ECE 534 Information Theory - MIDTERM 2

11/06/2013, LH 207.

- This exam has 4 questions, each of which is worth approximately 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 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: SOLUTIONS

Your UIN: ______________________

Your signature: ______________________

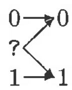
The exam has 4 questions, for a total of 100 points.

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

1. *Short calculations.*

   (a) (5 points) Consider three independent random variables U, V, W with entropies H_U, H_V, H_W. Let $X = (U, V)$ and $Y = (V, W)$. What is $H(X, Y)$? What is $H(X|Y)$? What is $I(X; Y)$?

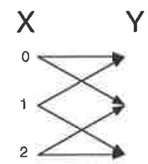
   (b) (5 points) Find the capacity of the following channel.



$$P(y=0\,|\,x=0) = 1; \quad P(y=0\,|\,x=?) = 1/2; \quad P(y=0\,|\,x=1) = 0;$$
$$P(y=1\,|\,x=0) = 0; \quad P(y=1\,|\,x=?) = 1/2; \quad P(y=1\,|\,x=1) = 1.$$

   (c) (5 points) Show an example of distribution $f(x)$ whose differential entropy $h(X)$ is negative.

   (d) (5 points) What is the capacity of the following channel?



   (e) (5 points) Find the differential entropy of $Y = 2X + 5$ if X is Gaussian with mean -1 and variance 10.

(a) $H(X,Y) = H(U,V,V,W) = H(U,V,W) = H(U) + H(V|U) + H(W|V,U)$

$= H(U) + H(V) + H(W)$ since independent

$= H_U + H_V + H_W$

$H(X|Y) = H(U,V|V,W) = H(U|V,W) \underset{\text{independent}}{=} H(U) = H_U$

$I(X;Y) = H(X) - H(X|Y) = H(U,V) - H(U) = H(U) + H(V) - H(U) = H(V) = H_V$

(b) By inspection, if ignore (don't use) the ? symbol, can transmit reliably at 1 bit/channel use. Clearly this is the maximum we can obtain as $C \le H(X)$ and $C \le H(Y) = 1$ bit.

(c) A random variable uniform on $[a,b]$ (denote as $\mathcal{U}[a,b]$) has differential entropy

$h(\mathcal{U}[a,b]) = \log_2(b-a)$ (bits).

Taking $\mathcal{U}[0, 1/2]$ gives $h(\mathcal{U}[0,1/2]) = \log_2(1/2) = -1$

(d) The transition probability matrix is given by

$$\begin{array}{c} \\ X=0 \\ X=1 \\ X=2 \end{array} \begin{array}{c} \begin{array}{ccc} Y=0 & Y=1 & Y=2 \end{array} \\ \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix} \end{array}$$

This is symmetric and hence

$C = \log|\mathcal{Y}| - H(\text{row of this matrix})$

$= \log(3) - H((1/2, 1/2, 0))$

$= \log(3) - \log(2) = \log(3/2)$

(e) $h(Y) = h(2X+5)$ translation does not affect diff. entropy

$= h(2X)$

$= h(X) + \log|2|$

$= \frac{1}{2}\log(2\pi e \cdot 10) + \log(2)$

2. *Maximum entropy distributions*

(a) (9 points) Find the distribution $f(x)$ on support $\mathcal{S}_X = [0, \infty)$ that maximizes the differential entropy $h(X)$ subject to a constraint on the mean, $E[X] = \mu$ for a non-negative constant μ. *HINT: hopefully you can recognize the form.*

(b) (8 points) Find the discrete probability mass function on alphabet $\{-3, -1, 0, 1, 3\}$ that maximizes the entropy $H(X)$ subject to a constraint on the mean, $E[X] = 0$.

(c) (8 points) Find the distribution $f(x)$ of random variable X on support $\mathcal{S}_X = (-\infty, \infty)$ that maximizes the mutual information $I(X;Y)$ if $Y = 2X + 5N$, where $N \sim \mathcal{N}(0, 100)$ subject to the constraint that $E[|X|^2] \leq 1000$, where X and N ~~may be arbitrarily correlated~~ are independent.

(a) $\bullet\ S = [0, \infty)$

$\bullet\ \int_0^\infty f(x)\,dx = 1$

$\bullet\ \int_0^\infty \underbrace{x}_{r_1(x)} f(x)\,dx = \underbrace{\mu}_{\alpha_1}$

Our Theorem gives us that the form of the maximizing distribution is

$$f(x) = e^{\lambda_0 + \lambda_1 \cdot x} \quad \text{on} \quad x \in [0, \infty)$$

for some constants λ_0, λ_1 which must be found to satisfy 2 constraints.

By inspection, recognize this as an exponential RV and since we want its mean to be μ, we have

$$f(x) = \frac{1}{\mu} e^{-\frac{1}{\mu} x} \quad \text{for } x \in [0, \infty)$$

$\left(e^{\lambda_0} = \frac{1}{\mu},\ \lambda_1 = -\frac{1}{\mu} \right)$. Can also calculate this, but takes a long time!

(b) ~~Luckily~~ Uniforms maximize discrete entropy over a given alphabet of and luckily, our mean constraint will be satisfied if we take

$$p(x) = 1/5 \text{ for } x = \{-3, -1, 0, 1, 3\}.$$

This will clearly be entropy maximizing and have entropy $\log_2(5)$.

Points earned: ________ out of a possible 25 points

(c) This is a Gaussian channel; as seen in class so we will want to take $2X$ to be Gaussian, ie. X to be zero-mean Gaussian satisfying the $E[|X|^2] = 1000$, or

$$X \sim \mathcal{N}(0, 1000).$$

The factors do not alter this.

3. *Short answers*

   (a) (~~5~~ 6 points) What does it mean for a codebook to be "randomly generated"?

   (b) (~~5~~ 6 points) What is the difference between an achievability proof and a converse proof?

   ~~(c) (5 points) Encode the bits 0110 using the (7,4) Hamming code.~~ Explain how you can detect and or correct ~~errors~~ errors using a Hamming ccode.

   (d) (~~5~~ 7 points) ~~For the codeword obtained in part 3), introduce errors in the 1st and the 2nd bits. Explain whether you can detect and/or correct these errors~~ and if so, how.

   (e) (~~5~~ 6 points) How is Fano's inequality used in the channel coding theorem?

(a) It means that each of the $1, 2, \ldots, 2^{nR}$ codewords of length n has each ~~element of the~~ symbol of the codeword generated i.i.d. according to some distribution $p(x)$.

(b) In an achievability proof, we demonstrate an encoding/decoding technique and show that as long as the rate is below a certain value, we can drive $P_e \to 0$ (P_e = probability of error).

In a converse, we show that if we have a code whose $P_e \to 0$, then necessarily, the rate of this code has to be bounded by a certain value.

(c) ~~If the Hamming code has parity check matrix~~ (can define a few)

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \Rightarrow \text{~~book gives us the codewords~~}$$

The Hamming code always has $d_{min} = 3$, meaning it can detect up to 2 errors and correct up to 1 error.

To see how, say we receive a vector Y which we know has 1 error. You can multiply this by the parity check matrix and if this is non-zero, the result $Y \cdot H$ ~~given~~ indicates where the error is.

If y is not a codeword ~~and have 2 errors~~, can detect up to 2 errors

(d) Fano's inequality (or a lemma thereof) is used in the converse of the channel coding Theorem to relate rates, Probability of error, and entropy/mutual information as follows:

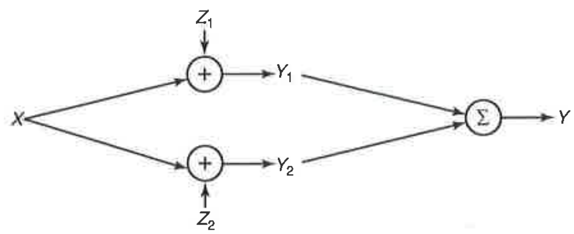
$$nR = nH(W)$$
$$= n[H(W|\hat{W}) + I(W;\hat{W})]$$
$$\le n[1 + P_e^{(n)} \cdot nR + I(W;\hat{W})]$$

Fano $H(W|\hat{W}) \le 1 + P_e^{(n)} \cdot nR$

4. *Gaussian channels*

   (a) (12 points) In class we looked at the capacity of an additive white Gaussian noise channel $Y = X + Z$, where $N \sim \mathcal{N}(0, N)$ and X, Z are independent subject to an *input* power constraint $E[|X|^2] \leq P$. Find the capacity of the same channel if we have an *output* power constraint $E[|Y|^2] \leq P$ instead of the input power constraint $E[|X|^2] \leq P$.

   (b) (13 points) Now consider a Gaussian noise channel with input power constraint P as in the figure below, where the signal takes two different paths and the noisy signals are added together at the destination (that is, $Y = Y_1 + Y_2$).



   Find the capacity of this channel if Z_1 and Z_2 are jointly Gaussian with covariance matrix, for $-1 \leq \rho \leq 1$:

$$K_Z = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

(a)
$$C = \max_{f(x):\, E((X+Z)^2) \leq P} I(X;Y)$$

$$= \max_{f(x):\, E((X+Z)^2) \leq P} h(Y) - h(Y|X)$$

$$= \max_{f(x):\, E((X+Z)^2) \leq P} h(Y) - h(Z)$$

Given a constraint on the output power of Y, the maximal differential entropy is achieved by a Gaussian of that power, which can

$Y \sim N(0, P)$

in turn be achieved by takin $X \sim \mathcal{N}(0, P-N)$ as $Y = X + Z$, and $Z \sim N(0, N)$.

(b) Let's write Y directly in terms of X:

$$Y = Y_1 + Y_2 = X + Z_1 + X + Z_2 = 2X + (Z_1 + Z_2)$$

The power constraint on the new input $2X$ is $4P$ ($Var(2X) = 4 Var X \leq 4P$.)

$Z_1 + Z_2$ is again Gaussian with mean 0 and variance $Var(Z_1+Z_2) = E[(Z_1+Z_2)^2] = E[Z_1^2] + 2E[Z_1 Z_2] + E[Z_2^2]$

$$= \sigma^2 + 2\rho\sigma^2 + \sigma^2$$ from $K_Z = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$

$$= 2\sigma^2 + 2\rho\sigma^2$$

Thus,

$$Y = 2X + \underbrace{(Z_1 + Z_2)}_{Z'}$$

$$\Rightarrow \quad Y = 2X + Z'$$

with $E[(2X)^2] \leq 4P$, $Z' \sim N(0, 2\sigma^2 + 2\rho\sigma^2)$

The capacity of this Gaussian channel is, as seen in class, given by

$$C = \frac{1}{2}\log\left(1 + \frac{4P}{2\sigma^2(1+\rho)}\right) = \frac{1}{2}\log\left(1 + \frac{2P}{\sigma^2(1+\rho)}\right)$$