CIE6020

Mid-term Examination
SSE, CUHK(SZ)

March 21, 2018

Answer all the questions in the Answer Book. No questions on this page!

1. (17 points) Choose one answer for each of the following questions.

   (a) (4 points) Consider a random variable X over the alphabet $\mathcal{X}$, and an arbitrary function f with domain $\mathcal{X}$. Which of the following inequalities is NOT always correct?

   A. $H(f(X)) \leq H(X)$ B. $H(Y|f(X)) \leq H(Y|X)$
   C. $I(f(X);Y) \leq I(X;Y)$ D. $H(f(X)|Y) \leq H(X|Y)$

   Solution: B.

   (b) (4 points) Let X be a uniformly distributed random variable on $\{0,1,2\}$ and Z be a uniformly distributed random variables on $\{0,1\}$. Suppose X and Z are independent. Which of the following choices is correct?

   A. $H(X+Z) = H(X) + H(Z)$ B. $H(X+Z|X) = H(Z|X)$
   C. $H(X+Z|X) = I(X+Z;Z)$
   D. $I(X;Z|X+Z) \geq I(X;Z)$

   Solution: B or D.
   First, $H(X) = \log 3$, $H(Z) = 1$, $H(X+Z) = \frac{1}{3} + \log 3$, and $I(X;Z) = 0$. Then, $H(X+Z|X) = 1$ and $I(X+Z;Z) = \frac{1}{3}$.

   (c) (3 points) Which of the following codes is a Huffman code for certain probability distribution?

   A. $\{0,10,11\}$ B. $\{00,01,10,110\}$ C. $\{01,10\}$

   Solution: A.

   (d) (3 points) Which of the following binary codes is not uniquely decodable?

   A. $\{0,10,11\}$ B. $\{0,1,01,10\}$ C. $\{01,10\}$

   Solution: B.

   (e) (3 points) Consider an arbitrary discrete memoryless source (DMS) with distribution p. Which of the following statements is NOT correct?

   A. By using block codes, we can achieve zero error and rate arbitrarily close to $H(p)$.
   B. By using variable-length codes, we can achieve zero error and rate arbitrarily close to $H(p)$.
   C. If we allow small error probability, variable-length codes can achieve rates lower than $H(p)$.

   Solution: A.

2. (8 points) For a postive integer n, define a random variable X_n with alphabet $\{1,2,\ldots,n\}$

and probability mass function

$$p(k) = \begin{cases} \frac{1}{2^k} & k = 1, 2, \ldots, n-1, \\ \frac{1}{2^{n-1}} & k = n. \end{cases}$$

Calculate $H(X_n)$ and $\lim_{n\to\infty} H(X_n)$.

Solution: First,

$$\begin{aligned} H(X_n) &= \frac{1}{2}\log 2 + \frac{1}{2^2}\log 2^2 + \cdots + \frac{1}{2^{n-1}}\log 2^{n-1} + \frac{1}{2^{n-1}}\log 2^{n-1} \\ &= \frac{1}{2^{n-1}}\left[2^{n-2} + 2\cdot 2^{n-3} + \cdots + (n-2)\cdot 2 + (n-1) + (n-1)\right]. \end{aligned}$$

Let $S_n = 2^{n-2} + 2\cdot 2^{n-3} + \cdots + (n-2)\cdot 2 + (n-1)$. We have $2S_n - S_n = 2^{n-1} + 2^{n-2} + \cdots + 2 - (n-1) = 2^n - 2 - (n-1)$. Therefore, $H(X_n) = \frac{2^n-2}{2^{n-1}}$, and $\lim_{n\to\infty} H(X_n) = 2$.

3. (8 points) (Huffman coding) Consider the random variable

$$X = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{bmatrix}$$

(a) Find a binary Huffman code for X.

(b) Find the expected code length for the above encoding.

Solution:

(a)

$$\begin{aligned} x_1 &\to 0 \\ x_2 &\to 10 \\ x_3 &\to 110 \\ x_4 &\to 11100 \\ x_5 &\to 11101 \\ x_6 &\to 11110 \\ x_7 &\to 11111 \end{aligned}$$

(b) The expected code length is $1\cdot 0.49 + 2\cdot 0.26 + 3\cdot 0.12 + 5\cdot(0.04 + 0.04 + 0.03 + 0.02) = 2.02$ bits.

4. (8 points) (Lempel-Ziv coding) Give the tree-structured Lempel-Ziv (LZ78) parsing and encoding of

$$AAAAAABBABABAAAAABBABA.$$

Solution: parsing: $A, AA, AAA, B, BA, BAB, AAAA, AB, BABA$.

encoding: $(0, A), (1, A), (2, A), (0, B), (4, A), (5, B), (3, A), (1, B), (6, A)$.

5. (27 points) Consider a channel with the input and output alphabet $\{0, 1\}$. The ith input X_i and the ith output Y_i, $i = 1, 2, \ldots$ are related by

$$Y_i = X_i + U_i$$

where the addition is modulo 2 and U_i has distribution $\Pr\{U_i = 1\} = 1 - \Pr\{U_i = 0\} = q$. Here U_j and $(X_i, i = 1, \ldots)$ are independent.

(a) (7 points) When $U_i, i = 1, 2, \ldots$ and $(X_j, j = 1, \ldots)$ are independent, show the channel is a memoryless binary symmetric channel and give its capacity.

(Hint: show that for any integer $n > 0$,

$$\Pr\{Y_i = y_i, i = 1, \ldots, n | X_i = x_i, i = 1, \ldots, n\} = \prod_{i=1}^{n} \Pr\{Y_i = y_i | X_i = x_i\},$$

i.e., the channel is memoryless.)

Solution: First,

$$\begin{aligned}
\Pr\{Y_i = y | X_i = x\} &= \Pr\{U_i = y - x | X_i = x\} \\
&= \Pr\{U_i = y - x\} \\
&\triangleq W(y|x).
\end{aligned}$$

We know that $W(1|0) = W(0|1) = q$ and $W(0|0) = W(1|1) = 1 - q$.

Write

$$\begin{aligned}
&\Pr\{Y_i = y_i, i = 1, \ldots, n | X_i = x_i, i = 1, \ldots, n\} \\
&= \frac{\Pr\{U_i = y_i - x_i, X_i, i = 1, \ldots, n\}}{\Pr\{X_i = x_i, i = 1, \ldots, n\}} \\
&= \Pr\{U_i = y_i - x_i, i = 1, \ldots, n\} \\
&= \prod_{i=1}^{n} \Pr\{U_i = y_i - x_i\} \\
&= \prod_{i=1}^{n} W(y_i | x_i).
\end{aligned}$$

Therefore, the channel is memoryless and binary symmetry.

The capacity of the channel is $1 - H(q)$.

(b) (4 points) When $U_i = U_{i+1}, i = 1, 3, 5, \ldots$, and $U_i, i = 1, 3, 5, \ldots$ and $(X_j, j = 1, \ldots)$ are independent, show that the channel is not memoryless.

(Hint: calculate $\Pr\{Y_1 = y_1, Y_2 = y_2 | X_1 = x_1, X_2 = x_2\}$ and show that it is not the same as $\Pr\{Y_1 = y_1 | X_1 = x_1\} \Pr\{Y_2 = y_2 | X_2 = x_2\}$.)

Solution: Let

$$\begin{aligned} W_2(y_1, y_2 | x_1, x_2) &\triangleq \Pr\{Y_1 = y_1, Y_2 = y_2 | X_1 = x_1, X_2 = x_2\} \\ &= \Pr\{U_1 = y_1 - x_1, U_2 = y_2 - x_2\}. \end{aligned}$$

As $U_1 = U_2$, we have for $x, y \in \{0, 1\}$,

$$\begin{aligned} W_2(x, y | x, y) &= 1 - q, \\ W_2(x + 1, y + 1 | x, y) &= q. \end{aligned}$$

As $W^2 \neq W_2$, the channel is not memoryless.

(c) (6 points) Under the condition of (b), the channel can be equivalent to a DMC by combining two consecutive uses of the channel. Give the transition matrix of this DMC, and calculate its capacity.

Solution: W_2 as the transition matrix. Let X' and Y' be the input and output of this channel. We have

$$\begin{aligned} I(X'; Y') &= H(Y') - H(Y' | X') \\ &= H(Y') - H(q) \\ &\leq 2 - H(q). \end{aligned}$$

As the maximum can be achieved by the uniform distribution of X', the capacity of the channel is $2 - H(p)$.

(d) (10 points) Assume you are given a set of capacity achieving codes for the memoryless binary symmetric channel under the condition of (a). Using these codes, construct a capacity achieving code for the channel under the condition of (b).

Solution: Consider an (n, M) code C for the binary symmetric DMC. We can modify this code to one for DMC $\{W_2\}$ of the same error probability and rate $1 + \log M/n$. For each codeword $(x_1, x_2, \ldots, x_n)$ of C, and n bits $(y_1, y_2, \ldots, y_n)$, we form a new codeword for W_2, where the i-th input is $(x_i, x_i + y_i)$.
This code has $M2^n$ codewords. Suppose the channel input is $(x_i, x_i + y_i), i = 1, \ldots, n$ and the corresponding output is $(u_i, v_i), i = 1, \ldots, n$. Then $y_i = u_i + v_i$ and $(u_i, i = 1, \ldots n)$ can be used to decode $(x_i, i = 1, \ldots, n)$ using the decoding algorithm of C.

6. (10 points) For any distributions P and Q on a finite set $\mathcal{X}$, and any transition

matrix $W : \mathcal{X} \to \mathcal{Y}$, let PW be the distribution on $\mathcal{Y}$ defined as

$$PW(y) = \sum_{x \in \mathcal{X}} P(x)W(y|x).$$

Let QW be the distribution on $\mathcal{Y}$ similarly defined. Show that $D(PW||QW) \leq D(P||Q)$.

(Hint: log-sum inequality.)

Solution: By definition

$$D(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)},$$

and

$$D(PW||QW) = \sum_y \left[\sum_x P(x)W(y|x)\right] \log \frac{\sum_x P(x)W(y|x)}{\sum_x Q(x)W(y|x)}.$$

For any y, the log-sum inequality gives

$$\begin{aligned}\left[\sum_x P(x)W(y|x)\right] \log \frac{\sum_x P(x)W(y|x)}{\sum_x Q(x)W(y|x)} &\leq \sum_x P(x)W(y|x) \frac{P(x)W(y|x)}{Q(x)W(y|x)} \\ &= \sum_x P(x)W(y|x) \frac{P(x)}{Q(x)}.\end{aligned}$$

Therefore

$$D(PW||QW) \leq \sum_y \sum_x P(x)W(y|x) \frac{P(x)}{Q(x)} = \sum_x P(x) \frac{P(x)}{Q(x)} = D(P||Q).$$

7. (12 points) Let X be a random variable over a finite alphabet $\mathcal{X}$ and let $X^n = (X_1, X_2, \ldots, X_n)$ be a sequence of independent random variables following the same distribution of X. Show that for any $0 < \eta < 1$ and $A \subset \mathcal{X}^n$, if $\Pr\{X^n \in A\} \geq \eta$, then there exists $\epsilon_n \to 0$ as $n \to \infty$, such that $\frac{1}{n} \log |A| \geq H(X) - \epsilon_n$.

(Hint: Using the properties of strongly typical set $T^n_{[X]\delta}$: 1. If $\mathbf{x} \in T^n_{[X]\delta}$, then $2^{-n(H(X)+\eta)} \leq p(\mathbf{x}) \leq 2^{-n(H(X)-\eta)}$, where $\eta \to 0$ as $\delta \to 0$; 2. $\Pr\{X^n \notin T^n_{[X]\delta}\} = \frac{|\mathcal{X}|^2}{4n\delta^2}$.)

Solution: (Note that in the hint, it should be $\Pr\{X^n \notin T^n_{[X]\delta}\} \leq \sum_{a:p(a)>0} \frac{|\mathcal{X}|^2}{4n\delta^2}$. No penalty is given if you use the hint to prove.)

Let $\delta = n^{-1/4}$ and $T_n = T^n_{[X]\delta}$. On the one hand, we have

$$\begin{aligned} P(A \cap T) &= P(A) - P(A \cap T_n^c) \\ &\geq P(A) - P(T_n^c) \\ &\geq P(A) - \sum_{a \in \mathcal{X}: p(a)>0} \frac{|\mathcal{X}|^2}{4n\delta^2} \\ &\geq \eta/2, \end{aligned}$$

where the second last inequality is obtained using the similar steps of proving Strong AEP 2, and the last inequality holds when n is sufficiently large.

On the other hand, by Strong AEP I,

$$P(A \cap T_n) \leq |A \cap T_n| 2^{-n(H(X)-\eta_n)},$$

where $\eta_n > 0$ and $\eta_n \to \infty$ as $n \to 0$.

Together, we have

$$|A| \geq |A \cap T| \geq \frac{\eta}{2} 2^{n(H(X)-\eta_n)} = 2^{n(H(X)-\eta_n+\frac{\log \eta/2}{n})}.$$

The proof is completed by letting $\epsilon_n = \eta_n - \frac{\log \eta/2}{n}$.

8. (10 points) Consider the following part of a proof of the converse of the channel coding theorem. Please justify the equalities/inequalities in (1), (2), (3), (4), (5).

Let R be an achievable rate. By definition, for any $\epsilon > 0$ and all sufficiently large n, there exists (n, M) code (f, φ) such that $\frac{1}{n} \log M > R - \epsilon$ and $\lambda_{\max} < \epsilon$. Let U be the uniform distributed random variable over the message set $\{1, 2, \ldots, M\}$. The codeword we transmit for U is the random variable $\mathbf{X} = f(U)$. Let $\mathbf{Y}$ be the output of the channel for input $\mathbf{X}$, i.e., $(\mathbf{X}, \mathbf{Y}) \sim p_{\mathbf{X}}(\mathbf{x}) W_n(\mathbf{y}|\mathbf{x})$. Let $\hat{U} = \varphi(\mathbf{Y})$. We have a Markov chain

$$U \to \mathbf{X} \to \mathbf{Y} \to \hat{U}.$$

Hence,

$$
\begin{aligned}
\log M &= H(U) \\
&= H(U|\hat{U}) + I(U;\hat{U}) \\
&\leq H(U|\hat{U}) + I(\mathbf{X};\mathbf{Y}) && (1)\\
&\leq 1 + P_e \log M + I(\mathbf{X};\mathbf{Y}), && (2)
\end{aligned}
$$

To bound $I(\mathbf{X};\mathbf{Y})$, we write

$$
\begin{aligned}
I(\mathbf{X};\mathbf{Y}) &= H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) \\
&= H(\mathbf{Y}) - \sum_{i=1}^{n} H(Y_i|Y^{i-1},\mathbf{X}) && (3)\\
&= H(\mathbf{Y}) - \sum_{i=1}^{n} H(Y_i|X_i) && (4)\\
&\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) && (5)\\
&= \sum_{i=1}^{n} I(X_i;Y_i) \\
&\leq nC.
\end{aligned}
$$

Solution: (1): data processing inequality.
(2): Fano's inequality.
(3): chain rule.
(4): memoryless channel.
(5): independence bound.