ECE 534 Information Theory - MIDTERM

10/05/2018, LH 210.

- You will be given the full 50 minutes. **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: Solutions

Your UIN: ____________________

Your signature: ____________________

The exam has 7 questions, for a total of 100 points.

Question:	1	2	3	4	5	6	7	Total
Points:	14	14	10	12	20	10	20	100
Score:								

1. (14 points) *Entropy and source coding.*
   A discrete memoryless source X of alphabet size K has an entropy of $H(X)$ bits. A binary prefix code for this source is found, and has an average length of $E[l(x)] = \frac{H(X)}{\log_2(3)}$ bits per source symbol. Prove that each letter has a probability of the form 3^{-i} for i an integer that can depend on the symbol.

$E[l(x)] = \frac{H_2(x)}{\log_2 3}$ bits / source symbol $\rightarrow$ in base 2

The entropy in base 3 would be $H_3(x) = \frac{H_2(x)}{\log_2(3)}$, so $H_2(x) = \log_2 3\, H_3(x)$

Then, the expected codeword length ~~to~~ if using a ~~source~~ code in base D would be

$E[l(x)] = \frac{H_2(x)}{\log_2 3}$ bits $= \frac{\cancel{\log_2 3} \times H_3(X)}{\cancel{\log_2 3}} = H_3(x)$ ~~bits~~ ternary bits / source symbol

Since this code would hit the entropy lower bound exactly, the PMF must be 3-adic, i.e. each symbol has a probability of the form 3^{-i} for some i.

2. (14 points) Consider the binary multiplier channel $Y = XZ$, where X and Z are independent binary random variables that take on values 0 and 1, and Z is Bernoulli α, i.e. $P(Z = 1) = \alpha$. Calculate $I(X;Y)$ when $p(x = 1) = q$ and hence $p(x = 0) = 1 - q$ for some $0 < q < 1$.

$p(x=0)=1-q, \quad p(x=1)=q$

$p(z=0)=1-\alpha, \quad p(z=1)=\alpha$

$y=0$ if $(x\cdot z)=0$, i.e. if $(x=0, z=0)$ or $(x=1, z=0)$ or $(x=0, z=1)$

$y=1$ if $x \cdot z = 1$, i.e. if $(x=1, z=1)$

$$I(X;Y) = H(Y) - H(Y|X)$$

$$= H(Y) - \underbrace{H(Y|X=0)}\, p(x=0) + \underbrace{H(Y|X=1)}\, \underbrace{p(X=1)}$$

if $X=0$ then $Y=0$ automatically, so $H(Y|X=0)=0$

if $X=1$, then q.

$$H(Y|X=1) = H(\cancel{1}\cdot Z|X=1) = H(Z) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$$

To calculate $H(Y)$, notice that

$$p(y=0) = \underbrace{(1-q)(1-\alpha) + (1-q)\alpha + (1-\alpha)q}_{\text{call this } q^*}.$$

$$p(y=1) = \alpha \cdot q.$$

So,

$$I(X;Y) = \cancel{\ldots}$$

$$= -\alpha\cdot q \log(\alpha\cdot q) - q^* \log q^* + q\cdot(\cancel{-}\alpha \log \alpha + (1-\alpha)\log(1-\alpha))$$

3. (10 points) For X, Y, Z random variables (not necessarily independent). Determine whether $H(X,Y,Z) - H(X,Y) \leq H(X,Z) - H(X)$ is true or not. If true, find the conditions for equality.

$$H(X,Y,Z) - H(X,Y) = H(X,Z) + H(Y|X,Z) - H(X) - H(Y|X)$$

$$= H(X,Z) - H(X) + [H(Y|X,Z) - H(Y|X)]$$

$$= H(X,Z) - H(X) - \underbrace{I(Y;Z|X)}$$

since Mutual information is non-negative This inequality must be true.

Equality holds when $I(Y;Z|X) = 0$, i.e. when Y and Z are conditionally independent given X, which happens when

$Y \to X \to Z$ form a Markov chain.

4. Let X, Y, Z all be binary random variables (not necessarily independent).
   (a) (6 points) Find a joint distribution $p(x, y, z)$ such that $I(X;Y) = 0$ and $I(X;Y|Z) = 1$ bit.
   (b) (6 points) Find a joint distribution $p(x, y, z)$ such that $I(X;Y) = 1$ bit and $I(X;Y|Z) = 0$.

(a) As seen in class, take X, Y to be independent Bernoulli(1/2) (or coin tosses), and let $Z = X \oplus Y$. Then

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

and

$$I(X;Y|Z) = H(X|Z) - \underbrace{H(X|Y,Z)}_{0}$$

$$= \underbrace{H(X|Z=0)}_{1}\underbrace{p(Z=0)}_{1/2} + \underbrace{H(X|Z=1)}_{1}\underbrace{p(Z=1)}_{1/2}$$

$$= 1$$

(b) Let X be a coin toss, and let $Y = Z = X$.

Then, $I(X;Y) = H(X) - \underbrace{H(X|Y)}_{0} = H(X) = 1$

$$I(X;Y|Z) = \underbrace{H(Y|Z)}_{0} - \underbrace{H(Y|X,Z)}_{0} = 0$$

5. Correct the following incorrect statements.
   (a) (5 points) All typical sequences in $A_\epsilon^{(n)}$ have the same probability.
   (b) (5 points) The typical set $A_\epsilon^{(n)}$ is defined as the smallest set of sequences with $\Pr\{A_\epsilon^{(n)}\} \geq 1-\epsilon$.
   (c) (5 points) The number of sequences in $A_\epsilon^{(n)}$ may be bounded as $|A_\epsilon^{(n)}| \leq 2^{-n(H(X)-\epsilon)}$.
   (d) (5 points) Let $I(X;Y)|_{p(x,y)}$ denote the mutual information between X, Y when they share a joint distribution $p(x,y)$. Then, for fixed $p_1(x), p_2(x)$ and any $0 \leq \lambda \leq 1$

$$I(X;Y)|_{p(y|x)(\lambda p_1(x)+(1-\lambda)p_2(x))} \leq \lambda I(X;Y)|_{p(y|x)p_1(x)} + (1-\lambda) I(X;Y)|_{p(y|x)p_2(x)}$$

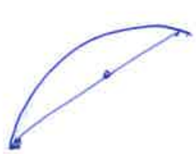
(a) False. $A_\varepsilon^{(n)}$ defined as $\{x^n : 2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)}\}$

So all typical sequences have approximately the same probability.

(b) False. As defined above. B_δ^{n} is the smallest set with $\Pr\{B_\delta^{(n)}\} \geq 1-\delta$.

(c) False. $|A_\varepsilon^{(n)}| \leq 2^{+n(H(X)+\varepsilon)}$

(d) False. $I(X;Y)$ is concave in $p(x)$ for fixed $p(y|x)$, so the inequality should be the other way.



$\geq$

Points earned: ____________ out of a possible 20 points

6. (10 points) State Jensen's inequality (formally, but no need to prove it.) State where we have used it (to prove what, the basis of what)?

Jensen's inequality:

If $f(x)$ is a convex function and X is a random variable,

$$E[f(X)] \geq f(E[X])$$

If f is strictly concave, then equality in the above implies that $X = E[X]$ with probability 1.

We used Jensen's inequality to show that $D(p\|q) \geq 0$, for any p, q which in turn was used to show that $I(X;Y) \geq 0$.

7. Your boss gives you a text document with letters distributed i.i.d. according to $p(x)$ and asks you to compress them as much as possible.
   (a) (10 points) How much can you compress them to, i.e. what is the fundamental limit of compression. Please be precise.
   (b) (10 points) If your boss does not care about hitting the limit exactly, but only wants to within 0.1 bits per source symbol of this limit, how would you go about designing a source code?

(a) If coding a block of n iid. symbols (a "super symbol"), Then we know that the minimal expected codeword length

$$\frac{1}{n} E[\ell(x_1, .. x_n)]$$

satisfies:

$$H(X) \le \frac{1}{n} E[\ell(x_1, .. , x_n)] \le H(X) + \frac{1}{n}$$

(b) To hit it within 0.1 bits per source symbol, use supersymbols of length at least 10. ~~For ea~~ For ~~each~~ the supersymbol, use either a Shannon code or a Huffman code to obtain a code that satisfies

$$\frac{1}{n} E[\ell(x_1, .., x_n] \le H(X) + \frac{1}{10}$$

0.1