MATH 4780 (MSSC 5780) Homework 1

Probaility and Statistics Review

Due Date: September 12, 2022 11:59 PM

1 Homework Instruction and Requirement

- Homework 1 covers course materials of Week 1 to 2.
- Please submit your work in **one PDF** file to **D2L > Assessments > Dropbox**. *Multiple files or a file that is not in pdf format are not allowed.*
- In your homework, please number and answer questions **in order**.
- There is no need to submit your work on R programming part. However, I highly recommend reviewing basic R syntax for R data structures and graphics if you haven't got familiar with it.
- It is your responsibility to let me understand what you try to show. If you type your answers, make sure there are no typos. I grade your work based on *what you show, not what you want to show.* If you choose to handwrite your answers, write them neatly. If I can't read your sloppy handwriting, your answer is judged as wrong.

2 R Programming

Please sharpen your R skill. **No need** to show your work on this part! :)

1. Please register RStudio Cloud or get R and RStudio installed in your laptop. Read the R and RStudio slides in Week 1 module for installation instruction and basic usage of RStudio.
2. If you are not familiar with basic R syntax and data types, please review my slides of MATH 3570 in Week 1 module.
3. You can test your understanding of R by doing the problems in `basic_r.pdf` that is actually Homework 1 of my MATH 3570 course Spring 2021.

3 Probability and Statistics Review

We will use some facts or properties discussed in MATH 4700 and 4710. Here we don't learn why (which should be taught in MATH 4700 and 4710), but just know what they are, and apply them for regression analysis.

1. Use the linearity of expected value $E(X+Y) = E(X) + E(Y)$ and $E(aX) = aE(X)$ where X and Y are random variables and a is a constant to show that for a random variable Y,

$$\mathrm{Var}(Y) = E(Y^2) - [E(Y)]^2$$

2. Let the two independent random variables be $Y_1 \sim N(3, 8)$ and $Y_2 \sim N(1, 4)$. What is the distribution of the variable $2Y_1 + 3Y_2$?

3. Suppose $Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$. Show that $\overline{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, where $\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$.

4. Let the sample variance be $S^2 = \frac{\sum_{i=1}^n (Y_i - \overline{Y})^2}{n-1}$. The two facts are

   i. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

   ii. $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{(n-1)S^2}{\sigma^2}$ are independent.

   Use i. and ii. and the fact that *If* $Z \sim N(0,1)$, $V \sim \chi^2_v$ *and* Z *and* V *are independent, then* $\frac{Z}{\sqrt{V/v}} \sim t_v$ to show that

$$\frac{\overline{Y} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

5. Suppose $Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$, with unknown μ and σ. The $100(1-\alpha)\%$ confidence interval (CI) for the population mean μ is $\left(\overline{y} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \overline{y} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$. Use simulation with $\alpha = 0.1$, $\mu = 4$ and $\sigma = 2$ to verify that such CIs contain μ about $100(1-\alpha)\%$ of times. Fill the percentage in the following table, and comment on your results. Attach your code at the end of your homework PDF file.

Simulation times	$n = 5$	$n = 30$	$n = 200$
20			
1000			
20000			

6. If U_1 and U_2 are independent and both are uniform variables over $[0,1]$ interval (https://en.wikipedia.org/wiki/Continuous_uniform_distribution), then X_1 and X_2 defined by

$$X_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2), \quad X_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2)$$

   are independent $N(0,1)$ variables. Draw 10,000 samples for U_1 and U_2 using the `runif()` function, and use the transformation to generate the samples of X_1 and X_2. Verify

- the standard normality of X_1 and X_2 by plotting their histogram with a superimposed standard normal density.
- the independence of X_1 and X_2 by plotting the scatterplot of X_1 and X_2 and computing their correlation coefficient.