ECE 534 Information Theory - MIDTERM 2

11/01/2017, LH 103.

- 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 6 questions, for a total of 100 points.

Question:	1	2	3	4	5	6	Total
Points:	12	18	21	22	15	12	100
Score:							

ECE534 Fall 2017 EXAM 2 Name:_____

1. *Hamming code.*

The $(7,4)$ Hamming code has parity check matrix

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

(a) (4 points) How many codewords does the $(7,4)$ Hamming code have?

(b) (8 points) ~~Decode the~~ following received strings: 1101011 and 1111111
(correct the single errors (if any) in the

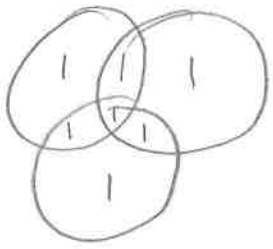
(a) Nullspace of H has dimension/rank 4 $\Rightarrow 2^4$ codewords.

(b) Let y denote the received string. $H \cdot y$ should indicate the column in error:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \Rightarrow$$ means 4th element in error, so decode as

1100011

For 2nd one, let us use an alternative technique, Venn diagrams:



$\rightarrow$ all parities check so correct!

1111111

Points earned: _____ out of a possible 12 points

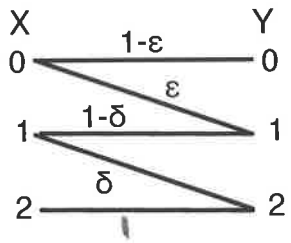
2. *Channel capacity.*

(a) (6 points) How is random coding used in the channel coding theorem?

Shannon used random coding to prove the existence of a code achieving a certain rate without having to construct it explicitly.

Randomly generate a code rather than construct one. If average code behavior is good (small prob. of error) => there exists a code with small prob. of error.

(b) (12 points) Find the equation to be solved to obtain the capacity of the following channel (simplify somewhat, but do not solve it!):



let $p(x=0)=p$, $p(x=1)=q$, $p(x=2)=1-p-q$.

We need an expression for $I(X;Y)$ in terms of $p, q, \varepsilon, \delta$ and will then seek to optimize this with respect to p and $q \in (0,1)$

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

$p(y=0) = p(1-\varepsilon)$, $p(y=1) = p\cdot\varepsilon + q(1-\delta)$, $p(y=2) = q\cdot\delta + 1-p-q$

$H(Y) = -p(y=0)\log p(y=0) - p(y=1)\log(p(y=1)) - p(y=2)\log(p(y=2))$

$H(Y|X)$ is obtained as:

$H(Y|X) = H(Y|X=0)p(x=0) + H(Y|X=1)p(x=1) + H(Y|X=2)p(x=2)$

$= H(\varepsilon)\cdot p + H(\delta)\cdot q + 0$

Now, write $I(X;Y)$ as a function of $p, q, \varepsilon, \delta$ and to optimize, set $\frac{\partial I(X;Y)}{\partial p} = \frac{\partial I(X;Y)}{\partial q} = 0$. (2 equations in 2 unknowns)

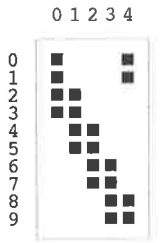
Points earned: _____ out of a possible 14 points

(c) (4 points) What is the capacity of the five-input, ten output channel whose transition probability matrix is given below?

inputs

outputs

$$\begin{bmatrix} 0.25 & 0 & 0 & 0 & 0.25 \\ 0.25 & 0 & 0 & 0 & 0.25 \\ 0.25 & 0.25 & 0 & 0 & 0 \\ 0.25 & 0.25 & 0 & 0 & 0 \\ 0 & 0.25 & 0.25 & 0 & 0 \\ 0 & 0.25 & 0.25 & 0 & 0 \\ 0 & 0 & 0.25 & 0.25 & 0 \\ 0 & 0 & 0.25 & 0.25 & 0 \\ 0 & 0 & 0 & 0.25 & 0.25 \\ 0 & 0 & 0 & 0.25 & 0.25 \end{bmatrix}$$



?

(check in same as Cover + Thomas)

This is a symmetric channel, hence

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} H(Y) - H(Y|X)$$

$$= \log(10) - \underbrace{\log(4)}_{H(\text{row})}$$

since uniform input ⇒ uniform output

$$= \log(2.5)$$

Points earned: _____ out of a possible 4 points

3. *Converses.*

(a) (4 points) What is the difference between an "achievability" proof and a "converse" proof?

(4 points)

(a) What is a memoryless channel?

A channel for which

$$p(y_n|y_1,\dots,y_{n-1},x_1,\dots,x_n) = p(y_n|x_n) \quad \text{for all } n.$$

(b) (5 points) Here is the channel coding converse for discrete memoryless channels without feedback. Justify all lettered steps (e.g. (a) follows from....)

$$\begin{aligned} nR &\overset{(a)}{=} H(W) \\ &\overset{(b)}{=} H(W|\hat{W}) + I(W;\hat{W}) \\ &\overset{(c)}{\le} 1 + P_e^{(n)} nR + I(W;\hat{W}) \\ &\overset{(d)}{\le} 1 + P_e^{(n)} nR + I(X^n;Y^n) \\ &\overset{(e)}{\le} 1 + P_e^{(n)} nR + nC \end{aligned}$$

(a) Messages uniform on $\{1,2,\dots,2^{nR}\}$

(b) Always true as $I(W;\hat{W}) = H(W) - H(W|\hat{W})$

(c) Fano's inequality: $W \to X^n \to Y^n \to \hat{W}$

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

(d) Data processing inequality $I(W;\hat{W}) \le I(X^n;Y^n)$

given the Markov chain $W \to X^n \to Y^n \to \hat{W}$

(e) $I(X^n;Y^n) \le nC$ shown in class, follows from memorylessness of channel

Points earned: _____ out of a possible 9 points

(c) (4 points) How does the definition of a channel code change when the encoder has access to perfect channel output feedback? Give the formal definition of a code with feedback.

NO-FB: $(2^{nR}, n)$ code → message $W \in \{1, 2, \ldots, 2^{nR}\}$ uniformly
→ encoding function of W only: $W \to x^n$
→ decoding function: $y^n \to \hat{W}$

FB: $(2^{nR}, n)$ code → message $W \in \{1, 2, \ldots, 2^{nR}\}$
→ encoding function at channel use i a function of W and Y^{i-1}
→ decoding function: $y^n \to \hat{W}$

(d) (8 points) Alter the converse in (b) to show a converse for discrete memoryless channels with perfect channel output feedback.

(5 points)

(d) Does feedback increase capacity of a channel? State why or why not (or for what types of channels it may).

Feedback does not increase the capacity of a discrete memoryless channel. It may increase capacity of channels with memory.

Points earned: _____ out of a possible 12 points

4. *Source coding.*

(a) (6 points) What does the source coding theorem say? Write it ~~mathematically, and~~ in your own words.

An iid source can be compressed down to its entropy. A source generating bits iid according to $p(x)$ can be represented with on average $H(X) = -\sum p(x) \log p(x)$ bits.

(b) (4 points) Is Huffman coding optimal? If so, in what precise sense? If not, explain why not.

Huffman coding is a compression method that yields the shortest expected length prefix code for a given distribution. Optimal in the the "shortest expected codeword length prefix code" sense.

(c) (4 points) State the Kraft inequality.

For any instantaneous (prefix) code over an alphabet of size D, the codeword lengths $l_1, l_2, \ldots, l_m$ must satisfy the inequality

$$\sum D^{-l_i} \leq 1$$

Conversely, given a set of codeword lengths satisfying the inequality, there exists a prefix code with these lengths.

(d) (4 points) Give a practical scenario in which one would use the Kraft inequality.

To check whether desired codeword lengths are even possible for a uniquely decodable code!

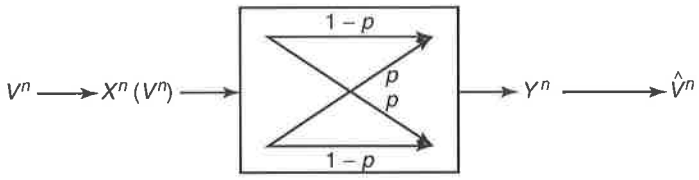
(e) (4 points) How did we use the Kraft inequality in class (to prove what?)?

Optimality of Huffman coding

Points earned: _____ out of a possible 22 points

5. (15 points) *Source channel separation.*

We wish to encode a source that outputs a string of bits V^n i.i.d. according to the Bernoulli(α) distribution, i.e. the probability of outputting a 1 is α and send this over a binary symmetric channel with crossover probability p, as shown below.



Find conditions on α and p so that the probability of error $P(\hat{V}^n \neq V^n)$ can be made to go to zero as $n \to \infty$.

Via source-channel separation, we can send V^n over this channel if the entropy rate of the source is below the channel's capacity.

Here, $H(\mathcal{V}) = \lim_{n\to\infty} \frac{1}{n} H(V_1, V_2, \ldots, V_n) = H(V) = \underbrace{H(\alpha)}_{\text{binary entropy function}}$

The channel capacity is given by $1 - H(p)$ (know that capacity of a BSC is $1 - H(p)$).

Thus, if

$$H(\alpha) < 1 - H(p)$$

we can transmit this source over this channel.

Points earned: _____ out of a possible 15 points

ECE534 Fall 2017 EXAM 2 Name:_____

6. Differential entropy.

(a) (6 points) Find the density that maximizes the differential entropy over all random variables X with non-negative support (i.e. can take on non-negative values) and with $E[X] = \mu$.

(form and equations to solve, no need to be explicit)

Via our theorem, the form is

$$f(x) = e^{\lambda_0 + \lambda_1 x}$$

where λ_0, λ_1 selected so as to satisfy

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

$$\int_0^\infty x f(x)\,dx = \mu.$$

(b) (6 points) Under what constraints does a Gaussian variable with zero mean and variance of 9 maximize the differential entropy (if ever)?

$N(0, 9)$ maximizes differential entropy under the constraint

$$E[X] = 0$$

$$E[X^2] \leq 9$$

(Gaussians max entropy under zero mean and variance constraints)

Points earned: _____ out of a possible 12 points