University of Toronto
November 8, 2000

Department of Electrical
& Computer Engineering

ECE1502S — Information Theory
Midterm Test

Instructions

- You have one hour and fifty minutes of "in-class" time, followed by five days of "take-home" time to complete this test.
- Complete as much as possible during the in-class time; your grade will be computed as a weighted average of your "in-class" grade and your "take-home" grade. (Weights to be determined later.)
- Answer **all** four [4] questions. All questions have equal value. Show all steps and present all results clearly.
- Take-home due date: Monday, November 13, 2000, 1:00 pm (in class), or earlier to the instructor's office, SY505. Please return a *complete* solution for the take home portion, even if you believe that you answered the question correctly in class.
- All aids are permitted during the take-home portion, but **all work is to be done independently**. (No mutual information!) Consultation with others is **not** permitted.
- Good luck!

1. (*Matching Distributions*) Let $P = \{p_1 \leq p_2 \leq \ldots \leq p_m\}$ be a probability distribution on m values, with the probability masses p_i sorted in non-decreasing order, and let Q be a set of m non-negative numbers that sum to unity, i.e., another distribution on m values.

   (a) Assuming all elements of Q are positive, show that the relative entropy $D(P||Q)$ is minimized when the probability masses of Q are also sorted into non-decreasing order, i.e., when $q_1 \leq q_2 \leq \ldots \leq q_m$.

   (b) Does this ordering for Q also minimize $D(Q||P)$? (Prove or give a counterexample.)

   (c) Find the distribution Q on $m-1$ values that minimizes $D(Q||P)$. (A distribution on $m-1$ values can be considered to be a distribution on m values with one probability mass constrained to be zero.)

2. (*Huffman Coding with Costs*) Given an alphabet $S = \{1, 2, \ldots, m\}$, and a set of positive weights $\{w_1, w_2, \ldots, w_m\}$, a binary prefix code with codeword lengths $l_1, \ldots, l_m$ is to be designed to minimize $\sum w_i l_i$.

   (a) Prove that (a trivial modification of) the Huffman procedure will still yield an optimal code, even though, in general, $\sum w_i \neq 1$.

   (b) In some situations, it may be costly to assign long codewords to certain messages (like Run! or Fire!); even though these message may occur infrequently.

   Suppose that $X = i$ with probability p_i, $i = 1, 2, \ldots, m$. Let l_i be the number of binary symbols in the codeword associated with $X = i$ and let c_i denote the cost per letter of the codeword when $X = i$. Thus the average cost C of the description of X is $C = \sum p_i c_i l_i$.

      i. Minimize C over all $l_1, l_2, \ldots, l_m$ such that $\sum 2^{-l_i} = 1$, ignoring any implied integer constraints on l_i. What are the minimizing l_i^* and what is the associated minimum average cost C^*?

      ii. How would you use the result of part (a) to minimize C over all uniquely decodable codes? Let C_{Huffman} denote this minimum.

      iii. Show that

$$C^* \leq C_{\text{Huffman}} < C^* + \sum_{i=1}^{m} p_i c_i.$$

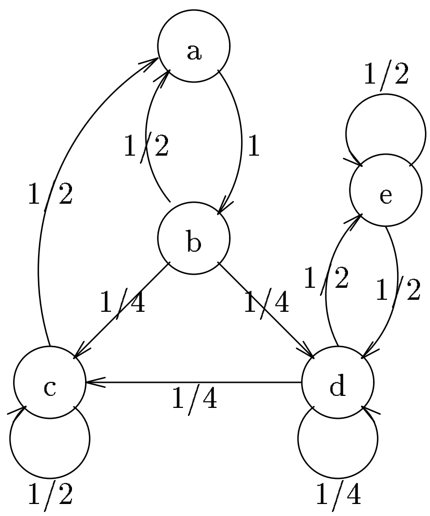

Figure 1: A five-state Markov chain, labeled with transition probabilities.

3. (*Coding a Markov Process*) Fig. 1 shows a five-state Markov chain, labeled with transition probabilities.

    (a) Write the transition matrix, and determine the stationary distribution for this Markov chain.

    (b) Determine the entropy rate for this Markov chain.

    (c) Using a binary representation alphabet, and assuming the initial state X_0 is chosen according to the stationary distribution, describe an efficient coding scheme for this Markov chain. Describe, explicitly, both the encoding algorithm and the decoding algorithm. Why do you consider your algorithm to be efficient?

    (d) Use your algorithm to encode the sequence bcababdedcc.

4. (*Diversity channels*) A binary symmetric channel (BSC) with crossover probability p is shown in Fig. 2(a). Let X denote the channel input, and Y the channel output.

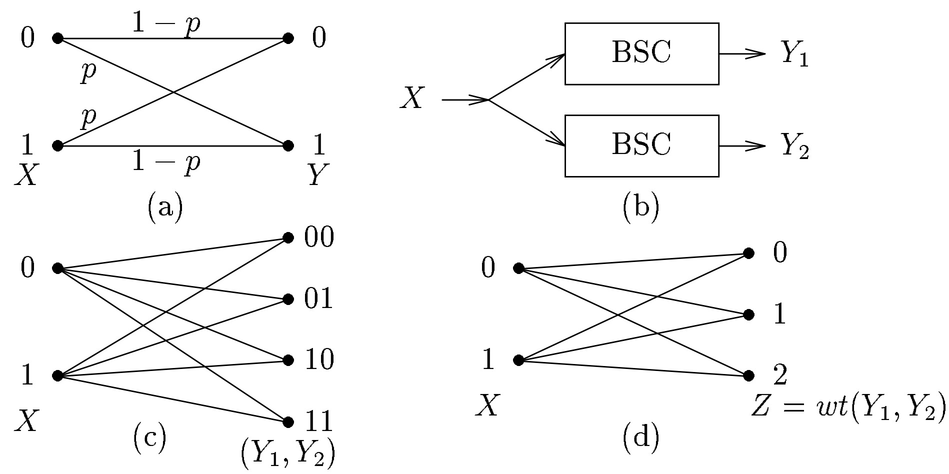

Figure 2: (a) A binary symmetric channel; (b) second-order diversity channe l; (c) equivalent channel; (d) proposed "combined" channel.

(a) Let $P_X(0) = q$. Find $I(X;Y)$ in terms of p and q. Which value of q maximizes $I(X;Y)$?

(b) In a diversity transmission system, the channel input X is transmitted simultaneously over several independent channels (called a diversity path) to the receiver. For example, two independent, identical binary symmetric channels (each with crossover probability p) provide a second-order diversity system, as shown in Fig. 2(b). Assuming $P_X(0) = 1/2$, find $I(X;(Y_1,Y_2))$. (Hint: the second-order diversity system is equivalent to the 2-input, 4-output channel shown in Fig. 2(c), for appropriate choice of crossover probabilities.)

(c) An Lth-order diversity system with channel input X uses L parallel independent binary symmetric channels, each with crossover probability p. Let Y_i be the output of the ith such channel. Assuming $P_X(0) = 1/2$, find an expression for $I(X;(Y_1,Y_2,\ldots,Y_L))$. What is $I(X;(Y_1,Y_2,\ldots,Y_L))$ when $p = 1/2$?

(d) It is proposed to combine the L channel outputs in the Lth-order diversity system of part (c) into a single variable Z, where $Z = wt(Y_1,Y_2,\ldots,Y_L)$ is the Hamming weight[1] of the vector $(Y_1,Y_2,\ldots,Y_L)$ of channel outputs. (When $L = 2$, for example, the 2-input, 3-output channel shown in Fig. 2(d) is obtained, for appropriate choice of crossover probabilities.) Compare $I(X;Z)$ with $I(X;(Y_1,\ldots,Y_L))$, and comment on the significance of your answer. Would your conclusion be different if the independent diversity paths did not all have the same cross-over probability?

[1]The Hamming weight of a binary L-tuple is defined as the number of nonzero components of that L-tuple. For example, when $L = 3$, we have $wt(000) = 0$, $wt(001) = wt(010) = wt(100) = 1$, $wt(011) = wt(101) = wt(110) = 2$, $wt(111) = 3$.