University of Toronto
October 24, 2023

ECE1502F — Information Theory
Midterm Test

Department of Electrical
& Computer Engineering

Instructions

- You have one hour and fifty minutes of "in-class time" followed by two days of "take-home" time to complete this test.
- The exam consists of four questions, each of equal value.
- Take-home due date: Thursday, October 26, 2023, 1:10pm (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 and Y be jointly distributed random variables taking values in the set $\{1, 2, 3, 4, 5\}$ according to the joint probability mass function $\Pr(X = i, Y = j) = p_{i,j}$, where $p_{i,j}$ is the entry in the ith row and jth column of the matrix

$$P = [p_{i,j}] = \frac{1}{32}\begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 4 & 1 & 2 & 0 & 1 \\ 4 & 2 & 4 & 2 & 4 \end{bmatrix}.$$

   (a) Determine the following quantities.

   2 marks i. $H(X)$.

   2 marks ii. $H(Y \mid X)$.

   2 marks iii. $I(X;Y)$.

   2 marks iv. $D(p_X || p_Y)$, where p_X and p_Y denote the marginal probability mass functions for X and Y, respectively.

   8 marks (b) Design a binary Huffman code for X, determine its average codeword length, and compare this with $H(X)$.

   9 marks (c) Suppose you use your code to describe X. Design a binary encoding scheme for Y that minimizes the expected number of *additional* bits required to describe Y given that X has already been encoded. Determine the average codeword length of the overall coding scheme, and compare this with $H(X, Y)$.

2. A probability mass function defined on a finite alphabet $\mathcal{X}$ is called "deterministic" if $p(x) = 1$ for some $x \in \mathcal{X}$. Let X and Y be jointly distributed random variables taking values in $\mathcal{X}$ and $\mathcal{Y}$, respectively. The random variable Y is said to be "a function of X with probability one" if, for all x with $\Pr(X = x) \neq 0$, the conditional probability mass function $\Pr(Y = y \mid X = x)$ is deterministic.

   4 marks (a) Show that $H(X) = 0$ if and only if X is deterministic.

   4 marks (b) Show that $H(Y \mid X) = 0$ if and only if Y is a function of X with probability one.

   7 marks (c) Let f be any function with domain $\mathcal{X}$. Show that $H(X) \geq H(f(X))$. When is equality achieved?

   10 marks (d) Let X and Y be independent random variables defined over the alphabet $\{0, 1, \ldots, m - 1\}$. Let $X \oplus Y$ denote the sum of X and Y reduced modulo m. Show that $H(X) \leq H(X \oplus Y) \leq H(X) + H(Y)$. Give (separate) examples, with $H(X) > 0$ and $H(Y) > 0$, (i) where $H(X) = H(X \oplus Y)$ and (ii) where $H(X \oplus Y) = H(X) + H(Y)$.

3. Let X be a random variable with alphabet $\mathcal{X} = \{1, 2, \ldots, 10\}$, and let p_X denote the probability mass function of X.

4 marks (a) Show that $1 \le E(X) \le 10$ and give conditions for when the two inequalities are achieved with equality.

2 marks (b) Which probability mass function for X has maximum entropy? What is $E(X)$ for the maximum entropy distribution?

4 marks (c) Let

$$q_X^{(s)}(x) = \frac{e^{sx}}{Z(s)}$$

be a family of probability mass functions indexed by the parameter $s \in \mathbb{R}$, where

$$Z(s) = \sum_{x \in \mathcal{X}} e^{sx}$$

is a normalizing constant. Suppose that X is distributed according to $q_X^{(s)}$. Denote $E(X)$ as $\mu(s)$. Determine

$$\lim_{s \to -\infty} \mu(s) \text{ and } \lim_{s \to \infty} \mu(s).$$

It can be shown that $\mu(s)$ is a continuous increasing function of s, and therefore $\mu(s)$ achieves every possible value between the two limits computed.

5 marks (d) Suppose that X is distributed according to $q_X^{(s)}$. Show that

$$H(X) = \ln Z(s) - s\mu(s),$$

when entropy is measured in nats.

10 marks (e) Again, let X be distributed according to $q_X^{(s)}$, and denote $H(X)$ as H_q. Let $p_X(x)$ be any probability mass function satisfying $\sum_{x \in X} x p_X(x) = \mu(s)$, and let $H_p = -\sum_{x \in X} p_X(x) \ln p_X(x)$ denote the corresponding entropy. Show that $H_q \ge H_p$.

4. Consider a memoryless Bernoulli source X producing sequences of letters drawn from the alphabet $\mathcal{X} = \{0, 1\}$, with $\Pr(X = 1) = p$, where $0 < p < 1$. Let $\mathcal{X}^n$ denote the set of binary n-tuples.

For any binary n-tuple $x \in \mathcal{X}^n$, let $w(x)$ denote the number of components of x that are nonzero. For example $w((0,0,0)) = 0$, $w((0,1,0)) = 1$, $w((1,0,1)) = 2$, etc. Clearly $0 \le w(x) \le n$ for every $x \in \mathcal{X}^n$. For $0 \le i \le n$, let $W_i^{(n)} = \{x \in \mathcal{X}^n : w(x) = i\}$.

Fix a positive integer n. For any nonnegative real number δ, let

$$T_\delta^{(n)} = \left\{ x \in \mathcal{X}^n : \left| \frac{w(x)}{n} - p \right| \le \delta \right\}.$$

As usual, for any nonnegative real number ϵ, let

$$A_\epsilon^{(n)} = \left\{ x \in \mathcal{X}^n : \left| -\frac{1}{n} \log \Pr((X_1, \ldots, X_n) = x) - H(X) \right| \le \epsilon \right\}$$

denote the set of typical sequences of length n produced by X.

3 marks (a) Show that if $W_i^{(n)} \cap T_\delta^{(n)} \ne \emptyset$, then $W_i^{(n)} \subseteq T_\delta^{(n)}$.

5 marks (b) Show that if $W_i^{(n)} \cap A_\epsilon^{(n)} \ne \emptyset$, then $W_i^{(n)} \subseteq A_\epsilon^{(n)}$.

10 marks (c) For a fixed ϵ determine $\{i : W_i^{(n)} \subseteq A_\epsilon^{(n)}\}$.

3 marks (d) Fix $\epsilon \ge 0$. Show that there exists a value of δ such that $T_\delta^{(n)} \subseteq A_\epsilon^{(n)}$.

4 marks (e) Fix $\delta \ge 0$. Find a specific example (i.e., a specific choice of n and p) where the reverse containment is not possible, i.e., where, in your example, there is no choice of ϵ for which $A_\epsilon^{(n)} \subseteq T_\delta^{(n)}$.