University of Toronto
October 26, 2021

ECE1502F — Information Theory
Midterm Test

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, October 29, 2021, 1:30pm (in class). 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. Let X be a memoryless Bernoulli source over $\mathcal{X} = \{0, 1\}$, where $p_X(0) = 2/3$ and $p_X(1) = 1/3$.
   (a) Compute the entropy of this source.
   (b) Determine the rate of a binary Huffman code (with 8 codewords, possibly of variable length) for the third extension of this source.
   (c) Determine the rate of a binary Tunstall code (with 8 codewords of fixed length) for this source.
   (d) Determine the probability of each of the eight phrases produced by the Tunstall parsing tree of (c). Is the fixed length code optimal for this distribution?

2. Let X be a discrete memoryless source with alphabet $\mathcal{X}$, and entropy $H(X)$ in bits. For a positive integer n and any $\epsilon > 0$, let

$$A_\epsilon^{(n)} = \left\{ (x_1, \ldots, x_n) \in \mathcal{X}^n : \left| -\frac{1}{n} \log p(x_1, \ldots, x_n) - H(X) \right| \le \epsilon \right\}$$

   be the typical set as defined in class, where $p(x_1, \ldots, x_n) = \prod_{i=1}^n p(x_i)$. For $0 < \delta \le 1$, let

$$B_\delta^{(n)} = \left\{ (x_1, \ldots, x_n) \in \mathcal{X}^n : \left| -\frac{1}{n} \log p(x_1, \ldots, x_n) - H(X) \right| \le \delta H(X) \right\}.$$

   Prove the following.
   (a) If $(x_1, \ldots, x_n) \in B_\delta^{(n)}$, then

$$2^{-n(1+\delta)H(X)} \le p(x_1, \ldots, x_n) \le 2^{-n(1-\delta)H(X)}.$$

   (b) $\Pr\left(B_\delta^{(n)}\right) \ge 1 - \delta$ when n is sufficiently large.
   (c) $|B_\delta^{(n)}| \le 2^{n(1+\delta)H(X)}$.
   (d) $|B_\delta^{(n)}| \ge (1 - \delta)2^{n(1-\delta)H(X)}$ when n is sufficiently large.
   (e) Show that either $A_\epsilon^{(n)} \subseteq B_\delta^{(n)}$ or $B_\delta^{(n)} \subseteq A_\epsilon^{(n)}$. Under what conditions are the two sets equal?

3. Let X be a random variable with range $\mathcal{X} = \{1, 2, 3, \ldots\}$ and expected value $E(X)$.
   (a) Show that $H(X) \leq E(X)$ when $H(X)$ is measured in bits.
   (b) When does equality occur?

   Now let Y be a random variable with range $\mathcal{Y} = \{0, 1, 2, 3, \ldots\}$ and expected value $E(Y)$.
   (c) Show that $H(Y) \leq 1 + E(Y)$ when $H(X)$ is measured in bits.
   (d) When does equality occur?

4. Random variables X and Y are jointly distributed over the finite alphabet $\mathcal{X} \times \mathcal{Y}$ according to the joint probability mass function $p_{X,Y}(x, y) = p_X(x)p_{Y|X}(y \mid x)$, where $p_X(x)$ denotes, for any $x \in \mathcal{X}$, the marginal probability mass function for X, and $p_{Y|X}(y \mid x)$ denotes, for any $x \in \mathcal{X}$ and any $y \in \mathcal{Y}$, the conditional probability mass function of Y given that $X = x$.

   The mutual information $I(X; Y)$ between X and Y, which is a function of p_X and $P_{Y|X}$, will be denoted as $I(X; Y) = I(p_X; p_{Y|X})$.
   (a) Show that $I(p_X; p_{Y|X})$ is **concave** in p_X for fixed $p_{Y|X}$, i.e., for any $\lambda \in [0, 1]$, and any probability mass functions p'_X and p''_X, the mutual information satisfies
   $$I(p^*_X; p_{Y|X}) \geq \lambda I(p'_X; p_{Y|X}) + (1 - \lambda)I(p''_X; p_{Y|X}),$$
   where $p^*_X = \lambda p'_X + (1 - \lambda)p''_X$. *Hint:* The inequality $A \geq B$ is equivalent to $A - B \geq 0$.
   (b) Show that $I(p_X; p_{Y|X})$ is **convex** in $p_{Y|X}$ for fixed p_X, i.e., for any $\lambda \in [0, 1]$, and any conditional probability mass functions $p'_{Y|X}$ and $p''_{Y|X}$, the mutual information satisfies
   $$I(p_X; p^*_{Y|X}) \leq \lambda I(p_X; p'_{Y|X}) + (1 - \lambda)I(p_X; p''_{Y|X}),$$
   where $p^*_{Y|X} = \lambda p'_{Y|X} + (1 - \lambda)p''_{Y|X}$. *Hint:* The inequality $A \leq B$ is equivalent to $B - A \geq 0$.

5. In a coin-tossing game, a player must try to guess the outcome of a fair coin toss. The guess must be made while the coin is in the air, i.e., after it is flipped, but before it is caught. A one dollar bet returns one additional dollar if the guess is correct. In other words, if C_0 denotes a player's total capital *before* the bet, and a player bets B dollars, then the player's total capital C_1 *after* the bet is given by
   $$C_1 = \begin{cases} C_0 + B & \text{if the guess is correct;} \\ C_0 - B & \text{if the guess is incorrect.} \end{cases}$$
   After years of patient observation, a player has learned to partially predict the outcome of the coin toss. By using this information, they hope to gain a tidy sum for their retirement.

   Let X denote the outcome of a given coin toss, and let Y denote the player's prediction. Assume that $P[X = Y] = p > 1/2$, and that successive coin tosses are independent.
   (a) Calculate $I(X; Y)$.
   (b) The player has an initial capital C_0. They have decided to wager a fraction q of their total capital on each successive coin toss, and withhold $1 - q$. After N trials, show that their total capital is the random variable
   $$C_N = C_0 \prod_{n=1}^{N} (1 + q)^{Z_n} (1 - q)^{1 - Z_n}$$
   where $Z_n = 1$ if the nth prediction is correct and 0 otherwise. Define the *growth rate* as
   $$R_N = \frac{1}{N} \log_2 \frac{C_N}{C_0}.$$
   Find the value of q that maximizes the expected growth rate $E(R_N)$. Also find the value of q that maximizes $E(C_N)$. Compare the maximum expected growth rate with $I(X; Y)$.
   (c) If you were the player, which value of q would you use and why? *Hint:* does the law of large numbers apply to either R_N or C_N as $N \to \infty$?