EGT2
ENGINEERING TRIPOS PART IIA

Tuesday 2 May 2017   9:30 to 11

Module 3F7

INFORMATION THEORY AND CODING

Answer not more than ***three*** *questions.*

All questions carry the same number of marks.

The ***approximate*** *percentage of marks allocated to each part of a question is indicated in the right margin.*

Write your candidate number ***not*** *your name on the cover sheet.*

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed
Attachment: Information Theory and Coding Data Sheet (3 pages).
Engineering Data Book

10 minutes reading time is allowed for this paper.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

1 (a) Let X,Y be jointly distributed discrete random variables each taking values in the set $\{1,2,3\}$. Their joint probability mass function is given in the following table.

	$P_{X,Y}$	$Y=1$	$Y=2$	$Y=3$
X	1	$\frac{1}{4}$	$\frac{1}{16}$	0
	2	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{16}$
	3	0	$\frac{1}{16}$	$\frac{1}{4}$

(i) Compute the entropy $H(X)$, and the joint and conditional entropies $H(X,Y)$ and $H(Y|X)$. [20%]

(ii) Let Z be formed from the product $X.Y$. Determine the probability mass function of Z, and compute its entropy $H(Z)$. [15%]

(iii) Compute the conditional entropies $H(Z|X)$ and $H(X|Z)$. [20%]
Hint: You can show that $H(Z|X)$ is equal to another quantity already computed above. To compute $H(X|Z)$, apply the chain rule to $H(X,Z)$ in two different ways.

(b) Let X be an exponential random variable with probability density function given by $f(x) = \frac{1}{\mu}e^{-x/\mu}$ for $x \geq 0$. Note that the mean of X equals μ.

(i) Compute the differential entropy $h(X)$ in bits. [20%]

(ii) Let Y be another continuous random variable taking values in $[0,\infty)$ with mean μ. Prove that $h(Y) \leq h(X)$, i.e. that the exponential random variable has the largest differential entropy among all continuous random variables taking values in $[0,\infty)$ with mean μ. [25%]
Hint: You may use the property that the relative entropy $D(g||f) \geq 0$, where f and g are suitably chosen probability density functions.

2 (a) Let X be a discrete random variable taking values in the set $\{a,b,c\}$ with probability mass function

$$P_X(a) = 0.4, \quad P_X(b) = 0.35, \quad P_X(c) = 0.25.$$

(i) Find an optimal prefix-free code for X, and compute its expected length. [15%]

(ii) Suppose that $X_1,\ldots,X_n$ is a sequence with each symbol drawn independently from P_X. What is the minimum expected length (in bits/symbol) of any code that assigns a unique binary codeword to each sequence $X_1,\ldots,X_n$? [5%]

(iii) With $X_1,\ldots,X_n$ drawn i.i.d. from P_X, briefly describe one practical technique to construct a prefix-free code that assigns a unique binary codeword to each sequence $X_1,\ldots,X_n$. Give an upper bound for the expected length of your code (in bits/symbol) that approaches the minimum value as n grows. [20%]

(b) Let X and Y be two independent discrete random variables, each with four equiprobable outcomes $\{1,2,3,4\}$. The sum $Z = X+Y$ is observed by Alice. Bob, who does not observe the sum, wishes to determine the sum by asking Alice a series of questions. Each question can only be answered with a 'yes' or a 'no' response.

(i) If Bob asked Alice the following sequence of questions, 'Is the sum 2? Is it 3? ... Is it 7?', stopping when he has determined the value of the sum, then what is the expected number of questions that Bob would ask? [25%]

(ii) Design an optimal strategy for Bob to learn the value of the sum. (An optimal strategy is one that minimizes the expected number of questions, and hence may be based on entropy coding concepts.) [20%]

(iii) Using the strategy in part b(ii) above, what is the expected number of questions? What is the sequence of questions that Bob would ask if the value of the sum is 3? [15%]

3 (a) Consider the binary *asymmetric* channel shown in Fig. 1, in which the 0 input has crossover probability $\frac{1}{4}$, and the 1 input has crossover probability $\frac{1}{2}$.

(i) If $P(X=0)=p$ for some $0 \leq p \leq 1$, compute the output distribution $\{P(Y=0), P(Y=1)\}$. [10%]

(ii) Find the capacity of the channel and the maximising input distribution. [40%]

Hint: Note that

$$\frac{d}{dp} H_2\left(\frac{1}{2}+\frac{p}{4}\right) = \frac{1}{4}\log_2\left(\frac{\frac{1}{2}-\frac{p}{4}}{\frac{1}{2}+\frac{p}{4}}\right),$$

where $H_2(\cdot)$ is the binary entropy function and $0 \leq p \leq 1$.

(iii) Describe briefly how you would construct a capacity achieving code for this channel. (Assume that you do not have to worry about the complexities of encoding, decoding, and storage.) [20%]

(b) A discrete-valued sequence $X^n = (X_1, \ldots, X_n)$ is passed through a discrete memoryless channel $P_{Y|X}$ to obtain the output sequence $Y^n = (Y_1, \ldots, Y_n)$. This is shown in Fig. 2. The capacity of the channel is denoted by $\mathcal{C}$.

Provide brief justifications for the relationships, labelled (a), (b), (c) and (d), in the following chain of steps which relate mutual information to channel capacity:

$$\begin{aligned} I(X^n;Y^n) &\overset{(a)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid Y_{i-1}, \ldots, Y_1, X^n) \\ &\overset{(b)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid X_i) \\ &\overset{(c)}{\leq} \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i \mid X_i) \\ &\overset{(d)}{\leq} n\mathcal{C}. \end{aligned}$$

[30%]

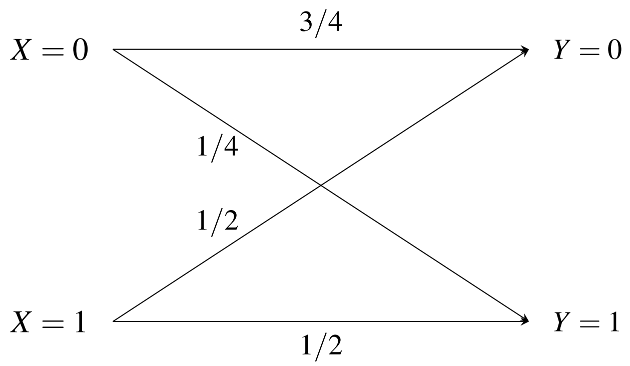

Fig. 1

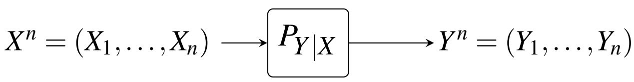

Fig. 2

4 (a) Consider a binary linear code with the following generator matrix.

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

(i) What are the dimension, block length, and rate of the code? [10%]

(ii) Find a parity check matrix for the code. [10%]

(iii) A codeword from the code is transmitted over a binary erasure channel, and the received sequence is $\underline{y} = [?,?,?,1,?,0,0]$ where '?' indicates an erasure. Find the transmitted codeword and the corresponding information sequence. [25%]

(b) A binary LDPC code has edge perspective polynomials

$$\begin{cases} \lambda(x) = 0.3x^2 + 0.4x^4 + 0.3x^5 \\ \rho(x) = x^5 \end{cases}$$

(i) What is the design rate of the code? [10%]

(ii) Briefly explain what the coefficients of the powers of x in $\lambda(x)$ and $\rho(x)$ signify. [10%]

(iii) This LDPC code is to be used over a channel with a binary input $X \in \{0,1\}$ and continuous-valued output $Y \in \mathbb{R}$. For $x \in \{0,1\}$, the conditional density describing the channel is

$$f_{Y|X}(y \mid x) = \frac{1}{2}\exp(-|y-x|), \quad y \in \mathbb{R}.$$

Compute and sketch the log-likelihood ratio $L(y)$, defined as [20%]

$$L(y) = \ln\left(\frac{f_{Y|X}(y \mid 0)}{f_{Y|X}(y \mid 1)}\right).$$

(iv) The LDPC code is used over the channel in part b(iii). A variable node j receives channel output $y_j = 0.1$, and is connected to three check nodes. In a belief-propagation decoding algorithm based on log-likelihood ratio rules, if the incoming messages received by the variable node along the three edges are $[-0.23, 0.53, 1.04]$, determine the outgoing message sent by the variable node along the *third* edge. [15%]

Note: The belief propagation message passing equations are given in the data sheet.

END OF PAPER