UNIVERSITY OF EDINBURGH

COLLEGE OF SCIENCE AND ENGINEERING

SCHOOL OF INFORMATICS

INFORMATION THEORY

Thursday 19 th December 2013

09:30 to 11:30

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

MSc Courses

Convener: B. Franke
External Examiners: A. Burns, S. Denham, P. Healey, T. Norman

THIS EXAMINATION WILL BE MARKED ANONYMOUSLY

You MUST answer this question.

1. (a) A small hard disk is designed to transfer a Gigabyte of data a day. Explain why an error probability per bit operation of 10^{-6} is likely to be insufficient for general computing use, whereas 10^{-20} might be good enough.

   According to the noisy channel coding theorem, error correction should allow arbitrarily low error probabilities. Explain under what conditions that theory holds, and why hard drive manufacturers don't tend to create drives with bit error probabilities of 10^{-100}. *[5 marks]*

   (b) For independent ensembles X and Y, entropy adds:

$$H(X,Y) = H(X) + H(Y).$$

   Demonstrate, with a simple example, that this relation does not hold if X and Y can be dependent. *[2 marks]*

   (c) For this part, you may state without proof that any of the following functions are strictly concave for $x \in [0,1]$:
   $\log x$, $-x \log x$, $-\exp(x)$, $H_2(x) = x \log \frac{1}{x} + (1-x) \log \frac{1}{1-x}$.
   A discrete ensemble with alphabet $\mathcal{A}_X = \{a_i\}$ has corresponding probabilities $\{p_i\}$. The probabilities are unknown. The empirical distribution after drawing N independent samples gives unbiased estimates:

$$\hat{p}_i = \frac{n_i}{N}, \qquad i = 1 \ldots |\mathcal{A}_X|,$$

   where n_i is the number of occurrences of the ith alphabet element.

   i. The information content of an outcome $h(a_i) = \log {}^1/_{p_i}$ is estimated as $\hat{h}(a_i) = \log {}^1/_{\hat{p}_i}$. Either show that $h(a_i) \leq \mathbb{E}[\hat{h}(a_i)]$ or show that $h(a_i) \geq \mathbb{E}[\hat{h}(a_i)]$, and state under what conditions, if any, $h(a_i) = \mathbb{E}[\hat{h}(a_i)]$. *[5 marks]*

   ii. The entropy of the ensemble H is also estimated by the entropy of the empirical distribution, $\hat{H}$. Either show that $H \leq \mathbb{E}[\hat{H}]$ or show that $H \geq \mathbb{E}[\hat{H}]$, and state under what conditions, if any, $H = \mathbb{E}[\hat{H}]$. *[5 marks]*

   iii. As a check, explicitly show which inequalities from the two previous parts hold for a single sample ($N=1$) from a uniform distribution over two alphabet elements. *[2 marks]*

   (d) Consider the following joint distributions over two binary variables x and y:

i)

$P(x,y)$	$x=0$	$x=1$
$y=0$	1	0
$y=1$	0	0

ii)

$P(x,y)$	$x=0$	$x=1$
$y=0$	$1/2$	0
$y=1$	0	$1/2$

iii)

$P(x,y)$	$x=0$	$x=1$
$y=0$	$1/4$	$1/4$
$y=1$	$1/4$	$1/4$

   For each case, what is the mutual information between these two variables? *[6 marks]*

2. **Source coding:**

   (a) Given a complete binary symbol code with codeword lengths $\{l_i\}$, explain why the ideal ensemble for this code has symbol probabilities $q_i = 2^{-l_i}$. *[3 marks]*

   (b) Given lengths $\{l_i\}$ from an incomplete code, $\sum_i 2^{-l_i} \neq 1$, the implied distribution requires normalizing, $q_i = \frac{1}{Z} 2^{-l_i}$.

      i. Write down an expression for Z.

      ii. Write down an expression for the expected length of the encoding per symbol, when encoding symbols that are drawn with probabilities q_i.

      iii. By comparing the expected length to its optimal value, derive a bound on the lengths of the codewords in a uniquely-decodable symbol code. *[5 marks]*

We are interested in compressing a stream of symbols. The marginal distribution of an individual symbol, taken uniformly at random from the stream, has distribution $\mathcal{P}_X = \{p_i\}$ with entropy $H(X)$.

   (c) Given a stream of N independent symbols, each sampled with distribution $\mathcal{P}_X$, what is the worst-case expected number of bits per symbol required to represent the stream with an arithmetic coder? *[2 marks]*

   (d) The same arithmetic coder, with a model of independent symbols, is used to encode a stream of dependent symbols with the same marginal distribution. Carefully justify whether the worst-case expected encoding length will be the same or different as in the previous part. *[4 marks]*

Consider the ensemble of strings of $N = 10^6$ independent Bernoulli outcomes with probability of a 1 equal to $f = 0.01$. For reference: $H_2(f) = 0.08$ bits (1sf), $\log_2 1/f = 7$ bits (1sf), and $\log_2 1/(1-f) = 0.01$ bits (1sf).

   (e) Sketch the distribution of the number of ones observed in a string from this ensemble. Include approximate numerical indications of both the location and width of the distribution. *[5 marks]*

   (f) What is the information content of the most probable string to one significant figure? *[1 mark]*

   (g) What is the average information content of strings sampled at random from the ensemble, to one significant figure? *[1 mark]*

   (h) What is the information content of the majority of strings drawn from the ensemble, to one significant figure? Justify your answer carefully. *[4 marks]*

3. **Noisy channel coding:**

(a) Consider the Binary Erasure Channel (BEC) with erasure probability f, and the Binary Symmetric Channel (BSC) with flip probability f.

i. Explain which of the two channels has the higher capacity for $f = 0.01$.

ii. Explain which channel has the higher capacity for $f = 0.99$. [*4 marks*]

A channel has ternary inputs $x \in \{1, 2, 3\}$ and ternary outputs $y \in \{1, 2, 3\}$, with transition probabilities $P(y = i \mid x = j) = Q_{ij}$ given by the matrix:

$$Q = \begin{pmatrix} 1-f & f/2 & f/2 \\ f/2 & 1-f & f/2 \\ f/2 & f/2 & 1-f \end{pmatrix}.$$

(b) Show that the conditional entropy of the output is:
$H(Y \mid X) = H_2(f) + f$ bits. [*3 marks*]

(c) Hence or otherwise, find the capacity of the channel. Show your working. If you use the optimal input distribution as part of your working, be careful to derive or carefully justify what it is. [*5 marks*]

A block code for this channel allows the user to choose K ternary symbols $x_k \in \{1, 2, 3\}$, $k = 1 \ldots K$, and deterministically computes $N - K$ redundant symbols to protect the user's message.

(d) Using this block code, how many bits of information can each ternary block of length N carry? [*2 marks*]

(e) Bound the ratio K/N for which, in the limit of large N, block codes exist with negligible probabilities of error. [*3 marks*]

(f) Explain how a large standard binary file could be converted to a ternary format using close to the minimum possible number of ternary symbols. [*4 marks*]

(g) Consider a family of block codes, such as low density parity check codes, that work at rates close to the capacity of the channel at negligible probabilities of error for large block lengths N. For a particular large N, explain whether i) the rate will be smaller or greater than that of a repetition code with N repetitions; ii) the bit-error probability will be smaller or greater than that of a repetition code with N repetitions. [*4 marks*]