EGT2
ENGINEERING TRIPOS PART IIA

Thursday 28 April 2022 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.

You may not remove any stationery from the Examination Room.

1 (a) Consider two sets $\mathcal{A} = \{1, 2, 3\}$ and $\mathcal{B} = \{3, 4\}$. We flip a fair coin (heads and tails equally likely). If the coin lands heads, we pick one of the three numbers in $\mathcal{A}$ with equal probabilty. If the coin lands tails, we pick one of the two numbers in $\mathcal{B}$ with equal probability. Let $X \in \{\text{Heads}, \text{Tails}\}$ denote the outcome of the coin flip, and $Y \in \{1, 2, 3, 4\}$ denote the number picked.

(i) Determine the probability mass function of Y and the entropy $H(Y)$. [15%]

(ii) Compute the conditional entropy $H(Y \mid X)$. [15%]

(iii) Compute the conditional entropy $H(X \mid Y)$. [15%]

(b) A source emits 4 symbols from a 2-character alphabet $\{a, b\}$. The first symbol X_1 is equally likely to be a or b. The successive symbols X_2, X_3, X_4 are generated as follows. For $k \geq 2$, the symbol X_k is the same as X_{k-1} with probability 0.7, and different from X_{k-1} with probability 0.3. An arithmetic encoder is used to compress the sequence (a, a, b, b) produced by the source.

(i) Determine the interval produced by the arithmetic encoder and the corresponding binary codeword. [25%]

(ii) Compute the minimum expected codelength of *any* prefix-free code for (X_1, X_2, X_3, X_4). [20%]

(iii) Give an upper bound for the expected length for an arithmetic codeword corresponding to (X_1, X_2, X_3, X_4) and comment on the usefulness of this bound. [10%]

2 (a) There are six bottles of juice, of which it is known that exactly one bottle has gone bad (tastes terrible). The probability p_i that the ith bottle is the bad one is given by

$$(p_1, p_2, p_3, p_4, p_5, p_6) = \left(\frac{8}{23}, \frac{6}{23}, \frac{4}{23}, \frac{2}{23}, \frac{2}{23}, \frac{1}{23}\right).$$

Tasting will determine the bad bottle.

(i) Suppose we taste the bottles one at a time, and stop once we taste the bad one. Choose the order of tasting to minimize the expected number of tastings required to determine the bad bottle. What is the expected number of tastings required? (Note that if the first five bottles pass the test, you don't have to taste the last.) [20%]

(ii) Now suppose that we are allowed to mix samples from different bottles and taste the mixture. The mixture will taste bad if it contains a sample from the bad bottle. We proceed, mixing and tasting, stopping when the bad bottle has been determined.

What is the minimum expected number of tastings required to determine the bad bottle? Also specify which mixtures will be tasted in your optimal strategy when the bad bottle is the fourth one. [30%]

(b) Consider a channel whose input X and output Y belong to a ternary alphabet $\{1, 2, 3\}$. The channel has the following transition matrix:

	$P_{Y\|X}$	Y = 0	1	2
X	0	0.7	0.2	0.1
	1	0.1	0.7	0.2
	2	0.2	0.1	0.7

(i) Compute the capacity of the channel. [20%]

(ii) A single X symbol, drawn from the probability distribution $P_X(0) = \frac{2}{5}$, $P_X(1) = \frac{1}{5}$, $P_X(2) = \frac{2}{5}$, is transmitted over the channel. We wish to estimate X from the output Y. Compute a lower bound on the probability of error of *any* such estimator. [30%]

3 (a) Consider a discrete memoryless channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transition probabilities given by the conditional distribution $P_{Y|X}$.

(i) Describe the components of an (n, k) code to convey k information bits in n channel transmissions. In particular, briefly describe the codebook, the encoder, and the decoder. [15%]

(ii) Give the formula for the channel capacity and state the channel coding theorem. [15%]

(b) A channel has a 4-bit input

$$X \in \{0000,\ 0001,\ 0010,\ 0011,\ \ldots,\ 1100,\ 1101,\ 1110,\ 1111\}$$

and a 3-bit output $Y \in \{000,\ 001,\ 010,\ 011,\ 100,\ 101,\ 110,\ 111\}$. Given an input X, the output Y is generated by *deleting* exactly one of the four input bits, selected at random. For example, if the input is $X = 1010$, then $P(Y \mid X)$ is $1/4$ for each of the outputs 010, 110, 100, 101; if the input is $X = 0001$, then $P(Y = 001 \mid X)$ is $3/4$ and $P(Y = 000 \mid X)$ is $1/4$.

(i) With a uniform input distribution over the set of 4-bit sequences, show that the conditional entropy $H(Y \mid X)$ satisfies $H(Y \mid X) < 2$ bits. [15%]

(ii) Determine the entropy $H(Y)$ with a uniform input distribution, and use this to argue that the capacity $\mathcal{C}$ of the channel satisfies $\mathcal{C} > 1$ bit. [20%]

(iii) Consider an input distribution that uses just four of the sixteen inputs:

$$P(X) = \begin{cases} 1/4 & \text{if } X \in \{0000,\ 0011, 1100,\ 1111\} \\ 0 & \text{otherwise.} \end{cases}$$

With this input distribution, evaluate $H(X \mid Y)$, and hence show that the capacity $\mathcal{C}$ satisfies $\mathcal{C} \geq 2$ bits. [20%]

(iv) Do you think that the capacity of the channel is exactly 2 bits, or is there a better input distribution than the one above? Briefly explain your reasoning, but further calculations are not expected. [15%]

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

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

(a) Specify the dimension and the rate of the code. [5%]

(b) Obtain a systematic parity check matrix for the code. [10%]

(c) Specify a systematic generator matrix and list all the codewords. [15%]

(d) Draw the factor graph for the parity check matrix **H** given in Equation (1). [10%]

(e) A codeword from the code is transmitted over a binary symmetric channel with crossover probability 0.2. The received sequence is $\underline{y} = [1\ 1\ 1\ 1\ 1\ 1]$.

(i) What is the maximum-likelihood decoded codeword? [5%]

(ii) Given the received sequence $\underline{y}$, determine the ratio of the posterior probabilities of the *third* bit of the codeword being a 1 versus a 0, assuming that each information symbol is equally likely to be zero or 1. With a bit-wise decoder, would this bit be decoded to a zero or a one? [25%]

(iii) Now consider belief propagation decoding of $\underline{y}$. Suppose that we stop the belief propagation decoder after one complete iteration of message passing, i.e., after one round of variable-to-check and check-to-variable messages. For the BSC with crossover probability 0.2, compute the final log-likelihood ratio for the third code-bit, and state whether it will be decoded to a zero or a one. [20%]

(iv) Compare the log-likelihood ratios corresponding to part (ii) with the one computed in (iii), and comment on why they are different. [10%]

END OF PAPER

THIS PAGE IS BLANK