Homework 2

Due: 13:10, 10/17, 2024 (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. (Mixture of random processes) [14]

In this problem we look at different ways to generate mixtures of random processes, and the entropy rate of the mixture of random processes. Consider two stationary random processes $\{X_0[i] \,|\, i \in \mathbb{N}\}$ and $\{X_1[i] \,|\, i \in \mathbb{N}\}$ taking values in disjoint alphabets $\mathcal{X}_0$ and $\mathcal{X}_1$ respectively. The two processes are independent from each other, that is, $\{X_0[i]\} \perp\!\!\!\perp \{X_1[i]\}$, and they have entropy rates $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively. Let $\{\Theta_i \,|\, i \in \mathbb{N}\}$ be a **stationary** Bernoulli random process, independent of everything else.

a) Let $\Theta_i = \Theta$ for all $i \in \mathbb{N}$, where $\Theta \sim \mathrm{Ber}(q)$. Is the random process $\{X_{\Theta_i}[i]\}$ stationary? What is its entropy rate? [6]

b) Let $\{\Theta_i\}$ be Markov with a probability transition matrix

$$\mathsf{P}_{\Theta_2|\Theta_1} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}, \text{ for } \alpha, \beta \in (0,1).$$

Suppose that both $\{X_0[i]\}$ and $\{X_1[i]\}$ are i.i.d. processes in this problem. Is the random process $\{X_{\Theta_i}[i]\}$ stationary? What is its entropy rate? [8]

2. (Binary hypothesis testing) [16]

Let $X_1, X_2, ...$ 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:

$$\mathcal{H}_0 : \quad p = p_0$$
$$\mathcal{H}_1 : \quad p = p_1$$

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]

3. (Mixture of information divergences) [8]

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:

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

where $\mathcal{P}(\mathcal{X})$ denotes the collection of probability distributions over $\mathcal{X}$, $\sum_{i=1}^{m} \lambda_i = 1$, and $\lambda_i > 0$ for $i = 1, 2, \ldots, m$. Show that $\sum_{i=1}^{m} \lambda_i P_i$ is a minimizer to the above problem.

4. (Rényi's divergence) [12]

Alfréd Rényi introduced the following generalization of information divergence called *Rényi's divergence of order* α (for simplicity, only deal with the discrete case):

$$\mathrm{D}_\alpha(\mathsf{P}\|\mathsf{Q}) := \frac{1}{\alpha - 1} \log \left(\sum_{a \in \mathcal{X}} \mathsf{P}(a)^\alpha \mathsf{Q}(a)^{1-\alpha} \right), \quad \alpha \in (0,1) \cup (1,\infty),$$

where P, Q are both probability distributions over a finite alphabet $\mathcal{X}$, and $\mathsf{supp}\,\mathsf{P} \subseteq \mathsf{supp}\,\mathsf{Q}$.

a) (Non-negativity) Show that $\mathrm{D}_\alpha(\mathsf{P}\|\mathsf{Q}) \ge 0$, with equality if and only if $\mathsf{P} = \mathsf{Q}$. [4]

b) (Relation with KL divergence) Show that $\mathrm{D}_\alpha(\mathsf{P}\|\mathsf{Q}) \ge \mathrm{D}(\mathsf{P}\|\mathsf{Q})$ for $\alpha > 1$ and $\mathrm{D}_\alpha(\mathsf{P}\|\mathsf{Q}) \le \mathrm{D}(\mathsf{P}\|\mathsf{Q})$ for $\alpha < 1$. Furthermore, $\lim_{\alpha \to 1} \mathrm{D}_\alpha(\mathsf{P}\|\mathsf{Q}) = \mathrm{D}(\mathsf{P}\|\mathsf{Q})$. [4]

c) (Data processing) Show that $\mathrm{D}_\alpha(\mathsf{P}\|\mathsf{Q})$ satisfies the data processing inequality. [4]