

# B3 - Information Theory : End-Semester Exam

*Yogeshwaran D.*

May 4, 2026.

Time : 2.00 - 5.00 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.

- Let  $X, Y$  be random variables. Which of the following are always true ?
  - $H(X, Y) = H(Y) + H(e^X | e^Y)$ .
  - $H(X, Y) = H(e^{-X}) + H(Y)$  if  $X, Y$  are independent.
  - $I(X; e^{-Y}) = I(X; Y)$ .
  - $I(X; Y | Y^2) = I(X; Y^2 | Y)$
- Let  $\{X_{11}, X_{12}, X_{13}, \dots\}$  be i.i.d. Bernoulli(1/4) random variables and let  $\{X_{21}, X_{22}, X_{23}, \dots\}$  be i.i.d. Bernoulli(3/4) random variables independent of previous sequence. Let  $\theta_1, \theta_2, \dots$  be i.i.d. uniform random variable in  $\{1, 2\}$  independent of the previous two sequences. Which of the following are true about the channel with above parameters ?
  - $\lim n^{-1} H(X_{\theta_1 i} : 1 \leq i \leq n) < H(1/2)$ .
  - $\lim n^{-1} H(X_{\theta_1 i} : 1 \leq i \leq n) > H(1/2)$ .
  - $\lim n^{-1} H(X_{\theta_1 i} : 1 \leq i \leq n) < \lim n^{-1} H(X_{\theta_2 i} : 1 \leq i \leq n)$ .
  - $\lim n^{-1} H(X_{\theta_1 i} : 1 \leq i \leq n) > \lim n^{-1} H(X_{\theta_2 i} : 1 \leq i \leq n)$ .
- Consider discrete memoryless channel with  $\mathcal{X} = \{-1, 0, 1\}$ ,  $\mathcal{Y} = \{0, 1\}$  and transition matrix  $P(1 | 1) = P(-1 | -1) = 1, P(1 | 0) = P(-1 | 0) = 1/2$ . Which of the following are true ?
  - The input distribution achieving the capacity has full support.
  - The input distribution achieving the capacity is supported on  $\{-1, 1\}$ .
  - Ber(1/4) input on  $\{-1, 1\}$  achieves the capacity.
  - Uniform input on  $\mathcal{X}$  achieves the capacity.

4. Let  $X$  have standard Laplace distribution ( $\frac{1}{2}e^{-|x|}$ ,  $x \in \mathbb{R}$ ) and  $Z$  be standard Normal random variable. Which of the following are true ?
- (A)  $H(|X|) < H(X)$ .  
 (B)  $H(|X|) = H(X)$ .  
 (C)  $H(X) + H(|X|) < H(Z)$ .  
 (D)  $H(X) + H(|X|) > H(Z)$ .
5. For a discrete memoryless channel, let  $R(D)$  and  $D(R)$  be respectively *rate distortion* and *distortion rate* functions. Which of the following are true ?
- (A)  $R(D)$  is non-decreasing in  $D$ .  
 (B)  $R(D)$  is convex in  $D$ .  
 (C)  $D(R)$  is non-increasing in  $R$ .  
 (D)  $D(R)$  is concave in  $R$ .

## PART B : 30 Points.

Answer any three 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. Consider a discrete memoryless channel with transition probability kernel  $P(y | x)$ . Show the following
- (a)  $I(X; Y)$  is a convex function of the transition probability kernel  $P(y | x)$ . (2)  
 (b)  $I(X; Y)$  is a concave function of the input pmf  $p_X$ . (2)  
 (c) If  $p_X^*$  is an optimal input distribution (i.e., achieves the channel capacity) then  $p_Y^*(y) = \sum_x p_X^*(x)P(y | x) > 0$  for all  $y$  such that  $P(y | x) > 0$  for some  $x$ . (3)  
 (d) For any optimal input distribution  $q_X$ , we have that  $q_Y \equiv p_Y^*$  with  $q_Y$  defined similar to  $p_Y^*$  as above. (3)
2. Find the capacity  $n^{-1} \max_{X^n} \{I(X^n; Y^n) : X_i \text{ independent}\}$  and an achieving input distribution for the following two channels.
- (a) A continuous memoryless channel where  $Y_i = X_i + Z_i$ ,  $1 \leq i \leq n$  where  $Z_i$  are i.i.d. Exponential random variables with mean  $\mu$  and we have power constraint on the mean i.e.,  $\mathbb{E}(X_i) \leq P$ . (5)  
 (b) A continuous memoryless channel  $Y_i = X_i + Z_i$ ,  $1 \leq i \leq n$  where  $\mathbb{E}(X_1^2) \leq nP$ ,  $\mathbb{E}(X_i^2) = 0$  for  $i > 1$  and  $Z_i$  are i.i.d.  $N(0, \sigma^2)$  random variables. (5)

3. Consider the following distortion measure  $d$  for  $\{0, 1\} \times \{a, b\}$ :  $d(0, a) = d(1, b) = \frac{1}{2}$  and  $d(0, b) = d(1, a) = 1$ . Let the input distribution be  $\text{Ber}(p)$  with  $p \leq \frac{1}{2}$ . Compute the rate distortion function  $R(D)$  for all  $D \geq 0$ . *Hint : Start with  $R(\frac{1+p}{2}), R(\frac{1}{2})$ .*
4. Let  $X_1, X_2, \dots$  be i.i.d. random variables drawn according to the geometric distribution

$$\Pr\{X = k\} = p^{k-1}(1 - p), \quad k = 1, 2, \dots$$

Find good estimates (to first order in the exponent) of

- (a)  $\Pr\left\{\frac{1}{n} \sum_{i=1}^n X_i \geq \epsilon\right\}$ . (4)
- (b)  $\Pr\left\{X_1 = k \mid \frac{1}{n} \sum_{i=1}^n X_i \geq \epsilon\right\}$ . (4)
- (c) Evaluate (a) and (b) for  $p = \frac{1}{2}$ ,  $\epsilon = 4$ . (2)

5. **Each of the parts carry 5 points.**

- (a) Show for a parametric family of pmfs  $\{p_\theta(x)\}$  on a finite set  $\mathcal{X}$  that

$$\lim_{\theta' \rightarrow \theta} \frac{1}{(\theta - \theta')^2} D(p_\theta \| p_{\theta'}) = \frac{1}{\ln 4} J(\theta),$$

where  $J(\theta) = \text{Var}_\theta(V(X))$  is the *Fisher information* with the score function defined as  $V(x) = \frac{\partial \log p_\theta(x)}{\partial \theta}$ .

- (b) Let  $X \sim f(x; \theta)$ , a parametric family of pdfs and let  $T(X)$  be an estimator for  $\theta$ . Let  $b_T(\theta) = \mathbb{E}_\theta[T] - \theta$  be the bias of the estimator. Show that

$$\mathbb{E}_\theta[(T - \theta)^2] \geq \frac{[1 + b'_T(\theta)]^2}{J(\theta)} + b_T(\theta)^2.$$

## COURSE RECAP

Relative entropy/ KL Divergence:  $D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$ .

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

Let  $X_1, \dots, X_n$  be i.i.d.  $\sim p$ . **Typical set.** For  $\epsilon > 0$ ,  $A_\epsilon^{(n)} = \left\{x^n : \left|-\frac{1}{n} \log p(x^n) - H(X)\right| \leq \epsilon\right\}$ .

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

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

## Optimal Code Length

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

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

**Channel Coding Theorem (with and without feedback.)** If  $R < C$ , there exist codes with  $\lambda^{(n)} \rightarrow 0$ . If  $R > C$ , then  $P_e^{(n)} \not\rightarrow 0$  for any code

**Rate Distortion theory:**  $R(D) = \min_{p(\hat{x}|x)} \{I(X; \hat{X}) : \mathbb{E}(d(X, \hat{X})) \leq D\}$ .  
 $D(R) = \min_{p(\hat{x}|x)} \{\mathbb{E}(d(X, \hat{X})) : I(X; \hat{X}) \leq R\}$

For  $R > R(D)$ , there exist  $(2^{nR}, n)$ -code achieving distortion  $D$  and conversely, if there exist  $(2^{nR}, n)$ -code achieving distortion  $D$  then  $R \geq R(D)$ .

For Binary symmetric channel with input  $\text{Ber}(p)$ ,  $R(D) = H(\text{Ber}(p)) - H(\text{Ber}(D))$  for  $D \leq \min\{p, 1-p\}$  and else 0.

**Sanov's theorem:** Let  $X^n$  be i.i.d  $\sim Q$  on  $\mathcal{X}$ . If  $\bar{E} = E \in \mathcal{P}(\mathcal{X})$  then  $n^{-1} \log Q^n(E) \rightarrow -D(P^* || Q)$  where  $P^* = \text{argmin}_{P \in \mathcal{E}} D(P || Q)$  and  $Q^n(\cdot) = \mathbb{P}(P_{X^n} \in \cdot)$ , a probability distribution on  $\mathcal{P}(\mathcal{X})$ .