Problem Set #7

Econ 103

Part I – Problems from the Textbook

Chapter 6: 1, 3, 5, 7
Chapter 7: 1, 3, 5, 9, 13, 17, 18, 19

The answer in the back of the book for 7-19 is wrong. *I will provide full solutions to 7-13 since it's hard, 7-18 since it's even-numbered, and 7-19 since the book is wrong.*

Part II – Additional Problems

1. In this question you will replicate some of the density plots from Lecture 13.
   (a) Plot a standard normal pdf on the same graph as a $t(1)$ pdf on the interval $[-5, 5]$. How do the pdfs compare? Explain.
   (b) Plot a standard normal pdf on the same graph as a $t(100)$ pdf on the interval $[-5, 5]$. How do the pdfs compare? Explain.
   (c) Plot a χ^2 pdf with degrees of freedom equal to 4 on the interval $[0, 20]$.
   (d) Plot an F pdf with numerator degrees of freedom equal to 4 and denominator degrees of freedom equal to 40 on the interval $[0, 5]$.
2. In this question you will verify the empirical rule both directly using `pnorm` and by simulation using `rnorm`.
   (a) Draw 100000 iid observations from a standard normal distribution and store your results in a vector called `sims`.
   (b) What proportion of the observations in `sims` lie in the range $[-1, 1]$?
   (c) What proportion of the observations in `sims` lie in the range $[-2, 2]$?
   (d) What proportion of the observations in `sims` lie in the range $[-3, 3]$?
   (e) Use `pnorm` to calculate the exact probabilities for a standard normal pdf that correspond to the above simulation experiments. How accurate were your simulations?

3. In this question you will replicate the Monte Carlo Experiment from Lecture 14. In particular, you will use R to study the sampling distribution of the sample mean where we take as our population the heights of all students in Econ 103.

   (a) Load the class survey data used in R Tutorial # 2, extract the height column and assign it to a variable called `height`. Use `!is.na` to remove all missing values from `height`.

   (b) Make a histogram of height and calculate the mean height for students in the class. For the purposes of this exercise, these correspond to the *population*.

   (c) Write a function that takes n as its only input and returns the sample mean of an iid random sample of size n drawn from the vector `height`. Call this function `x.bar.draw`. [Hint: use `sample` with `replace = TRUE`.]

   (d) Test the function you wrote for part 4 by running it with `n = 10000`. What value do you get? Your answer should be approximately equal the population mean you calculated above. If it isn't, something is wrong with your code.

   (e) Using the R function `replicate`, run your function `x.bar.draw` 10000 times with $n = 5$. Store the result in a vector called `x.bar.5`. Do the same for $n = 10, 20$ and 50 and store the results as vectors `x.bar.10`, `x.bar.20` and `x.bar.50`

   (f) Calculate the mean and variance of `x.bar.5`,`x.bar.10`, `x.bar.20` and `x.bar.50` and plot a histogram of each, being sure to label them.

4. In this question you will replicate the Law of Large Numbers (LLN) visualization from lecture 15, in which we plotted "running" sample means as we kept adding more and more simulations from a $N(\mu = 0, \sigma^2 = 100)$ distribution. Your plot won't look exactly like the one from class since this is a random experiment, but it will show the same qualitative behavior.

   (a) The R command for a "running" or "cumulative" sum is `cumsum`. Look at the help file for this command and test it out on a vector of ten ones and another containing the integers from one to ten to make sure you understand what it does.

   (b) Replicate the plot on slide 30/35 of lecture 15. First you'll need to draw 10,000 iid samples from a $N(\mu = 0, \sigma^2 = 100)$ distribution. Then you'll need to calculate the running means. You'll need to figure out how `cumsum` can be used to accomplish this. Finally, plot your results along with a dashed red line at the value to which the sample mean is converging. Make sure to label your axes.

   (c) Repeat the previous part but, rather than drawing $N(\mu = 0, \sigma^2 = 100)$ simulations, draw from a Student-t distribution with one degree of freedom. How do your results differ? Use what you know about the Student-t distribution to guess why our proof that the sample mean is consistent for the population mean doesn't work here.