# ECE 534 Information Theory - MIDTERM 1

09/29/2010, LH 305.

- This exam has 4 questions, each of which is worth 25 points.
- You will be given the full 1.25 hours. **Use it wisely!** Many of the problems have short answers; try to find shortcuts. Do questions that you think you can answer correctly first.
- You may bring and use one 8.5x11" double-sided crib sheet.
- No other notes or books are permitted.
- No calculators are permitted.
- Talking, passing notes, copying (and all other forms of cheating) is forbidden.
- Make sure you explain your answers in a way that illustrates your understanding of the problem. Ideas are important, not just the calculation.
- Partial marks will be given.
- Write all answers directly on this exam.

Your name: ____________________

Your UIN: ____________________

Your signature: ____________________

The exam has 4 questions, for a total of 100 points.

| Question: | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|
| Points: | 25 | 25 | 25 | 25 | 100 |
| Score: | | | | | |

1. *A dog looking for a bone.*

A dog walks on the integers, possibly reversing direction at each step with probability $p = 0.1$. Let $X_0 = 0$. The first step is equally likely to be positive or negative. A typical walk might look like this:

$$(X_0, X_1, X_2, \cdots) = (0, -1, -2, -3, -4, -3, -2, -1, 0, 1, 2, 1, 0, \cdots)$$

(a) (10 points) Find $H(X_1, X_2, \cdots X_n)$ for any $n$.

**Solution:** Cover+Thomas question 4.12.

(a) By the chain rule,

$$\begin{aligned} H(X_0, X_1, \ldots, X_n) &= \sum_{i=0}^{n} H(X_i|X^{i-1}) \\ &= H(X_0) + H(X_1|X_0) + \sum_{i=2}^{n} H(X_i|X_{i-1}, X_{i-2}), \end{aligned}$$

since, for $i > 1$, the next position depends only on the previous two (i.e., the dog's walk is 2nd order Markov, if the dog's position is the state). Since $X_0 = 0$ deterministically, $H(X_0) = 0$ and since the first step is equally likely to be positive or negative, $H(X_1|X_0) = 1$. Furthermore for $i > 1$,

$$H(X_i|X_{i-1}, X_{i-2}) = H(.1, .9).$$

Therefore,

$$H(X_0, X_1, \ldots, X_n) = 1 + (n-1)H(.1, .9).$$

(b) (10 points) Find the entropy rate of this browsing dog.

**Solution:**

(b) From a),

$$\begin{aligned} \frac{H(X_0, X_1, \ldots X_n)}{n+1} &= \frac{1 + (n-1)H(.1, .9)}{n+1} \\ &\to H(.1, .9). \end{aligned}$$

(c) (5 points) What is the expected number of steps the dog takes before reversing direction?

**Solution:**

(c) The dog must take at least one step to establish the direction of travel from which it ultimately reverses. Letting $S$ be the number of steps taken between reversals, we have

$$\begin{aligned} E(S) &= \sum_{s=1}^{\infty} s(.9)^{s-1}(.1) \\ &= 10. \end{aligned}$$

Starting at time 0, the expected number of steps to the first reversal is 11.

Points earned: ________ out of a possible 25 points

.

Points earned: ____________ out of a possible 0 points

2. *AEP.* Let $X_i$ be i.i.d. distributed according to $p(x)$ for $x \in \{1, 2, \cdots, m\}$. Let $\mu = U[X]$, and $H = -\sum_x p(x) \log p(x)$. Let $A^n$ be the (weakly) typical set as defined in class. Let $B^n$ be the set $B^n = \{x^n \in \mathcal{X}^n : |\frac{1}{n}\sum_{i=1}^n X_i - \mu| \leq \epsilon\}$.

(a) (4 points) Rigorously define the typical set $A^n$.

**Solution:** Cover+Thomas question 3.4.
$A^n = \{x^n \in \mathcal{X}^n : |-\frac{1}{2}\log p(x^n) - H(X)| \leq \epsilon\}$.

(b) (3 points) Does $\Pr\{X^n \in A^n\} \to 1$? If so, sketch a proof, if not, show why not.

**Solution:**

(a) Yes, by the AEP for discrete random variables the probability $X^n$ is typical goes to 1.

(c) (6 points) Does $\Pr\{X^n \in A^n \cap B^n\} \to 1$? If so, sketch a proof, if not, show why not.

**Solution:**

(b) Yes, by the Strong Law of Large Numbers $Pr(X^n \in B^n) \to 1$. So there exists $\epsilon > 0$ and $N_1$ such that $Pr(X^n \in A^n) > 1 - \frac{\epsilon}{2}$ for all $n > N_1$, and there exists $N_2$ such that $Pr(X^n \in B^n) > 1 - \frac{\epsilon}{2}$ for all $n > N_2$. So for all $n > \max(N_1, N_2)$:

$$\begin{aligned} Pr(X^n \in A^n \cap B^n) &= Pr(X^n \in A^n) + Pr(X^n \in B^n) - Pr(X^n \in A^n \cup B^n) \\ &> 1 - \frac{\epsilon}{2} + 1 - \frac{\epsilon}{2} - 1 \\ &= 1 - \epsilon \end{aligned}$$

So for any $\epsilon > 0$ there exists $N = \max(N_1, N_2)$ such that $Pr(X^n \in A^n \cap B^n) > 1 - \epsilon$ for all $n > N$, therefore $Pr(X^n \in A^n \cap B^n) \to 1$.

(d) (6 points) Show that $|A^n \cap B^n| \leq 2^{n(H+\epsilon)}$, for all $n$.

**Solution:**

(c) By the law of total probability $\sum_{x^n \in A^n \cap B^n} p(x^n) \leq 1$. Also, for $x^n \in A^n$, from Theorem 3.1.2 in the text, $p(x^n) \geq 2^{-n(H+\epsilon)}$. Combining these two equations gives $1 \geq \sum_{x^n \in A^n \cap B^n} p(x^n) \geq \sum_{x^n \in A^n \cap B^n} 2^{-n(H+\epsilon)} = |A^n \cap B^n| 2^{-n(H+\epsilon)}$. Multiplying through by $2^{n(H+\epsilon)}$ gives the result $|A^n \cap B^n| \leq 2^{n(H+\epsilon)}$.

(e) (6 points) Show that $|A^n \cap B^n| \geq \frac{1}{2} 2^{n(H-\epsilon)}$ for $n$ sufficiently large.

**Solution:**

(d) Since from (b) $Pr\{X^n \in A^n \cap B^n\} \to 1$, there exists $N$ such that $Pr\{X^n \in A^n \cap B^n\} \geq \frac{1}{2}$ for all $n > N$. From Theorem 3.1.2 in the text, for $x^n \in A^n$, $p(x^n) \leq 2^{-n(H-\epsilon)}$. So combining these two gives $\frac{1}{2} \leq \sum_{x^n \in A^n \cap B^n} p(x^n) \leq \sum_{x^n \in A^n \cap B^n} 2^{-n(H-\epsilon)} = |A^n \cap B^n| 2^{-n(H-\epsilon)}$. Multiplying through by $2^{n(H-\epsilon)}$ gives the result $|A^n \cap B^n| \geq (\frac{1}{2}) 2^{n(H-\epsilon)}$ for $n$ sufficiently large.

Points earned: ________ out of a possible 25 points

.

Points earned: ________ out of a possible 0 points

3. *Short answers.*

(a) (6 points) A source produces a character $x$ from the alphabet $\mathcal{A} = \{0, 1, 2, \cdots 9, a, b, c, \cdots, y, z\}$; with probability 1/3, $x$ is a numeral $\{0, 1, 2 \cdots, 9\}$, with probability 1/3, $x$ is a vowel $\{a, e, i, o, u\}$; and with probability 1/3 it's one of the 21 consonants. All numerals are equiprobable, and the same goes for vowels and consonants. Determine the entropy of X.

> **Solution:** $\log(3) + \frac{1}{3}(\log(10) + \log(5) + \log(21))$

(b) (7 points) Three squares have average area $A = 100m^2$. The average of the lengths of their sides is $l = 10m$. What can be said about the size of the largest of the three squares? *(HINT: Use Jensen's inequality.)*

> **Solution:** Let $x$ be the length of the side of a square, and let the probability of $x$ be $(1/3, 1/3, 1/3)$ over the three lengths $(l_1, l_2, l_3)$. Then the information that we have is that $E[x] = 10$ and $E[f(x)] = 100$, where $f(x) = x^2$ is the function mapping lengths to areas. This is a strictly convex function. We notice that the equality $E[f(x)] = f(E[x])$ holds, therefore $x$ is a constant, and the three lengths must all be equal. The area of the largest square is $100m^2$.

(c) (6 points) State and rigorously prove the chain rule for entropy (for 2 random variables only) without using the definition of mutual information.

> **Solution:**

$$H(X, Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x, y) \log p(x, y) \tag{2.15}$$

$$= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x, y) \log p(x)p(y|x) \tag{2.16}$$

$$= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x, y) \log p(x) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x, y) \log p(y|x) \tag{2.17}$$

$$= -\sum_{x\in\mathcal{X}} p(x) \log p(x) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x, y) \log p(y|x) \tag{2.18}$$

$$= H(X) + H(Y|X). \tag{2.19}$$

(d) (6 points) Provide an example of random variables $X, Y, Z$ such that $I(X; Y) < I(X; Y|Z)$.

> **Solution:**
> This example is also given in the text. Let $X, Y$ be independent fair binary random variables and let $Z = X + Y$. In this case we have that,

$$I(X; Y) = 0$$

and,

$$I(X; Y \mid Z) = H(X \mid Z) = 1/2.$$

So $I(X; Y) < I(X; Y \mid Z)$. Note that in this case $X, Y, Z$ are not markov.

Points earned: ______________ out of a possible 25 points

.

Points earned: ____________ out of a possible 0 points

4. *Coding theory.*

(a) (10 points) Can $(1, 2, 2)$ be the codeword lengths of a binary Huffman code? What about $(2, 2, 3, 3)$? Explain why/why not.

**Solution:** Cover+Thomas 5.36

Codeword lengths of a binary Huffman code must satisfy the Kraft inequality with equality, i.e. $\sum_i 2^{-l_i} = 1$. An easy way to see this is the following: every node in the tree has a sibling (property of an optimal binary code) and if we assign each node a "weight", namely $2^{-l_i}$, then $2 \times 2^{-l_i}$ is the weight of the father/mother node. Thus, collapsing the tree back, we see that $\sum_i 2^{-l_i} = 1$.

So, we see that Huffman codewords must satisfy the Kraft inequality with equality. So, for this problem all we need to do is check whether these codeword lengths satisfy the Kraft inequality with equality. (1,2,2) does, so these can be the codeword lengths of a Huffman code, while (2,2,3,3) do not, so these cannot be Huffman codeword lengths.

(b) (5 points) Find the Shannon code of the distribution $(0.5, 0.25, 0.125, 0.125)$.

**Solution:** Cover+Thomas 5.28

We build the following table

| Symbol | Probability | $F_i$ in decimal | $F_i$ in binary | $l_i$ | Codeword |
|---|---|---|---|---|---|
| 1 | 0.5 | 0.0 | 0.0 | 1 | 0 |
| 2 | 0.25 | 0.5 | 0.10 | 2 | 10 |
| 3 | 0.125 | 0.75 | 0.110 | 3 | 110 |
| 4 | 0.125 | 0.875 | 0.111 | 3 | 111 |

The Shannon code in this case achieves the entropy bound (1.75 bits) and is optimal.

(c) (10 points) Find the codeword lengths of the optimal binary encoding of the distribution $p = (\frac{1}{100}, \frac{1}{100}, \cdots, \frac{1}{100})$. Note – just find the optimal codeword lengths (how many codewords of various lengths) – no need to specify the actual code.

**Solution:** Cover+Thomas 5.44

Since the distribution is uniform the Huffman tree will consist of word lengths of $\lceil \log(100) \rceil = 7$ and $\lfloor \log(100) \rfloor = 6$. There are 64 nodes of depth 6, of which (64-$k$) will be leaf nodes; and there are $k$ nodes of depth 6 which will form $2k$ leaf nodes of depth 7. Since the total number of leaf nodes is 100, we have

$$(64 - k) + 2k = 100 \Rightarrow k = 36.$$

So there are 64 - 36 = 28 codewords of word length 6, and $2 \times 36 = 72$ codewords of word length 7.

Points earned: ____________ out of a possible 25 points

.

Points earned: ________ out of a possible **??** points