EE 276 - Information Theory
Final
March 19, 2024

1. There are a total of 5 questions. You have 3 hours to take the exam. Questions vary in difficulty and number of points. There are a total of 100 points.
2. Please write all answers in the designated area underneath the question. If you need more room for your answer, please indicate that under the question, and continue your response elsewhere.
3. Scratch paper will be provided and collected at the end of the exam, but will **not** be graded.
4. Answers should be justified, unless otherwise stated.
5. You are allowed to use non-electronic notes and material.
6. Calculators or any other electronic devices are not allowed.

Good luck!

Name:

SUID:

1. (*14 points*) Answer true or false to each of the following. You do not need to provide an explanation. A correct answer gets 2 points. An incorrect answer gets 0 points. Leaving the answer blank gets 1 point.

T F

(a) Shannon and Huffman coding both assume knowledge of the source distribution. ☐ ☐

(b) The discrete entropy of a random variable is invariant under one-to-one transformations of the random variable but the differential entropy is generally not. ☐ ☐

(c) If the mutual information $I(X;Y) = 0$, then for any Z we have $I(X;Y|Z) = 0$. ☐ ☐

(d) The noisy channel coding theorem guarantees that we can achieve a probability of error 0 for any block length n, as long as the rate is below the capacity of the channel. ☐ ☐

(e) Relative entropy between two distributions satisfies $D(p\|q) = D(q\|p)$. ☐ ☐

(f) The entropy rate of any stationary process $X_1, X_2, \ldots$ is never larger than that of the i.i.d. process $Y_1, Y_2, \ldots$, if X_1 and Y_1 have the same distribution. ☐ ☐

(g) Recall $\mathbb{P}^n$ denotes the set of PMFs that are empirical distributions of sequences of length n with components in a given finite alphabet. If $p, q \in \mathbb{P}^n$, and $p \neq q$, then $T(p) \cap T(q) = \emptyset$, where $T(p)$ is the type class of p. ☐ ☐

2. (*17 points*) Consider a PMF p over an alphabet $\mathcal{X}$ of size 4 given by the following table

x	a	b	c	d
$p(x)$	1/2	1/4	1/8	1/8

(a) Suppose random variable X is drawn from p but not observed. To discover the value of X, you're allowed to ask a sequence of 'yes/no' questions that will be truthfully answered. What is the minimum expected number of such questions that you could ask in order to discover the selected symbol? What is the sequence of questions achieving the minimum? Explain your reasoning.

(b) What is the entropy of X from the previous part?

A code $C(\cdot)$ for the alphabet $\mathcal{X}$ is *non-singular* if for all $x_1, x_2 \in \mathcal{X}$,

$$x_1 \neq x_2 \quad \text{implies} \quad C(x_1) \neq C(x_2).$$

In words, it doesn't assign the same codeword to two different alphabet symbols.

(c) For the alphabet $\mathcal{X}$ above, provide an example of a non-singular code which is not uniquely decodable. Explain why your answer satisfies the requirements.

(d) Construct a code which is uniquely decodable, but not a prefix code. Explain why it satisfies these requirements.

(e) Construct a uniquely decodable prefix-free code which is optimal in the sense of expected code length for the source above. Explain why it satisfies these requirements.

(f) Relate your prefix code to the 'yes/no' questioning scheme you proposed in (a).

3. (*14 points*) Let x^n be a sequence of n elements each taking values in a finite alphabet $\mathcal{X}$. Let $\epsilon > 0$, and p, q be two pmfs on $\mathcal{X}$. We say the sequence x^n is "*divergence ϵ-typical for p relative to q*" if

$$\left| \frac{1}{n} \log \frac{p(x^n)}{q(x^n)} - D(p\|q) \right| \leq \epsilon.$$

We'll use $\mathcal{A}_\epsilon^{(n)}(p|q)$ to denote the set of all such sequences.

(a) Find

$$\lim_{n\to\infty} P(\mathcal{A}_\epsilon^{(n)}(p|q)),$$

where $P(\mathcal{A})$ denotes the probability of the event $\{X^n \in \mathcal{A}\}$ when X_i are drawn IID from p.

(b) Show that

$$(1-\epsilon)2^{-n(D(p\|q)+\epsilon)} \leq Q(\mathcal{A}_\epsilon^{(n)}(p|q)) \leq 2^{-n(D(p\|q)-\epsilon)}$$

for n sufficiently large, where $Q(\mathcal{A})$ denotes the probability of the event $\{X^n \in \mathcal{A}\}$ when X_i are drawn IID from q.

4. (*35 points*) For $0 \le q \le 1$, let

$$\phi(q) = \max H(W),$$

where the maximization is over all probability mass functions of the random variable W with alphabet $\mathcal{W} = \{0, 1, 2\}$ and satisfying $P(W \neq 0) \le q$.

(a) Evaluate and qualitatively plot the function $\phi(q)$. Hint: for $0 \le q \le 2/3$ show that the maximum is achieved by the random variable W_q distributed as

$$W_q = \begin{cases} 0 & w.p. \;\; 1-q \\ 1 & w.p. \;\; q/2 \\ 2 & w.p. \;\; q/2 \end{cases} . \tag{1}$$

(b) Consider the rate distortion function $R(D)$ of a ternary memoryless source U under Hamming distortion, i.e. $\mathcal{U} = \mathcal{V} = \{0, 1, 2\}$ and

$$d(u, v) = \begin{cases} 0 & if \;\; u = v \\ 1 & if \;\; u \neq v \end{cases}$$

for all $u \in \mathcal{U}, v \in \mathcal{V}$. Show that for $D \ge 0$:

$$R(D) \ge H(U) - \phi(D).$$

(c) Show that in the setting of the previous part, when U is the uniform ternary source

$$P(U = 0) = P(U = 1) = P(U = 2) = 1/3, \tag{2}$$

the lower bound is achieved with equality, i.e. $R(D) = \log 3 - \phi(D)$ for $D \geq 0$.

(d) Recall the random variable W_q defined in (1). Consider now communication over a memoryless ternary-input ternary-output channel with mod-3 additive noise distributed as W_q, for some $0 \leq q \leq 2/3$. I.e., $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$, and the relationship between the channel input and output is given by

$$Y = (X + W_q) \mod 3, \tag{3}$$

where X and W_q are independent. Show that the capacity of this channel is $C = \log 3 - \phi(q)$.

(e) Consider now a joint-source-channel-coding (JSCC) scenario where the uniform source defined in (2) is to be communicated over the channel defined in (3). What is the maximum achievable communication rate if the source is to be communicated losslessly?

(f) Consider again the JSCC scenario as in the previous part. What is the minimum achievable distortion for communication at rate $\rho = 1$ source symbols per channel use?

(g) Suggest a concrete simple scheme for the setting of the previous part which achieves the optimum performance, that is, communicates at a rate of 1 source symbols per channel use and attains the minimum achievable distortion you found in the previous part. Explain your reasoning.

5. (*20 points*)

(a) Express the set

$$\left\{ u^n \in \{0,1\}^n : \frac{1}{n} \sum_{i=1}^{n} u_i \geq \gamma \right\}$$

as a union of types.

(b) For $U_i \overset{i.i.d}{\sim} \texttt{Ber}(p)$, show that

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} U_i \geq \gamma \right) \doteq 2^{-nD\left(\texttt{Ber}(\gamma) \middle\| \texttt{Ber}(p)\right)} \quad \text{for any} \quad \gamma \in [p, 1].$$

Hint: note that for any $\gamma \in [p, 1]$

$$\min_{q \in [\gamma, 1]} D\left(\texttt{Ber}(q) \middle\| \texttt{Ber}(p)\right) = D\left(\texttt{Ber}(\gamma) \middle\| \texttt{Ber}(p)\right).$$

Now consider communicating one bit $X \sim \texttt{Ber}(1/2)$ via n uses of a BSC(p) with a repetition code. The ith channel output is thus given by

$$Y_i = X \oplus Z_i,$$

where $\{Z_i\}_{i\geq 1}$ are IID $\sim \texttt{Ber}(p)$, independent of X. Assume in what follows that $p < \frac{1}{2}$.

(c) What is the optimal decoding rule in the sense of minimizing the probability of error?

(d) Let $P_e(n,p)$ be the probability of error of the optimal decoder from the previous part. Find the exponential decay rate of $P_e(n,p)$, i.e., what is

$$\lim_{n\to\infty} -\frac{1}{n}\log P_e(n,p)?$$