ECE 534 Information Theory - Midterm 2

Nov.4, 2009. 3:30-4:45 in LH103.

- 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 two 8.5x11" double-sided crib sheets.
- 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 65 points.

Question:	1	2	3	4	Total
Points:	18	17	12	18	65
Score:					

1. *A sum channel.* Let $\mathcal{X} = \mathcal{Y} = \{A, B, C, D\}$ be the input and output alphabets of a discrete memoryless channel with transition probability matrix $p(y|x)$, for $0 \le \epsilon, \delta \le 1$ given by

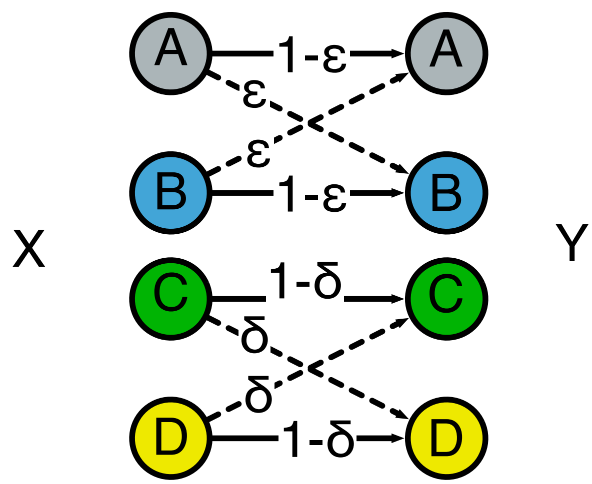
$$p(y|x) = \begin{bmatrix} 1-\epsilon & \epsilon & 0 & 0 \\ \epsilon & 1-\epsilon & 0 & 0 \\ 0 & 0 & 1-\delta & \delta \\ 0 & 0 & \delta & 1-\delta \end{bmatrix}.$$

Notice that this channel with 4 inputs and outputs looks like the sum or "union" of two parallel sub-channels with transition probability matrices

$$p_1(y|x) = \begin{bmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{bmatrix}, \qquad p_2(y|x) = \begin{bmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{bmatrix},$$

with alphabets $\mathcal{X}_1 = \mathcal{Y}_1 = \{A, B\}$ and $\mathcal{X}_2 = \mathcal{Y}_2 = \{C, D\}$ respectively.

(a) (2 points) Draw the transition probability diagram of this channel.



Solution:

(b) (3 points) Find the capacity of this channel if $\epsilon = \delta = 1/2$.

Solution: If $\epsilon = \delta = 1/2$ we have a symmetric channel, whose capacity we know is achieved by a uniform input distribution and has capacity

$$\begin{aligned} C &= \log_2 |\mathcal{Y}| - H(\text{ a row of the transition probability matrix}) \\ &= \log_2(4) - H(1/2, 1/2, 0, 0) \\ &= 2 - 1 = 1 \text{ (bit per channel use)} \end{aligned}$$

(c) (5 points) Let $p(x)$ be the probability mass function on $\mathcal{X}$ and let

$$p(A) + p(B) = \alpha, \qquad p(C) + p(D) = 1 - \alpha.$$

Show that the mutual information between the input X and the output Y may be expressed as

$$I(X;Y) = H(\alpha) + \alpha I(X;Y|X \in \{A, B\}) + (1-\alpha) I(X;Y|X \in \{C, D\}).$$

Points earned: ____________ out of a possible 10 points

Solution: Let θ be a random variable with the following probability mass function:

$$\theta * \begin{cases} 1 & \text{if } x \in \{A,B\} \Rightarrow p(1) = \alpha \\ 0 & \text{if } x \in \{C,D\} \Rightarrow p(0) = 1-\alpha \end{cases}$$

We can then express the mutual information between X and Y as

$$\begin{aligned} I(X;Y) &= I(X;\theta) + I(X;Y|\theta) \\ &= H(\theta) - H(\theta|X) + I(X;Y|\theta) \\ &= H(\alpha) - 0 + p(\theta=1)I(X;Y|\theta=1) + p(\theta=0)I(X;Y|\theta=0) \\ &= H(\alpha) + \alpha I(X;Y|X \in \{A,B\}) + (1-\alpha)I(X;Y|x \in \{C,D\}) \end{aligned}$$

(d) (2 points) Let C_1 and C_2 be the capacities of the subchannels described by $p_1(y|x)$ and $p_2(y|x)$. **Argue** why

$$\max_{p(x)} I(X;Y) = \max_{\alpha}[H(\alpha) + \alpha C_1 + (1-\alpha)C_2].$$

Solution: Notice that

$$C_1 = \max_{p_1(x):x\in\{A,B\}} I(X;Y), \quad C_1 = \max_{p_2(x):x\in\{C,D\}} I(X;Y).$$

Then since $x \in \{A,B\} \to y \in \{A,B\}$ and $x \in \{C,D\} \to y \in \{C,D\}$, we see that

$$\begin{aligned} \max_{p(x):x\in\{A,B,C,D\}} I(X;Y) &= \max_{p(x):x\in\{A,B,C,D\},0\le\alpha\le 1} [H(\alpha) + \alpha I(X;Y|X \in \{A,B\}) + (1-\alpha)I(X;Y|X \in \{C,D\}) \\ &= \max_{0\le\alpha\le 1} [H(\alpha) + \max_{p(x):x\in\{A,B\}} I(X;Y|X \in \{A,B\}) + \max_{p(x):x\in\{C,D\}} I(X;Y|X \in \{C,D\}) \\ &= \max_{0\le\alpha\le 1} [H(\alpha) + \alpha C_1 + (1-\alpha)C_2] \end{aligned}$$

(e) (6 points) Find the capacity C of the sum channel in terms of the capacities C_1 and C_2 of the sub-channels and NO other parameters.

Solution: Problem (d) makes obtaining the capacity significantly easier since we know the capacities of the two binary symmetric channels are $C_1 = 1 - H(\epsilon)$ and $C_2 = 1 - H(\delta)$. Then finding the capacity of the "sum" channel amounts to a 1-D optimization over α, which may be achieved by setting the derivative of $f(\alpha) := H(\alpha) + \alpha C_1 + (1-\alpha)C_2$ to zero to solve for α:

$$\begin{aligned} \frac{df(\alpha)}{d\alpha} &= H'(\alpha) + C_2 - C_1 \\ &= \log_2\left(\frac{1-\alpha}{\alpha}\right) + C_2 - C_1 \\ &= 0 \Rightarrow \alpha^* = \frac{2^{C_1}}{2^{C_2} + 2^{C_1}}, \ \ (1-\alpha^*) = \frac{2^{C_2}}{2^{C_2} + 2^{C_1}} \end{aligned}$$

Substituting this optimal value of α^* back into $f(\alpha)$, we obtain, after simplification that

$$C = \log_2\left(2^{C_1} + 2^{C_2}\right)$$

Points earned: __________ out of a possible 8 points

.

2. *Gaussian channels with interference.* Consider a channel with 2 independent transmitters and a single two-antenna receiver: user 1 transmits X_1 which is independent of the signal X_2 transmitted by user 2. The signals of the two users are received at antenna 1 and 2 as Y_1 and Y_2 respectively as:

$$Y_1 = X_1 + 2X_2 + Z_1,$$
$$Y_2 = X_1 + Z_2,$$

where Z_1 and Z_2 are i.i.d $\mathcal{N}(0, \sigma^2)$ additive white Gaussian noise. The signals of the two users X_1 and X_2 are independent and distributed as

$$X_1 \sim \mathcal{N}(0, P_1), \qquad X_2 \sim \mathcal{N}(0, P_2).$$

Our goal will be to determine the how much X_1 can reliably communicate with the 2-antenna receiver while treating X_2 as noise / interference - which depends on how the receiver processes the signals from the two receive antennas.

Solution: All these questions rely on being able to calculate the different covariance matrices - al of which will be denoted by K with the appropriate subscripts, and using the fact that for random variable (vector) X with covariance matrix K_X:

$$h(X) = \frac{1}{2}\log_2\left((2\pi e)|K_X|\right)$$

(a) (5 points) Compute $I(X_1; Y_1, Y_2)$.

Solution: Here we need K_{Y_1,Y_2} and $K_{Y_1,Y_2|X}$ which can be obtained as

$$\begin{aligned}
K_{Y_1,Y_2} &= \begin{bmatrix} E[Y_1^2] & E[Y_1Y_2] \\ E[Y_2Y_1] & E[Y_2^2] \end{bmatrix} \\
&= \begin{bmatrix} E[(X_1+2X_2+Z_1)^2] & E[(X_1+2X_2+Z_1)(X_1+Z_2)] \\ E[(X_1+Z_2)(X_1+2X_2+Z_1)] & E[(X_1+Z_2)^2] \end{bmatrix} \\
&= \begin{bmatrix} P_1+4P_2+\sigma^2 & P_1 \\ P_1 & P_1+\sigma^2 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
K_{Y_1,Y_2|X} &= \begin{bmatrix} E[Y_1^2|X_1] & E[Y_1Y_2|X_1] \\ E[Y_2Y_1|X_1] & E[Y_2^2|X_1] \end{bmatrix} \\
&= \begin{bmatrix} E[(X_1+2X_2+Z_1)^2|X_1] & E[(X_1+2X_2+Z_1)(X_1+Z_2)|X_1] \\ E[(X_1+Z_2)(X_1+2X_2+Z_1)|X_1] & E[(X_1+Z_2)^2|X_1] \end{bmatrix} \\
&= \begin{bmatrix} 4P_2+\sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}
\end{aligned}$$

Then

$$\begin{aligned}
I(X_1; Y_1, Y_2) &= h(Y_1, Y_2) - h(Y_1, Y_2|X_1) \\
&= \frac{1}{2}\log_2((2\pi e)^2|K_{Y_1,Y_2}|) - \frac{1}{2}\log_2((2\pi e)^2|K_{Y_1,Y_2|X_1}|) \\
&= \frac{1}{2}\log_2\left(\frac{(P_1+4P_2+\sigma^2)(P_1+\sigma^2)-P_1^2}{(4P_2+\sigma^2)\sigma^2}\right) \\
&= \frac{1}{2}\log_2\left(1+\frac{P_1}{\sigma^2}\frac{4P_2+2\sigma^2}{4P_2+\sigma^2}\right)
\end{aligned}$$

Points earned: ____________ out of a possible 5 points

(b) (5 points) The receiver now decides to process the received signals Y_1 and Y_2 and see if/how it affects the optimal communication rate with X_1. Let $Y_b = Y_1 + Y_2$ (the receiver sums the received signals). Compute $I(X_1; Y_b)$.

Solution: Here we need $K_{Y_1+Y_2}$ and $K_{Y_1+Y_2|X}$ which can be obtained as

$$\begin{aligned} K_{Y_1+Y_2} &= E[(Y_1+Y_2)^2] \\ &= E[(2X_1+2X_2+Z_1+Z_2)^2] \\ &= 4P_1+4P_2+2\sigma^2 \end{aligned}$$

$$\begin{aligned} K_{Y_1+Y_2|X_1} &= E[(Y_1+Y_2)^2|X_1] \\ &= E[(2X_1+2X_2+Z_1+Z_2)^2|X_1] \\ &= E[(2X_2+Z_1+Z_2)^2] \\ &= 4P_2+2\sigma^2 \end{aligned}$$

Then

$$\begin{aligned} I(X_1; Y_1+Y_2) &= h(Y_1+Y_2) - h(Y_1+Y_2|X_1) \\ &= \frac{1}{2}\log_2((2\pi e)|K_{Y_1+Y_2}|) - \frac{1}{2}\log_2((2\pi e)|K_{Y_1+Y_2|X_1}|) \\ &= \frac{1}{2}\log_2\left(\frac{4P_1+4P_2+2\sigma^2}{4P_2+2\sigma^2}\right) \\ &= \frac{1}{2}\log_2\left(1+\frac{P_1}{\sigma^2}\frac{2\sigma^2}{2P_2+\sigma^2}\right) \end{aligned}$$

(c) (4 points) Is Y_b a sufficient statistic for decoding X_1?

Solution: Y_b is a sufficient statistic if $I(X_1; Y_1, Y_2) = I(X_1, Y_b)$ (and we know that by the data processing inequality that $I(X_1; Y_b) \leq I(X_1; Y_1, Y_2)$ - sufficient statistics lose no information! However, it is not a sufficient statistic since

$$\frac{4P_2+2\sigma^2}{4P_2+\sigma^2} - \frac{2\sigma^2}{2P_2+\sigma^2} = \frac{8P_2}{(4P_2+\sigma^2)(2P_2+\sigma)^2} > 0, \quad \forall P_2 > 0$$

Hence $I(X_1; Y_b) < I(X_1; Y_1, Y_2)$, strictly and Y_b is NOT a sufficient statistic.

(d) (2 points) The receiver now decides to try and decode X_1 using only Y_2, ignoring Y_1. Compute $I(X_1; Y_2)$.

Solution:

$$\begin{aligned} I(X_1; Y_2) &= h(Y_2) - h(Y_2|X_1) \\ &= \frac{1}{2}\log_2((2\pi e)|K_{Y_2}|) - \frac{1}{2}\log_2((2\pi e)|K_{Y_2|X_1}|) \\ &= \frac{1}{2}\log_2\left(\frac{P_1+\sigma^2}{\sigma^2}\right) \\ &= \frac{1}{2}\log_2\left(1+\frac{P_1}{\sigma^2}\right) \end{aligned}$$

(e) (1 point) Find an example of powers P_1 and P_2 for which $I(X_1; Y_2) > I(X_1; Y_b)$.

Solution: For the solution in part (d) to be larger than that in part (c) we need

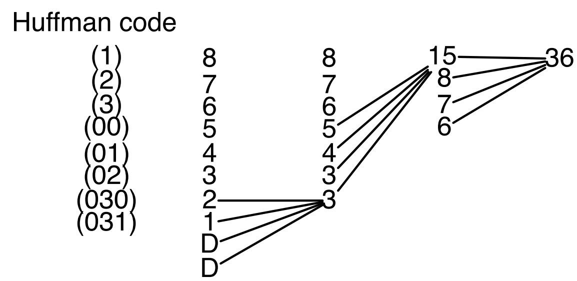
$$\frac{2\sigma^2}{2P_2 + \sigma^2} < 1 \Rightarrow P_2 > \frac{\sigma^2}{2}$$

Points earned: ________ out of a possible 1 points

3. *True or false (T/F) and short answer.*

(a) (2 points) Find the 4-ary Huffman code ($D = 4$) for the source with probability mass function $(\frac{8}{36}, \frac{7}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}, \frac{1}{36})$.

Solution: For this one, the key is to realize that the optimal Huffman code will have 2 Dummy (0 probability) extra symbols, resulting in the following Huffman tree:



(b) (2 points) Describe the meaning and use of the rate-distortion function in two sentences - I'm looking for meaning and intuition rather than formulas.

Solution: Each rate-distortion pair (R,D) on the rate-distortion function R(D) describes the minimal achievable rate (number of bits per source symbol) needed to represent the source under consideration to within an expected distortion of D. The R(D) function is useful in lossy (non-perfect) compression or sources.

(c) (2 points) Compare the capacities C_1 and C_2 of the channels where for the first channel, $Y_1 = (X_1 \mod 10)$ for $\mathcal{X}_1 = \{1, 2, \cdots, 100\}$ and for the second channel $Y_2 = (X_2 \mod 9)$ for $\mathcal{X}_2 = \{1, 2, \cdots, 90\}$.

Solution: By symmetry, uniform inputs will achieve uniform outputs and so the capacity $C_1 = \log|\mathcal{Y}_1| = \log(10)$ which is greater than the capacity $C_2 = \log|\mathcal{Y}_2| = \log(9)$.

(d) (1 point) T/F: perfect feedback can increase the capacity of a discrete channel with memory.

Solution: True.

(e) (2 points) T/F: the differential entropy $h(X_2)$ of a continuous random variable X_2 which is uniform on $[2a, 2b]$ is twice the differential entropy $h(X_1)$ of a continuous random variable X_1 which is uniform on $[a, b]$.

Solution: False. $h(X_2) = \log(2b-2a) = \log(2(b-a)) = 1+\log(b-a)$, while $h(X_1) = \log(b-a)$. This statement is only true if $(b-a) = 1$ but not in general.

(f) (3 points) Outline, in about 3-4 sentences, the main points/techniques used in the achievability proof of the channel coding theorem.

Solution: The main proof techniques are random coding, joint typicality decoding and bounding the probability of error using what we know about how likely it is for two independently chosen sequences to be jointly typical. The achievability proof follows these main lines:

- Generate a random codebook, which consists of 2^{nR} sequences of length n, were each element is generated i.i.d. according to $p(x)$.
- Encoding of message w takes place by looking up the w-th sequence in the codebook and sending that.
- The receiver decodes w using a *joint typicality decoder* whereby it declares w was sent if there exists one and only one w such that $x^n(w)$ and the received y^n re jointly typical. Otherwise it declares an error.
- Analyze the probability of error in this scheme - we need to show that the maximal probability of error decays to 0 as the blocklength n tends to infinity. This will follow from arguing that the probability that independently generated x^n and y^n have a probability of $2^{-n(I(X;Y)-\epsilon)}$ of being jointly typical - and there are maximally $2^{nR}-1$ independently generated typical x^n which are no the correct one.
- Last part of the proof involves showing that if the *average* probability of error decaying to 0 over all randomly chosen codebooks implies that there exists a codebook whose maximal probability of error will also decay to zero.

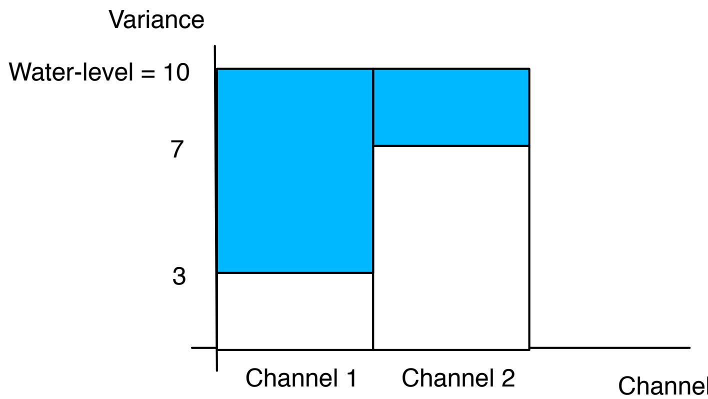
4. *Capacity of several simple channels.* Find the capacity of the following channels AND the input distribution which achieves this capacity.

   (a) (4 points) Consider 2 parallel Gaussian channels $Y_1 = X_1 + Z_1$ and $Y_2 = X_2 + Z_2$ where Z_1 and Z_2 are independent, zero mean additive white Gaussian noise $Z_1 \sim \mathcal{N}(0, \sigma_1^2 = 3)$ and $Z_2 \sim \mathcal{N}(0, \sigma_2^2 = 7)$ subject to a total power constraint of $P = 10$.

Solution: This is the classical, simplest waterfilling solution. Draw it out and see that we will "fill" 7 units of power into channel 1 and 3 units of power into channel 2, thereby obtaining a capacity of

$$C = \frac{1}{2}\log_2\left(1 + \frac{7}{3}\right) + \frac{1}{2}\log_2\left(1 + \frac{3}{7}\right),$$

which is achieved by an input distribution $X_1 \sim \mathcal{N}(0, 7)$ and $X_2 \sim \mathcal{N}(0, 3)$, and X_1, X_2 indepen-



dent.

   (b) (6 points) Again consider 2 parallel Gaussian channels, but now you're *constrained* to send the same signal X on both channels, i.e. $Y_1 = X + Z_1$ and $Y_2 = X + Z_2$ where Z_1 and Z_2 are independent, zero mean additive white Gaussian noise $Z_1 \sim \mathcal{N}(0, \sigma_1^2 = 3)$ and $Z_2 \sim \mathcal{N}(0, \sigma_2^2 = 7)$ subject to a total power constraint of $P = 10$.

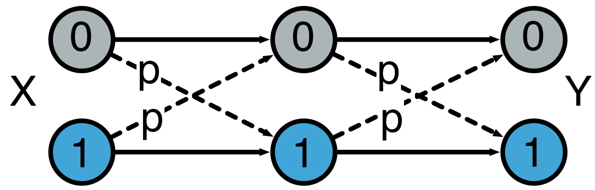
Solution: You can proceed with waterfilling, but it's easier if you just go directly from the capacity since due to out power constraint and due to the fact that we are forced to send the same signal on both channels - we know that $P_1 = P_2 = P/2 = 5$. Then, keeping things symbolic as long as possible for illustration, and noting that we are inputing Gaussian random variables (so the output are Gaussian as well)

$$K_{Y_1,Y_2} = \begin{bmatrix} \frac{P}{2} + \sigma_1^2 & \frac{P}{2} \\ \frac{P}{2} & \frac{P}{2} + \sigma_2^2 \end{bmatrix}, \quad K_{Y_1,Y_2|X} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

$$\begin{aligned} \max_{p(x)} I(X; Y_1, Y_2) &= h(Y_1, Y_2) - h(Y_1, Y_2|X) \\ &= \frac{1}{2}\log_2\left(\frac{(\frac{P}{2} + \sigma_1^2)(\frac{P}{2} + \sigma_2^2) - \frac{P^2}{4}}{\sigma_1^2\sigma_2^2}\right) \\ &= \frac{1}{2}\log_2\left(1 + \frac{\frac{P}{2}(\sigma_1^2 + \sigma_2^2)}{\sigma_1^2\sigma_2^2}\right) \end{aligned}$$

So, $C = \frac{1}{2}\log_2\left(1 + \frac{5(3+7)}{3\cdot 7}\right)$, achieved by $X \sim \mathcal{N}(0, 5)$.

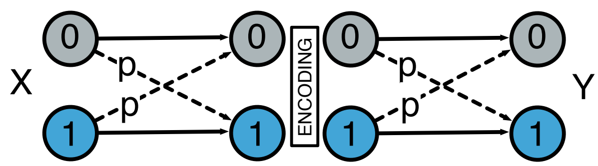
(c) (4 points) Cascade of two binary symmetric channels with crossover probability p *without* encoding between stages:



Solution: A cascade of binary symmetric channels is again a binary symmetric channel with a new corssover probability $p' = (1-p)\cdot p + p\cdot(1-p) = 2p(1-p)$. Then the optimal input distribution is uniform $p(x) = \{1/2, 1/2\}$ and the capacity is

$$C = 1 - H(2p(1-p).$$

(d) (4 points) Cascade of two binary symmetric channels with crossover probability p *with* encoding between stages:



Solution: When we can re-encode the symbols after the first BSC, the capacity becomes $C = \min(C_1, C_2\}$, where C_1 and C_2 are the capacities of the first and second BSCs, respectively. So, we see that, by symmetry, $C = 1 - H(p)$, achieved for $p(x) = \{1/2, 1/2\}$, uniform.