Homework 3

Due: 16:30, 10/19, 2023 (in class)

Homework Policy: (Read before You Start to Work)

- Copying from other students' solution is not allowed. If caught, all involved students get 0 point on that particular homework. Caught twice, you will be asked to drop the course.
- Collaboration is welcome. You can work together with **at most one partner** on the homework problems which you find difficult. However, you should write down your own solution, not just copying from your partner's.
- Your partner should be the same for the entire homework.
- Put your collaborator's name beside the problems that you collaborate on.
- When citing known results from the assigned references, be as clear as possible.

1. (Binary hypothesis testing) [16]

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. Bernoulli p random variables, that is,

$$\mathsf{Pr}\{X_i = 1\} = 1 - \mathsf{Pr}\{X_i = 0\} = p.$$

Based on the observations so far, the goal is of a decision maker to determine which of the following two hypotheses is true:

$$\begin{aligned} \mathcal{H}_0 &: \quad p = p_0 \\ \mathcal{H}_1 &: \quad p = p_1 \end{aligned}$$

where $0 < p_0 < p_1 \leq 1/2$.

a) (Warm-up) Consider the problem of making the decision based on X_1.
Draw the optimal $(\pi_{1|0}, \pi_{0|1})$ trade-off curve. [4]

b) Suppose the decision maker waits until an 1 appears and makes the decision based on the whole observed sequence. Sketch the optimal $(\pi_{1|0}, \pi_{0|1})$ trade-off curve. [4]

c) Now suppose the decision maker waits until in total n 1's appear and makes the decision based on the whole observed sequence. Let $\varpi^*_{0|1}(n, \epsilon)$ denote the minimum type-II error probability subject to the constraint that the type-I error probability is not greater than ϵ, $0 < \epsilon < 1$. Does $\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\varpi^*_{0|1}(n,\epsilon)}$ exist? If so, find it. Otherwise, show that the limit does not exist. [8]

2. (Asymptotic behavior of posterior probability [12])

Consider a binary hypothesis testing problem

$$\begin{cases} \mathcal{H}_0 : \; X_i \overset{\text{i.i.d.}}{\sim} \mathsf{P}_0, \; i = 1, 2, \ldots, n \\ \mathcal{H}_1 : \; X_i \overset{\text{i.i.d.}}{\sim} \mathsf{P}_1, \; i = 1, 2, \ldots, n \end{cases}.$$

Under a Bayes setup, the unknown binary parameter Θ is assumed to be *random* and follow a prior distribution defined by the *prior probabilities*

$$\pi_0^{(0)} := \mathsf{Pr}\left\{\Theta = 0\right\}, \; \pi_1^{(0)} := \mathsf{Pr}\left\{\Theta = 1\right\}.$$

Let the *posterior probabilities* be the conditional distribution of Θ given $X^n = x^n$:

$$\pi_0^{(n)}\left(x^n\right) = \mathsf{Pr}\left\{\Theta = 0 | X^n = x^n\right\}, \; \pi_1^{(n)}\left(x^n\right) = \mathsf{Pr}\left\{\Theta = 1 | X^n = x^n\right\}.$$

a) Derive the expressions of $\pi_0^{(n)}\left(x^n\right)$ and $\pi_1^{(n)}\left(x^n\right)$ in terms of $\pi_0^{(0)}, \pi_1^{(0)}, \mathsf{P}_0, \mathsf{P}_1$. [4]

b) Consider $\pi_0^{(n)}\left(X^n\right)$ and $\pi_1^{(n)}\left(X^n\right)$ as random variables, because they are functions of the random sequence X^n. Use the Strong Law of Large Numbers to show that if $\mathcal{H}_0$ is true, then with probability 1,

$$\pi_0^{(n)}\left(X^n\right) \to 1, \; -\frac{1}{n}\log \pi_1^{(n)}\left(X^n\right) \to \mathrm{D}\left(\mathsf{P}_0 \| \mathsf{P}_1\right) \quad \text{as } n \to \infty. \qquad [8]$$

3. Minimizing information divergence) [22]

a) Let $\mathcal{P}(\mathbb{N})$ denote the collection of all probability distributions over $\mathbb{N}$ and $\mathrm{G}(p) \in \mathcal{P}(\mathbb{N})$ be a geometric distribution with parameter $p \in (0, 1)$:

$$X \sim \mathrm{G}(p) \iff \mathsf{Pr}\{X = n\} = (1-p)p^{n-1}, \; n \in \mathbb{N} = \{1, 2, \ldots\}.$$

Under the constraint that $\mathsf{P} \in \mathcal{P}(\mathbb{N})$ and $\mathsf{E}_{X\sim\mathsf{P}}[X] = \sum_{x=1}^{\infty} x\mathsf{P}(x) = \mu > 1$, find the minimum value of $\mathrm{D}\left(\mathsf{P} \| \mathrm{G}(p)\right)$ and a minimizing distribution. [12]

b) For m discrete probability distributions $\mathsf{P}_1, \mathsf{P}_2, \ldots, \mathsf{P}_m$ with the same support $\mathcal{X}$, consider the following minimization problem:

$$\min_{\mathsf{Q} \in \mathcal{P}(\mathcal{X})} \sum_{i=1}^{m} \mathrm{D}\left(\mathsf{P}_i \| \mathsf{Q}\right),$$

where $\mathcal{P}(\mathcal{X})$ denotes the collection of probability distributions over $\mathcal{X}$. Find a minimizer to the above problem. [10]