COMP2610 - Information Theory: Assignment 3

Australian National University

Lecturers: Mark Reid and Aditya Menon

Handed out: 1 October 2013. Due: 18 October 2013 **by 4pm**.
Updated: 16 October 2013 (Q.II (b))

Instructions

Submission: You should submit a typed or scanned copy of your assignment by email to mark.reid@anu.edu.au or a hard copy to the Research School of Computer Science (RSCS) Student Administration at CSIT (108) building Room N340. Handwritten paper submissions are acceptable if the handwriting is neat and legible. Note that only Portable Document Format (PDF) files are acceptable for the electronic submission.

Late Penalty: 20% cumulative per working day overdue, after which a mark of 0 will be given (in the absence of medical evidence or other special permission).

Cheating and Plagiarism: All assignments must be done individually. Remember that *plagiarism is a university offence* and will be dealt with according to university procedures. Please refer to the the corresponding ANU policies:
`http://academichonesty.anu.edu.au/UniPolicy.html`.

I Arithmetic and Stream Coding [40 points]

1. Consider the ensemble X with alphabet $\mathcal{A} = \{\texttt{a}, \texttt{b}, \texttt{c}, \texttt{d}\}$ and uniform probabilities $\mathbf{p} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.

   (a) Compute the cumulative distribution function F for this ensemble and each of the binary intervals correponding to each of the symbols in $\mathcal{A}$.

   (b) Construct the Shannon-Fano-Elias code for X and encode the message `abab`.

   (c) How does the length of a Huffman codeword for the same ensemble and message compare with the one you just determined?

2. Consider the Dirichlet multinomial model with $\mathbf{m} = (1, 1, 1, 1)$ for the alphabet $\mathcal{A} = \{\texttt{a}, \texttt{b}, \texttt{c}, \texttt{d}\}$. Compute the following probabilities under this model:

   (a) $P(x = \texttt{b})$

   (b) $P(x = \texttt{a}|\texttt{baaa})$

   (c) $P(x = \texttt{b}|\texttt{ac})$

3. Using an arithmetic coder with the above Dirichlet multinomial model, code the sequence `acb`, showing the intervals and probabilities at each step.

4. Consider an ensemble X consisting of a uniform probability distribution over an alphabet consisting of the capital letters A through to Z, a space character, and punctuation characters '.', ',', '?', '!' and '"'. Suppose we wanted to compress the message $\mathbf{x} = AAA \ldots A$ consisting of 1000 repetitions of the character A.

   Estimate (or determine exactly) how many bits would be required to code the message $\mathbf{x}$ using:

   (a) A Huffman code for X.

   (b) An adaptive arithmetic code for X with a Dirichlet multinomial model with:

      i. uniform model parameters $\mathbf{m} = (1, 1, \ldots, 1)$.

      ii. model parameters $\mathbf{m} = (1000, 1, 1, \ldots, 1)$.

   (c) The Lempel-Ziv algorithm LZ78.

   Note: You do not have to actually code the message but do explicitly state any assumptions or estimates you make when calculating the above quantities.

II Noisy-Channel Coding [40 points]

1. You are put in charge of optimising communications over a complex noisy channel Q that takes as input one of 1,000 symbols from $\mathcal{X}$ and outputs a symbols from $\mathcal{Y}$ that also contains 1,000 symbols. A salesman calls you up offering a function $f : \mathcal{Y} \to \mathcal{Y}$ that can be added to Q to form a new channel Q' with increased capacity by mapping each y that Q outputs to $y' = f(y)$.

(a) By appealing to the definition of channel capacity and an inequality from the lectures, explain why the salesperson must be wrong.

(b) The salesperson responds by saying, "Oh! Did I say the function was $f : \mathcal{Y} \to \mathcal{Y}$? I meant $f : \mathcal{X} \to \mathcal{X}$." Appealing to the definition of capacity again, carefully explain why using such a function to pre-process symbols before sending them through the channel Q will also not increase its capacity.

2. Let $\mathcal{X} = \{\mathtt{a}, \mathtt{b}\}$ and $\mathcal{Y} = \{\mathtt{a}, \mathtt{b}\}$ be the input and output alphabets for the following two channels:

$$Q = \begin{bmatrix} 1 & 0.5 \\ 0 & 0.5 \end{bmatrix}. \quad \text{and} \quad Q' = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 1 \end{bmatrix}$$

(a) Describe the behaviour of Q and Q' when each of the symbols $\mathtt{a}$ and $\mathtt{b}$ are used as input.

(b) For an arbitrary distribution $\mathbf{p} = (p, 1-p)$ with $p \in [0, 1]$ over $\mathcal{X}$, write down the following expressions for Q:

i. The probabilities $P(y = \mathtt{a})$ and $P(y = \mathtt{b})$ in terms of p.

ii. The entropy of $H(Y)$ in terms of the probability $p = P(X = \mathtt{b})$ for X.

iii. The mutual information $I(X; Y)$ in terms of p.

(c) Using the previous results or some other argument determine the channel capacity for Q.

(d) Argue why the channel Q' must have the same capacity as Q.

(e) Suppose you used the channels Q and Q' to send messages by first flipping a fair coin and sending a symbol through Q if it landed heads and through Q' if it landed tails.

i. Construct a matrix Q^* that represents the channel defined by this process.

ii. Determine the capacity of this new channel Q^*. (Note: You do not have to derive this from the definition of capacity, you can appeal to result discussed in lectures or the textbook).

(f) Describe a way of using the channels Q and Q' to communicate without error.

III Combining Noisy Channels [20 points]

Suppose we have two arbitrary noisy channels Q and Q' where channel Q has inputs $\mathcal{X} = \{a_1, \ldots, a_N\}$ and outputs $\mathcal{Y} = \{b_1, \ldots, b_M\}$ while channel Q' has inputs $\mathcal{X}' = \{a'_1, \ldots, a'_{N'}\}$ and outputs $\mathcal{Y}' = \{b'_1, \ldots, b'_{M'}\}$. We will assume that the inputs of the channels and the outputs do not share any common symbols. That is, $\mathcal{X} \cap \mathcal{X}' = \emptyset$ and $\mathcal{Y} \cap \mathcal{Y}' = \emptyset$.

The *product channel* $Q \otimes Q'$ uses Q and Q' to send and receive pairs of symbols in parallel: its input alphabets is the product alphabet $\mathcal{V} = \mathcal{X} \times \mathcal{X}' = \{(a_1, a'_1), (a_1, a'_2), \ldots, (a_N, a'_{N'})\}$ (with $N \times N'$ elements) and its output alphabet is $\mathcal{W} = \mathcal{Y} \times \mathcal{Y}' = \{(b_1, b'_1), \ldots, (b_M, b'_{M'})\}$ (with $M \times M'$ elements). When $P(y = b | x = a)$ and $P(y' = b' | x' = a')$ are the transition probabilities in Q and Q' respectively, the transition probabilities in $Q \otimes Q'$ are given by their product, that is,

$$P(w = (b, b')|v = (a, a')) = P(y = b|x = a).P(y' = b'|x' = a')$$

for all $(a, a') \in \mathcal{X} \times \mathcal{X}'$ and $(b, b') \in \mathcal{Y} \times \mathcal{Y}'$.

The *sum channel* $Q \oplus Q'$ has input alphabet $\mathcal{V} = \mathcal{X} \cup \mathcal{X}'$ and output alphabet $\mathcal{W} = \mathcal{Y} \cup \mathcal{Y}'$ and sends symbols from $\mathcal{X}$ through Q and symbols from $\mathcal{X}'$ through Q'. The transition probabilities for $Q \oplus Q'$ are therefore

$$P(w|v) = \begin{cases} P(y = b|x = a) & \text{if } w = b \in \mathcal{Y} \text{ and } v = a \in \mathcal{X} \\ P(y' = b'|x' = a') & \text{if } w = b' \in \mathcal{Y}' \text{ and } v = a' \in \mathcal{X}' \\ 0 & \text{otherwise.} \end{cases}$$

1. For input alphabets $\mathcal{X} = \{\texttt{a}, \texttt{b}\}$, and $\mathcal{X}' = \{\texttt{c}, \texttt{d}\}$ and output alphabets $\mathcal{Y} = \{1, 2\}$ and $\mathcal{Y}' = \{3, 4\}$, compute the sum and product channels for

$$Q = \begin{bmatrix} 1 & 1/4 \\ 0 & 3/4 \end{bmatrix}. \quad \text{and} \quad Q' = \begin{bmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}$$

   That is, write down the input and output alphabets and transition matrices for both the sum channel $Q \oplus Q'$ and product channel $Q \otimes Q'$.

2. Prove that for arbitrary channels Q and Q' that the capacity of $Q \otimes Q'$ is the sum of the capacities of Q and Q'. That is, show

$$C(Q \otimes Q') = C(Q) + C(Q').$$

3. (*Challenging!*) Prove that for arbitrary channels Q and Q' that

$$C(Q \oplus Q') = \log_2 \left(2^C + 2^{C'} \right)$$

   where $C = C(Q)$ and $C' = C(Q')$ are the capacities of Q and Q', respectively.