Homework 5

Due: 16:30, 11/23, 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. (Information divergence) [10]

Consider two probability density functions $\mathsf{f}(\cdot)$ and $\mathsf{g}(\cdot)$. Let μ_1 and μ_2 denote the mean of f and g respectively. Let σ_1^2 and σ_2^2 denote the variance of f and g respectively.

a) Compute $\mathrm{D}(\mathsf{f}\|\mathsf{g})$ in the following cases: (1) both f and g are Gaussian; (2) both f and g are Laplace. [6]

b) If $\mu_1 = \mu_2$, which of the above cases gives the largest/smallest KL divergence? Your answer may depend on σ_1, σ_2. [2]

c) If $\sigma_1 = \sigma_2$, which of the above cases gives the largest/smallest KL divergence? Your answer may depend on σ_1, σ_2. [2]

2. (Differential entropy) [10]

a) Consider a Laplace random variable $X \sim \mathsf{Lap}(\mu, b)$, that is, the probability density function of X is $\mathsf{f}_X(x) = \frac{1}{2b}e^{-\frac{|x-\mu|}{b}}$, $x \in \mathbb{R}$. Compute its differential entropy $\mathrm{h}(X)$. [4]

b) Consider a problem of maximizing differential entropy $\mathrm{h}(X)$ subject to the constraint that $\mathsf{E}[|X|] \le \mathsf{B}$. Find the maximum differential entropy and show that a zero-mean Laplace distributed X attains the maximum value. [6]

3. (Channel Coding with Input-Output Cost Constraint) [10]

In this problem we explore channel coding with input and output cost constraint.

a) Consider a DMC $\left(\mathcal{X}, \mathsf{P}_{Y|X}, \mathcal{Y}\right)$. Let $b: \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ be an input-output cost function. Suppose the channel coding has to satisfy the following average cost constraint: for each codeword x^n,

$$\tfrac{1}{n} \textstyle\sum_{i=1}^{n} \mathsf{E}_{Y_i}[b(x_i, Y_i)] \le \mathsf{B}.$$

Note that Y_i follows distribution $\mathsf{P}_{Y|X}(\cdot|x_i)$.

Argue that the problem is equivalent to another channel coding problem with a properly defined input-only cost function. Show that the capacity-cost function is

$$\mathrm{C}(\mathsf{B}) = \max_{\mathsf{P}_X : \mathsf{E}_{\mathsf{P}_X \mathsf{P}_{Y|X}}[b(X,Y)] \le \mathsf{B}} \mathrm{I}(X;Y)\,. \qquad [6]$$

Hint: Consider the input-only cost function $\tilde{b}(x) := \mathsf{E}[b(x, Y)]$, and check that the steps in the proof of DMC with input cost in the lecture are still valid.

b) Using discretization techniques, the above DMC result can be extended to continuous memoryless channels. With the extension (no need to prove it here), let us consider an AWGN channel with *average output power constraint*

$$\tfrac{1}{n} \textstyle\sum_{i=1}^{n} \mathsf{E}\left[Y_i^2\right] \le \mathsf{B}.$$

where $Y = X + Z$, $Z \perp\!\!\!\perp X$, and $Z \sim \mathrm{N}(0, \sigma^2)$.

Evaluate the channel capacity $\mathrm{C}(\mathsf{B})$. [4]

4. (Compression with guarantee on the cross-entropy loss) [20]

Consider a discrete memoryless source $S \sim \pi$ with a finite alphabet $\mathcal{S} = \{1, 2, ..., k\}$, $|\mathcal{S}| = k < \infty$. The encoder aims to compress the source so that the decoder can give good estimates of the source sequence. In many applications, however, the decoder may not want to give a deterministic estimate. Instead, for each symbol s_i in a length-n sequence s^n, its goal is to produce a *probability vector* $\boldsymbol{q}_i$ in the k-dimensional probability simplex $\mathcal{P}_k$, where the l-th coordinate, $q_i(l)$, stands for the probability of $s_i = l$ that the decoder believes in based on what it receives from the encoder. A standard way to quantify the loss is the empirical cross entropy loss

$$\ell_{\mathrm{CE}}(s^n, \boldsymbol{q}^n) = \sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{q_i(s)}.$$

Note that it can be viewed as the average distortion per symbol when the distortion function is set to be

$$d : \mathcal{S} \times \mathcal{P}_d \to [0, \infty),\ (s, \boldsymbol{q}) \mapsto d(s, \boldsymbol{q}) = \log \frac{1}{q(s)}.$$

Hence, one can study a lossy source coding problem to understand how to represent a memoryless source with the smallest rate so that the decoder can declare an estimation probability vector with the empirical cross entropy loss not greater than a prescribed level D. By the lossy source coding theorem, the rate is given by the following rate distortion function:

$$\mathrm{R}(\mathsf{D}) = \inf_{(S,\boldsymbol{Q})} \left\{ \mathrm{I}(S; \boldsymbol{Q}) \,\middle|\, \mathsf{E}\left[\log \frac{1}{Q(S)}\right] \le \mathsf{D} \text{ and } S \sim \pi \right\}$$

a) Show that for the lossy source coding problem, $\mathsf{D}_{\min} = 0$ and $\mathsf{D}_{\max} = \mathrm{H}(\pi)$.

b) Show that for any jointly distributed $(S, \boldsymbol{Q}) \sim P$,

$$\mathrm{H}(S|\boldsymbol{Q}) \le \mathsf{E}_{(S,\boldsymbol{Q})\sim P}\left[\log \frac{1}{Q(S)}\right].$$

Then, argue that $\mathrm{R}(\mathsf{D}) \ge \mathrm{H}(\pi) - \mathsf{D}$, for $0 \le \mathsf{D} \le \mathrm{H}(\pi)$.

c) Show that for $0 \le \mathsf{D} \le \mathrm{H}(\pi)$,

$$\mathrm{R}(\mathsf{D}) \le \min_{(S,\hat{S}),\ \hat{S}\in\mathcal{S}} \left\{ \mathrm{I}\left(S;\hat{S}\right) \,\middle|\, \mathrm{H}\left(S\middle|\hat{S}\right) \le \mathsf{D} \text{ and } S \sim \pi \right\}.$$

d) Show that for $0 \le \mathsf{D} \le \mathrm{H}(\pi)$,

$$\mathrm{R}(\mathsf{D}) = \min_{(S,\hat{S})} \left\{ \mathrm{I}\left(S;\hat{S}\right) \,\middle|\, \mathrm{H}\left(S\middle|\hat{S}\right) \le \mathsf{D} \text{ and } S \sim \pi \right\} = \mathrm{H}(\pi) - \mathsf{D}.$$

Hence, $\mathrm{R}(\mathsf{D}) = \max\{0, \mathrm{H}(\pi) - \mathsf{D}\}$.