EE376A: Midterm

Instructions:

- You have **two hours**, 7PM - 9PM
- The exam has 3 questions, totaling 100 points.
- Please start answering each question on a new page of the answer booklet.
- You are allowed to carry textbook, your own notes and other course related material with you. Electronic reading devices [including kindles, laptops, ipads, etc.] are allowed, provided they are used solely for reading pdf files already stored on them and not for any other form of communication or information retrieval..
- You are required to provide a detailed explanation of how you arrived at your answers.
- You can use previous parts of a problem even if you did not solve them.
- As throughout the course, entropy (H) and Mutual Information (I) are specified in bits.
- Throughout the exam 'prefix code' refers to a variable length code satisfying the prefix condition.
- Good Luck !

1. **Vin's Idea** *(30 points)*
Vinith is very excited about a new lossless compression idea, and he claims it can beat entropy. Albert is very skeptical, as Vinith got a pretty low grade when he took EE376A. Albert, however, didn't do too well in EE376A either, so now he needs your help to analyze Vinith's scheme.

   Vinith: "Suppose $X_1, X_2, \ldots$ is an i.i.d. Bernoulli-1/2 sequence. We can break up this sequence into its pattern of 'repeats'. For instance, 0001100001... begins with repeats (also known as 'run-lengths') '000', '11', and '0000'. If we let L_i be the length of the ith repeat, we can represent the sequence by $(X_1, L_1, L_2, \ldots)$. For example,

   - $1010\ldots$ would be represented by $(1, 1, 1, 1, \ldots)$
   - $11100111110\ldots$ by $(1, 3, 2, 5, \ldots)$ and
   - $001011111110\ldots$ by $(0, 2, 1, 1, 7, \ldots)$.

   In particular, I suggest we describe the sequence $X_1, X_2, \ldots, X_{\sum_{i=1}^{10} L_i}$ by describing $(X_1, L_1, L_2, \ldots, L_{10})$, which I'm sure would be a heavily compressed representation!!"

   (a) *(5 points)* What is the entropy of the first repeat length $H(L_1)$?

(b) *(5 points)* Describe an optimal prefix code for L_1. What is its expected code-length?

(c) *(5 points)* What is $H(X_1, L_1, \ldots, L_{10})$?

(d) *(5 points)* Describe an optimal uniquely decodable code for $(X_1, L_1, \ldots, L_{10})$. What is its expected code-length? Call it "Vinith's code".

(e) *(5 points)* What is the expected number of source symbols $E[\sum_{i=1}^{10} L_i]$ that Vinith's code encodes?

(f) *(5 points)* Comment, based on your answers to the previous two parts, on whether Vinith's code is "beating entropy" on average.

2. Non-i.i.d. Source *(30 points)*

Consider a second-order binary Markov process $\{X_i\}_{i\geq 1}$ characterized as follows:

- $P(X_1 = 0, X_2 = 0) = P(X_1 = 1, X_2 = 1) = \frac{1}{6}$ and $P(X_1 = 0, X_2 = 1) = P(X_1 = 1, X_2 = 0) = \frac{1}{3}$.
- For $n \geq 3$,
  - If $X_{n-1} = X_{n-2}$, then $X_n = 1 - X_{n-1}$.
  - If $X_{n-1} \neq X_{n-2}$, then X_n is drawn as a fair coin flip, independent of $\{X_i\}_{i=1}^{n-1}$.

(a) *(6 points)* Find an optimal prefix code for the pair (X_1, X_2), along with its expected code-length.

(b) *(6 points)* Show that the distribution of (X_n, X_{n+1}) is the same for all $n \geq 1$ (and, hence, the process is stationary).

(c) *(6 points)* Find the "entropy rate" of the process

$$\lim_{n\to\infty} \frac{H(X_1, X_2, \ldots, X_n)}{n}.$$

[Hint: can justify and use the facts that $H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X^{i-1})$, and $H(X_i|X^{i-1}) = H(X_i|X_{i-1}, X_{i-2}) = H(X_3|X_2, X_1)$ for $i \geq 3$]

(d) *(6 points)* For fixed $n \geq 2$, does there exist a uniquely decodable code for $(X_1, X_2, \ldots, X_n)$ whose expected code-length is $H(X_1, X_2, \ldots, X_n)$? If so, describe one. If not, explain why.

(e) *(6 points)* Describe a uniquely decodable code for $(X_1, X_2, \ldots, X_n)$ that attains the entropy rate. That is, a code with length function ℓ_n such that

$$\lim_{n\to\infty} \frac{E\left[\ell_n(X_1, X_2, \ldots, X_n)\right]}{n}$$

is equal to the entropy rate from part (c).

3. Entropy of a Sum and a Difference of I.I.D. Random Variables *(40 points)*
We will prove that if (Y, Y') are i.i.d. discrete random variables then:

$$H(Y - Y') - H(Y) \leq 2(H(Y' + Y) - H(Y)).$$

We will prove this inequality in the following steps.

(a) *Data Processing Inequality for Mutual Information* *(10 points)*
Let X_1, X_2 be discrete random variables. Also, let $Y_1 = F(X_1)$ and $Y_2 = G(X_2)$ for some functions, $F(\cdot), G(\cdot)$. Prove that:

$$I(X_1; X_2) \geq I(Y_1; Y_2).$$

(b) *Submodularity* *(10 points)*

Suppose that there exist functions F, G and R such that $X_0 = F(X_1) = G(X_2)$ and $X_{12} = R(X_1, X_2)$, where X_1, X_2 are discrete random variables. Use the previous part to prove that

$$H(X_{12}) + H(X_0) \leq H(X_1) + H(X_2).$$

(c) *Ruzsa Triangle Inequality* *(10 points)*
Let X, Y, Z be independent discrete random variables. Use Part (b) to prove that:

i. $H(X - Z) \leq H(X - Y) + H(Y - Z) - H(Y)$
ii. $H(X - Z) \leq H(X + Y) + H(Y + Z) - H(Y)$

[Hint: Use Part(b) with $X_1 = (X - Y, Y - Z)$, $X_2 = (X, Z)$, $X_{12} = (X, Y, Z)$ and $X_0 = X - Z$.]

(d) *Sum and Difference of Entropy* *(10 points)*
Use the previous part to conclude that for i.i.d (Y, Y') random variables

$$H(Y - Y') - H(Y) \leq 2(H(Y' + Y) - H(Y)).$$