MTHE - MATH 474/874 - Information Theory
Fall 2019

A Sample of Past Midterm Problems

1. Prove or disprove the following statements.

   (a) The set of codewords $\{10, 010, 101\}$ is uniquely decodable.

   (b) Conditioning increases entropy.

   (c) The mutual information $I(X;Y)$ is non-negative.

   (d) Suppose we wish to estimate a random variable X by observing a related random variable Y. Let $\hat{X}$ denote the estimate of X. Then we can use Fano's inequality to obtain a lower bound on the probability of estimation error $P_e \triangleq \Pr\{X \neq \hat{X}\}$.

   (e) If the codeword lengths of a variable length code satisfy Kraft's inequality, then it is uniquely decodable.

   (f) Given a source with 8 letters, there exists a set of probabilities for which the binary Huffman codewords associated with the source letters are of equal length.

   (g) For any source with 30 letters, the minimum rate required to encode reliably the source via binary fixed-length coding is at most 4 bits/source symbol.

   (h) Let X, Y and Z be three discrete random variables with alphabets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, respectively. If $|\mathcal{X}| = 8$, $|\mathcal{Y}| = 4$ and $|\mathcal{Z}| = 8$ and $X \to Y \to Z$, then $I(X;Z) \leq 2$.

   (i) Let $\{X_n\}_{n=1}^{\infty}$ be a stationary source with alphabet $\mathcal{X}$ and entropy rate $H(\mathcal{X})$. Then

   $$H(\mathcal{X}) \leq H(X_2).$$

   (j) Consider a discrete memoryless source $\{X_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{X} = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ and probability distribution

   $$[p_1, p_2, p_3, p_4, p_5, p_6] = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\right],$$

   where $p_i \triangleq Pr\{X = a_i\}$, $i = 1, \cdots, 6$. The Shannon-Fano code $f : \mathcal{X} \to \{0,1\}^*$ for the source is optimal.

(k) Let $\{X_n\}_{n=1}^{\infty}$ be a stationary Markov source with alphabet $\mathcal{X} = \{0, 1\}$ and $Pr\{X_{n+1} = 1|X_n = 0\} = Pr\{X_{n+1} = 0|X_n = 1\} = 0.5$. Then the source entropy rate satisfies

$$H(\mathcal{X}) < H(X_1).$$

2. Show that the divergence $D(p\|q)$ is convex in the distribution pair (p, q) defined on a common alphabet.

3. Answer the following questions.

(a) Let X, Y and Z be jointly distributed random variables. Show that

$$H(X, Y, Z) - H(X, Y) \leq H(X, Z) - H(X),$$

and state the necessary and sufficient condition for equality.

(b) Let $\{X_n\}_{n=1}^{\infty}$ be a random process representing the output of a discrete stationary (but not necessarily memoryless) source. Define h_n by

$$h_n \triangleq H(X_n|X_1, X_2, \cdots, X_{n-1}).$$

Prove that h_n is monotone non-increasing.

(c) State the *log-sum inequality* and use it to show that the entropy of a random variable X with finite alphabet $\mathcal{X}$ satisfies

$$H(X) \leq \log_2 |\mathcal{X}|.$$

Give a necessary and sufficient condition for equality.

(d) Let X be a discrete random variable on an alphabet of size five. Suppose the distribution of X is

$$\mathbf{p} = \left(\frac{1}{4}, \frac{1}{4}, p_3, p_4, p_5\right),$$

where the last three probabilities are unknown. Find the values of (p_3, p_4, p_5) such that $H(X)$ is maximized.

(e) Let $\mathcal{X} \triangleq \{X_n\}_{n=1}^{\infty}$ be a discrete i.i.d. source with alphabet $\mathcal{X}$. Let $\mathcal{Y} \triangleq \{Y_n\}_{n=1}^{\infty}$ be a discrete stationary source on the same alphabet $\mathcal{X}$ such that for all letters $a \in \mathcal{X}$,

$$Pr\{X_1 = a\} = \Pr\{Y_1 = a\}.$$

Show that $H(\mathcal{Y}) \leq H(\mathcal{X})$.

4. *The cost of miscoding.* Consider a discrete memoryless source $\{X_n\}_{n=1}^{\infty}$ with alphabet $\mathcal{X} = \{a, b, c, d, e\}$. Let $p_1(\cdot)$ and $p_2(\cdot)$ be two possible probability mass functions for the source that are described in the table below. Also, let $\mathcal{C}_1 = f_1(\mathcal{X})$ where $f_1 : \mathcal{X} \to \{0,1\}^*$, and $\mathcal{C}_2 = f_2(\mathcal{X})$ where $f_2 : \mathcal{X} \to \{0,1\}^*$, be two binary variable-length codes for the source as shown below.

Symbol $x \in \mathcal{X}$	$p_1(x)$	$p_2(x)$	$f_1(x)$	$f_2(x)$
a	1/2	1/2	0	0
b	1/4	1/8	10	100
c	1/8	1/8	110	101
c	1/16	1/8	1110	110
e	1/16	1/8	1111	111

(a) Calculate $H(p_1)$, $H(p_2)$, $D(p_1||p_2)$ and $D(p_2||p_1)$.

(b) Let $\overline{R}(\mathcal{C}_i, p_j)$ denote the average code rate of code $\mathcal{C}_i$ for the source under distribution p_j, $i, j = 0, 1$.

Calculate $\overline{R}(\mathcal{C}_1, p_1)$, $\overline{R}(\mathcal{C}_1, p_2)$, $\overline{R}(\mathcal{C}_2, p_1)$, and $\overline{R}(\mathcal{C}_2, p_2)$.

(c) Compare the quantities calculated in (b) to those in (a) and interpret them qualitatively.

5. State the Asymptotic Equipartition Property (AEP) for a discrete memoryless source and explain its significance vis-a-vis the fixed-length source coding theorem.

6. Answer the following questions.

   (a) State Jensen's inequality and use it to show that

$$\sum_{i=1}^{n} a_i \ln x_i \leq \ln \left[\sum_{i=1}^{n} a_i x_i \right],$$

   where $x_1, x_2, \cdots, x_n$ are arbitrary positive numbers, and $a_1, a_2, \cdots, a_n$ are positive numbers such that $\sum_{i=1}^{n} a_i = 1$.

   (b) Deduce from the above the inequality of the *arithmetic and geometric means*:

$$x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \leq \sum_{i=1}^{n} a_i x_i.$$

7. *Binary divergence.* Show that the binary divergence defined by

$$D_b(p\|q) \triangleq p \ln \left(\frac{p}{q} \right) + (1-p) \ln \left(\frac{1-p}{1-q} \right),$$

   where $0 < p < 1$, $0 < q < 1$ and $\ln(\cdot)$ denotes the natural logarithm, satisfies

$$D_b(p\|q) \leq \frac{(p-q)^2}{q(1-q)}.$$

   Give a sufficient and necessary condition for equality.

8. *Conditioning increases divergence.* Given random variables X and $\hat{X}$ with common alphabet $\mathcal{X}$, and random variable Z with alphabet $\mathcal{Z}$, show that

$$D(P_{X|Z} || P_{\hat{X}|Z}) \geq D(P_X || P_{\hat{X}}),$$

   where

$$D(P_{X|Z} || P_{\hat{X}|Z}) \triangleq \sum_{z \in \mathcal{Z}} \sum_{x \in \mathcal{X}} P_{X,Z}(x,z) \log_2 \frac{P_{X|Z}(x|z)}{P_{\hat{X}|Z}(x|z)},$$

   and

$$D(P_X || P_{\hat{X}}) \triangleq \sum_{x \in \mathcal{X}} P_X(x) \log_2 \frac{P_X(x)}{P_{\hat{X}}(x)}.$$

9. Let $\{p_1, p_2, \ldots, p_m\}$ be a set of positive real numbers with $\sum_{i=1}^{m} p_i = 1$. If $\{q_1, q_2, \ldots, q_m\}$ is any other set of positive real numbers with $\sum_{i=1}^{m} q_i = \alpha$, where $\alpha > 0$ is a constant, show that

$$\sum_{i=1}^{m} p_i \log \frac{1}{p_i} \leq \sum_{i=1}^{m} p_i \log \frac{1}{q_i} + \log \alpha.$$

Give a necessary and sufficient condition for equality.

10. Consider a source $\{X_i\}_{i=1}^{\infty}$ with finite alphabet $\mathcal{X}$.

    (a) Show that $\frac{1}{n} H(X^n)$ is bounded.

    (b) Show that if the source is stationary, then $\frac{1}{n} H(X^n)$ is decreasing with n.

    (c) If the source is stationary and Markov, what is the source entropy rate ?

11. *Converse of the Lossless Variable-Length Coding Theorem for DMS:* Show that if any binary variable-length code for a discrete memoryless source $\mathcal{X}$ has its average codeword length $\bar{l}$ satisfying

$$\bar{l} < H(X),$$

where $H(X)$ is the source entropy, then the code is *not* uniquely decodable.

12. Let $\{X_n\}_{n=1}^{\infty}$ be a discrete memoryless source (i.i.d.) with alphabet

$$\mathcal{X} = \{a, b, c, d, e, f, g, h\}$$

and probability mass function

$$\begin{aligned}
Pr\{X_n = a\} &= 1/4, \quad Pr\{X_n = b\} = 1/4, \\
Pr\{X_n = c\} &= 1/8, \quad Pr\{X_n = d\} = 1/8, \\
Pr\{X_n = e\} &= 1/16, \quad Pr\{X_n = f\} = 1/16, \\
Pr\{X_n = g\} &= 1/16, \quad Pr\{X_n = h\} = 1/16.
\end{aligned}$$

    (a) Compute the source entropy rate $H(\mathcal{X})$. Provide the correct unit.

(b) Compute the source redundancies ρ_D, ρ_M and ρ_T.

(c) Design an optimal (minimum expected length) binary first-order prefix code for the source. Determine the average codeword length $\bar{l}$ for this code. How does $\bar{l}$ compare to $H(\mathcal{X})$?

(d) By observing the results in part (c), can you deduce the average rate $\overline{R}$ of an optimal binary 50th-order prefix code for the source? Explain your answer.

13. Consider a discrete memoryless source $\{X_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{X} = \{a_1, a_2, a_3, a_4, a_5\}$ and probability distribution given by $p_1 = p_2 = \frac{2}{7}$ and $p_3 = p_4 = p_5 = \frac{1}{7}$ where $p_i \triangleq Pr\{X = a_i\}$, $i = 1, \cdots, 5$.

    (a) Compute the source entropy $H(X)$ and the total (fixed-length) redundancy and provide the units.

    (b) Find an optimal binary prefix code for the source and compute its average code rate.

    (c) Is the optimal code in (b) unique ?

14. State and prove the necessary conditions of optimality for a binary prefix code.

15. Let $p(\cdot)$ and $q(\cdot)$ be two different probability distributions on the alphabet

$$\mathcal{X} = \{a_1, \cdots, a_K\}$$

such that $q(a_i) > 0$ for all $i = 1, \cdots, K$. Show that

$$\sum_{i=1}^{K} \frac{p(a_i)^2}{q(a_i)} \geq 1.$$

[Hint: Use Jensen's inequality.]

16. State the *Data Processing Theorem* and prove it.

17. State the Fixed-Length Source Coding Theorem for discrete memoryless sources and outline the key ideas and facts needed to prove it.

18. Prove Fano's inequality using the Data Processing Theorem and the Fundamental Inequality for the logarithm.

19. Consider a discrete memoryless source $\{X_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{X}$ and generic distribution $p(\cdot)$. Let $\mathcal{C} = f(\mathcal{X})$ be a uniquely decodable binary code

$$f : \mathcal{X} \to \{0,1\}^*$$

that maps single source letters onto binary strings such that its average code rate $\overline{R}_{\mathcal{C}}$ satisfies

$$\overline{R}_{\mathcal{C}} = H(X) \qquad \text{bits/source symbol.}$$

In other words, $\mathcal{C}$ is *absolutely optimal*.

Now consider a second binary code $\mathcal{C}' = f'(\mathcal{X}^n)$ for the source that maps source n-tuples onto binary strings:

$$f' : \mathcal{X}^n \to \{0,1\}^*.$$

Provide a construction for the map f' such that the code $\mathcal{C}'$ is also absolutely optimal.

20. Let X and Y be jointly distributed discrete random variables. Show that

$$I(X;Y) \geq I(f(X);g(Y)),$$

where $f(\cdot)$ and $g(\cdot)$ are deterministic functions.

21. *Converse of the Variable-Length Lossless Coding Theorem for Stationary Sources:* Consider a stationary source $\{X_i\}_{i=1}^{\infty}$ with finite alphabet $\mathcal{X}$ and entropy rate $H(\mathcal{X})$.

    (a) Show that $\frac{1}{n}H(X_1, X_2, \cdots, X_n)$ is non-increasing in n and deduce that

    $$H(\mathcal{X}) \leq \frac{1}{n}H(X_1, X_2, \cdots, X_n) \qquad \forall\, n \geq 1.$$

    (b) Show that for any $n \geq 1$, *every* binary uniquely decodable code $f : \mathcal{X}^n \to \{0,1\}^*$ must have an average code rate $\overline{R}$ satisfying

    $$\overline{R} \geq H(\mathcal{X}).$$