# ECE 534 Information Theory - FINAL EXAM

Dec.6, 2017, 4-5:15pm in LH 107.

- 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 three 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: | 20 | 30 | 25 | 25 | 100 |
| Score: | | | | | |

ECE534 Fall 2017 FINAL EXAM Name:______________________________

1. *Multiple access channels.*

   Consider the following multiple access channel, with inputs $X_1 \in \{2, 4\}$, $X_2 \in \{1, 2\}$ and output

$$Y = X_1^{X_2}.$$

   (a) (15 points) Find the capacity region of this multiple access channel.

   (b) (5 points) Suppose the range of $X_1$ is $\{1, 2\}$ instead. Does the capacity region get smaller? Why or why not? → then $\begin{array}{c|cc} Y & X_2{:}\ 1 & 2 \\ \hline X_1{:}\ 1 & 1 & 1 \\ 2 & 2 & 4 \end{array}$ → see next page.

(a)

$$\begin{array}{c|cc} Y & X_2{:}\ 1 & 2 \\ \hline X_1{:}\ 2 & 2 & 4 \\ 4 & 4 & 16 \end{array}$$

$R_1 \le I(X_1; Y|X_2) = I(X_1; X_1^{X_2}|X_2) = H(X_1) = 1$

$R_2 \le I(X_2; Y|X_1) = I(X_2; X_1^{X_2}|X_1) = H(X_2) = 1$

$R_1 + R_2 \le I(X_1, X_2; Y) = H(Y) - H(Y|X_1, X_2) = H(Y) = 1.5$

where the $R_1+R_2$ bound follows ~~gon~~ when one user transmits at 1 bit/use and the other user treats the channel as a binary erasure channel with capacity $1 - p_{erasure} = 1 - \frac{1}{2} = \frac{1}{2}$ bits/channel use.

To show that no larger rate is achievable we can try to optimize $H(Y)$ for arbitrary input distribution $p(x_1)p(x_2)$, parametrized by:

$X_1 = \begin{cases} 2 & \text{prob. } p \\ 4 & \text{prob. } 1-p \end{cases}$ $X_2 = \begin{cases} 1 & \text{prob. } q \\ 2 & \text{prob. } 1-q. \end{cases}$

Then $P(Y=2) = p \cdot q$, $P(Y=16) = (1-p)(1-q)$, $P(Y=4) = p(1-q) + (1-p) \cdot q$

Then

$H(Y) = -pq\log(pq) - (1-p)(1-q)\log(1-p)(1-q) - [p(1-q) + (1-p) \cdot q]\log[p(1-q)+(1-p)\cdot q]$

By symmetry, see that $p = q$ must optimize this and furthermore that $p = q = 1/2$ does so, for

Points earned: ________________ out of a possible 20 points

$H(Y) = +\frac{1}{4}\log 4 + \frac{1}{4}\log 4 + \frac{1}{2}\log 2 = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{2}$

(b)

| $X_1$ | $X_2$ | $Y$ |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 1 | 2 |
| 1 | 2 | 1 |
| 2 | 2 | 4 |

Note when $X_1 = 1$, $X_2$ has no effect on $Y$ and can not be recovered given $X_1$ and $Y$. If $X_1 \sim Br(\alpha)$ and $X_2 \sim Br(\beta)$ then:

$$\begin{aligned}
R_1 &\leq I(X_1; Y|X_2) = I(X_1; X_1^{X_2}|X_2) = H(\alpha) \\
R_2 &\leq I(X_2; Y|X_1) = H(Y|X_1) - H(Y|X_1, X_2) = H(Y|X_1) \\
&= p(X_1 = 1)H(Y|X_1 = 1) + p(X_1 = 2)H(Y|X_1 = 2) \\
&= \alpha H(\beta) \\
R_1 + R_2 &\leq I(X_1, X_2; Y) = H(Y) - H(Y|X_1, X_2) = H(Y) \\
&= H(\alpha\beta, \overline{\alpha}\overline{\beta}, 1 - \alpha\beta - \overline{\alpha}\overline{\beta}) = H(\alpha) + \alpha H(\beta)
\end{aligned}$$

We may choose $\beta = \frac{1}{2}$ to maximize the above bounds, giving

$$\begin{aligned}
R_1 &\leq H(\alpha) \\
R_2 &\leq \alpha \\
R_1 + R_2 &\leq H(\alpha) + \alpha
\end{aligned}$$

Above, we plot the region for $X_1 \in \{2, 4\}$ (solid line) against that when $X_1 \in \{1, 2\}$ (dotted). What we find is that, surprisingly, the rate region from the first case is not reduced in the second. In fact, neither region contains the other, so for each version of this channel, there are achievable rate pairs which are *not* achievable in the other.

2. *Short answers.*

(a) (4 points) How does the expected codeword length produced by a Huffman code for a source $X$ with probability mass function $p(x)$ compare with its entropy $H(x)$?

For symbol by symbol compression, the optimal expected codeword length $E(L)$ is bounded as

$$H(X) < E(L) < H(X) + 1$$

(b) (5 points) When compressing a source, when is useful to group together source symbols and code over larger blocks?

In general it is better (more efficient) to code over larger blocks, as $H(X) < E(L) < H(X) + \frac{1}{n}$ for blocks of length $n$.

The exception is when your source is Dyadic (probabilities are of the form $\left(\frac{1}{D}\right)^i$ for $i$ integer) in which case $E(L) = H(X)$ even for 1 symbol
discrete

(c) (4 points) Suppose I give you a $(2^{nR_1}, 2^{nR_2}, n)$ code for the multiple access channel with inputs $X_1, X_2$ and output $Y$, whose probability of error is less than some $\epsilon > 0$. What can we say about the conditional entropy $H(X_1^n|Y^n, X_2^n)$? Justify any steps.

$$H(X_1^n|Y^n, X_2^n) \leq H(X_1^n, X_2^n|Y^n) \leq H(P_e) + P_e \log |\mathcal{X}_1||\mathcal{X}_2| = H(P_e) + P_e \log n(R_1+R_2) \Rightarrow \boxed{\text{small}}$$

Since $H(X_1^n, X_2^n|Y^n) = \underbrace{H(X_2^n|Y^n)}_{\geq 0 \text{ since discrete entropy}} + H(X_1^n|X_2^n, Y^n)$

This is small!

(d) (4 points) True or false (no need to justify unless you are unsure):

(i) All typical sequences in $A_\epsilon^{(n)}$ have the same probability. F

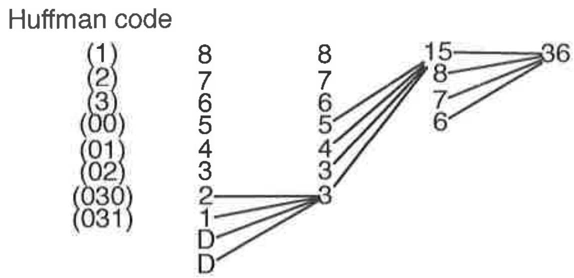
(ii) The typical set $A_\epsilon^{(n)}$ is defined as the smallest set of sequences with $\Pr\{A_\epsilon^{(n)}\} \geq 1 - \epsilon$. F

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

(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 $\prod_{i=1}^n p(x_i)p(y_i)$. T

(e) (5 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:




(f) (4 points) Consider all probability mass functions of the form $p(u, x, y) = p(u, x)p(y|x)$. Is $I(U; Y) \leq I(X; Y)$? If true give a short derivation, if not give a counter-example why not.

**Solution:** Since $U \to X \to Y$ (Markov chain), by the data-processing inequality, $I(U; Y) \leq I(X; Y)$. Sine $p(u,x,y) = p(u,x)\,p(y|x) = p(u)p(x|u)p(y|x)$

(g) (4 points) Prove the following inequality and find conditions for equality.

$$I(X; Z|Y) \geq I(Z; Y|X) - I(Z; Y) + I(X; Z).$$

**Solution:**

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

$$\Updownarrow$$

$$I(X; Z|Y) + I(Y; Z) = \overbrace{I(Y; Z|X)}^{=I(Z;Y|X)} + I(X; Z)$$

$$\Updownarrow$$

$$I(X; Z|Y) = I(Z; Y|X) - I(Y; Z) + I(X; Z)$$

So, $I(X; Z|Y) \geq I(Z; Y|X) - I(Z; Y) + I(X; Z)$ and the equality always holds.

3. *Bandlimited channel capacity.*

   (a) (5 points) Consider the discrete time AWGN channel $Y = gX + Z$ for $Z$ i.i.d. Gaussian noise with 0 mean and variance $\sigma^2$, $g \in \mathbb{R}$ *a known constant,* and $X$ the input is subject to the constraint $E[|X|^2] \leq P$. What is the capacity of this channel? What input achieves capacity?

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

   (b) (5 points) Consider the following continuous-time channel $Y(t) = h(t) * (gX(t) + Z(t))$ where $Z(t)$ is the noise waveform from a white Gaussian noise process with power spectral density $N_0/2$, and $h(t)$ is the impulse response of an ideal bandpass filter, which cuts out all frequencies greater than $W$. Suppose the input has a power constraint $P$. $g \in \mathbb{R}$ is a known constant. What is the capacity of this channel? (No need to derive it)

$$C = W \log\left(1 + \frac{|g|^2 P}{N_0 \cdot W}\right)$$

   (c) (5 points) As the bandwidth $W \to \infty$, what does the capacity tend to? How can you interpret this result?

$C = \frac{P}{N_0} \log_2 e$, it increases linearly with power rather than logarithmically with power at infinite bandwidth!

Points earned: ________ out of a possible 15 points

(d) (5 points) Now consider an additive white Gaussian noise *multiple access channel* defined by $Y(t) = h(t) * (X_1(t) + X_2(t) + Z(t))$ for $Z(t)$ the noise waveform from a white Gaussian noise process with power spectral density $N_0/2$, and $h(t)$ is the impulse response of an ideal bandpass filter, which cuts out all frequencies greater than $W$. Suppose input 1 has a power constraint $P_1$ and input 2 has a power constraint $P_2$. Find the capacity region of this channel.

$$R_1 \leq W \log\left(1 + \frac{P_1}{N_0 W}\right)$$

$$R_2 \leq W \log\left(1 + \frac{P_2}{N_0 W}\right)$$

$$R_1 + R_2 \leq W \log\left(1 + \frac{(P_1 + P_2)}{N_0 \cdot W}\right)$$

(e) (5 points) As the bandwidth $W \to \infty$, what does the capacity region tend to? How can you interpret this result?

$$R_1 < \frac{P_1}{N_0} \log_2 e$$

$$R_2 < \frac{P_2}{N_0} \cdot \log_2 e$$

$$R_1 + R_2 < \frac{P_1 + P_2}{N_0} \log_2 e$$ } Redundant! At $W \to \infty$ there is ~~no~~ "interference" between the 2 users!

4. *Rate-distortion.*

(a) (5 points) In 2 sentences describe the rate-distortion function (meaning not just equation) and when it may be useful.

**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 to within an expected distortion of D. The R(D) function is useful in lossy (non-perfect) compression or sources.

(b) (5 points) What is the rate-distortion function for a $\mathcal{N}(\mu, \sigma^2)$ Gaussian source? *(note the mean $\mu$!)*

**Solution:** From page 311 of the text,

$$R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D}, & 0 \le D \le \sigma^2 \\ 0, & D > \sigma^2 \end{cases}$$

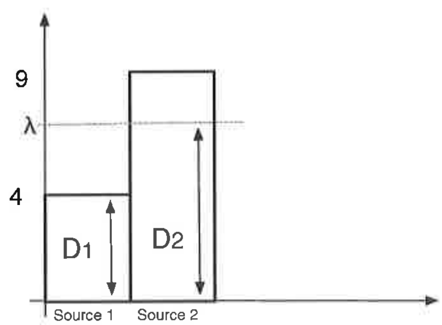
The mean does not affect the rate-distortion function.

(c) (15 points) Now let us consider 2 independent Gaussian sources which we hope to compress: $X_1 \sim \mathcal{N}(0,4)$ and $X_2 \sim \mathcal{N}(0,9)$. We will determine a few specific points on the rate-distortion function for this parallel Gaussian source. *RECALL: The rate is the sum of the rates on the individual channels and the distortion is the sum of the distortions on the individual channels. For what values of D do we only assign bits to one of the sources?*

**Solution:** All these solutions involve reverse water-filling according to the following diagram, where $\lambda$ is the water-level, and we have the important relationships

$$D_i = \begin{cases} \lambda & \text{if } \lambda < \sigma_i^2 \\ \sigma_i^2 & \text{if } \lambda \ge \sigma_i^2 \end{cases}, \text{ and } R = \sum_{i=1}^{2} \frac{1}{2} \log_2 \frac{\sigma_i^2}{D_i},$$

where $\lambda$ the water-level is reverse water-filled so that $D_1 + D_2 = D$.




$4 \le D \le 9$