M/STAT 501 Fall 2022: Homework 3

Due Wednesday, Sep 28 by 5:00pm in Gradescope

Directions: For this homework, there are three different types of problems: 1) *Graded for Accuracy Problems*, 2) *Graded for Completion Problems*, and 3) *Extra Practice Problems*. Please turn in your work for the *Graded for Accuracy Problems* and the *Graded for Completion Problems*, as well as any *Sources Used* by Wednesday, September 28 by 5:00pm via Gradescope. Show work and/or provide justification for credit—neatness counts! *The Extra Practice Problems* do not need to be turned in, and the solutions will be posted for you to check and compare your work.

You are encouraged to use R Markdown for your homework, but it is not required.

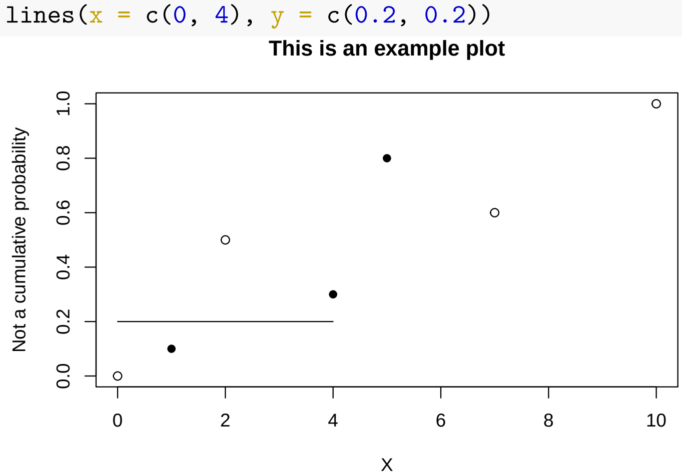
Graded for Accuracy Problems (turn in)

1. Suppose a random variable X has a cumulative distribution function

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.10 & 0 \leq x < 1 \\ 0.50 & 1 \leq x < 2 \\ 0.75 & 2 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$

a. Is this a discrete, continuous, or mixture random variable? Why?
b. Find the pmf/pdf of X.
c. Use R to plot a graph of the cdf of X, and a graph of the pmf/pdf of X. *Hint:* One option for plotting in R is to set up an empty plot window using the `plot` function, and then add points and lines using the `points` and `lines` functions. For example:

```
plot(x = c(0, 10), y = c(0, 1),
     xlab = "X", ylab = "Not a cumulative probability",
     main = "This is an example plot")
points(x = c(1, 4, 5), y = c(0.1, 0.3, 0.8), pch = 16) # pch = point character
points(x = c(2, 7), y = c(0.5, 0.6), pch = 1)
lines(x = c(0, 4), y = c(0.2, 0.2))
```



2. In a genetics study, it was found that the distance Y between mutations in a certain strand of DNA had a probability density function given by

$$f(y) = \begin{cases} ce^{-y/5} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

in kb.

a. Find the value of c.
b. Find the probability that the distance is between 4 and 8 kb.
c. What is $P(Y = 5)$?
d. Find the cdf $F(y)$. Show that $F(y)$ satisfies the three conditions of Theorem 1.5.3.

3. A random variable X has pdf

$$f_X(x) = \begin{cases} x^2 & 0 < x \leq 1 \\ \frac{2}{3} & 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

a. Verify that $f_X(x)$ is a pdf.
b. The median of a distribution is any number m such that $P(X \leq m) \geq \frac{1}{2}$ and $P(X \geq m) \geq \frac{1}{2}$. Find the median of X.
c. Find the CDF of X and plot the CDF using R. Plot the position of the median found in part b. on the graph.

Graded for Completion Problems (turn in)

Use the following information to explore a real-life application of CDFs. (Note: The point of this question is not mastery of the specific topic (item response theory), but instead exposure to a real-life use of the concepts we are discussing.)

Item response theory (IRT) is a model-based measurement paradigm in which individuals' latent trait estimates (i.e., estimates of unobservable ability or trait, such as "intelligence" or "extroversion") depend on both individuals' responses to items on an instrument and the properties of the items that were administered. Item response functions (IRFs) give the probability that an individual with a given level of the latent trait (i.e., ability) will answer a particular item correctly. These IRFs are constructed based on either the normal probability distribution function or the logistic function.

Find a resource or set of resources that describe and compare normal ogive and logistic IRFs. If desired, you may use the following online resources:

- Castro, S. *Psychometrics: Item Response Theory*. RPubs. Online: https://rpubs.com/castro/irt1 (Note that this resource has some awesome Shiny apps that allow you to explore different IRFs and how they change for items with different properties!)
- Wikipedia. Item response theory. Online: https://en.wikipedia.org/wiki/Item_response_theory.

For fixed item properties, briefly explain how the cumulative distribution function (CDF) of the standard normal distribution is used to construct an IRF.

Sources Used (turn in)

At the end of your assignment, list and cite all sources used when working on this assignment, including individuals with whom you discussed the problems. If you did not use any sources besides our textbook or the course notes, please note this.

Extra Practice Problems (do not turn in, solutions posted in D2L)

1. For each of the following, determine the value of c that makes $f_X(x)$ a pdf or pmf:
   a. $f_X(x) = ca^x I_{\{0,1,2,\ldots\}}(x)$, where $0 < a < 1$.
   b. $f_Y(y) = cxe^{-2x} I_{[0,\infty)}(x)$

2. Are the following functions CDFs? Prove or disprove.
   a. $G(x) = (1 - e^{-x}) I_{[-1,\infty)}(x)$
   b. $G(x) = \dfrac{e^x}{2} I_{(-\infty,0)}(x) + \left(1 - \dfrac{e^{-x}}{2}\right) I_{[0,\infty)}(x)$

3. For each found CDF in problem 2, find its corresponding pdf.