EE4-40
EE9SO20

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2012

MSc and EEE/ISE PART IV: MEng and ACGI

Corrected Copy

G2

INFORMATION THEORY

Thursday, 3 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

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".

The Questions

1.

a) Let the entropy function $H(p) = -(1-p)\log_2(1-p) - p\log_2 p$ denote the entropy of a Bernoulli random variable with probability mass vector $\mathbf{p} = \begin{bmatrix}1-p & p\end{bmatrix}$. Prove the following properties of this function:

i) $H'(p) = \log_2(1-p) - \log_2 p$ [2]

ii) $H''(p) = \dfrac{-\log_2 e}{p(1-p)}$ [2]

iii) $H(p) \geq 2\min(p, 1-p)$ [3]

iv) $H(p) \leq 1 - 2(\log_2 e)\cdot(p-½)^2$ [3]

b) X_i is a sequence of i.i.d. Bernoulli random variables with $p(X_i = 1) = p$ where p is unknown. We want to find a function f that converts n samples of X into a smaller number, K, of i.i.d. Bernoulli random variables, Z_i, with $p(Z_i=1)=½$. Thus $Z_{1:K}=f(X_{1:n})$ where K can depend on the values X_i.

i) Show that the following mapping for n=4 satisfies the requirements and find the expected value of K, $E[K]$.

0000,1111→ignore;	1010→0;	0101→1;	0001,0011,0111→00;
0010,0110,1110→01;	0100,1100,1101→10;		1000,1001,1011→11

[10]

ii) Justify the steps in the following bound on $E(K)$, where (1), (2), … are the step numbers.

$$nH(p) \overset{(1)}{=} H(X_{1:n}) \overset{(2)}{\geq} H(Z_{1:K}, K) \overset{(3)}{=} H(K) + H(Z_{1:K} \mid K)$$
$$\overset{(4)}{=} H(K) + E[K] \overset{(5)}{\geq} E[K]$$

[5]

2.

a) Markov chain. Let $X \to Y \to Z$ be a Markov chain. Justify each step of the following derivations.

i) X and Z are conditionally independent given Y:

$$p(x,z \mid y) \overset{(1)}{=} \frac{p(x,y)p(z \mid y)}{p(y)} \overset{(2)}{=} p(x \mid y)p(z \mid y).$$ [2]

ii) Markov chain is symmetrical:

$$p(x \mid y) \overset{(3)}{=} \frac{p(x,z \mid y)p(y)}{p(y,z)} \overset{(4)}{=} p(x \mid y,z)$$ [2]

iii) Data processing theorem:

$$I(X;Y) \overset{(5)}{=} I(X;Z) + I(X;Y \mid Z)$$

$$\overset{(6)}{\Rightarrow} I(X;Y) \geq I(X;Z) \text{ and } I(X;Y) \geq I(X;Y \mid Z)$$

[5]

b) Lossless source coding.

Clarification

'4-ary → 4 code symbols.

12:16.

i) Use the Kraft inequality to show that it is possible to construct a 4-ary instantaneous code with lengths {1, 1, 2, 2, 2, 2, 2, 2, 2}. [3]

ii) Construct such an instantaneous code for eight symbols that take the values *A*, *B*, …, *H*, *I* with probabilities {0.15, 0.15, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1}.

[4]

c) Reverse water filling. Let X_1, X_2, X_3 be independent zero-mean Gaussian information sources, with different variances $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$. Find the rate-distortion function $R(D)$ and corresponding range of D, for the following cases:

i) All the three sources are encoded. [3]

ii) Only X_2, X_3 are encoded. [3]

iii) Only X_3 is encoded.

[3]

3.

a) Consider the Gaussian channel with feedback shown in Fig. 3.1.

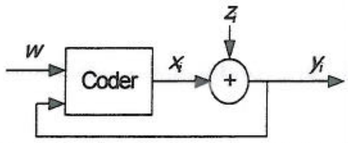


Fig. 3.1. Gaussian channel with feedback.

Justify each step of the following proof that feedback increases the capacity by ½ bits at most.

$$I(W;\mathbf{Y}) = h(\mathbf{Y}) - h(\mathbf{Y}\,|\,W) = h(\mathbf{Y}) - \sum_{i=1}^{n} h(Y_i\,|\,W, Y_{1:i-1})$$

$$\overset{(1)}{=} h(\mathbf{Y}) - \sum_{i=1}^{n} h(Y_i\,|\,W, Y_{1:i-1}, X_{1:i}, Z_{1:i-1})$$

$$\overset{(2)}{=} h(\mathbf{Y}) - \sum_{i=1}^{n} h(Z_i\,|\,W, Y_{1:i-1}, X_{1:i}, Z_{1:i-1})$$

$$\overset{(3)}{=} h(\mathbf{Y}) - \sum_{i=1}^{n} h(Z_i\,|\,Z_{1:i-1}) = h(\mathbf{Y}) - h(\mathbf{Z}) \overset{(4)}{\le} \frac{1}{2}\log_2 \frac{|\mathbf{K_y}|}{|\mathbf{K_z}|}$$

$$C_{n,FB} \overset{(5)}{=} \max_{\operatorname{tr}(\mathbf{K_x})\le nP} \tfrac{1}{2} n^{-1} \log_2 \frac{|\mathbf{K_y}|}{|\mathbf{K_z}|} \le \max_{\operatorname{tr}(\mathbf{K_x})\le nP} \tfrac{1}{2} n^{-1} \log_2 \frac{|2(\mathbf{K_x}+\mathbf{K_z})|}{|\mathbf{K_z}|}$$

$$\overset{(6)}{=} \max_{\operatorname{tr}(\mathbf{K_x})\le nP} \tfrac{1}{2} n^{-1} \log_2 \frac{2^n |\mathbf{K_x}+\mathbf{K_z}|}{|\mathbf{K_z}|} \overset{(7)}{=} \tfrac{1}{2} + \max_{\operatorname{tr}(\mathbf{K_x})\le nP} \tfrac{1}{2} n^{-1} \log_2 \frac{|\mathbf{K_x}+\mathbf{K_z}|}{|\mathbf{K_z}|} \overset{(8)}{=} \tfrac{1}{2} + C_n$$

[8]

b) Slepian-Wolf coding. The achievable region of Slepian-Wolf coding is given by

$$R_1 \ge H(X\,|\,Y)$$
$$R_2 \ge H(Y\,|\,X)$$
$$R_1 + R_2 \ge H(X,Y)$$

which corresponds to the following picture:

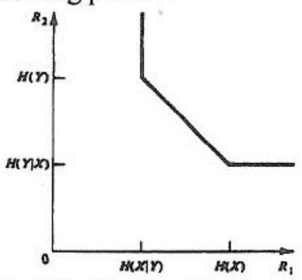


Fig. 3.2. Slepian-Wolf region.

Consider the corner point $R_1 = H(X)$, $R_2 = H(Y|X)$. Describe a joint typicality-based coding scheme that achieves the rates R_1 and R_2 at this corner point.

[7]

c) Max-flow min-cut.

i) Explain the max-flow min-cut bound for an information network. Is it achievable in general? [3]

ii) Show that max-flow min-cut bound is achievable for the two-user multiple access channel. [3]

iii) Show that max-flow min-cut bound is achievable for the Gaussian relay channel. Hint: the capacity of the Gaussian relay channel is given by

$$C = \max_{0 \le \alpha \le 1} \min \left\{ C\left(\frac{P + P_1 + 2\sqrt{(1-\alpha) P P_1}}{N_1 + N_2} \right), C\left(\frac{\alpha P}{N_1} \right) \right\}.$$ [4]

4.

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 $p_{xy}(x,y)$. Let $p_x(x)$ and $p_y(y)$ be the probability distribution functions of X_i and Y_i respectively. The jointly typical set is the set of vector pairs satisfying the following conditions:

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

where, since the sequences are i.i.d., the probability $p_x(\mathbf{x}) = \prod_{i=1}^n p_x(x_i)$ and $p_y(\mathbf{y})$ and $p_{xy}(\mathbf{x},\mathbf{y})$ can be written in a similar fashion. Now consider joint typicality arising from the binary symmetric channel $Y = X \oplus Z$ shown in Fig. 4.1, where the cross-over probability is $p = P(Z = 1) = 0.1$. Let the input distribution be the uniform distribution, i.e., $p_x(x) = (0.5, 0.5)$.

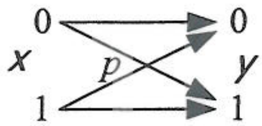


Fig. 4.1. Binary symmetric channel.

a) Calculate $H(X)$, $H(Y)$, $H(X, Y)$ and $I(X ; Y)$. [5]

b) Of the 2^n possible input sequences of length n, how many of them are typical, i.e., member of $T_\varepsilon^{(n)}(\mathbf{x}) = \{x : \left|-n^{-1}\log_2 p_x(x) - H(X)\right| < \varepsilon\}$ for $\varepsilon = 0.2$? How many are the typical sequences in $T_\varepsilon^{(n)}(\mathbf{y})$? [5]

c) It can be derived (details omitted) that $p(\mathbf{x},\mathbf{y}) = 2^{-n}(1-p)^{n-k}p^k$ where k is the number of places where the two sequences $\mathbf{x}$ and $\mathbf{y}$ differ. Derive the condition that $\mathbf{z}$ is typical, and show that $(\mathbf{x}, \mathbf{y})$ being jointly typical is equivalent to the condition that $\mathbf{x}$ is typical and $\mathbf{z}$ is typical. [5]

d) When $n = 25$, one obtains the table in the following page. Determine the size and probability of the jointly typical set $J_\varepsilon^{(n)}$ for $\varepsilon = 0.2$. [5]

k	$\binom{n}{k}$	$\sum_{j\le k}\binom{n}{j}$	$p(x^n)=p^k(1-p)^{n-k}$	$\binom{n}{k}p^k(1-p)^{n-k}$	Cumul. pr.	$-\frac{1}{n}\log p(x^n)$
0	1	1	7.178975e-02	0.071790	0.071790	0.152003
1	25	26	7.976639e-03	0.199416	0.271206	0.278800
2	300	326	8.862934e-04	0.265888	0.537094	0.405597
3	2300	2626	9.847704e-05	0.226497	0.763591	0.532394
4	12650	15276	1.094189e-05	0.138415	0.902006	0.659191
5	53130	68406	1.215766e-06	0.064594	0.966600	0.785988
6	177100	245506	1.350851e-07	0.023924	0.990523	0.912785
7	480700	726206	1.500946e-08	0.007215	0.997738	1.039582
8	1081575	1807781	1.667718e-09	0.001804	0.999542	1.166379
9	2042975	3850756	1.853020e-10	0.000379	0.999920	1.293176
10	3268760	7119516	2.058911e-11	0.000067	0.999988	1.419973
11	4457400	11576916	2.287679e-12	0.000010	0.999998	1.546770
12	5200300	16777216	2.541865e-13	0.000001	0.999999	1.673567

e) Assume that 2^{nR} codewords are chosen uniformly over the 2^n possible binary sequences of length n. The decoder uses jointly typical decoding, i.e., it looks at the received sequence and tries to find a codeword in the code that is jointly typical with the received sequence. What is the decoding error probability?

[5]

EE4-40 1.

Information Theory EE9SO20

Solutions 2012

B-Bookwork, E-New example, A-New application, T-New theory

1.

a)

(i) and (ii) are straightforward calculus.

[4B]

For the others, assume $½ < p < 1$ for convenience (other half follows by symmetry).

(iii) Since $H''(p) < 0$, $H(p)$ is concave and so lies above the straight line $2 - 2p$.

[3E]

(iv) At $p = ½$ the bound on the right has the same value and first two derivatives as $H(p)$. For $½ < p < 1$ its second derivative is greater than $H''(p)$ and so the bound follows.

[3E]

b)
(i) The probability of any given value of $x_{1:4}$ depends on the number of 1's and 0's. We create four subsets with equal probabilities to generate a pair of bits and two other subsets to generate one bit only.
Since the probabilities are equal, the pairs 00, 01, 10, 11 also have equal probabilities.
For the same reason, bits 0 and 1 have equal probabilities.
Therefore, the bits are i.i.d. Bernoulli random numbers.

[5A]

The expected number of bits generated is

$$E[K] = \underbrace{1 \times 2p^2(1-p)^2}_{\text{one bit}} + \underbrace{2 \times 4p(1-p)^3 + 2 \times 4p^2(1-p)^2 + 2 \times 4p^3(1-p)}_{\text{two bits}}$$
$$= 8p(1-p)^3 + 10p^2(1-p)^2 + 8p^3(1-p)$$

[5A]

(ii) i.i.d entropies add, (b) functions reduce entropy, (c) chain rule, (d) z_i are i.i.d. with entropy of 1 bit, (e) entropy is positive.

[5A]

2.

a) i) (1) $p(x,z|y) = \dfrac{p(x,z,y)}{p(y)} = \dfrac{p(x,y)\,p(z|x,y)}{p(y)}$ [9 B]

$= \dfrac{p(x,y)\,p(z|y)}{p(y)}$ Markov chain

(2) $\dfrac{p(x,y)}{p(y)} = p(x|y)$ probability

ii) (3) $p(x|y) = \dfrac{p(x,y)}{p(y)} = \dfrac{p(x,y)\,p(z|y)}{p(y)\,p(z|y)} = \dfrac{p(x,y)\,p(z|x,y)}{p(y)\,p(z|y)}$

$= \dfrac{P(x,y,z)}{P(y,z)} = \dfrac{P(x,z|y)\,p(y)}{P(y,z)}$ Markov Chain

(4) $= p(x|y,z)$

iii) (5) definition

(6) definition

(7) Conditioned on Y, X and Z are independent

(8) algebra

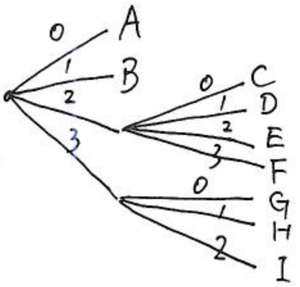
(9) Mutual infomation is non-negative.

b) i) Kraft's inequality

$$\sum D^{-l_i} \le 1$$ [3 E]

$D=4$, $\Rightarrow 2\times4^{-1} + 7\times4^{-2} = \frac{15}{16} \le 1$

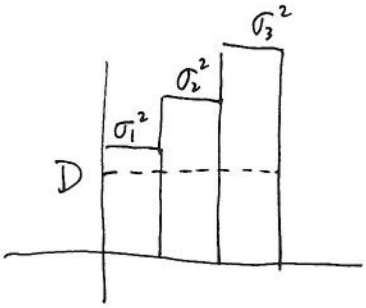
ii) Use a 4-ary tree



Codewords [4 E]

0
1
20
21
22
23
30
31
32

c) i)

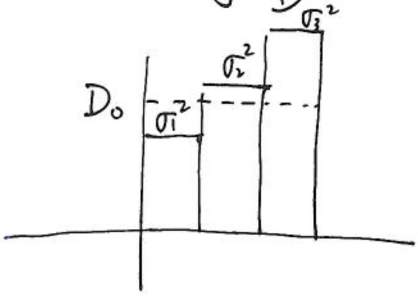


[3 A]

$0 < D < \sigma_1^2$

$$R(D) = \sum \frac{1}{2} \log \frac{\sigma_i^2}{D}$$

ii)



[3 A]

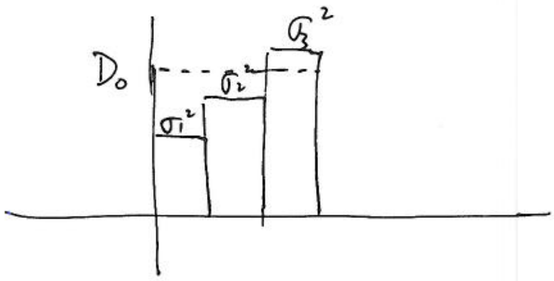
$\sigma_1^2 < D_0 < \sigma_2^2$

$$\sigma_1^2 + 2D_0 = 3D \Rightarrow D_0 = \frac{1}{2}(3D - \sigma_1^2)$$

$$\Rightarrow \sigma_1^2 < D < \frac{\sigma_1^2 + 2\sigma_2^2}{3}$$

$$R(D) = \frac{1}{2} \log \frac{\sigma_2^2}{D} + \frac{1}{2} \log \frac{\sigma_3^2}{D}$$

iii)



[3 A]

$\sigma_2^2 < D_0 < \sigma_3^2$

$$\sigma_1^2 + \sigma_2^2 + D_0 = 3D \Rightarrow D_0 = 3D - \sigma_1^2 - \sigma_2^2$$

$$\Rightarrow \frac{\sigma_1^2 + 2\sigma_2^2}{3} < D < \frac{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}{3}$$

$$R(D) = \frac{1}{2} \log \frac{\sigma_3^2}{D}$$

3. a) (1) $X_i = f_i(W, Y_{1:i-1})$, $Z_j = Y_j - X_j \quad 1 \le j \le i-1$ [8B]

(2) $Y_i = X_i + Z_i$ Shift doesn't change entropy

(3) given $Z_{1:i-1}$, Z_i and $(W, Y_{1:i-1}, X_{1:i})$ are conditionally independent.

(4) $h(Y) \le \frac{1}{2}\log|2\pi e K_Y|$ Gaussian has maximum entropy.

$h(Z) = \frac{1}{2}\log|2\pi e K_Z|$

(5) definition

(6) $|2K| = 2^n |K|$ for matrix K

(7) algebra

(8) $C_n = \max \frac{1}{2n}\log\frac{|K_X + K_Z|}{|K_Z|}$

b) At corner point $R_1 = H(X)$, $R_2 = H(Y|X)$, [7B]

We encode X as usual, at rate $H(X)$;

We consider joint typicality: associated each x, there is a jointly typical fan of size approximately $2^{nH(Y|X)}$. We use $2^{nH(Y|X)}$ colors to represent them.

We encode Y as its color at rate $H(Y|X)$.

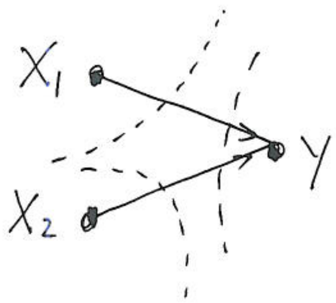
The receiver sees ~~(X, Y)~~ both codewords and will be able to recover X and Y based on joint typicality.

c)

i) Max-flow min-cut: Maximum flow through a network is upper-bounded by the minimum sum capacity of cut edges. In general, it is not achievable.

[2B]

ii)



MAC channel

[4T]

We calculate the bounds using three cuts

$$R_1 \le I(X_1; Y | X_2)$$

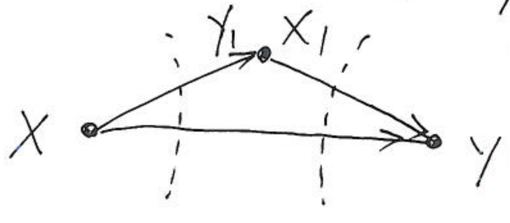
$$R_2 \le I(X_2; Y | X_1)$$

$$R_1 + R_2 \le I(X_1, X_2; Y)$$

These bounds coincide with the capacity region of the MAC.

iii)

Relay channel



[4T]

There are two cuts

$$C \le \min\{I(X, X_1; Y),\ I(X; Y, Y_1 | X_1)\}$$

4\. a) $H(X) = H(Y) = 1$

[5E]

$$H(X,Y) = H(X) + H(Y|X) = H(X) + H(p)$$
$$= 1 - 0.9\log 0.9 - 0.1\log 0.1 = 1.469$$

$$I(X;Y) = H(Y) - H(Y|X) = 0.531$$

b) Since X is uniformly distributed, every sequence has probability 2^{-n}, hence for every sequence,

[5 A]

$$-\frac{1}{n}\log p(X) = 1 = H(X)$$

Therefore, all the 2^n sequences are in $T_\varepsilon^{(n)}(X)$.

The same is true for Y.

c) (X,Y) is jointly typical

$\Rightarrow$ both X and Y are typical (from the first two conditions)

The third condition

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

(X,Y) is jointly typical $\Rightarrow$

$$\left|1 - \frac{k}{n}\log p - \frac{n-k}{n}\log(1-p) - H(X,Y)\right| < \varepsilon$$
$$\left|-\frac{k}{n}\log p - \frac{n-k}{n}\log(1-p) - H(p)\right| < \varepsilon$$

$\Rightarrow$ Z is typical

[5T]

d) Since every x is typical, we only need to be concerned with z. Notice that $H(z) = H(p) = 0.469$.

For $\varepsilon = 0.2$, the table show that the following rows and typical: $(0.269 \sim 0.669)$

$k = 1, 2, 3, 4$ [5T]

Size: $15276 - 1 = 15275$

probability: $0.902006 - 0.071790 = 0.83$

e) Decoding error $\Leftarrow$ (X, Y) are not jointly typical.

$\Rightarrow P_e = 1 - 0.83 = 0.17$

[5T]