B3 - Information Theory : Mid-Semester Exam

Yogeshwaran D.

February 28, 2026. **Time : 2.00 - 4.30 PM.** **Max. points : 30.**

There are two parts to the question paper - PART A and PART B. Read the instructions for each section carefully.

PART A : MULTIPLE-CHOICE QUESTIONS - 10 Points.

Please write only the correct choice(s) (for ex., (A), (B) et al.) in your answer scripts. No explanations are needed. Write PART A answers in a separate page.

Some questions will have multiple correct choices. Answer all questions. Each question carries 2 points. 1 point will be awarded if only some correct choices are chosen and no wrong choices are chosen.

1. Let X, Y be independent random variables. Which of the following are always true ?

   (A) $H(X^3, Y^2) = H(X) + H(Y)$.

   (B) $I(X^{28}; Y^2) = 0$.

   (C) $H(Y^2 \mid X) = H(Y)$.

   (D) $H(e^X, e^{-Y}) = H(X) + H(Y)$.

2. Let X, Y be random variables. Which of the following are always true ?

   (A) $H(X, Y) = H(Y) + H(X \mid e^Y)$.

   (B) $I(X; e^{-Y}) = I(X; Y)$.

   (C) $I(X; Y \mid Y^2) = I(X; Y^2 \mid Y)$

   (D) $H(X, Y \mid Y^3) = H(X, Y^3 \mid Y)$

3. Consider the binary code for the source alphabet $\{a, b, c\}$: $a \to 0$, $b \to 01$, $c \to 11$. Which of the following statements are true?

   (A) The code is prefix.

   (B) The code is uniquely decodable.

   (C) The Kraft inequality holds.

   (D) The code is optimal for some pmf.

4. Which of the following codes cannot be Huffman codes for any pmf ?

   (A) $\{0, 10, 11\}$

(B) $\{00, 01, 10, 110\}$

(C) $\{01, 10, 110\}$

(D) $\{00, 01, 10, 110, 111\}$

5. Let $X_n, n \geq 0$ be a stationary process. Which of the following is true ?

   (A) $H(X_0 \mid X_n) \geq H(X_n \mid X_{2n})$ for all $n \geq 1$.

   (B) $I(X_0; X_n) \leq I(X_0; X_{n-1})$ for all $n \geq 1$.

   (C) $H(X_n) > H(X_{n+1})$ for some $n \geq 1$.

   (D) If X_n is Markovian then $H(X_n \mid X_0) \geq H(X_{n-1} \mid X_0)$ for all $n \geq 1$.

PART B : 20 Points.

Answer any two questions only. All questions carry 10 points.
Give necessary justifications and explanations for all your arguments. If you are citing results from the class, mention it clearly.

1. Write down the Huffman code and Shannon-Fano-Elias code for the following pmfs - $(0.3, 0.2, 0.15, 0.15, 0.1, 0.05, 0.05)$ and $(\frac{1}{32}, \frac{1}{16}, \frac{1}{64}, \frac{1}{128}, \frac{1}{128}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2})$. Compute the entropy and average code lengths for both the codes as well.

2. Let $\{X_{11}, X_{12}, X_{13}, \dots\}$ be i.i.d. Bernoulli(p_1) random variables and let $\{X_{21}, X_{22}, X_{23}, \dots\}$ be i.i.d. Bernoulli(p_2) random variables. Let θ be a uniform random variable in $\{1, 2\}$ and let $Y_i = X_{\theta i}$, $i = 1, 2, \dots$, be the observed stochastic process. Thus Y observes the process $\{X_{1i}\}$ or $\{X_{2i}\}$. Define $Z_i = (Y_i, Y_{i+1}), i \geq 1$ to be the process of transitions. Define l_n and l'_n be the optimal coding lengths for Y^n and Z^n respectively.

   Compute the entropy of the processes $H((Y_i)_{i\geq 1})$ and $H((Z_i)_{i\geq 1})$ as well as $\lim_{n\to\infty} n^{-1} l_n$ and $\lim_{n\to\infty} n^{-1} l'_n$.

3. Below we give two examples of channels $(\mathcal{X}, p(y \mid x), \mathcal{X})$ and source message from $\mathcal{W}$. Find the rate of transmission and channel capacity in each of the example. Also in both examples find encoding scheme $\mathcal{W} \to \mathcal{X}$ and decoding scheme $\mathcal{X} \to \mathcal{W}$ such that the probability of error for a uniformly trasmitted word W from $\mathcal{W}$ is at most 0.2 i.e., $\mathbb{P}(\hat{W} \neq W) \leq 0.2$.

   (a) Consider Binary Symmetric Channel on $\mathcal{X} = \{0, 1\}^3$ with crossover (error) probability $\alpha = 0.1$ for each bit i.e., each bit is independently flipped with probability α. Let $\mathcal{W} = [4]$ i.e., W is a uniformly chosen message from $[4]$. **(5)**

   (b) Consider Binary Symmetric Channel on $\mathcal{X} = \{0, 1\}^3$ with erasure probability $\alpha = 0.1$ for each bit i.e., each bit is independently erased with probability α. Let $\mathcal{W} = [8]$ i.e., W is a uniformly chosen message from $[8]$. **(5)**

COURSE RECAP

Basic Definitions

$H(X) = -\sum_x p(x)\log p(x)$. $H(X,Y) = -\sum_{x,y} p(x,y)\log p(x,y)$, $\quad H(X|Y) := H(X,Y) - H(Y)$.

The relative entropy (KL divergence) is $D(p\|q) = \sum_x p(x)\log\frac{p(x)}{q(x)}$.

Mutual information: $I(X;Y) = H(X) - H(X|Y) = D(p_{XY}\|p_X p_Y)$.

Asymptotic Equipartition Property (AEP) and Entropy Rate

Let $X_1, \dots, X_n$ be i.i.d. with pmf $p(x)$. **AEP.** $-\frac{1}{n}\log p(X^n) \xrightarrow{a.s.} H(X)$.

Typical set. For $\epsilon > 0$, $A_\epsilon^{(n)} = \left\{x^n : \left|-\frac{1}{n}\log p(x^n) - H(X)\right| \le \epsilon\right\}$.

- $\Pr\{X^n \in A_\epsilon^{(n)}\} \to 1$; $|A_\epsilon^{(n)}| \le 2^{n(H+\epsilon)}$.
- **Source Coding theorem:** For every $\epsilon > 0$, there exists encoding and decoding $\mathcal{X}^n \to \{0,1\}^{n(H+\epsilon)} \to \mathcal{X}^n$ such the error probability (i.e., $\mathbb{P}(X^n \ne \hat{X}^n)$) vanishes asymptotically.

Entropy rate of a stationary stochastic process $(X_i)_{i\ge1}$ is $H((X_i)_{i\ge1}) := \lim n^{-1}H(X^n) = \lim H(X_n \mid X^{n-1})$.

Optimal Code Length

- A source code maps symbols x to binary strings of length $l(x)$.
- Optimal code length - $L^* := \min\{\sum_x l(x)p(x) \,:\, \sum_x 2^{-l(x)} \le 1,\, l(x) \in \mathbb{N}\}$.
- $H(X) \le L^* < H(X) + 1$,
- Shannon coding: $l(x) = \lceil -\log p(x)\rceil$; Huffman coding achieves optimal prefix codes

Channel Capacity

A **discrete memoryless channel** is specified by $(\mathcal{X}, p(y \mid x), \mathcal{Y})$, where $p(\cdot \mid \cdot)$ is a transition matrix. **Capacity.** $C = \max_{p_X(x)} I(X;Y)$.

Channel Coding Theorem.

- If $R < C$, there exist codes with $\lambda^{(n)} \to 0$
- If $R > C$, then $P_e^{(n)} \not\to 0$ for any code
- Above claims hold also for Feedback capacity.