Homework 1

Due September 19 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. (True or False) Prove the following statements or provide a counterexample. Let A, B, and C be events in a probability space.
   (a) If A and B are independent, then so are A^c and B.
   (b) If A and B are conditionally independent given C, then they are also conditionally independent given C^c.
   (c) Events in a partition cannot be independent (assume that every event in the partition has nonzero probability).
   (d) If $\mathrm{P}(A|B) = 1$ then $\mathrm{P}(B^c|A^c) = 1$.
   (e) $\mathrm{P}(B|A \cup B) \geq \mathrm{P}(B|A)$. (Hint: Use the facts that $\mathrm{P}(B \mid A \cup B) = \mathrm{P}(B \cap A \mid A \cup B) + \mathrm{P}(B \cap A^c \mid A \cup B)$ and $\mathrm{P}(B \mid (A \cup B) \cap A) = \mathrm{P}(B \mid A)$.)
2. (Probability spaces)
   (a) Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space. Let A be an event in the σ-algebra $\mathcal{F}$, such that $\mathrm{P}(A) \neq 0$, on which we want to condition. We define a collection of events $\mathcal{F}_A$ as the collection of the intersection of A with all the events in $\mathcal{F}$:

   $$\mathcal{F}_A = \{A \cap F : F \in \mathcal{F}\}.$$

   If we consider a new sample space $\Omega_A := A$, prove that $\mathcal{F}_A$ is a valid σ-algebra, and also that the conditional probability measure

   $$\mathrm{P}_A(S \cap A) := \frac{\mathrm{P}(S \cap A)}{\mathrm{P}(A)}, \tag{1}$$

   where $S \in \mathcal{F}$, is a valid probability measure on $\mathcal{F}_A$.

   (b) Suppose we have a sample space $\Omega = \{1, \ldots, M\}$ with σ-algebra $\mathcal{F} := 2^\Omega$, the power set of Ω. To determine P, the probability measure, we employ the following empirical procedure:

i. Collect N data points taking values in Ω (e.g., N rolls of an M-sided die). Call these observations $x_1, \ldots, x_N$.

ii. For each $S \subseteq \Omega$, define

$$\mathrm{P}(S) := \frac{\text{number of } i\text{-values such that } x_i \in S}{N}.$$

As an example, suppose $M = 2$ and we flip a coin $N = 10$ times getting 6 heads and 4 tails, where 1 denotes head and 2 denotes tail. Then

$$\mathrm{P}(\emptyset) = 0, \quad \mathrm{P}(\{1\}) = 0.6, \quad \mathrm{P}(\{2\}) = 0.4, \quad \text{and} \quad \mathrm{P}(\{1, 2\}) = 1.$$

If P is defined using the above procedure, will it always result in a valid probability measure? Either prove that it will, or give a counterexample.

3. (Testing) A company with 10 employees decides to test them for COVID-19 before they go back to work in person. From available data, they determine that the probability of each employee being ill is 0.01. The employees have not been in contact with each other for a while, so the events *Employee i is ill*, for $1 \leq i \leq 10$, are modeled as independent. If an employee is ill, the test is positive with probability 0.98. If they are not ill, the test is positive with probability 0.05.

   (a) Is it reasonable to model the events *Test i is positive*, for $1 \leq i \leq 10$, as independent? From now on model them as independent whether you think it is reasonable or not.

   (b) The company tests all employees. What is the probability that there is at least one positive test?

   (c) If there is at least one positive test, what is the probability that nobody is ill? If you make any independence or conditional independence assumptions, please justify them.

4. (Streak of heads) In this problem we consider the problem of testing whether a randomly generated sequence is truly random. A certain computer program is supposed to generate independent fair coin flips. When you try it out, you are surprised that it contains long streaks of 1s. In particular, you generate a sequence of length 200, which turns out to contain a sequence of 8 heads in a row.

   (a) Compute the probability that the longest streak of heads that you observe has length x for $x \in \{1, 2, 3, 4, 5\}$ when you flip a fair coin 5 times, and the flips are independent.

   (b) Complete the script *streaks.py* to estimate these probabilities using Monte Carlo simulation. Compare it to your answer in the previous question. The script will also apply your code to estimate the probability of streaks of heads with different lengths for 200 flips. Include your code in the answer as well as the figures generated by the script.

   (c) What is the estimated probability that the longest streak of heads has length 8 or more for 200 flips? Is the sequence of 8 ones evidence that the program may not be generating truly random sequences?