Exam

Time: 12:50 – 15:10

November 24, 2022

Name: ______________ Student ID: ______________

Policy: (Read before You Start to Work)

- The exam is **closed book**. However, you are allowed to bring **four A4-size cheat sheets (single-sheet, two-sided)**.
- If you access to any other materials such as books, computing devices, internet connected devices, etc., it is regarded as cheating, and the exam will not be graded. Moreover, we will file the case to the University Office.
- No discussion is allowed during the exam. Everyone has to work on his/her own.
- Please turn in this copy (exam sheets) when you submit your solution sheets.
- Please follow the seat assignment when you are seated.
- Only those written on the solution sheets will be graded. Those written on the exam sheets will not be graded.
- You can use Mandarin or English to write your solutions.

Note: (Read before You Start to Work)

- Part of the points will be given even if you cannot solve the problem completely. Write down your derivation and partial solutions in a clear and systematic way.
- You can make any additional reasonable assumptions that you think are necessary in answering the questions. Write down your assumptions clearly.
- You should express your answers as explicit and analytic as possible.
- You can reuse any known results from our lectures (**restricted to materials from the lecture slides L1–L6**) and homework problems (**HW1–HW4**) without reproving them. Other than those, you need to provide rigorous arguments, unless the problem mentions specifically.

Total Points: 100. Good luck!

1. (True or False) [36]

A puzzled student makes the following claims. For each claim, either prove it or disprove it.

a) For any continuous random variable X and a function $g(\cdot) : \mathbb{R} \to \mathbb{R}$ such that both differential entropies $\mathrm{h}(X)$ and $\mathrm{h}(g(X))$ exist, it is always true that

$$\mathrm{h}(X) \geq \mathrm{h}(g(X)). \qquad [8]$$

b) For any jointly distributed random variables $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ and a function $g(\cdot)$ on $\mathcal{X}$, it is always true that

$$\mathrm{I}(X; Y | g(X)) \leq \mathrm{I}(X; Y). \qquad [8]$$

c) Consider two data processing systems $\mathsf{W}^{(1)}_{Y|X}$ and $\mathsf{W}^{(2)}_{Y|X}$. For the first one, the input $X \sim \mathsf{P}_X$ and the output $Y \sim \mathsf{P}_Y$. For the second one, $X \sim \mathsf{Q}_X$ and the output $Y \sim \mathsf{Q}_Y$. While the two data processing system may not be identical, data processing cannot increase information, and hence it is always true that

$$\mathrm{D}(\mathsf{P}_X \| \mathsf{Q}_X) \geq \mathrm{D}(\mathsf{P}_Y \| \mathsf{Q}_Y). \qquad [8]$$

d) For a discrete memoryless source $S \sim \mathsf{P}_S$, at any length $n \in \mathbb{N}$ and $\delta > 0$, the sequence with the highest probability generated from the DMS is δ-weakly typical. [6]

e) For a lossy source coding problem of a discrete memoryless source $S \sim \mathsf{P}_S$, suppose that zero-distortion can be attained, that is, $\mathsf{D}_{\min} := \min_{\hat{s}(s)} \mathsf{E}[d(S, \hat{s}(S))] = 0$. Then, it is always true that $\mathrm{R}(0) = \mathrm{H}(S)$. [6]

2. (Hypothesis Testing with Rejection) [8]

In hypothesis testing, sometimes all the hypotheses are not convincing enough. In such situations, maybe it is better to reject all of them, that is, saying "I don't know." We consider the simple case with two hypotheses: for $\theta = 0, 1$,

$$\mathcal{H}_\theta : X_i \overset{\text{i.i.d.}}{\sim} P_\theta,\ i = 1, 2, ..., n.$$

A deterministic decision making algorithm is given by

$$\phi : \mathcal{X}^n \to \{\mathrm{r}, 0, 1\},$$

where "r" denotes the rejection option.

The performance metrics include:

- Total probability of rejection:

$$\pi_{\mathrm{r}}^{(n)}(\phi) := P_0^{\otimes n}\{\phi(X^n) = \mathrm{r}\} + P_1^{\otimes n}\{\phi(X^n) = \mathrm{r}\}.$$

- Total probability of error:

$$\pi_{\mathrm{e}}^{(n)}(\phi) := P_0^{\otimes n}\{\phi(X^n) = 1\} + P_1^{\otimes n}\{\phi(X^n) = 0\}.$$

Find

$$\min_{\phi:\mathcal{X}^n \to \{\mathrm{r},0,1\}} \left\{\pi_{\mathrm{r}}^{(n)}(\phi) + \pi_{\mathrm{e}}^{(n)}(\phi)\right\}.$$

Express your answer in terms of the total variation distance $\mathrm{TV}(P_0, P_1)$.

3. (Extremal information measures) [16]

a) Let X be a continuous random variable with a probability density function and $\mathrm{h}(X)$ exists. Under the constraint that $\mathsf{E}[X] = \mu$ and $\mathsf{E}[X^2] = r^2$ where $r^2 > \mu^2$, find the maximum value of $\mathrm{h}(X)$ and a maximizing distribution. [6]

b) 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 $P \in \mathcal{P}(\mathbb{N})$ and $\mathsf{E}_{X\sim P}[X] = \sum_{x=1}^{\infty} xP(x) = \mu > 1$, find the minimum value of $\mathrm{D}(P\|\mathrm{G}(p))$ and a minimizing distribution. [10]

4. (Sum Channel) [16]

Consider l DMC's

$$\left\{(\mathcal{X}^{(i)}, \mathsf{P}^{(i)}_{Y|X}, \mathcal{Y}^{(i)}) \,\middle|\, i = 1, 2, \ldots, l\right\},$$

where DMC $(\mathcal{X}^{(i)}, \mathsf{P}^{(i)}_{Y|X}, \mathcal{Y}^{(i)})$ has channel capacity $\mathrm{C}^{(i)}$, for $1 \le i \le l$. The channel input alphabets are disjoint, and so are the channel output alphabets, that is,

$$\mathcal{X}^{(i)} \cap \mathcal{X}^{(j)} = \mathcal{Y}^{(i)} \cap \mathcal{Y}^{(j)} = \emptyset, \ \forall\, i \neq j.$$

The **sum channel** $(\mathcal{X}^{\oplus}, \mathsf{P}^{\oplus}_{Y|X}, \mathcal{Y}^{\oplus})$ associated to these channels is defined as follows:

- Input alphabet is the union $\mathcal{X}^{\oplus} := \cup_{i=1}^{l} \mathcal{X}^{(i)}$ of the individual input alphabets.
- Output alphabet is the union $\mathcal{Y}^{\oplus} := \cup_{i=1}^{l} \mathcal{Y}^{(i)}$ of the respective output alphabets.
- At each time slot the transmitter chooses to use *one and only one* of the l channels to transmit a symbol, that is,

$$\mathsf{P}^{\oplus}(y|x) := \begin{cases} \mathsf{P}^{(i)}_{Y|X}(y|x), & \text{if } x \in \mathcal{X}^{(i)} \text{ and } y \in \mathcal{Y}^{(i)} \\ 0, & \text{otherwise} \end{cases}$$

a) Introduce a random variable I indicating which DMC is used in the sum channel, that is,

$$I = i \quad \text{if } X \in \mathcal{X}^{(i)},\ i = 1, 2, ..., l.$$

Show that for the sum channel $\mathsf{P}^{\oplus}_{Y|X}$, $\mathrm{I}(X;Y) = \mathrm{I}(X;Y|I) + \mathrm{H}(I)$. [4]

b) Find the capacity of the sum channel in terms of $\{\mathrm{C}^{(i)} \mid i = 1, 2, \ldots, l\}$. [6]

c) Find the optimal input probability distribution for the sum channel in terms of the optimal input probability distributions for the individual channels. [6]

5. (Compressing a Uniform Source) [24]

A DMS S is uniformly distributed over a finite set $\mathcal{S} = \{1, 2, \ldots, 2m\}$. Consider the following source coding problems and derive the optimal compression ratios (R^* for the lossless case and $\mathrm{R}(\mathsf{D})$ for the lossy case).

a) A lossless source coding problem. [6]

b) A lossy source coding problem with a reconstruction alphabet $\hat{\mathcal{S}} = \mathcal{S}$ and the Hamming distortion measure. [10]

c) A lossy source coding problem with a reconstruction alphabet $\hat{\mathcal{S}} = \mathcal{S}$ and the following distortion measure:

$$d(s, \hat{s}) = \mathbb{1}\{s \text{ and } \hat{s} \text{ have the same parity}\}.$$

Recall that the parity is the property of an integer of whether it is even or odd. [8]