University of Toronto
December 14, 2021

ECE1502F — Information Theory
Final Exam

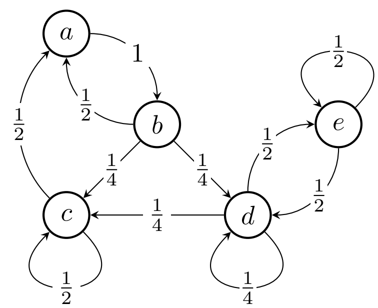
Department of Electrical
& Computer Engineering

Instructions

- You have one hour and fifty minutes of "in-class time" followed by three days of "take-home" time to complete this test.
- The exam consists of five questions, each of equal value.
- You may choose any **four of five** problems to complete during the in-class time. Please indicate clearly which problems you wish to have graded. You must answer **all five** questions during the take-home time.
- Take-home due date: Friday, December 17, 2021, 1:30pm either in person to the instructor (BA4132) or by Quercus PDF upload. Please return a *complete solution* for the take-home portion, even if you believe that you answered the question correctly in class.
- Your grade will be computed as a weighted average of your "in-class" grade and your "take-home" grade. (Weights to be determined later.)
- The exam is open book, and all aids are permitted, but **all work is to be done independently**. Consultation with others is **not** permitted. Good luck!

---

1. (*A Markov Source*) A source produces sequences of letters from the alphabet $\mathcal{X} = \{a, b, c, d, e\}$ according to the five-state Markov chain shown in the figure below. Transition probabilities between states are indicated in the figure.



   (a) Write the transition matrix, and determine the stationary distribution for this Markov chain. *Hint:* make use of Problem 4.1 from Problem Set 3.

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

   (c) Using a binary representation alphabet, and assuming the initial state X_1 is chosen according to the stationary distribution, describe an efficient coding scheme for this Markov source. Explicitly describe both the encoding algorithm and the decoding algorithm. Why do you consider your coding scheme to be efficient?

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

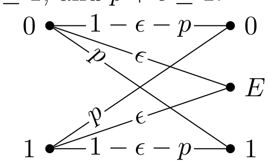
2. (*Maximum Entropy*)

   (a) Prove that if X is a random variable taking values in the natural numbers $\{0, 1, 2, \ldots, \}$ and having a mean $m > 0$, then $H(X) \le (m+1)\log(m+1) - m\log(m)$, with equality achieved if and only if $P(X = i) = \left(\frac{1}{m+1}\right)\left(\frac{m}{m+1}\right)^i$, i.e., when X is geometrically distributed with a mean of m.

   (b) Prove that, among all probability density functions defined on the interval $[a, b]$, where $a < b$, the uniform distribution has the largest differential entropy.

   (c) Let $X_1, \ldots, X_n$ be independent and uniformly distributed on $[a, b]$. Provide a geometrical description of the set $A_\epsilon^{(n)} = \left\{(x_1, \ldots, x_n) \in [a, b]^n \colon \left|-\frac{1}{n}\log f(x_1, \ldots, x_n) - h(X_1)\right| \le \epsilon\right\}$ of typical sequences, where f is the joint probability density function of $X_1, \ldots, X_n$.

3. (*Errors-and-Erasures Channel*) A binary "errors-and-erasures" channel with parameters p and ϵ is shown below. The input alphabet is $\mathcal{X} = \{0, 1\}$ and the output alphabet is $\mathcal{Y} = \{0, 1, E\}$. The channel parameters satisfy $0 \le p \le 1$, $0 \le \epsilon \le 1$, and $p + \epsilon \le 1$.



   (a) Find the capacity C for this channel in terms of p and ϵ.

   (b) Specialize to the binary erasure channel ($p = 0$).

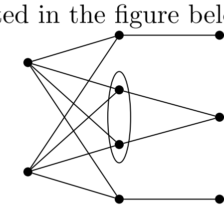
   (c) Specialize to the binary symmetric channel ($\epsilon = 0$).

   (d) Explain how a code designed for a particular binary symmetric channel can be adapted (with the help of a noiseless feedback channel between the transmitter and receiver) for use on this binary errors-and-erasures channel.

   (e) Find the capacity C of the q-ary errors-and-erasures channel. Here $q \ge 2$ is an integer, the input alphabet is $\mathcal{X} = \{0, 1, \ldots, q-1\}$, the output alphabet is $\mathcal{Y} = \mathcal{X} \cup \{E\}$, and the channel law is given as

$$p(y \mid x) = \begin{cases} \epsilon & y = E, \\ 1 - \epsilon - p & y = x, \\ p/(q-1) & \text{otherwise.} \end{cases}$$

4. (*Channel Reduction*) When output letters of a discrete memoryless channel are combined into a single letter, a *reduced channel* is said to result. Reduction is equivalent to processing the channel output by a "deterministic channel," as illustrated in the figure below.



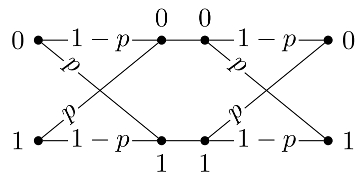
   (a) By the data processing inequality, reduction clearly cannot increase channel capacity, and may reduce it. Prove that, if the columns of the channel matrix that correspond to the letters to be combined into a single letter are *proportional* to one another, then capacity is not reduced. *Hint*: without loss of generality, consider a channel transition matrix whose first two columns are $(p_{11}, p_{21}, \ldots, p_{M1})^T$ and $(kp_{11}, kp_{21}, \ldots, kp_{M1})^T$, respectively.

   (b) Use this result to find the capacity of the discrete memoryless channel with transition matrix

$$\begin{bmatrix} \frac{5}{32} & \frac{3}{8} & \frac{5}{32} & \frac{5}{16} \\ \frac{7}{32} & \frac{1}{8} & \frac{7}{32} & \frac{7}{16} \end{bmatrix}.$$

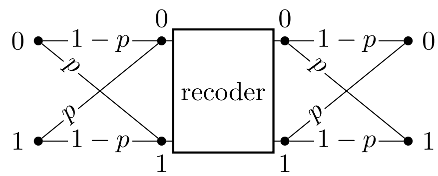
5. (*Composite Channels*) Suppose that two channels are connected end-to-end to form a composite channel. Each of the original channels is a binary symmetric channel with crossover probability p.

   (a) Suppose that the output of the first channel is connected directly to the input of the second, with no processing between:

   What is the capacity C_{cascade} of the composite channel?

   (b) Suppose that a decoder/encoder (a recoder) is allowed between the channels:

   What is the capacity, C_{recode}, under this arrangement?

   (c) Starting from the fact that all rates less than capacity are achievable for each of the two sub-channels, prove that all rates less than C_{recode} are achievable for the channel of part (b).

   (d) Suppose that the two sub-channels have (possibly different) crossover probabilities p_1 and p_2, respectively. Now what is the capacity C_{recode}?