Version RV/4

EGT2
ENGINEERING TRIPOS PART IIA

Thursday 26 April 2018 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) You are given four coins, and it is known that at least three of them have the same weight, while the remaining one may (or may not) be heavier. Assume that each of the five possibilities — all the coins have the same weight, coin number 1 is the heaviest, coin number 2 is the heaviest, etc. — are equally likely. A balance is available on which the weights of any two arbitrary groupings of coins may be compared. The operator will place any combination of coins on the balance that you request, and will answer a *yes/no* question regarding the relative weights of the coins. For example, your question might be: 'Are the coins on the right pan of the balance heavier than those on the left pan?' Or, 'Do the coins on the left pan of the balance have the same weight as those on the right pan?'

(i) Find a weighing strategy (i.e., a sequence of weighing steps and yes/no questions, where you are allowed to place any combination of coins of your choosing on the balance at each step) that will determine which of the four coins, if any, is heavier. The strategy you choose should be optimal in the sense of minimising the expected number of weighings required. Represent your strategy as a binary tree which specifies the question asked in each step depending on the answers of previous steps. [30%]

(ii) What is the expected number of weighings required when using your strategy? [10%]

(b) Let X and Y be random variables that take on real values in $\{x_1, \ldots, x_r\}$ and $\{y_1, \ldots, y_s\}$, respectively. Let $Z = X + Y$.

(i) Show that $H(Z|X) = H(Y|X)$. [10%]

(ii) Show that if X and Y are independent, then [20%]

$$H(Y) \leq H(Z), \text{ and } H(X) \leq H(Z).$$

That is, the addition of independent random variables can only increase uncertainty.

(iii) Give an example of random variables for which [10%]

$$H(X) > H(Z) \text{ and } H(Y) > H(Z).$$

Note that X, Y cannot be independent when these inequalities are satisfied.

(iv) Under what conditions does $H(Z) = H(X) + H(Y)$? [20%]

2 (a) A sequence of m uniformly random bits has to be transmitted over a discrete memoryless channel. The channel has transition probabilities specified by the conditional distribution $P_{Y|X}$, where X and Y denote the channel input and output symbols, respectively. The capacity of the channel is denoted by $\mathcal{C}$.

Design a coding scheme to transmit the sequence of m bits using a code of rate $0.8\mathcal{C}$. You first need to specify the code length, and then briefly describe how the codebook is constructed, as well as the encoding and decoding operations. (Assume you do not have to worry about the complexity of encoding and decoding.) [25%]

(b) Consider the set-up shown in Fig. 1, where an input X is transmitted through two separate discrete memoryless channels (DMCs), with outputs labelled Y and Z, respectively.

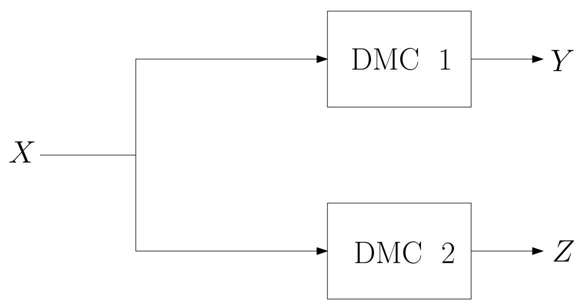

Fig. 1

The capacities of the two DMCs are $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. Let $\mathcal{C}$ denote the capacity of the *overall* channel, whose input is X and output is the pair (Y,Z).

(i) Prove that $\mathcal{C} \geq \max\{\mathcal{C}_1, \mathcal{C}_2\}$. [15%]

(ii) Prove that $\mathcal{C} \leq \mathcal{C}_1 + \mathcal{C}_2$. [20%]

Hint: Use the fact that Y, Z are conditionally independent given X.

(iii) If the two DMCs are binary symmetric channels with crossover probabilities 0.1 and 0.2, respectively, determine the exact value of $\mathcal{C}$. [40%]

3 (a) Let X be a discrete random variable with probability mass function P.

(i) Prove that the expected code length L of the Shannon-Fano code for X satisfies $L < H(X) + 1$. [15%]

(ii) Give an example of a random variable X for which the *optimal* code has expected code length very close to $1 + H(X)$. That is, for any $\varepsilon > 0$, you need to show that there exists a random variable X for which the expected code length of the optimal code for X is at least $1 + H(X) - \varepsilon$. [15%]

(iii) Part (a).(ii) shows that one cannot improve in general on the average code length bound of part (a).(ii) with any *symbol-by-symbol* code. Now consider a source producing symbols $X_1, X_2, \ldots$ that are independent and identically distributed according to P. If we construct a Shannon-Fano code over blocks of k source symbols, obtain an upper bound on its expected code length, expressed in terms of bits/source symbol. [15%]

(b) Consider a random variable X with continuous probability density function (pdf) $f(x)$. Suppose that the range of X is divided into bins of width Δ, as shown in Fig. 2.

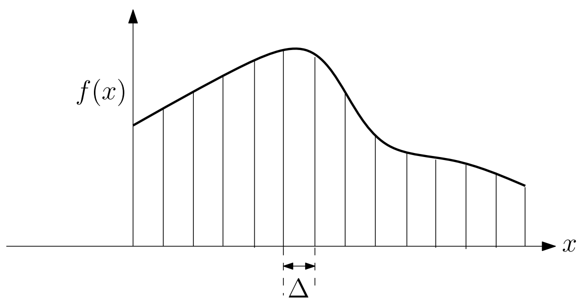

Fig. 2

Label the bins from left to right by $i = 0, 1, \ldots$. Let x_i be a value within bin i such that

$$f(x_i)\Delta = \int_{i\Delta}^{(i+1)\Delta} f(x)dx.$$

Then the quantised random variable X^Δ is defined as:

$$X^\Delta = x_i, \quad \text{if} \quad i\Delta \leq X \leq (i+1)\Delta.$$

(i) Let p_i be the probability that $X^\Delta = x_i$. Show that $p_i = f(x_i)\Delta$. [10%]

(ii) Show that the entropy of the quantised random variable can be expressed as [10%]

$$H(X^\Delta) = -\sum_i \Delta f(x_i) \log_2 f(x_i) - \sum_i f(x_i)\Delta \log_2 \Delta.$$

 (cont.

(iii) Show that $\sum_i f(x_i)\Delta = 1$, and use this to express $H(X^\Delta)$ as [10%]

$$H(X^\Delta) = -\sum_i \Delta f(x_i) \log_2 f(x_i) - \log_2 \Delta. \tag{1}$$

(iv) Let $h(X)$ denote the differential entropy of X. Using Eq. (1), it can be shown that (under appropriate conditions)

$$H(X^\Delta) + \log_2 \Delta \;\rightarrow\; h(X) \quad \text{as} \quad \Delta \rightarrow 0. \tag{2}$$

Taking $\Delta = 2^{-n}$, Eq. (2) says that the entropy of an n-bit quantisation of a continuous random variable X is approximately $h(X) + n$, when n is large. Use this result to show that if X is uniformly distributed in $[0, \frac{1}{8}]$, then X can be quantised to n-bit accuracy with only $(n-3)$ bits. (You are not required to prove Eq. (2).) [15%]

(v) Interpret the result in part (b).(iv) for the uniform $[0, \frac{1}{8}]$ distribution. [10%]

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

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

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

(b) A codeword from the code was transmitted over a binary erasure channel, and the received sequence was $\underline{y} = [?, 1, 1, ?, ?, 0]$ where ? are erasures. Find the information sequence, and the transmitted codeword. [15%]

(c) Find a systematic generator matrix for the code. [10%]

(d) Find a parity check matrix $\mathbf{H}$ for the code. [5%]

(e) What is the minimum distance of the code? [10%]

(f) A codeword from this code is transmitted over a binary symmetric channel (BSC), and the received sequence is $[0, 1, 0, 0, 0, 0]$. If we use a maximum-likelihood (minimum-distance) decoder, what is the decoded codeword? Why is this decoded codeword unique? [10%]

(g) Draw the factor graph corresponding to the parity check matrix $\mathbf{H}$ found in part (d). [10%]

(h) Now consider decoding the received sequence $[0, 1, 0, 0, 0, 0]$ using a belief propagation decoder, with messages in the log-likelihood ratio (LLR) format. The crossover probability of the BSC is 0.1. In the first iteration, what is the message sent by the second variable node to the check nodes connected to it? [10%]

(i) 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 0.1, compute the final LLR for the second code-bit, and state whether this bit will be correctly decoded. [20%]

END OF PAPER