UNIVERSITY OF EDINBURGH

COLLEGE OF SCIENCE AND ENGINEERING

SCHOOL OF INFORMATICS

INFORMATION THEORY

Thursday 10$^{\underline{\text{th}}}$ May 2012

09:30 to 11:30

MSc Courses

Convener: B. Franke
External Examiners: T. Attwood, R. Connor, R. Cooper, D. Marshall, M. Richardson

INSTRUCTIONS TO CANDIDATES

Answer QUESTION 1 and ONE other question.

Question 1 is COMPULSORY.

All questions carry equal weight.

CALCULATORS MAY NOT BE USED IN THIS EXAMINATION

1. **You MUST answer this question.**

(a) Give an advantage of digital communication over analogue communication. [*1 mark*]

(b) Explain why the probability of making an error when using many noisy channels, such as the binary symmetric channel, can never be *exactly* zero. Explain what is meant when it is claimed that error-correcting codes allow *reliable* communication. [*3 marks*]

(c) A system needs to store sets of 10,000 positive integers between 0 and 31 inclusive.

i. How many bits per set are required to guarantee storing these data? [*1 mark*]

ii. Now assume that the first integer is drawn uniformly from the possible integers, and subsequent integers are drawn from the following distribution:

$$P(x_{n+1} \,|\, x_n) = \begin{cases} 1/2 & x_{n+1} = x_n \\ 1/4 & x_{n+1} = (x_n + 1) \bmod 32 \\ 1/4 & x_{n+1} = (x_n - 1) \bmod 32. \end{cases}$$

Describe an optimal binary code for representing each sequence of 10,000 integers. Explain whether your code is instantaneously decodable. What is the expected length of your encoding? [*4 marks*]

(d) A system encounters two types of binary file: compressed files that appear to be bits drawn uniformly at random; and sparse files, with each bit set to one independently with probability $f = 1/10$. The sparse files are rarer, occurring with probability $1/5$.

Given that the first two bits of a file are ones, what is the probability that the next bit will also be a one, $P(x_3 = 1 \,|\, x_1 = 1, x_2 = 1)$? [*5 marks*]

(e) A system creates independent bits with the probability of generating a '`1`' equal to $f = 0.1$. Explain which of the following sequences of 50 bits has higher probability under the system.

i. `00000000000000000000000000000000000000000000000000`

ii. `00000100000000000000000100000000001011000010000000000`

One of these sequences was actually generated by the system. Which do you think it was, and why? [*3 marks*]

QUESTION CONTINUES ON NEXT PAGE

The remainder of this question concerns a distribution over a 20×20 binary image, a length-400 bit vector $\mathbf{x}$, defined through an 'energy function' $E(\mathbf{x})$. More probable images are given lower energy, with probabilities set equal to:

$$P(\mathbf{x}) = \frac{1}{Z} 2^{-E(\mathbf{x})},$$

where Z is the normalizing constant required to create a valid distribution.

A set of 'good' images $\mathcal{G}$ are each assigned energy $E_{\mathcal{G}} = -370$; there are $|\mathcal{G}| = 2^{40}$ such images. Each remaining image is assigned an energy of zero.

(f) What is the probability that an image drawn uniformly at random belongs to the set of 'good' images? [*1 mark*]

(g) Find an expression for $P(\mathbf{x} \notin \mathcal{G})$, the probability that an image drawn from $P(\mathbf{x})$, defined in terms of the energies above, does *not* belong to the good set. Show that this probability is ≈ 0.001, to one significant figure. [*2 marks*]

(h) Express $\log_2 Z$ numerically to 4 significant figures. Show your working. [*2 marks*]

(i) What is the entropy of this ensemble, in bits, to 1 significant figure? Give an interpretation of this result. (HINT: you might be able to write down a sufficiently accurate answer after considering the result from part g.) [*3 marks*]

2. **You should either answer this question or question 3.**

You may use the binary entropy function, $H_2()$, in your answers to this question:

$$H_2(p) \equiv p \log_2 \tfrac{1}{p} + (1-p) \log_2 \tfrac{1}{1-p}, \qquad p \in (0,1),$$

and $H_2(0) = H_2(1) = 0$. For reference: $\frac{\mathrm{d}H_2(p)}{\mathrm{d}p} = \log_2\left(\frac{1-p}{p}\right)$.

(a) Briefly define ensembles that have entropies of: i) $H_2(p)$, ii) $5H_2(p)$, and iii) $1+\frac{1}{2}H_2(p)$. Give expressions in terms of p. [*3 marks*]

Consider a discrete memoryless channel where both input, $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, and output, $y \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, take on one of three values. One value, `a`, is never corrupted. When `b` or `c` are used, they are confused with each other with probability f:

$$\begin{aligned}
&p(y{=}\mathtt{a} \,|\, x{=}\mathtt{a}) = 1, && \\
&p(y{=}\mathtt{b} \,|\, x{=}\mathtt{b}) = 1-f, \qquad && p(y{=}\mathtt{c} \,|\, x{=}\mathtt{b}) = f, \\
&p(y{=}\mathtt{c} \,|\, x{=}\mathtt{c}) = 1-f, \qquad && p(y{=}\mathtt{b} \,|\, x{=}\mathtt{c}) = f.
\end{aligned}$$

(b) Using a simple communication scheme with zero probability of error, explain why the capacity of this channel is lower bounded by 1 bit. [*2 marks*]

(c) Explain why the mutual information between the input and output ensembles, $I(X;Y)$, can be maximized with an input distribution of the form:

$$\begin{aligned}
p(x{=}\mathtt{a}) &= p_a \\
p(x{=}\mathtt{b}) &= (1-p_a)/2 \\
p(x{=}\mathtt{c}) &= (1-p_a)/2.
\end{aligned}$$

An argument using symmetry is expected. A careful explanation is required. [*3 marks*]

(d) What is p_a in the optimal input distribution when $f{=}0$, $f{=}1/2$, and $f{=}1$? Give explanations in terms of the simpler channels that result from these settings of f. Also state the capacities of the channel in these cases. [*5 marks*]

(e) Show that, for input distributions as in c), the mutual information between the input and output ensembles is $I(X;Y) = H_2(p_a) + (1-p_a)(1-H_2(f))$. [*4 marks*]

(f) Find an expression for the p_a that maximizes the mutual information in terms of a general $f \in [0,1]$. Justify why your answer is the maximum; that is, justify that no other p_a gives a higher mutual information. [*4 marks*]

(g) Describe how a randomly generated block code can (in theory) be constructed for this channel that would have a low probability of error at a rate close to the capacity. (You need not reproduce the proof that your code will have the desired properties.) State how, in theory, the code would be used to communicate a binary file, and the practical difficulty with doing so. [*4 marks*]

3. **You should either answer this question or question 2.**

This question concerns communicating over a binary symmetric channel with a very low flip/noise probability, such as $f = 10^{-8}$. The channel is going to be used many times, so this noise level could cause problems. However, reducing the bit error probability (the probability that a randomly selected source bit is decoded incorrectly) to 10^{-15} would be more than acceptable.

(a) State the rate of the repetition code R_3. Also state the bit error probability as a function of f, and numerically, for $f = 10^{-8}$, to 1 significant figure. [*3 marks*]

(b) Explain why the probability of incorrectly decoding a $[7, 4]$ Hamming code *block* is approximately $21f^2$. Find the *bit* error probability to 1 significant figure for $f = 10^{-8}$. Explain whether you would select the R_3 or $[7, 4]$ Hamming code in this situation.

(You may find it helpful to recall that two bit errors in a block decode to three bit errors, which are equally likely to be on message or parity bits.) [*5 marks*]

(c) Explain why neither repetition nor Hamming codes would be appropriate for use with noisier channels, such as the binary symmetric channel with $f = 0.1$. Name a code without these difficulties that could be used in practice. Give an advantage and disadvantage of replacing a Hamming code with the code that you propose for channels with very low noise levels. [*5 marks*]

If a reliable feedback channel were available, a simple check-bit code could be used. After sending a pair of source bits, their sum modulo two is sent as a third bit. If the receiver detects corruption of the triple, retransmission is requested.

(d) i) Explain how a receiver could decide whether to request retransmission of a triple of bits. Then express the probability that a retransmission will be requested, p_r, both ii) in terms of f, and iii) numerically, to 1 significant figure, for $f = 10^{-8}$. [*3 marks*]

(e) For simplicity, the receiver applies the same algorithm to request retransmission each time a triple is received (rather than attempting to combine all received copies). Express the expected number of channel uses, $\mathbb{E}[L]$, required to decode a pair of source bits in terms of the retransmission probability p_r. Thus, report a measure of the rate of the code: $2/\mathbb{E}[L]$. [*4 marks*]

(f) An alternative measure of the rate is $\mathbb{E}[2/L]$, the expected number of bits sent per channel use while sending a source pair. Is one of the performance measures always less than the other (if so, which)? Explain your reasoning. [*3 marks*]

(g) Give two real-world situations in which a code that relies on feedback would be less desirable than 'forward' error-correcting codes with no feedback. [*2 marks*]