Homework 5

Due October 31 at 11 pm

Unless stated otherwise, justify any answers you give. You can work in groups, but each student must write their own solution based on their own understanding of the problem.

When uploading your homework to Gradescope you will have to select the relevant pages for each question. Please submit each problem on a separate page (i.e., 1a and 1b can be on the same page but 1 and 2 must be on different pages). We understand that this may be cumbersome but this is the best way for the grading team to grade your homework assignments and provide feedback in a timely manner. Failure to adhere to these guidelines may result in a loss of points. Note that it may take some time to select the pages for your submission. Please plan accordingly. We suggest uploading your assignment at least 30 minutes before the deadline so you will have ample time to select the correct pages for your submission. If you are using LaTeX, consider using the minted or listings packages for typesetting code.

1. (Babysitter) A babysitter is taking care of a baby. She gives him some food and then puts him to sleep. We make the following assumptions:

   - The probability that the food is bad is 0.1.
   - If a baby eats food that is bad, they will wake up in the middle of the night. If the food is not bad, they may still wake up (with a probability that depends on whether they are good or bad sleepers).
   - All babies can be classified into *good sleepers* or *bad sleepers*. The probability that a baby that is a *good sleeper* wakes up is 0.1 (assuming the food is fine). The probability for a baby that is a *bad sleeper* is 0.8.
   - A baby is a *good sleeper* with probability 0.6. This is independent from the food.

   We model the problem by defining Bernoulli random variables $\tilde{b}$ indicating whether the baby is a good ($\tilde{b} = 1$) or bad sleeper ($\tilde{b} = 0$), $\tilde{w}$ indicating whether the baby wakes up in the middle of the night ($\tilde{w} = 1$) or not ($\tilde{w} = 0$), and $\tilde{x}$ indicating whether the food is bad ($\tilde{x} = 1$) or not ($\tilde{x} = 0$).

   (a) What is the probability that the baby wakes up in the middle of the night?

   (b) If the baby wakes up in the middle of the night, what is the probability that the food was bad?

   (c) Compute the probability that the food is bad conditioned on the baby waking up and being a good sleeper. Are $\tilde{b}$ and $\tilde{x}$ conditionally independent given $\tilde{w}$? Justify your answer mathematically and explain it intuitively.

2. (Earthquake) During a period of high seismic activity, a group of scientists is trying to predict the occurrence of earthquakes by measuring vibrations in the ground. They model the occurrence of an earthquake as a random variable $\tilde{e}$ ($\tilde{e} = 1$ if there is an earthquake, and $\tilde{e} = 0$ if there isn't), and the vibrations as a random variable $\tilde{v}$ ($\tilde{v} = 0$ if there are no vibrations, $\tilde{v} = 1$ if there are small vibrations, and $\tilde{v} = 2$ if there are large vibrations). The joint pmf of $\tilde{e}$ and $\tilde{v}$ is:

Vibrations

	$p_{\tilde{e},\tilde{v}}$	0	1	2
Earthquake	0	0.8	0.05	0
	1	0	0.05	0.1

The sensor reading is modeled as a Bernoulli random variable $\tilde{s}$ that is conditionally independent of the earthquake given the vibrations. If there are no vibrations, the reading is always 0, if there are small vibrations the reading is 1 with probability 0.5, and if there are large vibrations the reading is always 1.

(a) Derive the marginal pmf of $\tilde{s}$.

(b) What is the probability that there is an earthquake if the sensor reading equals 1?

(c) Are the random variables $\tilde{s}$ and $\tilde{e}$ independent? Justify your answer mathematically, but also explain it intuitively.

3. (Surgery) A hospital wants to evaluate two surgery procedures: A and B. There are two types of patients that receive the procedure, *mild* and *serious* cases. The truth is that procedure A is better. Mild cases recover with probability 0.9 if they receive A, and 0.8 if they receive B. Serious cases recover with probability 0.5 if they receive A, and 0.2 if they receive B.

   (a) The data shows that patients recover with probability 0.58 if they receive procedure A, and 0.68 if they receive B. How is this possible? Justify your answer mathematically. (Hint: Start by computing what fraction of patients receiving each procedure are mild or serious cases.)

   (b) Explain how to analyze the data in order to obtain an accurate conclusion about the surgery procedures. Under what assumption does this work?

   (c) Suggest how to design a follow-up study that would not require adjusting for confounding factors.

4. (Stock) The table in *pricedelta.csv* records daily share prices changes. For each stock, model whether the price goes up or down with a Bernoulli random variable. Estimate the following probabilities from the data.

   (a) What is the joint pmf of the three random variables representing *amzn*, *jpm*, *spy*? (One $2 \times 2 \times 2$ matrix)

   (b) What is the marginal pmf of each possible pair of random variables? (Three 2×2 matrices)

   (c) What is the conditional pmf of each possible pair given the remaining random variable? (Three $2 \times 2 \times 2$ matrices)