately referenced within the text.
4. When you have completed the assignment and have **checked** that your code both runs in the Console and knits correctly when you click Knit, rename the R Markdown file to `YourLastName_YourFirstName_lab5a.Rmd`, and knit it into a PDF. Submit the compiled PDF on Canvas.

In this lab, you will need access to the following R packages:

```
# Load some helpful libraries
library(tidyverse)
library(gridExtra)
```

Problem 1: Simulating Data in R

R can easily generate random samples from many different probability distributions. Here, you will use this functionality to explore the Central Limit Theorem by performing a simulation experiment.

Hint: You might find out more about the distribution functions in R using the help files, `?distributions`.

Step 1: Pick your favorite probability distribution.

- What distribution did you choose?

The F Distribution

- What are the parameters that characterize the distribution you chose?

The F distribution with $df_1 = n_1$ and $df_2 = n_2$ degrees of freedom has density

The F-distribution has two important properties:

It's defined only for positive values.

It's not symmetrical about its mean; instead, it's positively skewed.

$$f(x) = \frac{\Gamma((n_1 + n_2)/2)}{\Gamma(n_1/2) \Gamma(n_2/2)} \left(\frac{n_1}{n_2}\right)^{n_1/2} x^{n_1/2 - 1} \left(1 + \frac{n_1}{n_2} x\right)^{-(n_1 + n_2)/2}$$

for $x > 0$.

- Describe a situation in which you would expect to see this distribution in real-world data.

an example would be: positively skewed distribution of household incomes.

Step 2: Choose a value for each parameter in the distribution (e.g. the mean and variance for the Normal distribution). Use the random generation function for this distribution to construct 100 random samples of sample sizes $n = 10, 20, 50, 100, 500$. `df` gives the density, `pf` gives the distribution function, `qf` gives the quantile function, and `rf` generates random deviates.

```
# generating distribution for F Distribution
```

```
randomDist <- rf(10, 4, 5)
randomDist <- rf(20, 2, 3)
randomDist <- rf(50, 7, 8)
randomDist <- rf(100, 3, 4)
randomDist <- rf(500, 5, 6)
```

```
# par(mfrow=c(2,5)) ?gridExtra
```

Step 3: Compute the sample mean for each of the 100 random samples. Construct a visualization showing the distribution of the sample mean for each case (i.e. probability distribution and sample size pair). You might find the following code helpful for showing multiple plots at once.

- What is the true population mean for the distribution?
- What patterns do you see in the distribution of the sample mean as the sample size n increases?

Hint: Look at OpenIntro Statistics Chapter 3.

Hint: Each distribution function in R has an associated function to generate random deviates, e.g. `rbinom` for the Binomial distribution.

Hint: Most distributions are characterized by parameters related to the mean and variance.

- How does this simulation experiment demonstrate the Central Limit Theorem?