Exam

Time: 13:20 – 16:20

November 30, 2023

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 L0–L6**) and homework problems (**HW1–HW5**) without reproving them. Other than those, you need to provide rigorous arguments, unless the problem mentions specifically.

Total Points: 100. Good luck!

1. (Information measures) [28]

For two real-valued functions over the real line

$$f : \mathbb{R} \to \mathbb{R} \quad \text{and} \quad g : \mathbb{R} \to \mathbb{R},$$

recall that the *convolution* of f and g is

$$(f * g)(x) := \int_{-\infty}^{\infty} f(t)g(x-t)\ \mathrm{d}t.$$

a) (Warm up) Let f and g be two probability density functions. Moreover, let X and Y be two random variables with f and g being their PDFs respectively. Is it always true that $f * g$ is also a probability density function? If your answer is no, give a counter example. If your answer is yes, specify a joint distribution of (X, Y) so that there is a random variable Z that is a function of (X, Y) and has $f * g$ as its PDF. [4]

b) Let f and g be two probability density functions. Suppose f, g, and $f * g$ all have finite differential entropies. Consider the following chain of inequalities:

$$\max\left\{\mathrm{h}(f)\,, \mathrm{h}(g)\right\} \overset{(1)}{\leq} \mathrm{h}(f * g) \overset{(2)}{\leq} \mathrm{h}(f) + \mathrm{h}(g)\,.$$

Does (1) always hold? If so, prove it. If not, give a counter example. [8]

Does (2) always hold? If so, prove it. If not, give a counter example. [8]

c) Let f, g, and h be three probability density functions. Show that

$$\mathrm{D}(f \| g) \geq \mathrm{D}(f * h \| g * h)\,. \qquad [8]$$

2. (Joint source channel coding) [24]

Consider a two-terminal system consisting of a source terminal `Tx` and destination terminal `Rx`. The terminal `Tx` observes a source $\{S_i \mid i = 1, 2, \ldots, \}$ and is able to communicate to `Rx` via a memoryless channel, with channel input X and output Y. Coded information can hence be delivered over the channel so that `Rx` is able to reconstruct S^n either losslessly or within certain distortion on the average.

a) Suppose the source is a stationary and ergodic Bernoulli-q source, that is, $\mathsf{Pr}\{S_i = 1\} = q$. The memoryless channel is a binary symmetric channel with bit-flip probability p.

- Derive the maximum entropy rate over all possible stationary and ergodic Bernoulli-q sources. Which stationary and ergodic source attains the maximum? [8]
- Suppose $\{S_i\}$ is a Markov process satisfying $S^{i-1} - S_i - S_{i+1}$ for all $i = 1, 2, \ldots$ and $\mathsf{P}_{S_2|S_1}(1|0) = \alpha$. What is the maximum number of source symbols per channel use that can be losslessly reconstructed at the destination? [8]

b) Suppose the source is memoryless and comprises i.i.d. Gaussian random variables $\mathrm{N}(0, \sigma_S^2)$. The memoryless channel is an additive white Gaussian noise channel with an input power constraint B and noise variance σ_Z^2. The terminal `Rx` aims to reconstruct S^n at a rate R source symbols per channel use, so that that the average *squared-error* distortion is at most D as $n \to \infty$. What is the smallest distortion D that can be achieved? Express your answer in terms of $\sigma_S^2, \mathsf{B}, \sigma_Z^2, \mathrm{R}$. [8]

3. (Capacity of the permutation channel) [28]

A channel model in neural communication is the following:

- Input/ouput alphabet: $\mathcal{X} = \mathcal{Y} = \{0,1\}^d$
- Channel law:

$$\mathsf{P}_{Y|X}(\boldsymbol{y}|\boldsymbol{x}) = \begin{cases} 1/\binom{d}{\|\boldsymbol{x}\|_1}, & \text{if } \|\boldsymbol{y}\|_1 = \|\boldsymbol{x}\|_1 \\ 0, & \text{otherwise} \end{cases}$$

  Note that for a d-dimensional *binary* vector $\boldsymbol{x}$, its ℓ_1-norm is the number of 1's in $\boldsymbol{x}$:

$$\|\boldsymbol{x}\|_1 = \textstyle\sum_{i=1}^d \mathbb{1}\{x_i = 1\}.$$

In words, the channel permutes the length-d binary vector uniformly at random. In this problem, let us compute the capacity of this channel, namely, find

$$\mathrm{C} = \max_{\mathsf{P}_{\boldsymbol{X}}} \mathrm{I}(\boldsymbol{X};\boldsymbol{Y}).$$

a) Let $L := \|\boldsymbol{X}\|_1$. Show that

$$\mathrm{C} = \max_{\mathsf{P}_L} \left\{ \mathrm{H}(L) + \max_{\mathsf{P}_{\boldsymbol{X}|L}} \mathrm{I}(\boldsymbol{X};\boldsymbol{Y}|L) \right\}$$

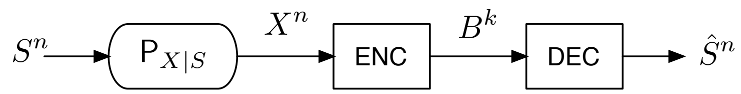
and compute the channel capacity C accordingly. What is the capacity achieving input distribution? [12]

b) Suppose now the channel coding problem is associated with an average input cost constraint as in our lecture. Let the input cost function be $b(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$. Compute the capacity-cost function $\mathrm{C}(\mathsf{B})$ of this channel. What is the capacity achieving input distribution? [8]

c) Let α be a constant between 0 and 1, that is, $0 < \alpha < 1$. In continuation of Part b), now suppose the channel delivers $\boldsymbol{x}$ noiselessly with probability $(1-\alpha)$, and permutes $\boldsymbol{x}$ uniformly at random with probability α (note: keeping $\boldsymbol{x}$ the same is also one possible permutation.). Compute the channel capacity C of this channel. What is the capacity achieving input distribution? [8]

4. (Compression of a remote source) [20]

In standard source coding problems, the encoder is able to observe the source directly. However, in many practical scenarios, the source is *remote* and corrupted by some noise when the encoder observes it. In this problem, we aim to establish the lossy source coding theorem for such *remote* sources.



Consider a DMS $S_i \overset{\text{i.i.d.}}{\sim} \mathsf{P}_S$ for all i, and $\{X_i\}$ be the *corrupted* DMS obtained by passing S_i through a DMC $\mathsf{P}_{X|S}$. The source encoder encodes the noisy observation X^n into bit sequences, and the source decoder would like to produce a reconstruction $\hat{S}^n$ so that the distortion, measured by $d(s, \hat{s})$, is not greater than D.

a) Define
$$\tilde{d}(x, \hat{s}) := \mathsf{E}_{S \sim \mathsf{P}_{S|X}(\cdot|x)} \left[d(S, \hat{s}) \,|\, X = x \right]$$
and consider an alternative lossy source coding problem with X being the source to be reconstructed (instead of S) and $\tilde{d}(\cdot, \cdot)$ being the distortion function. Argue that this lossy source coding problem and the remote lossy source coding problem are **equivalent** in the sense that any (R, D) pair that is achievable in one problem is also achievable in the other and vice versa. [8]

b) Based on Part a), invoke the lossy source coding theorem in our lecture to characterize the rate distortion function of the remote lossy source coding problem. Your characterization should be expressed as a minimization problem of some mutual information term involving X and $\hat{S}$ over $\mathsf{P}_{\hat{S}|X}$ subject to an expected distortion constraint on $d(S, \hat{S})$. [8]

Specify $\mathsf{D}_{\min}$ and $\mathsf{D}_{\max}$. [4]