

## Unit 9 - Week 8: Information theoretic lower bounds and Data Compression-1

Course outline
How does an NPTEL online course work?
Week 1: Information and probabilistic modeling
Week 2: Uncertainty, compression, and entropy
Week 3: Randomness and entropy
Week 4: Information and statistical inference-1
Week 5: Information and statistical inference-2
Week 6: Properties of measures of Information-1
Week 7: Properties of measures of Information-2 and Information Theoretic lower bounds
Week 8: Information theoretic lower bounds and Data Compression-1
<input type="radio"/> Information Theory Review 7 <input type="radio"/> Lower bound for random number generation <input type="radio"/> Strong converse <input checked="" type="radio"/> Lower bound for minmax statistical estimation <input type="radio"/> Variable length source codes <input type="radio"/> Quiz : Weekly Assignment 8 <input checked="" type="radio"/> Unit 8 notes <input type="radio"/> Solution 8
Week 9: Data Compression-1 (Unit 9) and Data Compression-2 (Unit 10)
Week 10: Data Compression-2 (Unit 10) and Data Compression-3 (Unit 11)
Week 11: Channel coding and capacity (Unit 12) and Shannon's channel coding theorem proof (Unit 13)
Week 12: Shannon's channel coding theorem proof (Unit 13) and Gaussian channels (Unit 14)
VIDEO DOWNLOAD
Live Session

## Weekly Assignment 8

The due date for submitting this assignment has passed.  
As per our records you have not submitted this assignment.

**Due on 2020-11-11, 23:59 IST.**

Note: All the questions below may have multiple correct answers. There is no negative marking for wrong choices. However, partial marking is considered only when none of the marked choices is wrong.

Unless stated otherwise, assume that probability distributions are defined over a finite alphabet, and  $d$  that  $\log$  is to the base 2. In addition to the lectures in Week 8, also review the lecture on 'Lower bound for hypothesis testing' from Week 7 for this assignment.

1) Let  $X$  be a random variable taking values in alphabet  $\mathcal{X}$ . Consider the binary hypothesis testing problem  $H_0: X \sim P, H_1: X \sim Q$ . Recall that  $\beta_\epsilon(P, Q)$  is defined as  $\beta_\epsilon(P, Q) = \min_{T: \mathcal{X} \rightarrow \{0,1\}} \{ \sum_{x \in \mathcal{X}} Q(x) T(0|x) : \sum_{x \in \mathcal{X}} P(x) T(1|x) \leq \epsilon \}$ . Define another quantity  $\hat{\beta}_\epsilon(P, Q)$  as follows:  $\hat{\beta}_\epsilon(P, Q) \triangleq \min_{A: \mathcal{X}} \{ Q(A) : P(A) \geq 1 - \epsilon \}$ . Which of the following are true? 1 point

- $\hat{\beta}_\epsilon(P, Q) \geq \beta_\epsilon(P, Q)$
- $\hat{\beta}_\epsilon(P, Q) \leq \beta_\epsilon(P, Q)$
- $\hat{\beta}_\epsilon(P, Q) = \beta_\epsilon(P, Q)$
- The two quantities are incomparable.

No, the answer is incorrect.

Score: 0

Accepted Answers:

$\hat{\beta}_\epsilon(P, Q) \geq \beta_\epsilon(P, Q)$

2) Let  $X$  be a random variable taking values in alphabet  $\mathcal{X}$ . Consider the binary hypothesis testing problem  $H_0: X \sim P, H_1: X \sim Q$ . Let  $T: \mathcal{X} \rightarrow \{0, 1\}$  be a test. Let  $P_{FA} = P(T(X) = 1)$  be the probability of false alarm, and let  $P_D = Q(T(X) = 1)$  be the probability of detection. Which of the following are true? (Ber( $p$ ) denotes Bernoulli distribution with parameter  $p$ .) 1 point

- $D(P||Q) \geq D(\text{Ber}(P_{FA})||\text{Ber}(P_D))$
- $D(P||Q) \geq D(\text{Ber}(P_{FA})||\text{Ber}(1 - P_D))$
- $D(P||Q) \leq D(\text{Ber}(P_{FA})||\text{Ber}(P_D))$
- $D(P||Q) \leq D(\text{Ber}(1 - P_{FA})||\text{Ber}(1 - P_D))$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$D(P||Q) \geq D(\text{Ber}(P_{FA})||\text{Ber}(P_D))$

3) Let  $X$  be a random variable taking values in alphabet  $\mathcal{X}$  with distribution  $P$ . Let  $f: \mathcal{X} \rightarrow \{0, 1\}^\ell$  and let  $Q$  be a distribution over  $\{0, 1\}^\ell$  defined as  $Q(y) = P(f(X) = y)$  for every  $y \in \{0, 1\}^\ell$ . Suppose it is given that  $Q(y) \in [(1 - \epsilon)2^{-\ell}, (1 + \epsilon)2^{-\ell}]$  for every  $y \in \{0, 1\}^\ell$ . Which of the following are true? 1 point

- $H(P) \leq H(Q)$
- $H(P) \geq H(Q)$
- $H(P) \geq \ell(1 - \epsilon) - 1$
- $H(Q) \geq \ell(1 - \epsilon) - 1$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$H(P) \geq H(Q)$

$H(P) \geq \ell(1 - \epsilon) - 1$

$H(Q) \geq \ell(1 - \epsilon) - 1$

4) Let  $X_1, \dots, X_n$  be positive-valued random variables. Which of the following are true? 1 point

- $\frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i \log X_i] \geq \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \log \mathbb{E}[X_i]$
- $\frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i \log X_i] \leq \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \log \mathbb{E}[X_i]$
- $\frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i \log X_i] \geq \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \right) \log \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \right)$
- $\frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i \log X_i] = \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \right) \log \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \right)$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i \log X_i] \geq \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \log \mathbb{E}[X_i]$

$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i \log X_i] \geq \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \right) \log \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \right)$

5) Let  $X, Y, Z$  be random variables. Let  $P_{XYZ}$  and  $Q_{XYZ}$  be two joint distributions of these random variables. Which of the following are true? 1 point

- $D(P_{XYZ}||Q_{XYZ}) = D(P_X||Q_X) + \mathbb{E}_X[D(P_{Y|X}||Q_{Y|X})] + \mathbb{E}_{X,Y}[D(P_{Z|X,Y}||Q_{Z|X,Y})]$
- $D(P_{XYZ}||Q_{XYZ}) = D(P_X||Q_X) + \mathbb{E}_Y[D(P_{X|Y}||Q_{X|Y})] + \mathbb{E}_Z[D(P_{XYZ}||Q_{XYZ})]$
- $D(P_{XYZ}||Q_{XYZ}) = D(P_X||Q_X) + \mathbb{E}_X[D(P_{YZ|X}||Q_{YZ|X})]$
- $D(P_{XYZ}||Q_{XYZ}) = D(P_Z||Q_Z) + \mathbb{E}_Z[D(P_{XYZ}||Q_{XYZ})]$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$D(P_{XYZ}||Q_{XYZ}) = D(P_X||Q_X) + \mathbb{E}_X[D(P_{Y|X}||Q_{Y|X})] + \mathbb{E}_{X,Y}[D(P_{Z|X,Y}||Q_{Z|X,Y})]$

$D(P_{XYZ}||Q_{XYZ}) = D(P_X||Q_X) + \mathbb{E}_X[D(P_{YZ|X}||Q_{YZ|X})]$

$D(P_{XYZ}||Q_{XYZ}) = D(P_Z||Q_Z) + \mathbb{E}_Z[D(P_{XYZ}||Q_{XYZ})]$

6) Suppose  $X$  is a random variable such that  $\mathbb{P}(|X| \geq 3) \geq 1/2$ . Which of the following are true? 1 point

- $\mathbb{E}[|X|] \geq 3/2$
- $\mathbb{E}[|X|] \leq 3/2$
- $\mathbb{P}(|X| \geq 1) \geq 2/3$
- $\mathbb{P}(|X| \geq 4) \geq 1/2$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$\mathbb{E}[|X|] \geq 3/2$

7) Consider the Gaussian mean estimation problem, where, given i.i.d. samples  $X_1, \dots, X_n \sim \mathcal{N}(\mu, 1)$ , the task is to estimate  $\mu \in \mathbb{R}$ . Suppose we have an estimator  $\hat{\mu}: (X_1, \dots, X_n) \rightarrow \mathbb{R}$  with the guarantee that  $\mathbb{E}_\mu[|\mu - \hat{\mu}(X_1, \dots, X_n)|] \leq 0.1$  for every  $\mu \in \mathbb{R}$ . 1 point

Suppose, now, that someone tells us that the samples we are observing are coming either from  $\mathcal{N}(0, 1)$  or from  $\mathcal{N}(1, 1)$ . Our task is to determine which of these hypotheses is true. Let  $H_0: X_1, \dots, X_n \sim \mathcal{N}(0, 1)$  and  $H_1: X_1, \dots, X_n \sim \mathcal{N}(1, 1)$ . Since we already have an estimator  $\hat{\mu}$  with us, we design a test  $T: (X_1, \dots, X_n) \rightarrow \{0, 1\}$  as follows:  $T(X_1, \dots, X_n) = 0$ , if  $|\hat{\mu}(X_1, \dots, X_n)| \leq 0.5$ ;  $T(X_1, \dots, X_n) = 1$ , if  $|\hat{\mu}(X_1, \dots, X_n)| > 0.5$ . Let the probability of error of this test be  $P_e(T) = \frac{1}{2}(\mathbb{P}(T = 1|H_0) + \mathbb{P}(T = 0|H_1))$ . Which of the following are true?

(You may use the fact that  $D(\mathcal{N}(0, 1)||\mathcal{N}(1, 1)) = \frac{1}{2 \ln 2}$ , where the  $D(\cdot||\cdot)$  is computed using  $\log_2$ . Furthermore,  $d(P, Q) \leq \sqrt{\frac{\ln 2}{2} D(P||Q)}$  for distributions  $P, Q$ .)

- $P_e(T) \geq 1/3$
- $P_e(T) \leq 1/5$
- $P_e(T) \geq \frac{1}{2} \left( 1 - \frac{\sqrt{n}}{2} \right)$
- $P_e(T) \leq \frac{1}{2} \left( 1 - \frac{\sqrt{n}}{2} \right)$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$P_e(T) \leq 1/5$

$P_e(T) \geq \frac{1}{2} \left( 1 - \frac{\sqrt{n}}{2} \right)$

8) Let  $X, Y$  be random variables with joint distribution  $P_{XY}$ , where  $X \in \mathcal{X}, Y \in \{0, 1\}^*$ , and  $X$  is distributed uniformly over  $\mathcal{X}$ . Suppose there is a function  $f: \{0, 1\}^* \rightarrow \mathcal{X}$  such that  $P_{XY}(X \neq f(Y)) \leq 1/4$ . Which of the following are true? 1 point

- $I(X \wedge Y) \leq \frac{3}{4} \log |\mathcal{X}| - 1$
- $I(X \wedge Y) \leq \frac{3}{4} \log |\mathcal{X}| - 1$
- $I(X \wedge Y) \leq H(Y)$
- $I(X \wedge Y) \leq \log |\mathcal{X}|$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$I(X \wedge Y) \leq \frac{3}{4} \log |\mathcal{X}| - 1$

$I(X \wedge Y) \leq H(Y)$

$I(X \wedge Y) \leq \log |\mathcal{X}|$

9) Let  $X, Y$  be random variables with joint distribution  $P_{XY}$ , where  $X \in \mathcal{X}, Y \in \{0, 1\}^*$ . Let  $Z = \text{len}(Y)$ , where  $\text{len}(y)$  denotes the length of the binary string  $y$ . Let  $\ell = \mathbb{E}[Z]$ . Which of the following are true? 1 point

- $H(Z|Y) = \ell$
- $H(Z|Y) = 0$
- $H(Z|Y) \leq \ell$
- $H(Y|Z) \leq \sum_{z=0}^{\infty} z P_Z(z)$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$H(Z|Y) = 0$

$H(Z|Y) \leq \ell$

$H(Y|Z) \leq \sum_{z=0}^{\infty} z P_Z(z)$

10) Let  $f(x) = x \log x$  for  $x > 0$ . Which of the following are true? (ln denotes the natural log.) 0 points

- $f''(x) \geq 0$  for every  $x > 0$
- $f''(x) \leq 0$  for every  $x > 0$
- $f(x) - f(x-1) \geq \frac{1}{\ln 2} (1 + \ln(x-1))$  for every  $x > 1$
- $f(x) - f(x-1) \leq \frac{1}{\ln 2} (1 + \ln(x-1))$  for every  $x > 1$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$f''(x) \geq 0$  for every  $x > 0$

$f(x) - f(x-1) \geq \frac{1}{\ln 2} (1 + \ln(x-1))$  for every  $x > 1$

$f(x) - f(x-1) \leq \frac{1}{\ln 2} (1 + \ln(x-1))$  for every  $x > 1$