EE376A - Information Theory
Midterm, Monday February 12th

Instructions:

- You have **two hours**, 6:00PM - 8:00PM
- The exam has 3 questions, totaling 100 points. (There are additional 20 points bonus)
- Please start answering each question on a new page of the answer booklet.
- You are allowed to carry the 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.
- Calculators are allowed for numerical computations.
- You are required to provide a sufficiently 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.
- log is taken in base 2.
- Throughout the exam 'prefix code' refers to a variable length code satisfying the prefix condition.
- Good Luck!

1. **Universal Prefix Codes** *(35 points)*
In this problem we consider binary prefix codes over the set of non-negative natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$. We do not know the probability distribution $P \equiv (p_j, j \in \mathbb{N})$, but we do know that it is a monotone distribution, i.e. $p_j \geq p_{j+1} \forall j \in \mathbb{N}$. We wish to construct prefix codes which perform *well* irrespective of the source probabilities. For a given code $c_j, j \in \mathbb{N}$ (where each c_j is a binary codeword), we denote the code lengths by $l_j^c, j \in \mathbb{N}$ and the expected code length by $\bar{L}_c := \sum_{j=1}^{\infty} p_j l_j^c$. Also, 0^j denotes a sequence of j zeros.

   (a) (*6 points*) Consider the code $u_j = 0^j 1$. Is this code prefix free? Justify.

   (b) (*Bonus, 5 points*) Find a monotone distribution P, such that $H(P) < \infty$, but $\bar{L}_u = \infty$. (it is fine to specify p_j up to a normalizing factor).

   (c) (*8 points*) Consider now the code b_j, which is the binary representation of j (Eg: $b_5 = 101$). Note that the codelength of b_j is given by: $l_j^b = \lfloor \log_2 j \rfloor + 1$. Is this code prefix free?

   (d) (*8 points*) For any monotone distribution P, show that the binary code b_j in (c) has expected code length $\bar{L}_b \leq H(P) + 1$.

   (e) (*8 points*) Now, consider the code $c_j = 0^{\lfloor \log_2 j \rfloor + 1} 1 b_j$ with $l_j^c = 2\lfloor \log_2 j \rfloor + 3$. Argue that this code is prefix free.

   (f) (*5 points*) For the code in (e), show that $\bar{L}_c \leq 2H(P) + 3$ for all monotone distributions P.

   (g) (*Bonus, 5 points*) Can you suggest prefix codes which improve on the performance of the code from part (e), i.e., achieve performance $\bar{L}_c \leq c_1 H(P) + c_2$, where $c_1 < 2$ (c_1 is a constant, c_2 is a lower-order term of $H(P)$)?

Solution to Problem 1

   (a) This is a prefix code. Different codes u_j have different number of zeros before 1.

   (b) Choose $p_j \propto (j+1)^{-2}$. This is well-defined since $\sum_{n=1}^{\infty} n^{-2} < \infty$. Also, $H(P) < \infty$ since $\sum_{j=1}^{\infty} (j+1)^{-2} \log(j+1) < \infty$ (the integral $\int_1^{\infty} \frac{\log x}{x^2} dx$ is finite). However,

$$\bar{L}_u = \sum_{j=1}^{\infty} p_j (j+1) \propto \sum_{j=1}^{\infty} \frac{1}{j+1} = \infty$$

   for the integral $\int_1^{\infty} \frac{dx}{x}$ diverges.

   (c) This code is not prefix-free. For example, $b_1 = 1$ is a prefix of $b_3 = 11$.

   (d) For monotone distributions, we have $p_j \leq \frac{1}{j} \sum_{k=1}^{j} p_k \leq \frac{1}{j}$ for any j. Hence,

$$\bar{L}_b \leq \sum_{j=1}^{\infty} p_j (\log_2 j + 1) \leq \sum_{j=1}^{\infty} p_j (\log_2 \frac{1}{p_j} + 1) = H(P) + 1.$$

(e) Assume by contradiction that c_j is a prefix of $c_{j'}$ for $j \neq j'$. Comparing the number of zeros in the front, we must have $\lfloor \log_2 j \rfloor = \lfloor \log_2 j' \rfloor$. Hence, b_j and $b_{j'}$ must have the same length, and the prefix assumption implies $b_j = b_{j'}$. Since b_j is the binary representation of j, we then have $j = j'$, a contradiction!

(f) Similar to (d), we have $jp_j \leq 1$. Hence,

$$\bar{L}_c \leq \sum_{j=1}^{\infty} p_j (2 \log_2 j + 3) \leq \sum_{j=1}^{\infty} p_j (2 \log_2 \frac{1}{p_j} + 3) = 2H(P) + 3.$$

(g) For $l_j^c = \lfloor \log_2 j + A \log_2 \log_2 j + B \rfloor$, since $\int_1^\infty \frac{dx}{x(\log x)^\alpha} < \infty$ for any $\alpha > 1$, suitable choices of A, B give $\sum_{j=1}^{\infty} 2^{-l_j^c} < 1$. By Kraft's inequality, there exist a prefix code c_j with codelength l_j^c. Using $jp_j \leq 1$ again, the average codelength for this code is

$$\begin{aligned}
\bar{L}_c &\leq \sum_{j=1}^{\infty} p_j \left(\log_2 j + A \log_2 \log_2 j + B \right) \\
&\leq \sum_{j=1}^{\infty} p_j \left(\log_2 \frac{1}{p_j} + A \log_2 \log_2 \frac{1}{p_j} + B \right) \\
&\leq H(P) + A \sum_{j=1}^{\infty} \log_2 \left(\sum_{j=1}^{\infty} p_j \log_2 \frac{1}{p_j} \right) + B \\
&= H(P) + A \log_2 H(P) + B.
\end{aligned}$$

2. Perfect Secrecy *(30 points)*

Alice wishes to communicate a message M to Bob, where M is chosen randomly from some alphabet $\mathcal{M}$. To prevent an eavesdropping adversary from reading the message, Alice encrypts the message using a deterministic function $C = E(K, M)$ to obtain the ciphertext $C \in \mathcal{C}$, where $K \in \mathcal{K}$ is a secret random key known to both Alice and Bob, and is independent of the message. Bob receives the ciphertext and decrypts it back to M using another deterministic function $M = D(K, C)$. We say that this system is *perfectly secure* if $I(M; C) = 0$.

(a) (*6 points*) Explain intuitively why a perfectly secure system is safe from an eavesdropping adversary.

(b) (*9 points*) Show that $H(M|C) \leq H(K|C)$ (under any system, secure or not).

(c) (*9 points*) Using part (b), show that $I(M; C) \geq H(M) - H(K)$.

(d) (*6 points*) Part (c) suggests that a perfectly secure system must have $H(K) \geq H(M)$. Do you think this is practical? Explain.

(e) (*Bonus, 5 points*) Now, assume that $\mathcal{M} = \mathcal{K} = \mathcal{C} = \{0, 1\}^n$ with M and K uniformly and independently distributed in $\{0, 1\}^n$. Can you suggest perfectly secure encryption and decryption functions $E(K, M)$ and $D(K, C)$?

Solution to Problem 2

(a) For a perfectly secure system, M and C are independent. Hence, an eavesdropper who observes the ciphertext C cannot infer any information of M from C.

(b) We have

$$\begin{aligned} H(M|C) &= H(M,K|C) - H(K|M,C) \\ &\le H(M,K|C) \\ &= H(K|C) + H(M|K,C) \\ &= H(K|C) \end{aligned}$$

where the last step follows from $H(M|K,C) = 0$, for $M = D(K,C)$ is a deterministic function of K, C.

(c) We have

$$I(M;C) = H(M) - H(M|C) \ge H(M) - H(K|C) \ge H(M) - H(K).$$

The first inequality follows from (b); the second inequality is due to the fact that conditioning reduces entropy.

(d) Under a perfectly secure system, $0 = I(M;C) \ge H(M) - H(K)$, thus $H(K) \ge H(M)$. This is not practical: usually the message is very long (i.e., $H(M)$ is large), but we need to transmit/store the key which is as long as the message in a perfectly secure system.

(e) Consider $E(K,M) = K \oplus M, D(K,C) = K \oplus C$, where $\oplus$ denotes the coordinate-wise modulo-2 sum. Clearly $D(K, E(K,M)) = M$. Moreover,

$$\begin{aligned} I(M;C) &= H(C) - H(C|M) = H(C) - H(K \oplus M|M) \\ &= H(C) - H(K|M) = H(C) - H(K) = 0 \end{aligned}$$

where the last step follows from the fact that both M, C are uniformly distributed on $\{0,1\}^n$. Hence, this is a perfectly secure system.

3. Mix of Problems *(35 points)*

(a) **Pairwise Independence** *(12 points)*
We say random variables $X_1, X_2, \ldots, X_n$ are pairwise independent if any pair of random variables (X_i, X_j), $j \neq i$ are independent.

i. Let X_1, X_2, X_3 be pairwise independent random variables, distributed identically as $Bern(0.5)$. Then:

A. (*6 points*) Show that: $H(X_1, X_2, X_3) \le 3$. When is equality achieved?

B. (*6 points*) Show that: $H(X_1, X_2, X_3) \ge 2$. When is equality achieved?

ii. (*Bonus, 5 points*) Let $Z_1, Z_2, \ldots, Z_k$ be i.i.d $Bern(0.5)$ random variables. Show that using the Z_i's, you can generate $2^k - 1$ pairwise independent random variables, identically distributed as $Bern(0.5)$.

(b) **Individual Sequences** *(12 points)*
Let x^n be a given arbitrary binary sequence, with n_0 0's and n_1 1's ($n_1 = n - n_0$). You are also provided a compressor C which takes in any arbitrary i.i.d distribution $q(x)$ as a parameter, and encodes x^n using:

$$\bar{L}_q = \frac{1}{n} \log \frac{1}{q(x^n)}$$

bits per symbol (ignoring integer constraints).

i. (*6 points*) Given the sequence x^n, what distribution $q(x)$ will you choose as a parameter (in terms of n_0, n_1) to the compressor C, so that $\bar{L}_q$ is minimized. Justify.

ii. (*6 points*) When compressing any given individual sequence x^n, we also need to store the parameter distribution $q(x)$ (as it is required for decoding). Show that you can represent the parameter distribution $q(x)$ using $\log(n+1)$ bits. Find the effective compression ratio.

(c) **AEP** *(11 points)*
Let $p(x)$ and $q(x)$ be two distinct distributions supported on the same alphabet $\mathcal{X}$.

i. (*5 points*) Let X^n be distributed i.i.d according to distribution $p(x)$. Then, for what distributons $p(x), q(x)$ is the following relationship satisfied for all $\epsilon > 0$?

$$P\left(\left\{x^n \in \mathcal{X}^n : \left|\frac{1}{n}\log\frac{1}{p(x^n)} - H(q)\right| < \epsilon\right\}\right) \to 1, \text{as } n \to \infty$$

ii. (*6 points*) Let X^n be distributed i.i.d according to distribution $p(x)$. Show that for any $\epsilon > 0$:

$$P\left(\left\{x^n \in \mathcal{X}^n : \left|\frac{1}{n}\log\frac{1}{q(x^n)} - (H(p) + D(p\|q))\right| < \epsilon\right\}\right) \to 1, \text{as } n \to \infty$$

Solution to Problem 3

(a) i. A. We have

$$\begin{aligned} H(X_1, X_2, X_3) &= H(X_1, X_2) + H(X_3|X_1, X_2) \\ &\le H(X_1, X_2) + H(X_3) \\ &= H(X_1) + H(X_2) - I(X_1; X_2) + H(X_3) = 3. \end{aligned}$$

Equality holds if and only if X_3 is independent of (X_1, X_2), which together with the pairwise independence implies that X_1, X_2, X_3 are mutually independent.

B. We have

$$\begin{aligned} H(X_1, X_2, X_3) &= H(X_1, X_2) + H(X_3|X_1, X_2) \\ &\ge H(X_1, X_2) \end{aligned}$$

$$= H(X_1) + H(X_2) - I(X_1; X_2) = 2.$$

Equality holds if X_3 is a deterministic function of (X_1, X_2). We also require X_3 to have the correct marginal distribution of $Bern(0.5)$, and satisfy pairwise independent properties. The only functions possible are: $X_3 = X_1 \oplus X_2$ and $X_3 = 1 \oplus X_1 \oplus X_2$.

ii. For any non-empty subset $S \subset \{1, \cdots, k\}$, we define a random variable $X_S = \sum_{i \in S} Z_i$. There are $2^k - 1$ random variables in total. To show $X_S \sim Bern(0.5)$, pick any $i_0 \in S$ and note that $X_S | (Z_i)_{i \neq i_0} \sim Bern(0.5)$. For pairwise independence, suppose $S \neq S'$ are two different non-empty subsets. By symmetry, assume that we can pick $i_0 \in S - S'$, then $X_S | (Z_i)_{i \neq i_0} \sim Bern(0.5)$ and $X_{S'}$ is a deterministic function of $(Z_i)_{i \neq i_0}$. This shows that X_S and $X_{S'}$ are independent.

(b) i. For $q(0) = 1 - q, q(1) = q$, we have

$$\bar{L}_q = \frac{1}{n} \log \frac{1}{(1-q)^{n_0} q^{n_1}} = -\frac{n_0}{n} \log(1-q) - \frac{n_1}{n} \log(q).$$

We see that $\bar{L}_q$ is convex in q, and taking derivative w.r.t q gives $q^* = \frac{n_1}{n}$.

ii. By the previous part, it suffices to store $n_1 \in \{0, 1, \cdots, n\}$ for full knowledge of $q(x)$. Hence, $\log(n+1)$ bits are enough. The effective compression ratio is

$$\bar{L}_q + \frac{\log(n+1)}{n} = H(\frac{n_1}{n}) + \frac{\log(n+1)}{n}.$$

(c) i. By AEP, $H(q) = H(p)$ suffices. This is also necessary, for a sequence of random variables cannot converge in probability to two different limits.

ii. By LLN, we have

$$\begin{aligned} \frac{1}{n} \log \frac{1}{q(x^n)} &= \frac{1}{n} \sum_{i=1}^{n} \log \frac{1}{q(x_i)} \\ &\to \mathbb{E}_P[\log \frac{1}{q(x)}] = \sum_x p(x) \log \frac{1}{q(x)} = H(p) + D(p \| q) \end{aligned}$$

in probability under P, which is exactly the desired statement.