SI 140A: Probability and Statistics for EECS
Spring 2025, Instructor: Xavier Lagorce, TA: Junye Wang, Chuyan Zhou

Homework 2

Due: 23:59 (CST), Mar 14, 2025
Full Mark: 50 points, 5 problems including 1 bonus.

Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- You are required to write down all the major steps towards making your conclusions; otherwise you may obtain limited points of the problem.
- Write your homework in English; otherwise you will get no points of this homework.
- Any form of plagiarism will lead to 0 point of this homework.
- If there are any descriptions unclear, we recommend you to ask on Piazza.

Problem 1

(Conditional Probability: Paradoxes 1 *2'+2'+4'+2'+4'***)**

Consider the following scenario, from Tversky and Kahneman: Let A be the event that before the end of next year, Peter will have installed a burglar alarm system in his home. Let B denote the event that Peter's home will be burglarized before the end of next year.

a) Intuitively, which do you think is bigger, $P(A|B)$ or $P(A|B^c)$? Explain your intuition.

b) Intuitively, which do you think is bigger, $P(B|A)$ or $P(B|A^c)$? Explain your intuition.

c) Show that for **any** events A and B (with probabilities not equal to 0 or 1), $P(A|B) > P(A|B^c)$ is equivalent to $P(B|A) > P(B|A^c)$.

d) If your intuition shows the exactly contrary proposition to (c), that is a frequent result, investigated by Tversky and Kahneman. What is a plausible explanation for why your intuition shows this proposition despite (c) showing that it is impossible for these inequalities both to hold?

You are told that a genetic test is extremely good: 100% sensitive (that is, it is always correct if you have the disease) and 99.99% specific (that is, it gives a false positive only 0.01% of the time).

e) A rare genetic disease of which only one in a million carriers is discovered. Using Bayes' Rule, explain why you might not want to take this genetic test.

Problem 2

(Recurrent Coins *5'+4'+5'***)**

A coin with a probability p of getting a head is tossed n times, and let the event Q_n be that the head never appears twice (or more) in a row. Let $q_n = P(Q_n)$. For convenience, let H_n be the event of getting a head.

a) Find q_n in terms of a recurrence relation when $n \geq 3$.

b) Find q_n in terms of n when $p = \frac{2}{3}$. If you determine to use characteristic equations, please explain why they are effective.

c) Generally, we can abstract our problems into this second order difference equation:

$$f_{n+1} = bf_n + af_{n-1}, \ \forall n \geq 1.$$

Please determine f_n given f_0 and f_1 known.

Problem 3

(Naive Parity Check over a Noisy Channel *3'+4'+4'***)**

A message is sent over a noisy channel. The message is a sequence $x_1, x_2, ..., x_n$ of n bits ($x_i \in \{0, 1\}$). Since the channel is noisy, there is a chance that any bit might be corrupted, resulting in an error (a 0 becomes a 1 or vice versa). Assume that the error events are independent. Let p be the probability that an individual bit has an error ($0 < p < \frac{1}{2}$). Let $y_1, y_2, \ldots, y_n$ be the received message (so $y_i = x_i$ if there is no error in that bit, but $y_i = 1 - x_i$ if there is an error there).

To help detect errors, the n-th bit is reserved for a parity check: x_n is defined to be 0 if $x_1 + x_2 + \ldots + x_{n-1}$ is even, and 1 if $x_1 + x_2 + \ldots + x_{n-1}$ is odd. When the message is received, the recipient checks whether y_n has the same parity as $y_1 + y_2 + \ldots + y_{n-1}$. If the parity is wrong, the recipient knows that at least one error occurred; otherwise, the recipient assumes that there were no errors.

a) For $n = 5, p = 0.1$, what is the probability that the received message has errors which go undetected?

b) For general n and p, write down an expression (as a sum) for the probability that the received message has errors which go undetected.

c) Give a simplified expression, not involving a sum of a large number of terms, for the probability that the received message has errors which go undetected.

Problem 4

(A/B Testing *4'+4'+3'***)**

A/B testing is a form of randomized experiment that is used by many companies to learn about how customers will react to different treatments. For example, a company may want to see how users will respond to a new feature on their website (compared with how users respond to the current version of the website) or compare two different advertisements. As the name suggests, two different treatments, Treatment A and Treatment B, are being studied. Users arrive one by one, and upon arrival are randomly assigned to one of the two treatments. The trial for each user is classified as "success" (e.g., the user made a purchase) or "failure".

The probability that the n-th user receives Treatment A is allowed to depend on the outcomes for the previous users. This set-up is known as a two-armed bandit. Many algorithms for how to randomize the treatment assignments have been studied. Here is an especially simple (but fickle) algorithm, called a "stay-with-a-winner" procedure:

- (i) Randomly assign the first user to Treatment A or Treatment B, with equal probabilities.
- (ii) If the trial for the n-th user is a success, stay with the same treatment for the $(n+1)$-st user; otherwise, switch to the other treatment for the $(n+1)$-st user.

Let a be the probability of success for Treatment A, and b be the probability of success for Treatment B. Assume that $a \neq b$, but that a and b are unknown (which is why the test is needed). Let p_n be the probability of success on the n-th trial and a_n be the probability that Treatment A is assigned on the n-th trial (using the above algorithm).

a) Find p_n and a_{n+1} in terms of a_n, where the latter is the recurrence relation for a_n.

b) Use the results in a) to find the recurrence relation for p_n.

c) Find $\lim_{n\to+\infty} p_n$.

Problem 5

(Bonus: Dependence Graph and Mutual Independence *10'***)**

Notice: We won't give partial credit for this problem.
Given these two statements:

- An event E_{n+1} is said **mutually independent** to the set of events $\{E_1, \ldots, E_n\}$ if for any subset $I \subseteq [1, n]$, we have

$$P\left(E_{n+1} \middle| \bigcap_{j \in I} E_j\right) = P(E_{n+1}).$$

- A **dependence graph** for the set of events $\{E_1, \ldots, E_n\}$ is a graph $G = (V, E)$ such that $V = \{1, \ldots, n\}$, and for $i = 1, \ldots, n$, event E_i is **mutually independent** to the events $\{E_j | (i, j) \notin E\}$

Assume there exist real numbers $x_1, \ldots, x_n \in [0, 1]$ such that, for any $i (1 \leq i \leq n)$,

$$P(E_i) \leq x_i \cdot \prod_{j:(i,j) \in E} (1 - x_j),$$

prove this following inequality:

$$P\left(\bigcap_{i=1}^{n} E_i^c\right) \geq \prod_{i=1}^{n} (1 - x_i).$$