IMPERIAL COLLEGE LONDON

E4.40
CS7.26
SO20
ISE4.51

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2009

MSc and EEE/ISE PART IV: MEng and ACGI

Corrected Copy

INFORMATION THEORY

Thursday, 14 May 10:00 am

Time allowed: 3:00 hours

No Correction

There are SIX questions on this paper.

Answer FOUR questions.

All questions carry equal marks

Any special instructions for invigilators and information for candidates are on page 1.

Examiners responsible First Marker(s) : C. Ling

Second Marker(s) : A. Manikas

Information for students

Notation:

(a) Random variables are shown in Tahoma font. x, $\mathbf{x}$, $\mathbf{X}$ denote a random scalar, vector and matrix respectively.

(b) The size of a set A is denoted by $|A|$.

(c) The normal distribution is denoted by $N(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

(d) $\oplus$ denotes the exclusive-or operation, or modulo 2 addition.

(e) "i.i.d." means "independent identically distributed".

(f) Entropy function for a binary source $H(p)=-p\log_2 p-(1-p)\log_2(1-p)$; its derivative $H'(p)=\log_2(1-p)-\log_2 p$.

(g) $C(x)=\frac{1}{2}\log_2(1+x)$ is the capacity function for the Gaussian channel in bits/channel use.

The Questions

1.

a) Let the joint distribution of two random variables X and Y be given by

$p(X,Y)$	$Y=0$	$Y=1$
$X=0$	1/3	1/3
$X=1$	0	1/3

Compute:

i) The entropies $H(X)$, $H(Y)$

ii) The conditional entropies $H(X|Y)$, $H(Y|X)$

iii) The joint entropy $H(X,Y)$

iv) The mutual information $I(X,Y)$

v) Draw a Venn diagram for the above quantities.

[10]

b) A fair coin is flipped until the first head occurs. Let X denote the number of flips required. Find the entropy $H(X)$ in bits. The following equalities may be useful.

$$\sum_{n=1}^{\infty} r^n = \frac{r}{1-r} \qquad \sum_{n=1}^{\infty} nr^n = \frac{r}{(1-r)^2} \qquad |r|<1.$$

[10]

c) Let $p(X,Y)$ be the joint probability distribution of random variables X and Y. Show that the mutual information $I(X,Y)$ is always nonnegative. State the condition when $I(X,Y) = 0$. You may assume without proof that the relative entropy

$D(\mathbf{p}\,||\,\mathbf{q}) = \sum_i p_i \log_2\left(\dfrac{p_i}{q_i}\right) \geq 0$ where $\mathbf{p} = [p_1, p_2, \ldots]^T$ and $\mathbf{q} = [q_1, q_2, \ldots]^T$ are two arbitrary probability mass vectors.

[5]

2.

a) Consider the source code {10, 01, 0010, 0111} of four symbols.

i) Is it non-singular? Why?

ii) Is it uniquely decodable? Why?

iii) Is it instantaneous? Why?

iv) Does it satisfy the Kraft inequality? Why? [10]

b) Consider the probability distribution of a random variable X:

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{pmatrix}$$

i) Find a binary Huffman code for X.

ii) Find the expected code length for this code. [10]

c) Lempel-Ziv coding. Give the LZ78 parsing and encoding of the following sequence:

000000110101000000110101

[Note: For this question, you will see less than 15 phrases; so ALWAYS use four bits to represent the location of a phrase. Do not worry about how to save such bits.] [5]

3.

a) Consider the binary erasure channel shown in Fig. 3.1.

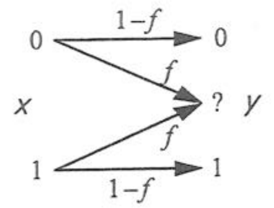


Fig. 3.1. Binary erasure channel.

Justify each step of the following derivation:

$$
\begin{aligned}
I(x;y) &\overset{(1)}{=} H(x) - H(x \mid y) \\
&\overset{(2)}{=} H(x) - p(y=0)\times 0 - p(y=?)H(x) - p(y=1)\times 0 \\
&\overset{(3)}{=} H(x) - H(x)f = (1-f)H(x) \\
&\overset{(4)}{\leq} 1-f
\end{aligned}
$$

What is the capacity of this channel and what is the input distribution achieving the capacity?

[10]

b) Calculate the capacity of the following channels with forward probability transition matrix

i)
$$Q = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \quad x, y \in \{0,1,2\}$$

ii)
$$Q = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix} \quad x, y \in \{0,1,2,3\}$$

[10]

c) Consider the channel $y = xz$ where x (the input) and z are independent binary random variables that take on values 0 and 1. z is Bernoulli(a), i.e. $P(z=1) = a$. Find the capacity of this channel and the corresponding distribution on x.

[5]

4. Consider the discrete-time additive noise channel of Fig. 4.1. X and Y are continuous signals discrete in time and the zero-mean noise Z is independent, identically distributed and is independent of X.

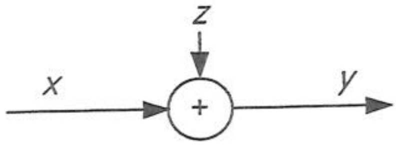


Figure 4.1 Discrete-time additive channel.

a) The power of X is P and the variance of Z is N. When the noise Z is Gaussian, justify each step of the following derivation.

$$
\begin{aligned}
I(X;Y) &\overset{(1)}{=} h(Y) - h(Y \mid X) \overset{(2)}{=} h(Y) - h(X+Z \mid X) \\
&\overset{(3)}{=} h(Y) - h(Z \mid X) \overset{(4)}{=} h(Y) - h(Z) \\
&\overset{(5)}{\le} \tfrac{1}{2}\log_2 2\pi e(P+N) - \tfrac{1}{2}\log_2 2\pi e N \\
&\overset{(6)}{=} \frac{1}{2}\log_2\left(1+\frac{P}{N}\right)
\end{aligned}
$$

And give the channel capacity C and the corresponding input distribution. [10]

b) Consider an expected output power constraint $E[Y^2] = P$. If the variance of Z is still N, find the channel capacity. [5]

c) Parallel channels and waterfilling. Consider the following three parallel Gaussian channels

$$
\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix}
$$

where

$$
\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}\right)
$$

with a power constraint $E(X_1^2 + X_2^2 + X_3^2) \le 3P$. Assume that $\sigma_1^2 \ge \sigma_2^2 \ge \sigma_3^2$. At what power does the channel behave like

i) a single channel with noise variance σ_3^2?

ii) a pair of channels with noise variances σ_3^2 and σ_2^2?

iii) three channels with noise variances σ_3^2, σ_2^2, and σ_1^2?

iv) find the channel capacities for cases i), ii), and iii). [10]

5.

a) Consider a two-user multiple access Gaussian channel with reference to Fig. 5.1.

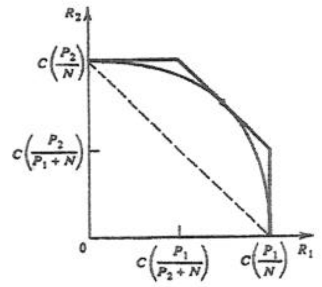


Fig. 5.1. Capacity region of multi-access channel.

i) Describe the capacity region of this channel. Interpret the corner points (i.e., why can one of the users achieve the full capacity of a single-user channel as if the other user were absent?)

ii) Verify the following equality for the corner point:

$$C\left(\frac{P_1}{N}\right)+C\left(\frac{P_2}{P_1+N}\right)=C\left(\frac{P_1+P_2}{N}\right)$$

where $C(x)$ is the capacity function.

[10]

b) Slepian-Wolf region for binary sources. Let X_i be i.i.d. where $P(X_i = 0) = p$ and $P(X_i = 1) = 1 - p$. Let Z_i be i.i.d. where $P(Z_i = 0) = 1 - r$ and $P(Z_i = 1) = r$, and let Z be independent of X. Finally, let $Y = X \oplus Z$. Let X be described at rate R_1 and Y be described at rate R_2: What region of rates allows recovery of X, Y with probability of error tending to zero? Sketch the Slepian-Wolf region.

[10]

c) Consider a two-user scalar Gaussian broadcast channel

$$Y_1 = X + Z_1$$
$$Y_2 = X + Z_2$$

where Z_1 and Z_2 are independent Gaussian random variables with power N_1 and N_2 ($N_1 < N_2$), respectively. The capacity region is given by

$$R_1 \le C\left(\frac{\alpha P}{N_1}\right), \qquad R_2 \le C\left(\frac{(1-\alpha)P}{\alpha P + N_2}\right), \qquad 0 \le \alpha \le 1.$$

Sketch the region. What is the maximum sum rate R_1+R_2? Interpret your result.

[5]

6. Consider discrete-valued random vectors $\mathbf{x}$ and $\mathbf{y}$ of length n where each pair (X_i, Y_i) is drawn i.i.d. from the joint probability distribution function $p_{xy}(x,y)$. The jointly typical set $J_\varepsilon^{(n)}$ is the set of vector pairs satisfying the following conditions:

$$J_\varepsilon^{(n)} = \Big\{\mathbf{x},\mathbf{y} : \left|-n^{-1}\log_2 p_x(\mathbf{x}) - H(X)\right| < \varepsilon,$$
$$\left|-n^{-1}\log_2 p_y(\mathbf{y}) - H(Y)\right| < \varepsilon,$$
$$\left|-n^{-1}\log_2 p_{xy}(\mathbf{x},\mathbf{y}) - H(X,Y)\right| < \varepsilon\Big\}$$

where $p_x(x)$ and $p_y(x)$ are the probability distribution functions of X_i and Y_i respectively. Since the sequences are i.i.d., the probability $p_x(\mathbf{x}) = \prod_{i=1}^n p_x(x_i)$ and $p_x(\mathbf{x})$ and $p_{xy}(\mathbf{x},\mathbf{y})$ can be written in a similar fashion.

a) Show the size of $J_\varepsilon^{(n)}$ satisfies

$$(1-\varepsilon)2^{n(H(X,Y)-\varepsilon)} \overset{n>N_\varepsilon}{<} \left|J_\varepsilon^{(n)}\right| \le 2^{n(H(X,Y)+\varepsilon)}$$

by justifying each step (1) to (5) in the following derivation:

$$1-\varepsilon \overset{n>N_\varepsilon}{<} \sum_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p(\mathbf{x},\mathbf{y}) \overset{(1)}{\le} \left|J_\varepsilon^{(n)}\right| \max_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p(\mathbf{x},\mathbf{y}) \overset{(2)}{\le} \left|J_\varepsilon^{(n)}\right| 2^{-n(H(X,Y)-\varepsilon)}$$

$$1 \overset{(3)}{\ge} \sum_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p(\mathbf{x},\mathbf{y}) \overset{(4)}{\ge} \left|J_\varepsilon^{(n)}\right| \min_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p(\mathbf{x},\mathbf{y}) \overset{(5)}{\ge} \left|J_\varepsilon^{(n)}\right| 2^{-n(H(X,Y)+\varepsilon)}$$

[10]

b) Suppose the joint distribution $p_{xy}(x,y)$ is given by

$p_{xy}(x,y)$	$y = 0$	$y = 1$
$x = 0$	0.45	0.05
$x = 1$	0.05	0.45

$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are drawn i.i.d from the above distribution.

i) Of the 2^n possible sequences $\mathbf{x}$ of length n, how many of them are in the typical set $A_\varepsilon^{(n)}(\mathbf{x}) = \{\mathbf{x} : \left|-n^{-1}\log_2 p_x(\mathbf{x}) - H(X)\right| < \varepsilon\}$ for $\varepsilon = 0.1$?

ii) Of the 2^n possible sequences $\mathbf{y}$ of length n, how many of them are in the typical set $A_\varepsilon^{(n)}(\mathbf{y}) = \{\mathbf{y} : \left|-n^{-1}\log_2 p_y(\mathbf{y}) - H(Y)\right| < \varepsilon\}$ for $\varepsilon = 0.1$?

iii) Explain why $p(\mathbf{x},\mathbf{y}) = 2^{-n}(1-p)^{n-k}p^k$ where k is the number of places where the two sequences $\mathbf{x}$ and $\mathbf{y}$ differ, and $p = 0.1$.

iv) Now suppose $n = 10$. Determine the size and probability of the jointly typical set $J_\varepsilon^{(n)}$ for $\varepsilon = 0.1$.

[15]

Information Theory

Solutions

(6.4.09)

B - bookwork
E - new example
A - new application

E4.40
ISE 4.51
C57.26
S020

1. a) Distribution: $p(X=0) = \frac{2}{3}$ $\quad p(X=1) = \frac{1}{3}$

$p(Y=0) = \frac{1}{3}$ $\quad p(Y=1) = \frac{2}{3}$

i) $H(X) = \frac{2}{3}\log\frac{3}{2} + \frac{1}{3}\log 3 = 0.918 \text{ bits} = H(Y)$ [2E]

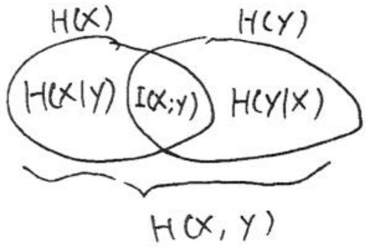
ii) $H(X|Y) = \frac{1}{3}H(X|Y=0) + \frac{2}{3}H(X|Y=1)$ [2E]

$= \frac{1}{3}\times 0 + \frac{2}{3}\times 1 = \frac{2}{3} = 0.667 \text{ bits} = H(Y|X)$

iii) $H(X,Y) = 3 \times \frac{1}{3}\log 3 = \log 3 = 1.585 \text{ bits}$ [2E]

iv) $I(X;Y) = H(X) - H(X|Y) = 0.918 - 0.667 = 0.251 \text{ bits}$ [2E]

v) [2B]



b) $X=n$ means that Tail occurs for the first $n-1$ flips, while Head occurs for the n-th flip. Thus [5E]

$$p(X=n) = \left(\tfrac{1}{2}\right)^{n-1}\tfrac{1}{2} = \left(\tfrac{1}{2}\right)^n$$ [5E]

~~i)~~ $H(X) = \sum_{n=1}^{\infty} 2^{-n}\log 2^n = \sum_{n=1}^{\infty} n \cdot 2^{-n} \cdot \log 2 = \frac{\frac{1}{2}}{(1-\frac{1}{2})^2} = 2 \text{ bits}$

~~ii) Ask if $X = 1, 2, 3, \cdots$ in turn, i.e.,~~ [~~5E~~]

~~Is $X=1$?~~

~~If not, is $X=2$?~~

~~If not, is $X=3$?~~

~~...~~

~~Expected number of questions $= \sum_{n=1}^{\infty} n\left(\frac{1}{2}\right)^n = 2$~~

c) $I(X;Y) = H(X) + H(Y) - H(X,Y)$ [5B]

$$= E\log\frac{p(X,Y)}{p(X)p(Y)} = D(P_{X,Y}\,\|\,P_X \otimes P_Y) \geq 0$$

$I(X;Y) = 0$ iff $P_{X,Y} = P_X \otimes P_Y$, i.e., X and Y are independent.

2. a)

i) It is non-singular, because the codewords are different. [2 E]

ii) It is uniquely decodable, because the strings of codewords are unique. [3 E]

iii) It is instantaneous, because no codeword is a prefix of other codewords. [2 E]

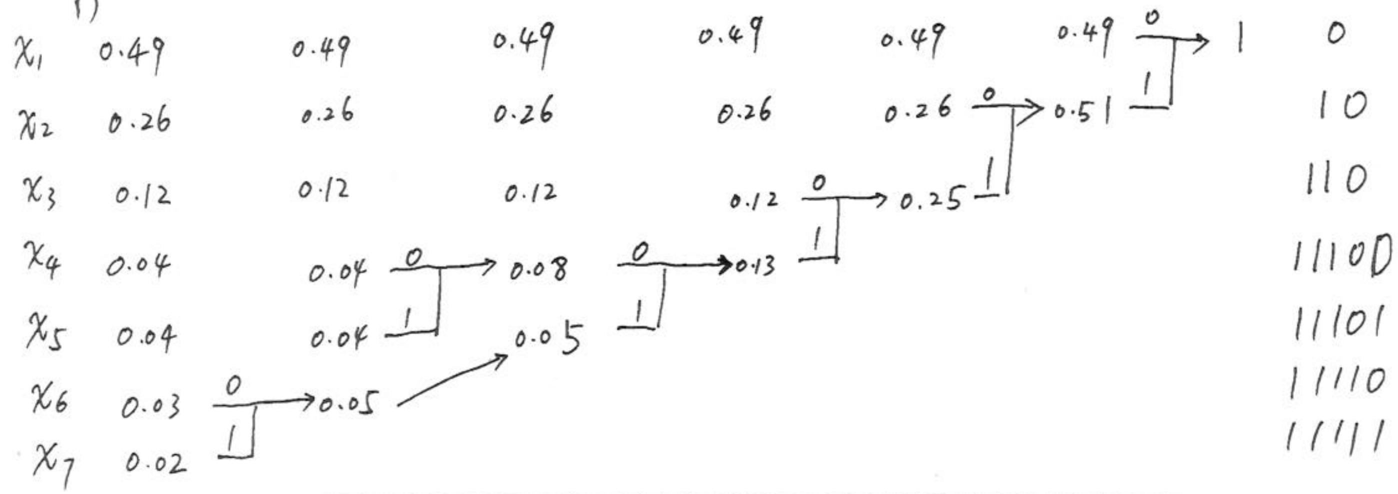
iv). Yes. [3 E]

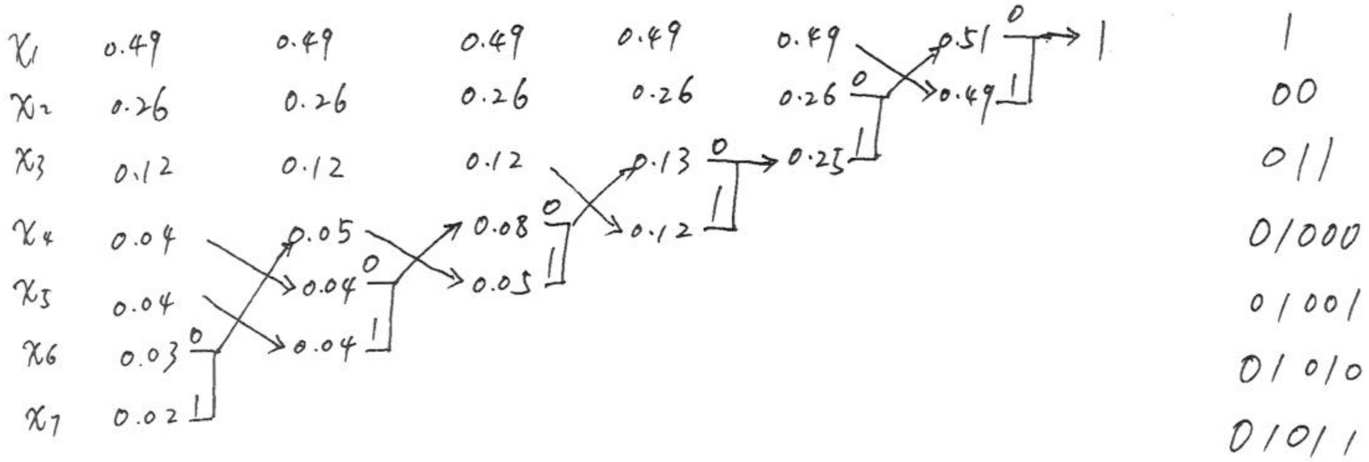
$$2^{-2} + 2^{-2} + 2^{-3} + 2^{-3} = \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = \frac{3}{4} < 1$$

[8 E]

b) Both are correct:

i)





ii) $L = \Sigma p(x_i) l(x_i) = 1 \times 0.49 + 2 \times 0.26 + 3 \times 0.12 + 5 \times 0.13$ [2 E]

$= 2.02$

Q 2

c) Parsing: 0,00,000,1,10,101,0000,01,1010,1 [5E]

[1]

There are 10 phrases, so we need 4 bits to represent the locations.

[2]

Encoding: (0000, 0), (0001, 0), (0010, 0), (0000, 1), (0100, 0)

(0101, 1), (0011, 0), (0001, 1), (0110, 0), (0000, 1)

[3]

3 a) (1) definition [2B]

(2) $H(X|Y) = \sum_i H(X|y=i)$ average row entropy [2B]

(3) algebra [2B]

(4) $H(X) \le 1$ [2B]

$\therefore \; C = 1 - f$ [2B]

This is achieved if $H(X)=1$, i.e., X is uniformly distributed.

b)

i) Since the channel is symmetric, [5E]

$$C = \log|Y| - H(Q_{1,:})$$
$$= \log 3 - H(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3})$$
$$= \log 3 - 3 \times \tfrac{1}{3} \log 3$$
$$= 0$$

ii) Again, this is a symmetric channel, Thus [5E]

$$C = \log|Y| - H(Q_{1,:})$$
$$= \log 4 - H(\tfrac{1}{2}, \tfrac{1}{2}, 0, 0)$$
$$= \log 4 - H(\tfrac{1}{2})$$
$$= 1$$

c) Let $p(X=1)=p$ Then $P(Y=1) = p(X=1, Z=1)$ [5A]

$= p(X=1)p(Z=1) = ap$ [1]

$p(Y=0) = 1-ap$

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

$= H(ap) - H(Y|X=0)p(X=0) - H(Y|X=1)p(X=1)$

$= H(ap) - 0 - H(XZ|X=1)p(X=1)$ $\quad Y=0$ if $X=0$

$= H(ap) - H(Z|X=1)p(X=1)$

$= H(ap) - H(Z)p(X=1)$ $\quad X$ and Z indepedent

[2]

Therefore,

$$I(X;Y) = H(ap) - pH(a)$$

$$\boxed{\begin{aligned} H(p) &= -p\log p - (1-p)\log(1-p) \\ H'(p) &= \log(1-p) - \log p \end{aligned}}$$

$$\frac{\partial I}{\partial p} = \frac{\partial H(ap)}{\partial p} - H(a)$$

$$= a \cdot [\log(1-ap) - \log ap] - H(a)$$

[2]

$$= a \cdot \log\left(\frac{1}{ap} - 1\right) - H(a) = 0$$

$$\log\left(\frac{1}{ap^*} - 1\right) = \frac{H(a)}{a}$$

p^*: optimum value

$$\frac{1}{ap^*} - 1 = 2^{\frac{H(a)}{a}}$$

$$p^* = \frac{1}{a\left(2^{\frac{H(a)}{a}} + 1\right)}$$

$$C = H(ap^*) - p^* H(a)$$

$$= H\left(\frac{1}{2^{\frac{H(a)}{a}} + 1}\right) - \frac{1}{a\left(2^{\frac{H(a)}{a}} + 1\right)} H(a)$$

4. a)

(1): definition [1B]

(2) $Y = X + Z$ [1B]

(3) translation doesn't change differential entropy [2B]

(4) X . Z independent [2B]

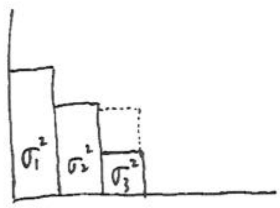
(5) Given the power, Gaussian distribution has maximum entropy; entropy of Gaussian r.v. [2B]

(6) algebra [2B]

b) In this case, the power of x is $P - N$. [5E]

$$C = \frac{1}{2}\log\left(1 + \frac{P-N}{N}\right) = \frac{1}{2}\log\frac{P}{N}$$

c) [3A]



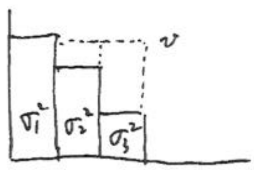
i) Single channel is when

$$3P \leq \sigma_2^2 - \sigma_3^2$$

Capacity

$$C = \frac{1}{2}\log\left(1 + \frac{3P}{\sigma_3^2}\right)$$

[3A]



ii) A pair of channel is when

$$\sigma_2^2 - \sigma_3^2 < 3P \leq \sigma_1^2 - \sigma_2^2 + \sigma_1^2 - \sigma_3^2$$ [1]
$$= 2\sigma_1^2 - \sigma_2^2 - \sigma_3^2$$

$$3P = v - \sigma_2^2 + v - \sigma_3^2 \Rightarrow v = \frac{3P + \sigma_2^2 + \sigma_3^2}{2}$$

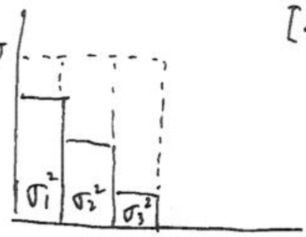
$$P_2 = v - \sigma_2^2 = \frac{3P - \sigma_2^2 + \sigma_3^2}{2}$$ [1]

$$P_3 = v - \sigma_3^2 = \frac{3P + \sigma_2^2 - \sigma_3^2}{2}$$

$$C = \frac{1}{2}\log\left(1 + \frac{P_2}{\sigma_2^2}\right) + \frac{1}{2}\log\left(1 + \frac{P_3}{\sigma_3^2}\right)$$ [1]
$$= \frac{1}{2}\log\left(1 + \frac{3P - \sigma_2^2 + \sigma_3^2}{2\sigma_2^2}\right) + \frac{1}{2}\log\left(1 + \frac{3P + \sigma_2^2 - \sigma_3^2}{2\sigma_3^2}\right)$$

iii) Three channels is when [4A]

$$3P > 2\sigma_1^2 - \sigma_2^2 - \sigma_3^2$$



[2]

$$3P = v - \sigma_1^2 + v - \sigma_2^2 + v - \sigma_3^2$$

$$\Rightarrow v = \frac{3P + \sigma_1^2 + \sigma_2^2 + \sigma_3^2}{3} = P + \frac{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}{3}$$

$$P_1 = v - \sigma_1^2 = P + \frac{\sigma_2^2 + \sigma_3^2 - 2\sigma_1^2}{3}$$

$$P_2 = v - \sigma_2^2 = P + \frac{\sigma_1^2 + \sigma_3^2 - 2\sigma_2^2}{3}$$

$$P_3 = v - \sigma_3^2 = P + \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_3^2}{3}$$

[2]

$$C = \frac{1}{2}\log\left(1 + \frac{P_1}{\sigma_1^2}\right) + \frac{1}{2}\log\left(1 + \frac{P_2}{\sigma_2^2}\right) + \frac{1}{2}\log\left(1 + \frac{P_3}{\sigma_3^2}\right)$$

5. a)

i) Capacity region [5 B]

$$R_1 < C(\frac{P_1}{N})$$

$$R_2 < C(\frac{P_2}{N})$$ [3]

$$R_1 + R_2 < C\left(\frac{P_1+P_2}{N}\right)$$

At the corner point, the decoder decodes one user first, treating the other user as noise. Thus, it achieves rate $R_1 = C(\frac{P_1}{P_2+N})$. After that, the decoder subtracts off user 1, meaning user 2 is only subject to noise. Thus, it can achieve rate $R_2 = C(\frac{P_2}{N})$. This strategy is called successive interference cancellation or "Onion peeling". [2]

ii) $C(\frac{P_1}{N}) + C(\frac{P_2}{P_1+N})$ [5 E]

$= \frac{1}{2}\log(1+\frac{P_1}{N}) + \frac{1}{2}\log(1+\frac{P_2}{P_1+N})$ [1]

$= \frac{1}{2}\log\left(\frac{P_1+N}{N}\cdot\frac{P_1+P_2+N}{P_1+N}\right)$ [1]

[1]

$= \frac{1}{2}\log\left(\frac{P_1+P_2+N}{N}\right)$ ~~[1]~~

$= \frac{1}{2}\log\left(1+\frac{P_1+P_2}{N}\right)$ [1]

[1]

$= C\left(\frac{P_1+P_2}{N}\right)$

b) Slepian-Wolf region [5A]

$$R_1 > H(X|Y)$$

$$R_2 > H(Y|X)$$ [2]

$$R_1 + R_2 > H(X,Y)$$

We need to calculate the entropies.

Q5

$X = \text{Bernoulli}(p) \Rightarrow H(X) = H(p)$ [3]

$Y = X \oplus Z, \quad Z = \text{Bernolli}(r) \Rightarrow Y = \text{Bernolli}(p * r)$

where $p * r = p(1-r) + r(1-p)$

$$H(Y) = H(p*r)$$ [3]

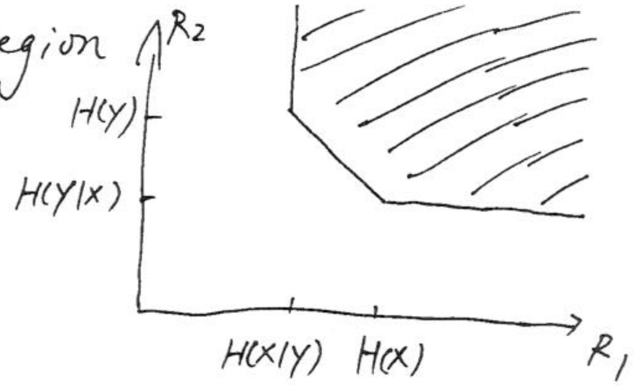
$H(X,Y) = H(X, X\oplus Z) = H(X,Z) = H(X) + H(Z)$ independence

$= H(p) + H(r).$ [5A]

$H(Y|X) = H(X\oplus Z|X) = H(Z|X) = H(Z) = H(p)$

$H(X|Y) = H(X,Y) - H(Y) = H(p) + H(r) - H(p*r).$

Slepian-Wolf region [2]



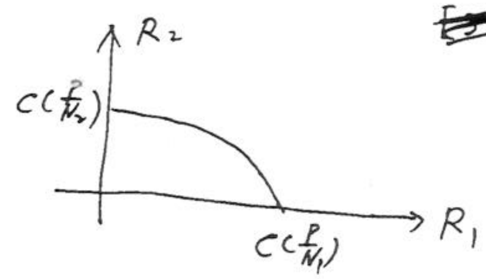
c) ~~Gaussian relay channel~~

$\cancel{C = \max_{0\le\alpha\le 1} \min\left\{ C\left(\frac{P+P_1+2\sqrt{(1-\alpha)PP_1}}{N_1+N_2}\right), C\left(\frac{\alpha P}{N_1}\right)\right\}}$

Capacity region:

$$R_1 \le C\left(\frac{\alpha P}{N_1}\right)$$

$$R_2 \le C\left(\frac{(1-\alpha)P}{\alpha P + N_2}\right)$$



Sum rate

$$R_1 + R_2 \le C\left(\frac{\alpha P}{N_1}\right) + C\left(\frac{(1-\alpha)P}{\alpha P + N_2}\right)$$ [3A]

$$= \frac{1}{2}\log\left(\frac{\alpha P + N_1}{N_1} \cdot \frac{\alpha P + N_2 + P - \alpha P}{\alpha P + N_2}\right)$$

$$= \frac{1}{2}\log\left(\frac{P+N_2}{N_1} \cdot \frac{\alpha P + N_1}{\alpha P + N_2}\right)$$ [2A]

$$\le \frac{1}{2}\log\left(\frac{P+N_2}{N_1} \cdot \frac{P+N_1}{P+N_2}\right)$$ maximum when $\alpha = 1$ since $N_1 < N_2$

$$= \frac{1}{2}\log\left(1 + \frac{P}{N_1}\right)$$

Put all power to User 1, the better user.

6

a)

(1) $p(x,y) \le \max_{J_\varepsilon^{(n)}} p(x,y)$ [2B]

(2) $\max_{J_\varepsilon^{(n)}} p(x,y) \le 2^{-n(H(X,Y)-\varepsilon)}$ [2B]

(3) Total probability couldn't be larger than 1 [2B]

(4) $p(x,y) \ge \min_{J_\varepsilon^{(n)}} p(x,y)$ [2B]

(5) $\min_{J_\varepsilon^{(n)}} p(x,y) \ge 2^{-n(H(x,y)+\varepsilon)}$ [2B]

b) From the joint distribution, we can derive that X and Y are i.i.d. sequences with distribution

$$p(X=0) = p(X=1) = \tfrac{1}{2}$$
$$p(y=0) = p(y=1) = \tfrac{1}{2}$$ [4 A]

i) $H(X) = 1$

The probability of a particular sequence X is given by

$$p(X) = \left(\tfrac{1}{2}\right)^m \left(\tfrac{1}{2}\right)^{n-m} = \left(\tfrac{1}{2}\right)^n \qquad m\text{: the number of ones}$$

Thus, $-\frac{1}{n}\log p(X) = -\frac{1}{n}\log\left(\frac{1}{2}\right)^n = 1 = H(X)$

Therefore, all 2^n sequences are in the typical set.

ii) $H(Y) = 1$ [3 A]

Similarly, all 2^n sequences Y are in the typical set.

Q6

iii) From the joint distribution, we deduce that [3A]

$$p(x,y) = 0.45^{n-k}\, 0.05^{k}$$

where k is the number of places where they differ. It can be rewritten as

$$p(x,y) = 2^{-n}(1-p)^{n-k}p^{k}$$

iv) $H(X,Y) = 1.469$ [5A]

$$-\frac{1}{n}\log p(x,y) = -\frac{1}{n}\log\left[2^{-n}(1-p)^{n-k}p^{k}\right] = 1-\frac{1}{n}\log\left[(1-p)^{n-k}p^{k}\right]$$

(x,y) is typical if $H(X,Y)-\varepsilon < -\frac{1}{n}\log p(x,y) < H(X,Y)+\varepsilon$, i.e., [1]

$$0.369 < -\frac{1}{n}\log\left[(1-p)^{n-k}p^{k}\right] < 0.569$$

k	$\binom{n}{k}$	$(1-p)^{n-k}p^{k}$	$-\frac{1}{n}\log[(1-p)^{n-k}p^{k}]$	prob.
0	1	0.3487	0.152	0.3487
1	10	0.0387	0.469	0.387
2	45	0.0043	0.786	
3	120	0.00048	1.103	
⋮				

[2]

k can take values in $0, 1, 2, \cdots, 10$. The number of such sequences is $\binom{n}{k}$.
The Table shows that only $0.469 \in [0.369, 0.569]$; All other sequences are atypical.

Therefore, $|J_\varepsilon^{(n)}| = 10$ [2]

$$P(J_\varepsilon^{(n)}) = 0.387$$