EE4-40
EE9CS7-2
EE9SO20

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2011

MSc and EEE/ISE PART IV: MEng and ACGI

Corrected Copy
Q4(c)

INFORMATION THEORY

Wednesday, 18 May 10:00 am

Time allowed: 3:00 hours

There are FOUR questions on this paper.

Answer ALL 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

© Imperial College London

Information for students

Notation:

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

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

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

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

© Imperial College London

The Questions

1.

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

$p(X,Y)$	$Y=0$	$Y=1$	$Y=2$
$X=0$	1/4	0	0
$X=1$	0	1/4	0
$X=2$	0	1/4	1/4

Compute:

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

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

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) Let X and Y be two independent discrete random variables taking integer values. X is uniformly distributed over {1, 2, 3, 4} and $P(Y=k)=2^{-k},\ k=1,2,3,\cdots$.

i) Find $H(X)$.

ii) Find $H(Y)$.

iii) Find $H(X+Y, X-Y)$.

[7]

c) Let X_1 and X_2 be discrete random variables drawn according to probability mass function $p_1(.)$ and $p_2(.)$ over the respective alphabets X_1 = {1, 2, …, m} and X_2 = {m+1, …, n}. Let

$$X=\begin{cases} X_1 & \text{with probability } \alpha \\ X_2 & \text{with probability } 1-\alpha \end{cases}$$

Find $H(X)$ in terms of $H(X_1)$, $H(X_2)$ and α.

[8]

© Imperial College London

2.

a) Consider the rate-distortion function $R(D) = \min I(X;\hat{X}),\ E_{x,\hat{x}} d(X,\hat{X}) \le D$, where $E_{x,\hat{x}}$ denotes the expectation with respect to $X,\hat{X}$. Justify each step in the following derivation of the rate-distortion function for a Bernoulli source $X = \{0,1\}$, $p_X = \{1-p, p\}$ $(p \le 1/2)$, and $d(x,\hat{x}) = x \oplus \hat{x}$. In the following, (1), (2), …, (6) are the step numbers.

$$\text{If } D \ge p \Rightarrow R(D) \overset{(1)}{=} 0;$$

$$\text{If } D < p \le 1/2,$$

$$I(X;\hat{X}) \overset{(2)}{=} H(X) - H(X \mid \hat{X}) \overset{(3)}{=} H(p) - H(X \oplus \hat{X} \mid \hat{X})$$

$$\overset{(4)}{\ge} H(p) - H(X \oplus \hat{X}) \overset{(5)}{\ge} H(p) - H(D) \Rightarrow R(D) \overset{(6)}{\ge} H(p) - H(D).$$

[9]

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

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ 0.05 & 0.05 & 0.25 & 0.2 & 0.15 & 0.3 \end{pmatrix}$$

i) Find a binary Huffman code for X.

ii) Find the expected code length for this code.

[6]

c) Lempel-Ziv coding. Consider the following all-zero sequence of length n:

$$X^n = 00000000000\ldots$$

i) Give the LZ78 parsing and encoding. You may simply use numbers 1, 2, 3, … to represent the locations.

ii) Show that the number of encoding bits per symbol for this sequence goes to zero as $n \to \infty$.

[10]

© Imperial College London

3.

a) Calculate the capacity of the following channels with probability transition matrices

i)
$$\mathbf{Q} = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} \quad x, y \in \{0,1,2\}$$

ii)
$$\mathbf{Q} = \begin{bmatrix} 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{bmatrix} \quad x \in \{0,1\},\ y \in \{0,1,2\}$$

[8]

b) Compute the capacity of the concatenated binary symmetric channel shown in Fig. 3.1, where the cross-over probabilities are p and q, respectively.

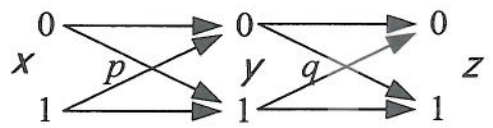


Fig. 3.1. Concatenated binary symmetric channel.

[7]

c) Fano's inequality. Consider the Markov chain shown in Fig. 3.2, where x is the channel input, y is the channel output, and the estimate of x is simply $\hat{x} = y$.

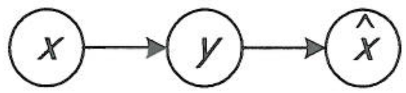


Fig. 3.2. Markov chain.

The input alphabet $X = \{1,2,3,4,5\}$, probability mass vector $\mathbf{p}_x = [0.35, 0.35, 0.1, 0.1, 0.1]$. The output alphabet $Y = \{1,2\}$; if $x \leq 2$, then $y = x$ with probability 6/7, while if $x > 2$, then $y = 1$ or 2 with equal probability.

i) Compute the actual error probability.

ii) Compute the conditional entropy $H(x \mid y)$.

iii) Using Fano's inequality $H(x \mid y) \leq 1 + p_e \log(|X| - 1)$ where P_e is the error probability, compute the Fano bound on the error probability, and compare with i).

[10]

© Imperial College London

4.

a) With reference to Fig. 4.1, justify each step in the following proof of the converse of the Gaussian channel coding theorem. That is, if the error probability $P_e^{(n)} \to 0$ and $n^{-1}\mathbf{x}^T\mathbf{x} < P$ for each $\mathbf{x}(w) = x_{1:n}$, then the rate $R \le ½\log\left(1+PN^{-1}\right)$.
Here P is the signal power, while N is the noise power. (1), (2), …, (9) are step numbers.

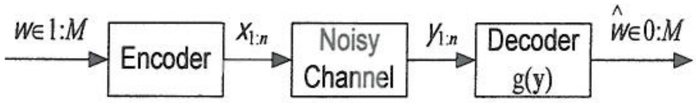


Fig. 4.1. Communication over a noisy channel. $M = 2^{nR}$.

$$nR \overset{(1)}{=} H(w) \overset{(2)}{=} I(w;y_{1:n}) + H(w \mid y_{1:n}) \overset{(3)}{\le} I(x_{1:n};y_{1:n}) + H(w \mid y_{1:n})$$
$$\overset{(4)}{=} h(y_{1:n}) - h(y_{1:n} \mid x_{1:n}) + H(w \mid y_{1:n}) \overset{(5)}{\le} \sum_{i=1}^{n} h(y_i) - h(z_{1:n}) + H(w \mid y_{1:n})$$
$$\overset{(6)}{\le} \sum_{i=1}^{n} I(x_i;y_i) + 1 + nRP_e^{(n)} \overset{(7)}{\le} \sum_{i=1}^{n} ½\log\left(1+PN^{-1}\right) + 1 + nRP_e^{(n)}$$
$$\Rightarrow R \overset{(8)}{\le} ½\log\left(1+PN^{-1}\right) + n^{-1} + RP_e^{(n)} \Rightarrow R \overset{(9)}{\le} ½\log\left(1+PN^{-1}\right) \text{ as } n \to \infty$$

[10]

b) Calculate the differential entropy of a zero-mean Gaussian random vector with correlation matrix

$$\mathbf{K} = \begin{bmatrix} 1 & 2 \\ 2 & 6 \end{bmatrix}.$$

[5]

c) Slepian-Wolf coding. Let (X, Y) have the joint probability mass function

$p(x,y)$	1	2	3
1	α	β	β
2	β	α	β
3	β	β	α

$\to 6$

where $4\beta + 3\alpha = 1$. (Note: this is a joint, not a conditional, probability mass function.)

i) Find the Slepian-Wolf rate region for this source.

ii) What is the rate region if $\alpha = 1/3$?

[10]

Info. Theory INFORMATION THEORY

2011 Answers 2011

EE4-40
EE9 CS7-2
EE9SO20

B – Bookwork
E – new example
A – new application

1/8

1. a) The maginal distributions

$p(x)$	
0	$\frac{1}{4}$
1	$\frac{1}{4}$
2	$\frac{1}{2}$

$p(y)$	
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

i) $H(x) = \frac{1}{4}\times 2 + \frac{1}{4}\times 2 + \frac{1}{2}\times 1 = 1.5$ bits [2E]

$H(y) = 1.5$ bits

ii) $H(y|x) = \sum_{k=0}^{2} H(y|x=k)\, p(x=k)$ [2E]

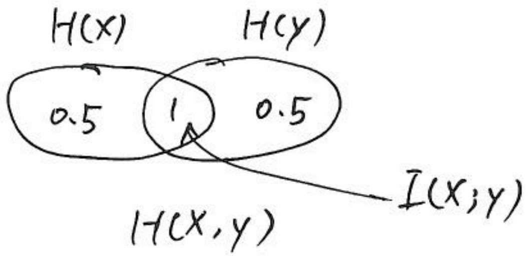
$= 0 + 0 + \frac{1}{2} = \frac{1}{2}$

$H(x|y) = 0 + \frac{1}{2} + 0 = \frac{1}{2}$

iii) $H(X,Y) = \log 4 = 2$ [2E]

iv) $I(X;Y) = H(X) - H(X|Y) = 1$ [2E]

v) [2E]



b) i) $H(X) = \log 4 = 2$ [2B]

ii) $H(Y) = -\sum_{k=1}^{\infty} p(y=k)\log p(y=k)$ [2E]

$= +\sum 2^{-k}\cdot k$

$= \frac{\frac{1}{2}}{(1-\frac{1}{2})^2} = 2$

iii) The mapping $(X,Y) \to (X+Y, X-Y)$ is one-to-one.

$\therefore H(X+Y, X-Y) = H(X,Y) = H(X) + H(Y) = 4$

[3A]

c) The probability mass vector for X is

[2A] $\alpha p_1(1)$ $\alpha p_1(2)$ $\cdots$ $\alpha p_1(m)$, $(1-\alpha)p_2(m+1)$, $\cdots$, $(1-\alpha)p_2(n)$

[2A]
$$H(X) = -\sum_{k=1}^{m} \alpha p_1(k) \log[\alpha p_1(k)] - \sum_{k=m+1}^{n} (1-\alpha)p_2(k) \log[(1-\alpha)p_2(k)]$$

[2A]
$$= -\sum_{k=1}^{m} \alpha p_1(k)[\log\alpha + \log p_1(k)] - \sum_{k=m+1}^{n} (1-\alpha)p_2(k)[\log(1-\alpha) + \log p_2(k)]$$

$$= -\alpha\log\alpha - (1-\alpha)\log(1-\alpha) + \alpha\sum_{k=1}^{m} p_1(k)\log p_1(k) + (1-\alpha)\sum_{k=m+1}^{n} p_2(k)\log p_2(k)$$

[2A]
$$= H(\alpha) + \alpha H(X_1) + (1-\alpha)H(X_2)$$

3.

2. a) (1) send nothing when $D \geq p$, as the distortion ~~is~~ p. [1B]

(2) definition of mutual information [1B]

(3) $H(X) = H(p)$ [1B]

$H(X|\hat{X}) = H(X \oplus \hat{X} | \hat{X})$

(4) conditioning reduces entropy [1B]

(5) $H(X \oplus \hat{X}) \leq H(D)$ [3B]

This is because $E_{x,\hat{x}}\, d(x,\hat{x}) \leq D$

$= 1 \cdot Pr((X \oplus \hat{X}) = 1) + 0 \cdot (Pr\ X \oplus \hat{X} = 0)$

$= Pr(X \oplus \hat{X} = 1) \leq D$

$H(X \oplus \hat{X}) \leq H(D)$ as $H(p)$ is monotonic when $p \leq \frac{1}{2}$

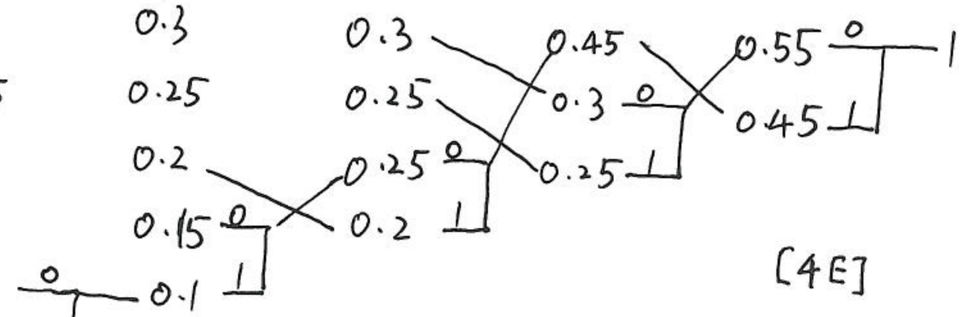
(6) $I(X;\hat{X}) \geq H(p) - H(D)$ [2B]

$\Rightarrow \min I(X;\hat{X}) \geq H(p) - H(D)$

$R(D) \geq H(p) - H(D)$

b) i)

x_0	00	0.3
x_3	01	0.25
x_4	11	0.2
x_5	100	0.15
x_1	1010	0.05
x_2	1011	0.05



[4E]

ii) expected length [2E]

$\bar{L} = 2 \times 0.75 + 3 \times 0.15 + 4 \times 0.1$

$= 2.35$

c)

i)

location → 1 2 3 4 5

0, 00, 000, 0000, 00000, ... [4E]

encoding 0, 10, 20, 30, 40, ...

ii) for a length-n sequence, the number k of phrases is given by the equation

$$1 + 2 + 3 + \cdots + k = n$$ [2A]

$$k(k+1) = n$$

$$k < \sqrt{n}$$

The number of encoded bits

$$< k \cdot (\log_2 k + 1)$$ [2A]

The number of bits / symbol

$$< \frac{k \cdot (\log_2 k + 1)}{n}$$

$$< \frac{\sqrt{n} \cdot (\log_2 \sqrt{n} + 1)}{n}$$ [2A]

$$< \frac{\log_2 \sqrt{n} + 1}{\sqrt{n}}$$

$$\rightarrow 0 \quad \text{as} \quad n \rightarrow \infty$$

3. a) Both are weakly symmetric channels

i) $C = \log|\mathcal{Y}| - H(Q_{1,:})$

$= \log 3 - (\frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{4} \times 2)$ [4 E]

$= \log 3 - \frac{3}{2} \simeq 0.08$

ii) $C = \log 3 - (\frac{1}{3} \times \log 3 + \frac{1}{6} \times \log 6 + \frac{1}{2} \times 1)$

$\simeq 0.12$ [4 E]

b) The transitional matrix of the concatenated channel is

$$\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \begin{pmatrix} 1-q & q \\ q & 1-q \end{pmatrix}$$ [2 A]

$$= \begin{pmatrix} 1-p-q+2pq & p+q-2pq \\ p+q-2pq & 1-p-q+2pq \end{pmatrix}$$ [2 A]

This is still a BSC

[3 A]

$C = 1 - H(p+q-2pq)$

c) i) Joint distribution

$Y=\hat{X}$ \ X	1	2	3	4	5
1	0.3	0.05	0.05	0.05	0.05
2	0.05	0.3	0.05	0.05	0.05

[2 E]

error probability is when $\hat{X} \neq X$

$$P_e = 0.4$$

[2 E]

ii) $H(X|Y)$ = average row entropy

$$= \cancel{-0.3\log 0.3 - 4\times 0.05 \log 0.05}$$

$$= -0.6\log 0.6 - 4\times 0.1 \log 0.1$$

[2 E]

$$= 1.771$$

iii)

$$P_e \geq \frac{H(X|Y)-1}{\log(|X|-1)}$$

[2 E]

$$= \frac{1.771-1}{\log 4}$$

$$= \frac{0.771}{2}$$

$$= 0.3855$$

This is a valid lower bound of the actual error probability. [2 E]

4.

a) (1) uniformly distributed [1 B]

(2) by definition [1 B]

(3) $I(W; Y_{1:n}) \le I(X_{1:n}; Y_{1:n})$ $\quad W \to X \to Y \to \hat{W}$ Markov chain [1 B]

(4) definition [1 B]

(5) indep. bound [1 B]

(6) Fano's inequality [1 B]

(7) $I(X_i; Y_i) \le \frac{1}{2}\log\left(1+\frac{P}{N}\right)$ Capacity is the maximum mutual information. [1 B]

(8) algebra [1 B]

(9) $P_e^{(n)} \to 0$, $n^{-1} \to 0$ as $n \to \infty$ [2 B]

b)

$$h(x) = \frac{1}{2}\log\left((2\pi e)^2 |K|\right)$$ [2 E]

$$= \frac{1}{2}\log\left((2\pi e)^2 \cdot 2\right)$$ [3 E]

$$= 4.57$$

c)

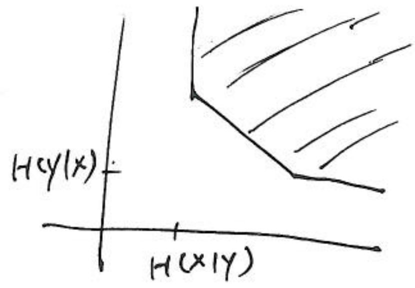
i) Slepian-Wolf region

[2 B]

$$R_X \geq H(X \mid Y)$$

$$R_Y \geq H(Y \mid X)$$

$$R_X + R_Y \geq H(X, Y)$$



$$H(X \mid Y) = H(3\alpha, 3\beta, 3\beta)$$

[2 A]

$$H(Y \mid X) = H(3\alpha, 3\beta, 3\beta)$$

$$H(X, Y) = H(X \mid Y) + H(Y)$$

$$= H(3\alpha, 3\beta, 3\beta) + \log 3$$

[2 A]

Both X and Y are uniformly distributed

$\Rightarrow H(X) = H(Y) = \log 3$

ii) If $\alpha = \frac{1}{3}$, $\beta = 0$

[2 A]

$$H(X \mid Y) = H(Y \mid X) = 0$$

$$H(X, Y) = \log 3$$

$$R_X + R_Y \geq \log 3$$

[2 A]

$$R_X \geq 0 \qquad R_Y \geq 0$$