ECE 534 Information Theory - MIDTERM

10/02/2013, LH 207.

- 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:	30	20	30	20	100
Score:					

1. *Entropy and the typical set*

Consider a random variable X which takes on the values $-1, 0$ and 1 with probabilities $p(-1) = 1/4, p(0) = 1/2, p(1) = 1/4$. For parts (c) and (d) and (e) and (f), we consider a sequence of 8 i.i.d. throws, or instances of this random variable.

(a) (4 points) Find the entropy in base 4 of this random variable.

Solution:

$$H_4(X) = \frac{1}{4}\log_4(4) + \frac{1}{2}\log_4(2) + \frac{1}{4}\log_4(4) = \frac{1}{4} + \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{4} = 0.75$$

(b) (4 points) Find the entropy in base 2 of this random variable.

Solution:

$$H_2(X) = \frac{1}{4}\log_2(4) + \frac{1}{2}\log_2(2) + \frac{1}{4}\log_2(4) = \frac{1}{4}\cdot 2 + \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 = 1.5$$

Alternatively, you could have done

$$H_2(X) = \log_2(4)H_4(X) = 2H_4(X) = 1.5$$

(c) (2 points) How many sequences of length 8 are there in total?

Solution: There are 3^8 possible sequences of length 8.

(d) (5 points) Give one example of a typical sequence of length $n = 8$, for $\epsilon = 0.1$.

Solution: Easiest way is to make sure it's strongly typical, which will imply it's also weakly typical. So we just need to ensure that we have the proper proportions. An example of such a sequence would be

$$-1, -1, 0, 0, 0, 0, 1, 1$$

(e) (5 points) Find upper and lower bounds (numbers) on the probability of any typical sequence in $A_{0.1}^{(8)}$.

Solution: If a sequence is in the typical set, then by definition we have

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, \cdots x_n) \leq 2^{-n(H(X)-\epsilon)}$$

which in this case becomes:

$$2^{-8(1.5+0.1)} \leq p(x_1, \cdots x_8) \leq 2^{-8(1.5-0.1)}$$

(f) (5 points) Approximately how many sequences are there in $A_{0.1}^{(8)}$?

Solution: There are approximately $2^{n(H(X))}$ sequences in the typical set, or about $2^{8(1.5)} = 2^{12}$ sequences in this set.

(g) (5 points) Using base 4 instead of base 2, give the definition of the typical set $A_\epsilon^{(n)}$.

Solution:

$$A_\epsilon^{(n)} = \{(x_1, \cdots x_n) : 4^{-n(H_4(X)+\epsilon)} \leq p(x_1, \cdots x_n) \leq 4^{-n(H_4(X)-\epsilon)}\}$$

Points earned: ____________ out of a possible 0 points

.

Points earned: ____________ out of a possible 0 points

2. *Concepts.* A few sentences, with equations as needed, will suffice. Point at the key ideas, do not take too much time or write too much.

(a) (5 points) Does compressing over blocks of length n and letting $n \to \infty$ *always* improve "performance" of a source code? If so, argue intuitively why, if not, show why not.

Solution: If the source is D-adic then we may obtain the lower bound on the expected codeword length $H_D(X) = E_X[l]$, where l are the codeword lengths of a particular code. If the source is not D-adic, then coding over blocks of lengths n, where n increases does improve the expected codeword length *bounds*, as we showed that

$$H(X) \leq E_X[l] \leq H(X) + 1$$

for single letter codes, but that

$$\frac{1}{n}H(X_1, \cdots X_n) \leq E_{X_1, \cdots X_n}[(l(X_1, \cdots X_n)] \leq \frac{1}{n}H(X_1, \cdots, X_n) + \frac{1}{n}.$$

Thus, as $n \to \infty$, the expected codeword length my compressing blocks of length n at a time will tend to the entropy of that source (and will not be 1 bit away).

(b) (5 points) We flip a fair coin twice. Let X denote the outcome of the 1st flip and Y denote the outcome of the second flip. Find $H(X,Y), H(X)$ and $H(X|Y)$.

Solution: Since X and Y are independent and identical, $H(X,Y) = H(X)+H(Y) = 2H(X) = 2 \cdot 1 = 2$. Then $H(X) = 1$ and $H(X|Y) = 1$ by independence of X and Y.

(c) (5 points) Your boss has a 10×10 image (10 pixels in each dimension) in which each pixel can either be black or white, with probability 0.5, and all pixels are independent. He wants to compress this to a 5×5 image (again, with black and white pixels). If this is possible, find a code or describe the procedure you would use to do so. If this is not possible, prove why not.

Solution: This is impossible – we can consider the 10x10 image to be a 100 dimension vector, and since each pixel is i.i.d. and Bernoulli(1/2), each pixel has entropy 1 bit and cannot be compressed. So, a 100 bit vector = 100x1 = n H(X) is the best we can compress to losslessly.

(d) (5 points) Find a prefix code over an alphabet of size $D = 3$ with codeword lengths $(1, 1, 2, 3)$ or explain why it does not exist.

Solution: This exists since the Kraft inequality is satisfied. Here is a possible code: $(0, 1, 20, 210)$.

.

Points earned: ___ out of a possible 0 points

3. *Short answers.* Be short!

(a) (5 points) If $Y_1 \to Y_2 \to Y_3 \to \cdots \to Y_m$ forms a Markov chain, what is $I(Y_1; Y_2, Y_3, \cdots, Y_m)$? (simplify as much as possible)

Solution: $I(Y_1; Y_2, \cdots, Y_m) = I(Y_1; Y_2) + I(Y_1; Y_3|Y_2) + \cdots I(Y_1; Y_m|Y_2, \cdots Y_{m-2})$. By the Markov property, the past and future are conditionally independent given the present and hence all terms except the first are zero. Therefore,

$$I(Y_1; Y_2, \cdots Y_m) = I(Y_1; Y_2).$$

(b) (5 points) True or False? Explain why or why not.

$$I(X; Y, Z) \geq I(X; Y|Z)$$

Solution: True, by the chain rule of mutual information, $I(X; Y, Z) = I(X; Z) + I(X; Y|Z)$ and since $I(X; Z) \geq 0$ always, the inequality $I(X; Y, Z) \geq I(X; Y|Z)$ holds.

(c) (5 points) Let X and Y be independent. Find $H(2X, -2Y)$ in terms of the entropies $H(X), H(Y)$ and $H(X, Y)$.

Solution: Since $X \to 2X$ and $Y \to -2Y$ are one-to-one mappings (just change the values X takes but not their probabilities), and $2X$ and $-2Y$ are still independent,

$$H(2X, -2Y) = H(2X) + H(-2Y) = H(X) + H(Y)$$

(d) (5 points) Why / how is the Asymptotic Equipartition Property useful?

Solution: The AEP is useful as it allows us to describe the properties of the typical set, i.e. it allows us to count the number of elements in the typical set and show that the probability of the typical set approaches 1. This is turn is useful for compression and later, channel capacity.

(e) (5 points) T/F (and explain why): The average length of a non-singular code for a source X must be greater than the entropy of that source, $H(X)$. If true give a short derivation, if not give a counter-example.

Solution: False. Take $|\mathcal{X}| = 4$ and let X be uniform over the four symbols. Then $H(X) = \log_2(4) = 2$ bits. Take the non-singular code $0, 1, 00, 11$. Then the average length of the non-singular code is $1/4 \cdot 1 + 1/4 \cdot 1 + 1/4 \cdot 2 + 1/4 \cdot 2 = 1.5 < H(X)$.

(f) (5 points) T/F (and explain why): The typical set $A_\epsilon^{(n)}$ is the smallest set of sequences of length n which has total probability greater that $1 - \epsilon$.

Solution: False. The typical set is not defined in this way, and there could be another set with less elements which has probability close to 1. The above definition is that we used for the "smallest probable set" $B_\epsilon^{(n)}$.

.

4. *Huffman and optimal codes.* Consider a source which outputs 5 symbols with probabilities (0.3,0.3,0.2,0.1,0.1).

   (a) (5 points) To how many (on average) bits per source symbol may this source be compressed? (an expression rather than a number is fine).

   Solution: The minimal average (expected) number of bits per source symbol we can compress this source to is given by the source's entropy $H(X)$, computed using log in base 2 (for bits). Let the source be X

$$H(X) = 2 \times \frac{3}{10}\log_2\left(\frac{10}{3}\right) + + \frac{2}{10}\log_2\left(\frac{10}{2}\right) + 2 \times \frac{1}{10}\log_2\left(\frac{10}{1}\right)$$

   (b) (5 points) Find the average codeword length of a binary *Huffman* code for this source.

   Solution: he binary Huffman code for this distribution looks like:

Codeword	X	Probability				
10	1	0.3	0.3	0.4	0.6	1
11	2	0.3	0.3	0.3	0.4	
00	3	0.2	0.2	0.3		
101	4	0.1	0.2			
011	5	0.1				

   The average codeword length is thus $(0.3+0.3+0.2)\times 2 + (0.1+0.1)\times 3 = 2.2$ bits/symbol.

   (c) (5 points) Find the average codeword length of an *optimal* quaternary (4 symbols) code for this source.

   Solution: The quaternary Huffman code (which is optimal) for this distribution looks like:

Codeword	X	Probability		
1	1	0.3	0.3	1
2	2	0.3	0.3	
3	3	0.2	0.2	
41	4	0.1	0.2	
42	5	0.1		
	Dummy	0		
	Dummy	0		

   The average codeword length is thus $(0.3+0.3+0.2)\times 1 + (0.1+0.1)\times 2 = 1.2$ quaternary symbols/source symbol.

   (d) (5 points) For what probability distribution (on the same five symbols, but not necessarily (0.3,0.3,0.2,0.1,0.1)) would the binary code you constructed in part 2 exactly achieve the minimal expected number of bits per source symbol?

   Solution: The codeword lengths in part 2 were $(2,2,3,3,3)$. These are optimal codeword lengths (achieving entropy lower bound) for a source with distribution $= (1/4, 1/4, 1/4, 1/8, 1/8)$.

.

Points earned: ____________ out of a possible 0 points