EGT2
ENGINEERING TRIPOS PART IIA

Friday 26 April 2019 9.30 to 11.10

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
Engineering Data Book

10 minutes reading time is allowed for this paper at the start of the exam.

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

1 (a) A die, for which the six possible outcomes $\{1, \ldots, 6\}$ are equally likely, is tossed. Let X denote the outcome of the die. If X is 1,2,3, or 4, a coin is tossed once. If X is 5 or 6, the coin is tossed twice. Assume heads and tails are equally likely in each coin toss. Let Y denote the number of heads obtained.

(i) Compute the probability mass function of Y. [20%]

(ii) Find the mutual information between X and Y. [15%]

(b) Let X be a continuous random variable with probability density function $f(x)$, where $f(x) \neq 0$ on the interval $[a, b]$, and is zero elsewhere. Show that the differential entropy $h(X)$ is maximum when X is uniformly distributed on $[a, b]$. [20%]

(c) Let $X \in \{0, 1\}$ and $Y \in \{a, b, c\}$ be jointly distributed random variables with the following joint probability mass function (pmf) P_{XY}:

	P_{XY}	Y: a	b	c
X	0	0	$\frac{1}{4}$	$\frac{1}{4}$
	1	$\frac{1}{4}$	$\frac{1}{4}$	0

Let P_X and P_Y denote the marginal distributions of X and Y, respectively. Define the following sets of length-n sequences:

$$A_n^X = \left\{ x^n : -\frac{1}{n} \log_2 P_X(x^n) = H(X) \right\},$$

$$A_n^Y = \left\{ y^n : -\frac{1}{n} \log_2 P_Y(y^n) = H(Y) \right\},$$

$$A_n^{XY} = \left\{ (x^n, y^n) : -\frac{1}{n} \log_2 P_{XY}(x^n, y^n) = H(X, Y) \right\}.$$

Here $P_X(x^n) = \prod_{i=1}^n P_X(x_i)$, $P_Y(y^n) = \prod_{i=1}^n P_Y(y_i)$, and $P_{XY}(x^n, y^n) = \prod_{i=1}^n P_{XY}(x_i, y_i)$.

(i) Describe precisely what the sequence pairs (x^n, y^n) in A_n^{XY} look like, in terms of n and $n_{1a}, n_{0b}, n_{1b}, n_{0c}$. Here n_{1a} denotes the number of occurrences of the symbol pair $(X = 1, Y = a)$ in (x^n, y^n), and n_{0b}, n_{1b}, n_{0c} are similarly defined. [15%]

(ii) Describe precisely what the sequences in A_n^X and A_n^Y look like, in terms of n and the number of occurrences of each symbol in the relevant alphabet. (Use n_0, n_1 for the number of occurrences of 0, 1, respectively, in x^n. Similarly define n_a, n_b, n_c for y^n.) [15%]

(iii) For $n = 8$, give an example of a sequence pair (x^n, y^n) such that $(x^n, y^n) \in A_n^{XY}$, but $y^n \notin A_n^Y$. [15%]

2 (a) Consider a source S with a five symbol alphabet with the symbol probabilities $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\}$. Without actually performing the Huffman encoding procedure, show that a Huffman code must achieve the entropy of this source exactly. [20%]

(b) Consider a source S' with probability mass function $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16} + \gamma, \frac{1}{16} - \gamma\}$, where $0 < \gamma \leq \frac{1}{32}$ is an unknown constant. A Huffman code designed for source S in part (a) is used to compress a long sequence of i.i.d. symbols produced by S'.

(i) Compute the redundancy of this code in terms of γ. (The redundancy of a code is the amount by which its expected code length exceeds the minimum possible value for the source.) [10%]

(ii) For which value of $\gamma \in (0, \frac{1}{32}]$ do you expect the redundancy to be the largest? Briefly justify your answer. (You are not required to compute the maximum.) [10%]

(c) A source X has alphabet $\{a_1, a_2, a_3, a_4, a_5\}$ with symbol probabilities

$$P(a_1) = \frac{1}{4},\ P(a_2) = \frac{1}{4},\ P(a_3) = \frac{5}{24},\ P(a_4) = \frac{1}{6},\ P(a_5) = \frac{1}{8}.$$

Let $\mathcal{B}$ be the set of prefix-free symbol codes for this source with codeword lengths $\ell_1 = 2$, $\ell_2 = 2$, $\ell_3 = 3$, and $\ell_5 = 4$, where ℓ_i is the length of the codeword for a_i. Notice that ℓ_4 has not been specified.

(i) Find the minimum value of ℓ_4 among the prefix-free codes in set $\mathcal{B}$, and the expected code length of the corresponding code. [15%]

(ii) Does the set $\mathcal{B}$ contain an optimal symbol code for X (i.e., one with the minimum expected code length)? Justify your answer. [10%]

(iii) Assume that you are no longer restricted to codes in the set $\mathcal{B}$. To compress a long sequence of i.i.d. symbols from X, which of the following would be a better choice: Huffman coding, or arithmetic coding? Justify your answer. [10%]

(d) Let X be a random variable that takes values in an m letter alphabet with probability mass function (pmf) $\{p_1, \ldots, p_m\}$. Suppose that $p_i > p_j$ for some $i, j \in \{1, \ldots, m\}$. Consider a random variable X' with pmf obtained by replacing p_i with $(p_i - \epsilon)$, and p_j with $(p_j + \epsilon)$, where $\epsilon > 0$ is such that $p_i - \epsilon > p_j + \epsilon$. The other probabilities are unchanged.

Prove that $H(X') > H(X)$. [25%]

3 (a) Prove the following inequalities:

(i) $I(X; g(Y)) \leq I(X; Y)$, where X, Y are jointly distributed discrete random variables and $g(Y)$ is a function of Y. [15%]

(ii) $\sum_{x \in \mathcal{X}} \frac{(P(x))^2}{Q(x)} \geq 1$, where P, Q are probability mass functions on the same alphabet $\mathcal{X}$. *Hint*: Start with $D(P\|Q) \geq 0$. [15%]

(b) Consider two discrete memoryless channels whose input alphabet, output alphabet, and transition probability matrix are $(\mathcal{X}_1, \mathcal{Y}_1, P_{Y_1|X_1})$ and $(\mathcal{X}_2, \mathcal{Y}_2, Q_{Y_2|X_2})$, respectively. In each time instant, we are allowed to transmit one input symbol over *either* channel 1 or channel 2, but not both. Assume that $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are distinct so that the receiver can tell which individual channel each output symbol came from; you may also assume that $\mathcal{X}_1$ and $\mathcal{X}_2$ are distinct.

We call this a union channel since its input and output alphabets are $\mathcal{X}_1 \cup \mathcal{X}_2$ and $\mathcal{Y}_1 \cup \mathcal{Y}_2$, respectively. Let X and Y denote the input and output symbols of the union channel.

(i) Define a random variable Z that is equal to 1 if channel 1 was used, and equal to 2 otherwise. Show that $I(X; Y, Z) = I(X; Y)$. [10%]

(ii) Any input distribution P_X for the union channel can be expressed as $P_X = \alpha P_{X_1} + (1-\alpha) P_{X_2}$, where P_{X_1} and P_{X_2} are input distributions for the two individual channels, and $\alpha \in [0, 1]$ is the probability with which channel 1 is used. For such an input distribution, show that

$$I(X; Y, Z) = \alpha I(X_1; Y_1) + (1-\alpha) I(X_2; Y_2) + H_2(\alpha),$$

where Z is as defined in part (b)(i), and $H_2(\alpha) = \alpha \log_2 \frac{1}{\alpha} + (1-\alpha) \log_2 \frac{1}{(1-\alpha)}$. [15%]

(iii) Show that the capacity $\mathcal{C}$ of the union channel satisfies

$$2^{\mathcal{C}} = 2^{\mathcal{C}_1} + 2^{\mathcal{C}_2},$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are the capacities of the two individual channels. [25%]
Hint: You may use the fact that $\frac{dH_2(\alpha)}{d\alpha} = \log_2 \frac{1-\alpha}{\alpha}$.

(iv) Compute the capacity and the maximising input distribution for the channel described by the following probability transition matrix. [20%]

	$P_{Y\|X}$	Y: 0	1	2
X	0	0.9	0.1	0
	1	0.1	0.9	0
	2	0	0	1

4 Consider a binary linear code with the following parity check matrix.

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

(a) What are the dimension and the rate of the code? [10%]

(b) A codeword from the code was transmitted over a binary erasure channel, and the received sequence was $\underline{y} = [1, ?, ?, ?, 1, 1]$ where the ? outputs are erasures. Find the transmitted codeword. [15%]

(c) Find the edge perspective degree polynomials $\lambda(x)$ and $\rho(x)$ of the parity check matrix. [5%]

(d) Consider another code of large block length with the same edge perspective polynomials as in part (c). This code is used over a binary erasure channel with erasure probability ϵ, with message passing decoding. Let p_t denote the probability that a variable-to-check message is an erasure after t iterations of message passing, for $t \geq 1$. Derive the following density evolution equation for p_t: [20%]

$$p_t = \epsilon(1 - (1 - p_{t-1})^2), \qquad \text{with } p_0 = \epsilon.$$

Hint: Let q_t denote the probability that a check-to-variable message is an erasure in iteration t. Express p_t in terms of q_t, and then q_t in terms of p_{t-1}.

(e) What is the key assumption in deriving the density evolution equation above, and when is it a reasonable one? [10%]

(f) Consider a channel with binary input $x \in \{0, 1\}$ and continuous-valued output y generated according to a conditional density $f(y \mid x)$ as follows: if $x = 0$, then $f(y \mid x = 0)$ is a uniform density in the interval $[-1, 1]$; if $x = 1$, then $f(y \mid x = 1)$ is a uniform density in the interval $[0, 2]$. Compute the likelihood ratio $\frac{f(y \mid x=0)}{f(y \mid x=1)}$ for $y \in [-1, 2]$. [10%]

(g) The code with the parity check matrix $\mathbf{H}$ defined above is used over the channel defined in part (f). If the received sequence is $\underline{y} = [0.1,\ 1.2,\ -0.6,\ 1.7,\ 0.5,\ 0.9]$, find the transmitted codeword. (You may use any appropriate decoder.) [15%]

(h) Find the capacity of the channel defined in part (f). [15%]

END OF PAPER

THIS PAGE IS BLANK