EE376A - Information Theory
Midterm, Tuesday February 10th Solutions

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 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.
- 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.
- log is taken in base 2.
- Throughout the exam 'prefix code' refers to a variable length code satisfying the prefix condition.
- Good Luck!

1. Mix of Questions *(40 points)*

You only need to answer **four out of the five** questions presented below. Each of them is worth 10 points.

1) Let $Z_1, Z_2, Z_3, \ldots$ be i.i.d. random variables that take values "0" and "1" with equal probability. Further, let

$$X_i = \sum_{j=1}^{i} Z_j, \text{ for } 1 \leq i \leq n. \tag{1}$$

Find $I(X_1; X_2, X_3, \ldots, X_n)$.

2) Let $U_1, U_2, U_3, \ldots$ be i.i.d. taking values A, B, C, D, E and F, with the following distribution:

Symbol	A	B	C	D	E	F
Probability	1/2	1/4	1/8	1/16	1/32	1/32

(a) Compute $H(U_1)$.

(b) What is the most probable sequence of a given length n? What is its probability?

(c) Recall the definition of the ϵ-typical set for a memoryless source U:

$$A_\epsilon^{(n)} = \left\{ u^n : \left| -\frac{1}{n} \log p(u^n) - H(U) \right| \leq \epsilon \right\}. \tag{2}$$

Does the sequence you found in part (b) belong to $A_\epsilon^{(n)}$ for $\epsilon = 0.1$? How about for $\epsilon = 1$?

3) Let (X_i, Y_i) be i.i.d. $\sim p(x, y)$. Find the limit in probability, as $n \to \infty$, of

$$\frac{1}{n} \log \frac{p(X^n, Y^n)}{p(X^n)p(Y^n)}. \tag{3}$$

4) Consider a source with five symbols u_1, u_2, u_3, u_4, u_5, with probabilities $p(u_1) \geq p(u_2) \geq p(u_3) \geq p(u_4) \geq p(u_5)$.

(a) Suppose $p(u_1) \geq p(u_2) = p(u_3) = p(u_4) = p(u_5)$. Find the minimum value of q such that $p(u_1) \geq q$ implies $n_1 = 1$. Here n_1 denotes the length of the codeword associated with symbol u_1 generated by a Huffman code applied to the source.

(b) Suppose $p(u_1) \geq p(u_2) \geq p(u_3) \geq p(u_4) > p(u_5) = 0$. Find the largest value of r such that $p(u_1) \leq r$ implies $n_1 > 1$. Here n_1 denotes the length of the codeword associated with symbol u_1 generated by a Huffman code applied to the source.

5) Consider a random variable X which takes on four possible values with probabilities $(1/3, 1/3, 1/4, 1/12)$.

(a) Construct a Huffman code for this random variable.

(b) Show that there exist two different sets of optimal lengths for the codewords, namely, show that codeword length assignments $(1, 2, 3, 3)$ and $(2, 2, 2, 2)$ are both optimal.

(c) Are there optimal codes with codeword lengths for some symbols that exceed the Shannon code length $\lceil \log \frac{1}{p(x)} \rceil$? (Hint: Check the codeword lengths from the previous part.)

Solution:

1)

$$\begin{aligned}
I(X_1; X_2, X_3, \ldots, X_n) =& H(X_2, X_3, \ldots, X_n) - H(X_2, X_3, \ldots, X_n | X_1) \\
=& H(X_2) + \sum_{i=3}^{n} H(X_i | X_2, \ldots, X_{i-1}) - \sum_{i=2}^{n} H(X_i | X_1, \ldots, X_{i-1}) \\
=& H(X_2) + \sum_{i=3}^{n} H(Z_i) - \sum_{i=2}^{n} H(Z_i) \\
=& H(X_2) - H(Z_2) \\
=& \frac{3}{2} - 1 \\
=& \frac{1}{2},
\end{aligned}$$

since X_2 takes value "0" with probability 1/4, value "1" with probability 1/2 and value "2" with probability 1/4, which gives $H(X_2) = 3/2$.

2) (a) $H(U_1) = 1/2 \log 2 + 1/4 \log 4 + 1/8 \log 8 + 1/16 \log 16 + 2/32 \log 32 = 31/16$.

(b) The most probable sequence is the symbol A repeated n times, that is, AAAAAAA.... Its probability is $(\frac{1}{2})^n$.

(c) Note that $-\frac{1}{n} \log p(u^n) = -\frac{1}{n} \log(\frac{1}{2})^n = 1$. We also have from part (a) that $H(U) = \frac{31}{16}$. Therefore,

$$\begin{aligned}
\left| -\frac{1}{n} \log p(u^n) - H(U) \right| =& \left| 1 - \frac{31}{16} \right| \\
=& \frac{15}{16} \\
& > 0.1 \text{ and } < 1
\end{aligned}$$

Thus the most typical sequence A^n belongs to $A_\epsilon^{(n)}$ for $\epsilon = 1$, but not for $\epsilon = 0.1$.

3) First, note that we have $p(X^n, Y^n) = \prod_{i=1}^n p(X_i, Y_i)$, $p(X^n) = \prod_{i=1}^n p(X_i)$ and $p(Y^n) = \prod_{i=1}^n p(Y_i)$.

$$\begin{aligned}
\frac{1}{n} \log \frac{p(X^n, Y^n)}{p(X^n)p(Y^n)} &= \frac{1}{n} \log \frac{\prod_{i=1}^n p(X_i, Y_i)}{\prod_{i=1}^n p(X_i) \prod_{i=1}^n p(Y_i)} \\
&= \frac{1}{n} \sum_{i=1}^n \log \frac{p(X_i, Y_i)}{p(X_i)p(Y_i)} \\
&\xrightarrow{n\to\infty} \mathbb{E}[\log \frac{p(X, Y)}{p(X)p(Y)}] \\
&= I(X; Y),
\end{aligned}$$

where we have used the Law of Large Numbers.

4) (a) Note that $p(u_2) = p(u_3) = p(u_4) = p(u_5) = (1 - p(u_1))/4$. Without loss of generality, the Huffman code will first combine symbols u_4 and u_5, creating a super-symbol with probability $(1 - p(u_1))/2$, which is bigger than $(1 - p(1))/4$. Thus the Huffman code will then combine symbols u_2 and u_3, creating another super-symbol with probability $(1 - p(u_1))/2$. For symbol u_1 to be assigned a codeword with length 1, we need the Huffman code to combine next the two super-symbols. Therefore, we must have $p(u_1) > (1 - p(u_1))/2$. Or, equivalently,

$$p(u_1) > \frac{1}{3}. \tag{4}$$

(b) In this case, the Huffman code will first combine symbols u_3 and u_4 into a super-symbol with probability $p(u_3) + p(u_4)$. For the length of the codeword assigned to symbol u_1 to be bigger than 1, we must have $p(u_1) < p(u_3) + p(u_4)$, which implies $p(u_2) < p(u_3) + p(u_4)$, since $p(u_2) > p(u_1)$. For $p(u_3) + p(u_4)$ to be as small as possible, we must have $p(u_1) = p(u_2)$, which implies $p(u_3) + p(u_4) = 1 - 2p(u_1)$. Thus $p(u_1)$ must satisfy $p(u_1) < 1 - 2p(u_1)$, which implies

$$p(u_1) < \frac{1}{3}. \tag{5}$$

5) (a) Applying the Huffman algorithm gives us the following table: which gives code-

Code	Symbol	Probability			
0	1	1/3	1/3	2/3	1
11	2	1/3	1/3	1/3	
101	3	1/4	1/3		
100	4	1/12			

word lengths of $1, 2, 3, 3$ for the different codewords.

(b) Both set of lengths $1, 2, 3, 3$ and $2, 2, 2, 2$ satisfy the Kraft inequality, and they both achieve the same expected length (2 bits) for the above distribution. Therefore they are both optimal.

(c) The symbol with probability 1/4 has an Huffman code of length 3, which is greater than $\lceil \log \frac{1}{p(x)} \rceil$. Thus the Huffman code for a particular symbol may be longer than the Shannon code for that symbol. But on the average, the Huffman code cannot be longer than the Shannon code.

2. Non-prefix Code *(30 points)*

Suppose $p(x)$ is a PMF over $\mathcal{X} = \{1, 2, \cdots, K\}$, with $p(1) > p(2) > \cdots > p(K)$. We want to encode a random variable $X \sim p$. We care about encoding only this one random variable, therefore we do not require Unique Decodability but merely that the code be one-to-one, i.e., a different codeword for each of the K source symbols. Note that even the zero length codeword is valid, i.e., sending nothing (the empty string) can represent one of the source symbols.

(a) *(5 points)* Construct a coding scheme $c(X)$ that has the minimum expected code length. Let $l(i)$ be the length of the codeword of symbol i. Show that $l(i) = \lfloor \log i \rfloor$, where $\lfloor a \rfloor$ is the greatest integer no bigger than a.

(b) *(10 points)* Prove that the coding scheme from Part (a) satisfies

$$l(i) \leq -\log p(i)$$

and conclude that the minimum expected code length is less than or equal to the entropy, i.e.,

$$\mathbb{E}[l(X)] \leq H(X).$$

[Hint : Note that $p(i)$ is the i-th largest value, and therefore, $P(i) \leq \frac{1}{i}$]

(c) *(15 points)* Show that

$$\mathbb{E}[l(X)] \geq H(X) - 1 - \log(1 + \log K).$$

That is, lossless codes, even if not Uniquely Decodable, cannot beat the entropy by much.

[Hint : You may want to use the fact that $\sum_{i=1}^{K} \frac{1}{i} \leq 1 + \log K$]

Solution: Non-prefix Code

(a) Assign the codewords in the following order

$$\phi, 0, 1, 00, 01, 10, 11, 000, 001, \cdots$$

This gives us $l(i) = \lfloor \log i \rfloor$.

(b) Note that $p(i)$ is the i-th largest value, therefore $p(i) \leq \frac{1}{i}$. Thus,

$$\begin{aligned} l(i) &= \lfloor \log i \rfloor \\ &\leq \log i \\ &\leq \log \frac{1}{p(i)} \\ &= -\log p(i) \end{aligned}$$

This implies $\mathbb{E}[l(i)] \leq \mathbb{E}[-\log p(X)] = H(X)$.

(c)

$$
\begin{aligned}
\mathbb{E}[l(X)] =& \sum_{i=1}^{K} p(i) \lfloor \log i \rfloor \\
\geq& \sum_{i=1}^{K} p(i)(\log i - 1) \\
=& \sum_{i=1}^{K} p(i)(-\log p(i) + \log ip(i) - 1) \\
=& H(X) - 1 - \sum_{i=1}^{K} p(i) \log \frac{1}{ip(i)} \\
\geq& H(X) - 1 - \log \left(\sum_{i=1}^{K} p(i) \frac{1}{ip(i)} \right) \\
\geq& H(X) - 1 - \log \left(\sum_{i=1}^{K} \frac{1}{i} \right) \\
\geq& H(X) - 1 - \log(1 + \log K)
\end{aligned}
$$

3. The prime number theorem *(30 points)*

Some time around 300 B.C., someone showed that there are infinitely many prime numbers – we know this because a proof appears in Euclid's famous *Elements*. In this problem, we will not only show that there are infinitely many prime numbers, but we will also give a lower bound on the rate of their growth using information theory.

Let $\pi(n)$ denote the number of primes no greater than n. Note that every positive integer n has a **unique** prime factorization of the form

$$n = \Pi_{i=1}^{\pi(n)} p_i^{X_i}, \tag{6}$$

where $p_1, p_2, p_3, \ldots$ are the primes, that is, $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, etc., and $X_i = X_i(n)$ is the non-negative integer representing the multiplicity of p_i in the prime factorization of n.

Let N be uniformly distributed on $\{1, 2, 3 \ldots n\}$.

(a) *(8 points)* Show that $X_i(N)$ is an integer-valued random variable satisfying

$$0 \leq X_i(N) \leq \log n. \tag{7}$$

[Hint : Try finding a lower and an upper bound for $p_i^{X_i(N)}$]

(b) *(22 points)* Show that

$$\log n = H(N) \leq \pi(n) \log(\log n + 1). \tag{8}$$

Thus, not only is $\pi(n) \to \infty$ but in fact $\pi(n) \geq \frac{\log n}{\log(\log n + 1)}$.

[Hint : Do $X_1(N), X_2(N), \ldots, X_{\pi(n)}(N)$ determine N? What does that say about the respective entropies?].

Solution:

(a) $0 \leq X_i(N)$ is trivial. Note also that $2^{X_i} \leq p_i^{X_i} \leq N \leq n$.
Thus, combining both results, $0 \leq X_i(N) \leq \log n$, as we wanted to show.

(b)

$$\begin{aligned} \log n &= H(N) && (9)\\ &= H(X_1, X_2 \ldots X_{\pi(n)}) && (10)\\ &= \sum_{i=1}^{\pi(n)} H(X_i | X_1, \ldots, X_{i-1}) && (11)\\ &\leq H(X_1) + H(X_2) + \ldots H(X_{\pi(n)}) && (12) \end{aligned}$$

$$\leq \pi(n) \log(\log n + 1), \tag{13}$$

where the first step follows because there is a one-to-one mapping between N and $X_1, X_2 \ldots X_{\pi(n)}$. The second step is by the chain rule for entropy. The next step is because conditioning reduces entropy, and the last one is because the distribution that maximizes entropy is the uniform one, there are $\pi(n)$ entropy terms, and the X_i's can take at most $\log n + 1$ different values.