UNIVERSITY OF CAMBRIDGE

EXAMINATION IN MATERIALS MODELLING FOR THE DEGREE OF MASTER OF PHILOSOPHY

Tuesday 23rd April 2002 **9 – 10**

MODELLING OF MATERIALS (3)

Answer **two** *of the three questions.*

Write on **one** *side of the paper only.*

The answer to **each question** *must be tied up* **separately**, *with its own cover-sheet.*

Write the relevant **question number** *in the square labelled 'Section' on each cover–sheet. Also, on* **each** *cover–sheet, list the numbers of* **all** *questions attempted from this paper.*

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

1 *Answer all three parts.*

(a) A binary source X has an alphabet of eleven characters

$$\{a, b, c, d, e, f, g, h, i, j, k\},$$

all of which have equal probability, 1/11.

Find an *optimal uniquely decodeable symbol code* for this source.

How much greater is the expected length of this optimal code than the entropy of X?

(b) In a clinical trial, drug A is tested on 1800 patients and it cured 600 of them. Drug B was tested on 1 patient, who was cured. The probability that drug A cures a patient is called 'the effectiveness of drug A', f_A. In the light of these data, and assuming a uniform prior probability on the effectiveness f_A,

$$P(f_A) = 1 \qquad 0 \leq f_A \leq 1,$$

and a uniform prior probability on the effectiveness f_B, *estimate* the probability that drug B has greater effectiveness than drug A.

(c) In a recent court case, a defendant's first two babies died suddenly aged 11 weeks and 8 weeks respectively; she stated that they died naturally. When she was put on trial, accused of murdering the babies, a prosecution witness advanced the following argument:

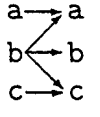
> 'In affluent non-smoking families like the defendant's, sudden infant death syndrome strikes one baby in 8,500. The probability of two babies both dying from sudden infant death is thus
>
> $$\frac{1}{8500} \times \frac{1}{8500} \simeq \frac{1}{72,000,000}.$$
>
> Because this probability is so small, the two deaths give strong evidence that the defendant is guilty of murdering them both.'

Comment on this argument.

2. *Answer both parts.*

(a) A channel with input $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ and output $y \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ has transition probability matrix:

$$Q = \begin{bmatrix} 1 & 1/3 & 0 \\ 0 & 1/3 & 0 \\ 0 & 1/3 & 1 \end{bmatrix}$$



(TURN OVER for continuation of Question 2)

Assuming an input distribution of the form $\{q, p, q\}$, where $q = \frac{1}{2}(1 \quad p)$, explain why the entropy of the output can be written in the form

$$H(Y) = H_2\left(\frac{p}{3}\right) + \left(1 - \frac{p}{3}\right) H_2\left(\frac{1}{2}\right)$$

where $H_2(x) = x \log_2(1/x) + (1 - x) \log_2(1/(1 - x))$.

Find the *conditional entropy* $H(Y|X)$ and the *mutual information* $I(X;Y)$ between the input and output, as a function of p.

Find an expression for the *optimal input distribution* of this channel.

[You may find the identity $\frac{d}{dx} H_2(x) = \log_2 \frac{1 - x}{x}$ helpful. Also, if

$$\log_2 \frac{1 - x}{x} = a \quad \text{then} \quad x = \frac{1}{1 + 2^a}.]$$

(b) Two spies wish to communicate with each other using a channel based on a pack of playing cards. One spy will use his message to determine the *order* of the 52 cards in the deck. The other spy will recover the message from the order.

Estimate the *capacity* of this channel in bits.

How would you design a practical method for conveying a *binary message* in one use of this channel? Remember, the spies would like to communicate lots of bits. Estimate how many bits your practical method conveys, and compare it with your estimate of the capacity, explaining why they differ, if they differ; or why not, if not.

The spies decide to introduce a 'noisy' protocol. The first spy will choose the order of the 52 cards, then, before the cards are handed over, the first spy will allow a third party to cut the deck, at random, once. The second spy receives the reordered deck. What is the capacity of this noisy channel, in bits? Discuss whether your communication scheme can be modified to allow reliable communication over this noisy channel.

[Explanation of what 'cutting the deck' means:

Cutting the deck at the eighth card, for example, reorders the deck such that the order (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ..., 52) becomes (8, 9, 10, ..., 52, 1, 2, 3, 4, 5, 6, 7).]

3. Write an essay discussing clustering and classification of unlabelled and labelled data.

END OF PAPER