University of Illinois at Chicago
Department of Electrical and Computer Engineering

ECE 534: Information Theory

Fall 2009

Midterm 1 - Solutions

NAME:

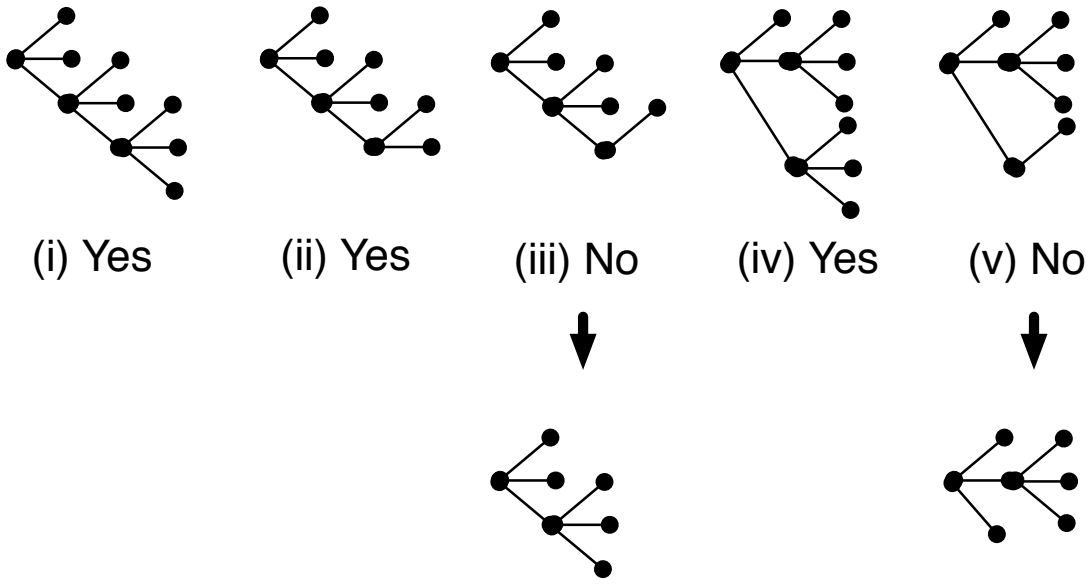
- This exam has 4 questions, each of which is worth 15 points.
- You will be given the full class time: 75 minutes. **Use it wisely!** Many of the problems have short answers; try to find shortcuts.
- You may bring and use one 8.5x11" double-sided crib sheet.
- No other notes or books are permitted.
- 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.

1. (15 points) **True of false and short answer. Brief explanations (rather than lengthy proofs) suffice.**

(a) (5 points) Which of the following sequences of codeword lengths cannot be the codeword lengths of a $3-ary$ (ternary, $D=3$) Huffman code?

(i) (1,1,2,2,3,3,3)
(ii) (1,1,2,2,3,3)
(iii) (1,1,2,2,3)
(iv) (1,2,2,2,2,2,2)
(v) (1,2,2,2,2)

Solution: The easiest way to see which can be ternary Huffman codeword lengths by trying to construct the Huffman tree. From the figure, we can see that (iii) and (v) are impossible and could be shortened/pruned to (1,1,2,2,2) and (1,1,2,2,2) respectively.



(b) (6 points) We have defined the mutual information $I(X;Y)$ between the two random variables X and Y. Let us *try* to define the mutual information between three random variables X, Y and Z as $I(X;Y;Z) = I(X;Y) - I(X;Y|Z)$.

(i) Is this definition symmetric in its arguments? Prove why or give an example of why not.
(ii) Is $I(X;Y;Z)$ positive? Prove why or give an example of why not.
(iii) True or false: $I(X;Y;Z) = H(X,Y,Z) - H(X) - H(Y) - H(Z) + I(X;Y) + I(Y;Z) + I(Z;X)$.
(iv) True of false: $I(X;Y;Z) = H(X,Y,Z) - H(X,Y) - H(Y,Z) - H(Z,X) + H(X) + H(Y) + H(Z)$.

Solution:

(i) Yes, it is symmetric, as can be seen from (iii) and (iv). Both expansions will be shown to be true and are symmetric in their arguments.

(ii) No, it is not necessarily positive. Take X and Y independent and identical coin flips and let $Z = X + Y$ mod 2. Then $I(X;Y) = 0$ but $I(X;Y|Z) = H(X|Z) - H(Y|X,Z) = H(X|Z) - 0 = 1$.

(iii) True.

$$\begin{aligned} I(X;Y) = I(X;Y|Z) &= H(X) - H(X|Y) - H(X|Z) + H(X|Y,Z) \\ &= H(X) - H(X,Y) + H(Y) - H(X,Z) + H(Z) + H(X,Y,Z) - H(Y,Z) \\ &= H(X,Y,Z) + (H(X) + H(Y) - H(X,Y)) + (H(X) + H(Z) - H(X,Z)) + (H(Y) + H(Z) - H(Y \\ &= H(X,Y,Z) + I(X;Y) + I(X;Z) + I(Y;Z) - H(X) - H(Y) - H(Z) \end{aligned}$$

(iv) True.

$$\begin{aligned} I(X;Y) = I(X;Y|Z) &= H(X) - H(X|Y) - H(X|Z) + H(X|Y,Z) \\ &= H(X) - H(X,Y) + H(Y) - H(X,Z) + H(Z) + H(X,Y,Z) - H(Y,Z) \end{aligned}$$

(c) (4 points) True or false:

(i) All typical sequences in $A_\epsilon^{(n)}$ have the same probability.
(ii) The typical set $A_\epsilon^{(n)}$ is defined as the smallest set of sequences with $\Pr\{A_\epsilon^{(n)}\} \geq 1 - \epsilon$.
(iii) The number of sequences in $A_\epsilon^{(n)}$ may be bounded as $|A_\epsilon^{(n)}| \leq 2^{-n(H(X)-\epsilon)}$.
(iv) If x^n is a *typical* sequence drawn i.i.d. according to $p(x^n) = \prod_{i=1}^n p(x_i)$ and y^n is a *typical* sequence drawn i.i.d. according to $p(y^n) = \prod_{i=1}^n p(y_i)$ then (x^n, y^n) are *jointly typical* according to $\prod_{i=1}^n p(x_i)p(y_i)$.

Solution:

(i) False. They have *approximately* the same probability. The *typical set* $A_\epsilon^{(n)}$ with respect to $p(x)$ is the set of sequences $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ with the property

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, x_2, \ldots, x_n) \leq 2^{-n(H(X)-\epsilon)}.$$

(ii) False. This is NOT its definition. The definition is given in (i) and is based strictly on the probabilities of the sequences in the set.

(iii) False. The number of sequences in $A_\epsilon^{(n)}$ may be bounded as $|A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$.

(iv) True. Note that in general x^n typical and y^n typical does NOT imply that (x^n, y^n) is jointly typical according to $p(x, y)$. However, we're only asking for then to be typical according to the $p(x)p(y)$, the product to the marginals, hence it is jointly typical.

2. (15 points) **Channels with memory.**

(a) (1 point) State what it means for a discrete channel to be *memoryless.*

(b) (1 point) Do you expect the capacity of a channel with memory (all else being equal) to be larger or smaller than its memoryless version? A sentence on your intuition why is sufficient.

Consider a binary symmetric channel with $Y_i = X_i + Z_i \mod 2$, and $X_i, Y_i, Z_i \in \{0, 1\}$. Suppose $\{Z_i\}$ has constant marginal probabilities $P(Z_i = 1) = p = 1 - P(Z_i = 0)$. We use the channel many times in sequence (for $i = 1, 2, \cdots n$ and $n \to \infty$).

(c) (1 point) What does it mean for this channel to be memoryless?

(d) (3 points) Suppose the channel is memoryless and determine the channel capacity. A short answer is enough.

(e) (5 points) Suppose now that the channel has *memory* of length K, which we define to mean that K consecutive instances of the noise Z_i are equal, i.e. $Z_1 = Z_2 = Z_3 = \cdots Z_K$, then the next K are independent of the first K and are again equal, $Z_{K+1} = Z_{K+2} = \cdots = Z_{2K}$ and so forth. We still have that $P(Z_1 = 1) = p = 1 - P(Z_1 = 0)$. Define the capacity of this channel with memory K to be

$$C^{(K)} = \max_{p(x_1,\ldots,x_K)} \frac{1}{K} I(X_1, \ldots, X_K; Y_1, \ldots, Y_K).$$

Determine the capacity of this channel and compare with the memoryless case.

(f) (2 points) Now consider the discrete memoryless channel $Y_i = Z_i X_i$ with input alphabet $X_i \in \{-1, 1\}$. What is the capacity of this channel when $\{Z_i\}$ is i.i.d. with $p(Z_i = 1) = p(Z_i = -1) = 0.5$?

(g) (2 points) Consider the same channel but with infinite memory: before transmission starts Z is randomly chosen and fixed for all time, i.e. $Y_i = ZX_i$ for $i = 1, 2, \cdots$. Find the capacity of this channel if $p(Z = 1) = p(Z = -1) = 0.5$ and compare to the memoryless case. *(HINT: there's a very intuitive/clever/simple way of figuring this out.)*

Solution:

(a) A discrete channel is memoryless if $p(y_n|x_1,\ x_2, \ldots, x_n,\ y_1,\ y_2, \ldots,\ y_n) = p(y_n|x_n)$, that is, the output at channel use n is depends only on the input at channel use n and not on any other inputs or outputs.

(b) We would it expect it to be larger as, intuitively, we will be able to exploit the memory, or structure of the noise to achieve higher rates. That is, in a channel with memory you can very roughly "predict" some of the noise, which is a good thing in terms of capacity.

(c) This channel is memoryless is Z_i's are i.i.d.

(d) This channel is a binary symmetric channel with crossover probability p. Thus, its capacity we say in class and in the textbook and is equal to $C = 1 - H(p)$.

(e)

$$\begin{aligned}
C^(K) &= \max_{p(x_1,x_2,\cdots,x_k)} \frac{1}{K} I(X_1,\cdots,X_K;Y_1,\cdots,Y_K) \\
&= \max_{p(x_1,x_2,\cdots,x_k)} \frac{1}{K} \left[H(X_1,X_2,\cdots,X_K)) - H(X_1,X_2,\cdots,X_K|Y_1,Y_2,\cdots Y_K)\right] \\
&= \max_{p(x_1,x_2,\cdots,x_k)} \frac{1}{K} \left[H(X_1,X_2,\cdots,X_K)) - H(Z)\right] \\
&= \max_{p(x_1,x_2,\cdots,x_k)} \frac{1}{K} \left[H(X_1,X_2,\cdots,X_K)) - H(p)\right] \\
&= \frac{1}{K}\left[K - H(p)\right] \\
&= 1 - \frac{H(p)}{K}
\end{aligned}$$

This is larger than the memoryless version of capacity $1 - H(p)$. How much larger depends on K.

(f) $max_{p(x)} I(X;Y) = max_{p(x)} H(X) - H(X|Y) = \max_{p(x)} H(X) - H(X) = 0$. Alternatively, this channel is a binary symmetric channel with cross-over probability 0.5, so we know its capacity is $1 - H(0.5) = 1 - 1 = 0$.

(g) Its channel capacity is 1 bit/channel use. To see why consider the following transmission scheme: during the 1st channel use, send 1. Then the receiver can determine Z without any ambiguity. Thus, for all future channel uses we know Z and so can transmit 1 bit/channel use. Thus, over n channel uses we can transmit $n-1$ bits, and thus the channel capacity is given by

$$C = \lim_{n\to\infty} \frac{n-1}{n} = 1.$$

3. (15 points) **Huffman coding.**

(a) (5 points) A source has an alphabet of 4 letters x_1, x_2, x_3, x_4, and we know that $p(x_1) > p(x_2) = p(x_3) = p(x_4)$. Let N_1, N_2, N_3, N_4 denote the lengths of the binary Huffman codewords for letters x_1, x_2, x_3, x_4 respectively. Find the smallest number q such that $p(x_1) > q$ implies that $N_1 = 1$.

(b) (2 points) Show that for the q you found in part (a) that if $p(x_1) = q$ then a Huffman code exists with $N_1 > 1$.

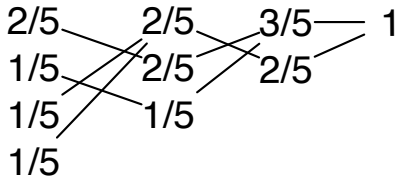
(c) (5 points) Now assume that the source has an arbitrary number of letters, $K \geq 1$, with $p(x_1) > p(x_2) \geq p(x_3) \geq p(x_4) \cdots \geq p(x_K)$. Show that $p(x_1) > q$ (the same q as before) still implies that $N_1 = 1$. *(HINT: induction.)*

(d) (3 points) Assume $p(x_1) \geq p(x_2) \geq \cdots \geq p(x_K)$ for $K \geq 3$. Find the largest q' such that $p(x_1) < q'$ implies that $N_1 > 1$.

Solution:

(a) Let $p = p(x_2) = p(x_3) = p(x_4)$. Then $p(x_1) = 1 - 3p$. To obtain $N_1 = 1$ we need $p(x_1)$ to remain the largest probability in the Huffman code construction until the last step. In the first phase we have the probabilities, in order $(1-3p, p, p, p)$. In the next phase we have $(1-3p, 2p, p)$. In order to guarantee $N_1 = 1$ we need $1 - 3p > 2p$, or $p > 1/5$. Thus, $q = 1 - 3p = 2/5$ will guarantee $N_1 = 1$.

(b) See figure.



(c) For $K = 1, K = 2$ we have $N_1 = 1$ regardless. For $K = 3$, the worst case distribution is $(2/5, 3/10, 3/10)$ which will also result in $N_1 = 1$. For $K = 4$ we proved it in (a). We now prove that $p(x_1) > 2/5 \rightarrow N_1 = 1$ for any $K > 4$. Assume the statement is true for $K = M$ (induction hypothesis). We now prove it for $K = M + 1$.

Assume $p_1 > p_2 \geq p_3 \geq \cdots p_{m+1}$. Combine the symbols x_{m+1} and x_m into a new symbol x'_m according to Huffman coding procedure. Then $p(x'_m) = p(x_m) + p(x_{m+1}) \leq \frac{2}{m}(1 - p(x_1)) < \frac{2}{3}(1 - p(x_1))$. Thus, if $p(x_1) > 2/5$ then $p(x_1) > p(x'_m)$ and by the induction hypothesis we obtain that $N_1 = 1$.

(d) Notice that the only difference here is that $p(x_1)$ can equal $p(x_2), p(x_3), \cdots p(x_K)$. The "worst case" scenario is if all probabilities are equal and $K = 3$. So then we have a distribution that looks like (p, p, p). For $N_1 > 1$ we need $p(x_1) < 1/3$, since then it will be combined in the first stage of the Huffman code, leading to $N_1 = 2$. For any smaller $p(x_1)$, a distribution can be created for which $N_1 = 2$.

4. (15 points) **Jamming.** A sender transmits a random variable X through a channel in which a jammer inserts a random variable Z, resulting in a channel output given by

$$Y = X + Z \mod 2.$$

The variables X and Z are independent and can only take on binary values 0 or 1 (i.e. $\mathcal{X} = \mathcal{Z} = \{0, 1\}$). The sender is limited in energy so that $E_X[X^2] \leq 1/2$ and the jammer is limited in energy so that $E_Z[Z^2] \leq 1/4$.

(a) (2 points) What do the energy constraints tell you about the probability mass functions (PMF) $p_X(x)$ of X and $p_Z(z)$ of Z?

(b) (2 points) Find an expression for the mutual information $I(X;Y)$.

The sender choses its PMF $p_X(x)$ to *maximize* the mutual information $I(X;Y)$, while the jammer choses its PMF $p_Z(z)$ to *minimize* the mutual information $I(X;Y)$.

(c) (4 points) Consider the problem from the jammer's perspective - it wishes to minimize the $I(X;Y)$ and knows that the sender will be trying to maximize $I(X;Y)$. Find its optimal jamming strategy.

(d) (4 points) Now consider the problem from the sender's perspective - it wishes to maximize the $I(X;Y)$ and knows that the jammer will be trying to minimize $I(X;Y)$. Find its optimal transmission strategy.

(e) (1 point) How many bits/channel use can the sender possibly reliably communicate?

(f) (2 points) How does the result of (e) change if the jammer's energy is allowed to equal the sender's energy? *Do not redo everything, just argue here.*

Solution:

(a) Since X and Z are binary random variables, they may be described by a single parameter: $p = p(X = 1)$ (thus $1 - p = p(X = 0)$ and $q = p(Z = 1)$ (thus $1 - q = p(Z = 0)$. Thus, $E[X^2] = 0^2(1-p) + 1^2(p) = p \leq 1/2$ and $E[Z^2] = 0^2(1-q) + 1^2(q) = q \leq 1/4$.

(b) To obtain an expression for the mutual information $I(X;Y)$, we need the distribution of Y in terms of the distributions of X and Z (i.e. in terms of p and q). So,

$$p(Y = 0) = p(X = 0, z = 0) + p(X = 1, Z = 1) = (1-p)(1-q) + pg = 1 + 2pq - p - q, \;\; p(Y = 1) = p + q - 2pq.$$

Define $r = 1 + 2pq - p - q$. Then, $I(X;Y) = H(Y) - H(Y|X) = H(r) - H(q) = H(1 + 2pq - p - q) - H(q)$.

(c) First off, notive that for fixed p, $I(X;Y)$ is convex in q and for fixed q, $I(X;Y)$ is concave in p. Thus, all maxima and minima we find will be global ones. The jammer wishes to minimize $I(X;Y)$ and knows that the sender will be trying to maximize $I(X;Y)$. Thus, its strategy may be obtained from the following optimization:

$$\min_{0 \leq q \leq 1/4} \max_{0 \leq p \leq 1/4} I(X;Y) = \min_{0 \leq q \leq 1/4} \max_{0 \leq p \leq 1/2} H(1 + 2pq - p - q) - H(q)$$

Since $\max_{0 \leq p \leq 1/2} H(1 + 2pq - p - q) - H(q)$ is achieved at $p = 1/2$ (just take the partial derivative with respect to p and set it to zero), we obtain

$$\begin{aligned} \min_{0 \leq q \leq 1/4} \max_{0 \leq p \leq 1/4} I(X;Y) &= \min_{0 \leq q \leq 1/4} \max_{0 \leq p \leq 1/2} H(1 + 2pq - p - q) - H(q) \\ &= \min_{0 \leq q \leq 1/4} 1 - H(q) \\ &= 1 - H(1/4) \end{aligned}$$

So, the optimal jamming strategy is to set $q = 1/4$, maximizing its transmit power.

(d) The sender wants to do the opposite, and its strategy may be obtained as the solution to the following optimization problem:

$$\max_{0\leq p\leq 1/2}\min_{0\leq q\leq 1/4} I(X;Y) = \max_{0\leq p\leq 1/2}\min_{0\leq q\leq 1/4} H(1+2pq-p-q) - H(q)$$

Consider $\min_{0\leq q\leq 1/4} H(1+2pq-p-q) - H(q)$. Taking the partial with respect to q, we obtain

$$\begin{aligned}
&\frac{\partial H(1+2pq-p-q) - H(q)}{\partial q} \\
&= \frac{\partial}{\partial q}\left[-(1+2pq-p-q)\log(1+2pq-p-q) - (p+q-2pq)\log(p+q-2pq) + q\log(q) + (1-q)\log(1-q)\right] \\
&= (1-2p)\log\left(\frac{1+2pq-p-q}{p+q-2pq}\right) + \log\left(\frac{q}{1-q}\right)
\end{aligned}$$

Since $\frac{\partial}{\partial q} I(X;Y) \leq 0$ for all $0 \leq p \leq 1/2$, the jammer will take $q = 1/4$ to minimize the $I(X;Y)$. Then, the sender will want to maximize the following:

$$\begin{aligned}
\max_{0\leq p\leq 1/2}\min_{0\leq q\leq 1/4} I(X;Y) &= \max_{0\leq p\leq 1/2}(1-2p)\log\left(\frac{1+2pq-p-q}{p+q-2pq}\right) + \log\left(\frac{q}{1-q}\right)\Bigg| q = 1/4 \\
&= I(X;Y)|p = 1/2, q = 1/4 \\
&= 1 - H(1/4)
\end{aligned}$$

(e) The maximal number of bits per channel use the power-limited sender can hope to communicate in the presence of a power-limited jammer is $1 - H(1/4)$ bits/channel use.

(f) If the jammer's energy is allowed to equal the sender's energy, the resulting channel will have capacity $C = H(\text{sender energy limit}) - H(\text{jammer energy limit}) = H(1/2) - H(1/2) = 0$. No bits may be reliably communicated!