

# Homework 1

Due September 22 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 L<sup>A</sup>T<sub>E</sub>X, 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  $P(A|B) = 1$  then  $P(B^c|A^c) = 1$ .
2. (Probability spaces)
  - (a) Let  $(\Omega, \mathcal{C}, P)$  be a probability space. Let  $A$  be an event in the collection  $\mathcal{C}$ , such that  $P(A) \neq 0$ , on which we want to condition. We define a collection of events  $\mathcal{C}_A$  as the collection of the intersection of  $A$  with all the events in  $\mathcal{C}$ :

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

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

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

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

- (b) Suppose we have a sample space  $\Omega = \{1, \dots, M\}$  with collection  $\mathcal{C} := 2^\Omega$ , the power set of  $\Omega$ . To determine  $P$ , the probability measure, we employ the following empirical procedure:
  - i. Collect  $N$  data points taking values in  $\Omega$  (e.g.,  $N$  rolls of an  $M$ -sided die). Call these observations  $x_1, \dots, x_N$ .

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

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

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

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

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

3. (Baby sleep) A babysitter is taking care of a baby. They give her some food and put her to sleep. Assume the following:

- 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 in the middle of the night is 0.1. The probability for a baby that is a *bad sleeper* is 0.8.
- A baby is a *good sleeper* with probability 0.6.

Answer the following questions indicating any (reasonable) assumptions you make about independence between events.

- (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 is bad?
  - (c) Compute the conditional probability that the food is bad given that a good sleeper wakes up.
  - (d) Under our assumptions, are the events *baby is a good sleeper* and *food is bad* conditionally independent given the event *wakes up in the middle of the night*? Justify your answer mathematically and explain it intuitively.
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?