ETH Zürich
Dept. Computer Science

Spring Semester 2015

Information theory

Final exam

13 August 2015

Time limit:	120 minutes
Number of pages:	4
Total points:	100 points

Please remember to try all the questions before the exam ends. If you are stuck in any part of a problem do not dwell on it, try to move on and attempt it later. Collaboration on the exam is strictly forbidden.

[1 point] Please fill in your name and student ID:

1 [25 points] Source Coding with Unknown Probabilities

We would like to compress a discrete memoryless source of alphabet $\{a, b, c, d, 1, 2, 3\}$, using the Huffman coding algorithm. We are not aware of the source probability distribution, but we are told that Latin characters are twice more probable than the numbers.

a) [8 points] Design the best prefix-free code that you can.

b) [7 points] Let us now assume that the true probability distribution (that you are not aware of) is given in Table 1. What is the average codeword length for the code that you designed in part (a)?

c) [10 points] How far are you from the best average code word length (that you could have, had you known the true distribution when designing the code)?

source symbol	probability
a	$\frac{1}{6}$
b	$\frac{1}{6}$
c	$\frac{1}{4}$
d	$\frac{1}{12}$
1	$\frac{1}{16}$
2	$\frac{1}{4}$
3	$\frac{1}{48}$

Table 1: True probability distribution of the source.

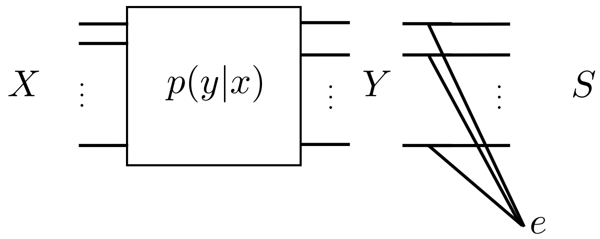

Figure 1: Y goes immediately through an erasure channel with erasure probability α.

2 [40 points] Channels with an Erasure Link

Consider a discrete memoryless channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transition probabilities $p(y|x)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Let C denote the capacity of this channel.

Suppose that this channel is cascaded immediately with an erasure channel, see Figure 1. That is, once the channel outputs a symbol $y \in \mathcal{Y}$, then either y stays the same with probability $1 - \alpha$, or it gets erased with probability α.

Thus, the overall channel has input alphabet $\mathcal{X}$, output alphabet $\mathcal{S} = \mathcal{Y} \cup \{e\}$, and transition probabilities

$$\Pr\{S = y \mid X = x\} = (1 - \alpha)p(y \mid x), \quad \text{for } y \in \mathcal{Y} \text{ and } x \in \mathcal{X},$$
$$\Pr\{S = e \mid X = x\} = \alpha, \quad \text{for } x \in \mathcal{X}.$$

Let the capacity of the overall channel (from $\mathcal{X}$ and $\mathcal{S}$) be denoted by C_{cascade}. In the first part of this problem we will compute C_{cascade} in terms of C.

a) [4 points] Show that the following holds no matter what the distribution on X is:

$$\Pr\{S = y\} = (1 - \alpha)\Pr(Y = y), \quad \text{for } y \in \mathcal{Y},$$
$$\Pr\{S = e\} = \alpha.$$

b) [4 points] Show that

$$H(S) = h_2(\alpha) + (1 - \alpha)H(Y),$$

where we define $h_2(\alpha) = -\alpha \log_2 \alpha - (1 - \alpha) \log_2(1 - \alpha)$.

c) [2 points] Show that

$$H(S|X) = h_2(\alpha) + (1 - \alpha)H(Y|X).$$

d) [4 points] Deduce from the above steps that

$$C_{\text{cascade}} = (1 - \alpha)C.$$

In the second part of this problem, we consider the channel in Figure 2. Let us denote its capacity by C'.

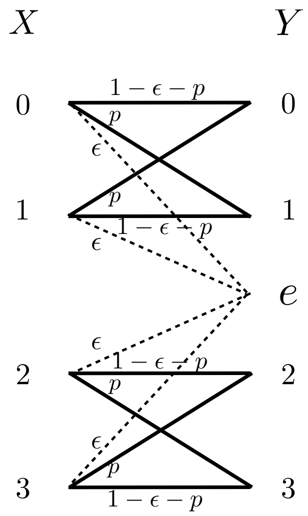

Figure 2: The channel for the second part of problem 2.

e) [10 points] Compute C'. Hint: You may use the result of part (d).

f) [8 points] Suppose that the input alphabet is equiprobable (i.e. $\Pr\{X = i\} = 1/4$ for $i = 0, 1, 2, 3$). We want to guess the input to the channel from the observed output. Design a guessing strategy that has the minimum possible probability of error. Describe your strategy in detail.

g) [8 points] Let p_e be the error probability of the guessing function you designed in part (f). Show that

$$p_e \geq \frac{1 - C'}{2}.$$

Hint: In class, we talked about an inequality that relates error probability to entropy.

3 [35 points] Plutkin Construction: $u, u \oplus v$ Codes

Suppose $\mathcal{C}_1$ and $\mathcal{C}_2$ are binary linear codes of blocklength n, that is $\mathcal{C}_i \subseteq \{0,1\}^n$ for $i = 1, 2$. Denote the number of codewords of $\mathcal{C}_i$ by M_i and the minimum distance of $\mathcal{C}_i$ by d_i. . For $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$ let $\langle \mathbf{u}|\mathbf{v} \rangle$ denote the concatenation of the two sequences, i.e.,

$$\langle \mathbf{u}|\mathbf{v} \rangle = (u_1, \ldots, u_n, v_1, \ldots, v_n).$$

Let $\mathcal{C}$ denote the binary code of blocklength $2n$ obtained from $\mathcal{C}_1$ and $\mathcal{C}_2$ as follows:

$$\mathcal{C} = \{\langle \mathbf{u}|\mathbf{u} \oplus \mathbf{v} \rangle : \mathbf{u} \in \mathcal{C}_1, \mathbf{v} \in \mathcal{C}_2\}.$$

a) [5 points] Is $\mathcal{C}$ a linear code?

b) [10 points] How many codewords does $\mathcal{C}$ have? Carefully justify your answer. What is the rate R of $\mathcal{C}$ in terms of the rates R_1 and R_2 of the codes $\mathcal{C}_1$ and $\mathcal{C}_2$?

c) [8 points] Show that the Hamming weight of $\langle \mathbf{u}|\mathbf{u} \oplus \mathbf{v} \rangle$ satisfies

$$w_H(\langle \mathbf{u}|\mathbf{u} \oplus \mathbf{v} \rangle) \geq w_H(\mathbf{v}).$$

d) [2 points] Show that the Hamming weight of $\langle \mathbf{u}|\mathbf{u} \oplus \mathbf{v} \rangle$ satisfies

$$w_H(\langle \mathbf{u}|\mathbf{u} \oplus \mathbf{v} \rangle) \geq \begin{cases} w_H(\mathbf{v}) & \text{if } \mathbf{v} \neq 0 \\ 2w_H(\mathbf{u}) & \text{else.} \end{cases}$$

e) [5 points] Show that the minimum distance d of $\mathcal{C}$ satisfies

$$d \geq \min\{2d_1, d_2\}.$$

f) [5 points] Show that $d = \min\{2d_1, d_2\}$.