UNIVERSITY OF EDINBURGH

COLLEGE OF SCIENCE AND ENGINEERING

SCHOOL OF INFORMATICS

INFORMATION THEORY

Wednesday 27 th April 2011

14:30 to 16:30

MSc Courses

Convener: C. Stirling
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) A discrete random variable can take one of seven values in the alphabet $\mathcal{A}_X = \{a_1, a_2, a_3, \ldots, a_7\}$. The probabilities of the first 3 values are $P(x=a_1) = P(x=a_2) = P(x=a_3) = 0.25$. Derive tight bounds, both upper and lower, on the entropy of the ensemble, $H(X)$, that could result from setting the remaining probabilities. *[4 marks]*

(b) Construct a uniquely and instantaneously decodable binary symbol code for the above ensemble, with expected length no worse than 2.5 bits/symbol, no matter how the remaining probabilities $\{p_4, p_5, p_6, p_7\}$ are set. Explain i) why your code always achieves the required performance; ii) for which settings of the remaining probabilities the average performance can be no better than 2.5 bits/symbol. *[4 marks]*

(c) For each of the following, state, with reasons, whether a symbol code using these codewords is: 1) uniquely decodable; 2) instantaneously decodable; 3) could result from an implementation of the Huffman algorithm.

i. $\{\mathtt{0}, \mathtt{11}, \mathtt{100}, \mathtt{101}\}$

ii. $\{\mathtt{0}, \mathtt{101}, \mathtt{11}\}$

iii. $\{\mathtt{1}, \mathtt{101}\}$

iv. $\{\mathtt{10}, \mathtt{11}, \mathtt{01}, \mathtt{000}, \mathtt{001}, \mathtt{100}\}$

[8 marks]

(d) Values $x^{(1)}$ and $x^{(2)}$ are drawn independently from a common probability distribution with entropy $H(X)$. (Not necessarily the same distribution as in previous parts of this question.) One of the following inequalities is correct:

$$P(x^{(1)}=x^{(2)}) \geq 2^{-H(X)} \qquad \text{or} \qquad P(x^{(1)}=x^{(2)}) \leq 2^{-H(X)}.$$

Use the fact that 2^z is a strictly convex function of z to prove the correct inequality. Which distributions have the equality $P(x^{(1)}=x^{(2)}) = 2^{-H(X)}$? *[4 marks]*

(e) A Binary Symmetric Channel (BSC), where the probability that a bit is flipped is $f = 0.1$, has capacity $C \approx 0.53$ bits. What is the probability of incorrectly decoding a bit sent with the repetition code R_3? How could this probability of error be reduced? Is it possible to communicate over this channel at a rate of 0.1 bits with an error probability of 10^{-6}? Explain whether this can be achieved with a repetition code and/or a Hamming code. *[5 marks]*

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

(a) Sketch the binary entropy function, $H_2(p)$ for probabilities $p \in [0,1]$ where:

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

$H_2(0) = \lim_{\epsilon\to 0} H_2(\epsilon)$ and $H_2(1) = \lim_{\epsilon\to 0} H_2(1-\epsilon)$. Briefly give an intuitive meaning of $H_2(p)$, and explain why the limiting values for $H_2(0)$ and $H_2(1)$ are used. *[4 marks]*

You may use the $H_2()$ function in expressing answers to the remainder of this question. For reference: $\frac{\mathrm{d}H_2(p)}{\mathrm{d}p} = \log_2\left(\frac{1-p}{p}\right)$.

A communication channel takes two bits of input, $x \in \mathcal{A}$, and gives two bits of output, $y \in \mathcal{A}$, where $\mathcal{A} = \{\texttt{00}, \texttt{01}, \texttt{10}, \texttt{11}\}$. Usually the bits are transmitted uncorrupted, $y = x$, but with probability f the bits are reversed:

$$\begin{array}{ll}
p(y=\texttt{00}\,|\,x=\texttt{00}) = 1, & \\
p(y=\texttt{01}\,|\,x=\texttt{01}) = 1-f, & p(y=\texttt{10}\,|\,x=\texttt{01}) = f, \\
p(y=\texttt{10}\,|\,x=\texttt{10}) = 1-f, & p(y=\texttt{01}\,|\,x=\texttt{10}) = f. \\
p(y=\texttt{11}\,|\,x=\texttt{11}) = 1, &
\end{array}$$

Assume that the input distribution satisfies $P(x=\texttt{00}) = P(x=\texttt{11}) = p_s$ and $P(x=\texttt{01}) = P(x=\texttt{10}) = p_d$. It can be shown that the optimal input distribution must have this symmetry for $f \in [0,1)$.

(b) Show that the entropy of the output satisfies $H(Y) = 1 + H_2(2p_d)$. *[2 marks]*

(c) Show that the conditional entropy satisfies $H(Y\,|\,X) = 2p_d H_2(f)$. *[2 marks]*

(d) Write down the mutual information $I(X;Y)$ for the channel and find p_d for the optimal input distribution. *[5 marks]*

(e) Write down the capacity of the channel when $f = 0$, $f = 0.5$ and $f = 1$. Explain how the results for these special cases are apparent without detailed calculation. *[4 marks]*

(f) For $f \in (0,1)$ how many non-confusable inputs can be selected? At what rate can one communicate reliably by using only these inputs? *[2 marks]*

(g) Explain when and how is it possible to communicate with very small probability of error at a rate greater than that found in the previous part when $f \in (0, 0.5)$. Include how a file might be processed into a form usable by the system, what the maximum possible performance is, and what practical considerations may reduce the actual performance. *[6 marks]*

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

(a) Compare and contrast the relative merits of using arithmetic coding and Huffman codes as part of a compression system for data such as text or digital images. [*6 marks*]

In the *Dasher* text-entry system, strings are assigned to non-overlapping intervals in $[0, 1)$. The height of the interval for string $\mathbf{x}$ is its probability as assigned by a model, just as in arithmetic coding. In one of *Dasher*'s input modes, text can be entered by pressing buttons to zoom its display into one of the four quarters of the previously-displayed interval. Mathematically, if the current interval is $[a, b)$, first initialized to $[0, 1)$, the buttons zoom in to one of the intervals

$$A = [a, c_1), \quad B = [c_1, c_2), \quad C = [c_2, c_3), \quad D = [c_3, b),$$

where $c_1 = a+(b-a)/4$, $c_2 = a+(b-a)/2$, and $c_3 = a+3(b-a)/4$. Once the displayed interval is completely inside the desired string's interval, the user can stop.

(b) On average, what is the highest number of bits that pressing one of four buttons can convey? Might the actual number be less than that? Explain your answer. [*3 marks*]

(c) Assume a user can reliably select and press one of four buttons A–D, each of which selects the corresponding quarter of the current interval. Derive the minimum and maximum number of button presses that might be needed to enter a string with probability $p(\mathbf{x})$. Express your final answers in terms of the information content of the string, $h(\mathbf{x})$, measured in bits. [*4 marks*]

Some users are only able to press a single button. In an input mode that allows them to enter text, the currently displayed part of $[0, 1)$ is split into N non-overlapping intervals. The computer sequentially highlights each interval and zooms into the highlighted interval whenever a button is pressed. Waiting for and selecting the nth interval takes n seconds. The fractions of the screen occupied by the intervals are given by $\mathbf{q} = \{q_1, q_2, \ldots, q_N\}$ such that $\sum_n q_n = 1$. When the four uniformly-sized intervals A–D with $q_1 = q_2 = q_3 = q_4 = 1/4$ are used, it takes 1, 2, 3 or 4 seconds to wait for the required interval and to press the button.

(d) A measure of the rate of text entry is the average information per button press, H, divided by the average time for each press, T. Evaluate this information rate, H/T, in bits/s when using the uniform intervals A–D. [*3 marks*]

(e) If the time taken to select an interval is to be proportional to the information conveyed about the desired string, show that the fractional heights have the form $q_n = 2^{-Rn}$. State what determines R. [*3 marks*]

QUESTION CONTINUES ON NEXT PAGE

(f) Use Gibbs' inequality, $\sum_n p_n \log 1/q_n \geq \sum_n p_n \log 1/p_n$, to show that the information rate using alternative fractional heights, $\mathbf{p}$, satisfies $H(\mathbf{p})/T(\mathbf{p}) \leq R$, where the optimum rate is achieved by using $p_n = 2^{-Rn}$. [*3 marks*]

(g) Show that by using intervals with non-equal heights, the rate of text entry can be better than in part (c) but no better than 1 bit per second. (Knowing that $\sum_{n=1}^{\infty} 2^{-n} = 1$ may be helpful.) [*3 marks*]