This problem set covers material from Week 4, dates 9/29 – 10/02.

Instructions: Write or type complete solutions to the following problems and submit answers to the corresponding Canvas assignment. Your solutions should be neatly-written, show all work and computations, include figures or graphs where appropriate, and include some written explanation of your method or process (enough that I can understand your reasoning without having to guess or make assumptions). A general rubric for homework problems appears on the final page of this assignment.

Monday 9/29

None – added to Wednesday section!

Wednesday 10/01

1. The American Community Survey (ACS) is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 ACS estimated that 14.6% of Americans live below the poverty line, 20.7% speak a language other than English (i.e. a foreign language) at home, and 4.2% fall into both categories.

   (a) Are living below the poverty line and speaking a foreign language at home disjoint? Why or why not?

   (b) Draw a Venn diagram that clearly summarizes the events and their associated probabilities. Be sure to complete the diagram by including a "bounding box".

   (c) What percent of Americans live below the poverty line and only speak English at home?

   (d) What percent of Americans live below the poverty line or speak a foreign language at home?

   (e) What percent of Americans live above the poverty line and only speak English at home?

   (f) Is the event that someone lives below the poverty line independent of the event that the person speaks a foreign language at home?

2. A school has three divisions: Social Sciences, Humanities, and Natural Sciences. A student belongs to exactly one division. The Social Sciences have 50 students, and 20% of them are female-identifying. The Humanities have 30 students, and 30% are female-identifying. The Natural Sciences have 20 students, and 40% are female-identifying. We will randomly sample one student from the entire school.

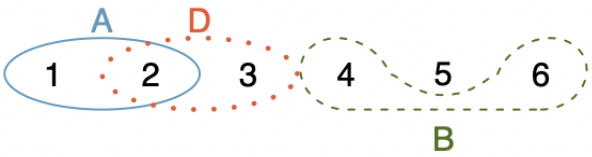
   (a) What is the probability that the randomly selected student is in the Humanities and identifies as a female?

   (b) Using the LoTP, what is the probability that the randomly selected student identifies as male?

(c) Given that the student identifies as female, what is the probability that they are in the Natural Sciences?

(d) Are division and gender independent? Why or why not?

3. Suppose our random process is rolling a possibly *unfair* die one time. The sample space is $S = \{1, 2, 3, 4, 5, 6\}$, as usual. From this process, we define three different events A, B and D as denoted by the coloring and circling:



(a) Which pairs of events A, B, D are disjoint?

(b) Let X be the random variable corresponding to the random process above. Also suppose that $\mathrm{P}(B) = 0.2$, $\Pr(D^c) = 0.6$, and events A and D are independent. Based on this information, create a table that represents a valid probability distribution for X. *Remember: distributions are define probabilities for outcomes!*

Thursday 10/02

4. The global coronavirus pandemic illustrates the need for accurate testing of COVID-19, as its extreme infection rate poses a significant public health threat. Due to the time-sensitive nature of the situation, the FDA enacted emergency authorization of a number of serological tests for COVID-19. Full details of these tests may be found on its website.

   Let D be the event that a patient has COVID-19, and let T be the event that a test is positive for COVID-19. We will define some probabilities that are commonly used in these medical testing settings:

   - Prevalence: $\mathrm{P}(D)$
   - Sensitivity: $\mathrm{P}(T|D)$
   - Specificity: $\mathrm{P}(T^c|D^c)$
   - Positive predictive value: $\mathrm{P}(D|T)$
   - Negative predictive value: $\mathrm{P}(D^c|T^c)$

   According to the website, the Immunodiagnostic Systems COVID-19 test has a sensitivity of 97.6% and specificity of 99.6%. Suppose the prevalence of COVID-19 in the general US population 2%. What are the positive and negative predictive values of the Immunodiagnostic Systems test?

5. To get to Middlebury College, a professor uses their car 30% of the time, walks 20% of the time, and bikes 50% of the time. They are late 5% when walking, 10% of the time when driving (because this is Vermont and people stop for all pedestrians), and 2% of the time when biking.

   (a) What is the probability the professor drove to work if they were late?

   (b) What is the probability the professor walked to work if they were on time?

6. The `R` problems in the associated `.qmd` document. `R` problems 1-3 will be graded as one problem, as will 4-5, and 6-7. **Note: I will start deducting points for code that is left in your submission but is not necessary for answering the question. Code that has been commented out is okay though!**

General rubric

Points	Criteria
5	The solution is correct *and* well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect.
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification.
0	Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information).
Notes:	For problems with multiple parts, the score represents a holistic review of the entire problem. Additionally, half-points may be used if the solution falls between two point values above.
Notes:	For problems with code, well-written means only having lines of code that are necessary to solving the problem, as well as presenting the solution for the reader to easily see. It might also be worth adding comments to your code.