UNIVERSITY OF EDINBURGH

COLLEGE OF SCIENCE AND ENGINEERING

SCHOOL OF INFORMATICS

INFORMATION THEORY

Mock Paper 2011

Time allowed: 2 hours

MSc Courses

Convener: ITO-Will-Determine
External Examiners: ITO-Will-Determine

INSTRUCTIONS TO CANDIDATES

***** The rubric in your real exam may differ. Always read the rubric. *****

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) Use the Huffman algorithm to find an optimal binary symbol code for the ensemble with alphabet $\mathcal{A}_X = \{\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d},\mathtt{e},\mathtt{f},\mathtt{g}\}$ and probabilities $\mathcal{P}_X = \{1/4, 1/4, 1/8, 1/8, 1/8, , 1/16, 1/16\}$. i) Find the expected number of bits per symbol used by your code. ii) Explain whether a stream from this ensemble might be compressed better with an arithmetic code. [*4 marks*]

(b) In another stream, using the same alphabet $\mathcal{A}_X$ as the previous part, the probability of a symbol depends on the context of preceding symbols. However, the marginal probability of a symbol, the probability of a single symbol taken randomly from the whole stream, is still the same $\mathcal{P}_X$. Will the answer to i) from the previous part be better, worse or the same for this new case? Re-answer question ii) from the last part. [*4 marks*]

(c) Suppose in yet another stream of independent symbols, one of the symbols appears more than half of the time on average, $p(\mathtt{a}) > 0.5$. Must a Huffman codeword for this symbol always be 1 bit long? Prove or disprove. [*3 marks*]

(d) Give two definitions of the *mutual information*, $I(X;Y)$ in terms of some of the standard entropies: $H(X)$, $H(Y)$, $H(X\,|\,Y)$, $H(Y\,|\,X)$, and $H(X,Y)$. Relate one definition to an intuitive explanation of one of the things that $I(X;Y)$ measures about the random variables X and Y. [*3 marks*]

(e) A very long stream of symbols in $\mathcal{A}_X = \{\mathtt{a},\mathtt{b}\}$ is drawn, with each symbol chosen independently with probabilities $\mathcal{P}_X = \{\frac{1}{3}, \frac{2}{3}\}$. An arithmetic coder is used to compress the stream, but due to a mistake uses a fixed probability distribution, $\mathcal{Q}_X = \{\frac{2}{3}, \frac{1}{3}\}$, with the probabilities reversed. Compute the average number of extra bits per symbol this compression system uses compared to an arithmetic coder using the correct $\mathcal{P}_X$ probabilities. [*3 marks*]

(f) Ben wishes to compute the information content of a file under a model formed by a mixture of two different models, with distributions P_1 and P_2:

$$P(\mathbf{x}) = \tfrac{1}{2}\,P_1(\mathbf{x}) + \tfrac{1}{2}\,P_2(\mathbf{x}).$$

The information content measured by the two models are 2023 and 2025 bits respectively. Ben writes the following computer code:

```
ln_prob1 = -2023 * log(2.0);
ln_prob2 = -2025 * log(2.0);
ln_prob = log(0.5*exp(ln_prob1) + 0.5*exp(ln_prob2));
inf_content = -ln_prob / log(2.0);
```

On Ben's computer, using C or MATLAB, this code reports an information content of "`Inf`".

What is the correct information content to 3 significant figures?

Rewrite the above code to avoid numerical problems. [*3 marks*]

QUESTION CONTINUES ON NEXT PAGE

(g) In a $[N, K]$ block error-correcting code, such as the [7,4] Hamming code, explain what the N and K refer to. Give a definition for the 'rate' of a code and how it relates to $[N, K]$. [*2 marks*]

(h) State how many bit-errors the $[7, 4]$ Hamming code can tolerate in a block, and write down an expression for the probability that a block is corrupted. Estimate this probability for the binary symmetric channel, with flip probability $f = 0.01$, to one significant figure. [*3 marks*]

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

A random variable Y with $\mathcal{A}_Y = \{0, 1, 2, \ldots, M\}$ is deterministically created by multiplying two independent random variables X and Z:

$$Y = XZ.$$

X takes on values from $\mathcal{A}_X = \{1, 2, \ldots, M\}$ with probabilities $\mathcal{P}_X = \{p_1, p_2, \ldots, p_M\}$ and Z is binary: $\mathcal{A}_Z = \{0, 1\}$ with probabilities $\mathcal{P}_Z = \{f, 1-f\}$.

(a) Write down the entropy of the resulting variable, $H(Y)$, in terms of $H(X)$ and f. [*2 marks*]

(b) What choices of probabilities $\mathcal{P}_X$ and $\mathcal{P}_Z$ maximize $H(Y)$? [*2 marks*]

(c) If Z is a source of noise with fixed probability f, find the capacity of the discrete memoryless channel with input X and output Y? [*3 marks*]

(d) Describe a scheme that can communicate at a rate equal to the capacity with zero probability of error with the use of a feedback channel. Briefly show that your scheme does achieve the required rate. [*3 marks*]

(e) Name an error correcting code that could be applied to this channel that can communicate at rates close to this capacity with very low probability of error without feedback. Briefly outline how practical encoding and decoding is performed for this code. [*3 marks*]

Another variable W with $\mathcal{A}_W = \{0, 1, \ldots, M-1\}$ is deterministically created by adding the two independent random variables X and Z modulo M:

$$W = X + Z \mod M.$$

(f) Identify a simple scheme to communicate using the discrete memoryless channel with input X and output W that has zero probability of error. Hence prove that the capacity of this channel is at least $\log_2 \lfloor M/2 \rfloor$ bits. [*3 marks*]

When $M = 2$ the scheme and resulting bound from the previous part are not useful. Indeed any single isolated use of the channel will be ambiguous for $f > 0$. In the course we reviewed a proof of Shannon's noisy channel coding theorem.

(g) How, and in what sense, is reliable communication possible on this type of channel, even when $M = 2$? [*2 marks*]

(h) What is the Nth extension of a channel, and how do its capacity and input, output and joint distributions relate to the original channel? [*3 marks*]

(i) A code can be constructed by choosing S codewords at random from the optimal input distribution for the Nth extension. How many codewords would be required to communicate at the capacity of the channel? [*1 mark*]

QUESTION CONTINUES ON NEXT PAGE

We use slightly fewer codewords and communicate at slightly less than the capacity. We can then prove that, averaging over all random codes, the probability of decoding a codeword incorrectly is less than some tolerance, 2δ say, which can be made arbitrarily small as $N \to \infty$. Therefore there exists a code with block error less than 2δ.

(j) The probability of error when using some codewords can be worse than others. What can be said about the 'maximal block error', the probability of error for the most confusable codeword? How can the code be modified to obtain a better bound on the block error, and why does this not significantly decrease the rate of the code in the proof? [*3 marks*]

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

   A real exam would have a third question here. This year (2011) you will only have to do one of questions 2 and 3, which will each be slightly more focussed in coverage than question 1.

   The structure could change in future years. When the time for your exam comes, always read the rubric for the paper that you are actually sitting.