

Master copy -
June 2008

Paper Number(s): **E4.40**
C5.27
SO20
ISE4.51

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE
UNIVERSITY OF LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2008

MSc and EEE/ISE PART IV: MEng and ACGI

INFORMATION THEORY

Monday, 12 May 10:00 am

There are SIX questions on this paper.

Answer FOUR questions.

All questions carry equal marks

Time allowed: 3:00 hours

Examiners responsible:

First Marker(s): D.M. Brookes

Second Marker(s): C. Ling

Information for Candidates:

- Notation:**
- (a) Random variables are shown in a sans serif typeface. Thus $x, \mathbf{x}, \mathbf{X}$ denote a random scalar, vector and matrix respectively. The alphabet of a discrete random scalar, \mathcal{X} , is denoted by \mathbf{X} and its size by $|\mathbf{X}|$.
 - (b) $\mathbf{x}_{1:n}$ denotes the sequence x_1, x_2, \dots, x_n .
 - (c) The normal distribution function is denoted by:
$$N(x; \mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp(-1/2(x - \mu)^2 \sigma^{-2})$$
 - (d) \oplus denotes the exclusive-or operation or, equivalently, addition modulo 2.
 - (e) $\log x = \frac{\ln x}{\ln 2}$ denotes logarithm to base 2.
 - (f) $P(\bullet)$ denotes the probability of the discrete event \bullet .
 - (g) “i.i.d.” denotes “independent identically distributed”

The Questions

1. (a) If \mathbf{p} is an arbitrary probability mass vector and \mathbf{q} is a uniform probability mass vector with the same number of elements, show that $H(\mathbf{p}) \leq H(\mathbf{q})$. You may assume without proof that $D(\mathbf{p} \parallel \mathbf{q}) = \sum_i p_i \log \left(\frac{p_i}{q_i} \right) \geq 0$. [3]

- (b) X and Y are Bernoulli random variables. They are added together to form $Z = X + Y$ which lies in the range 0 to 2.
 - (i) By considering the alternative expansions [5]

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

Show that if X and Y are independent, $H(Z) \geq H(Y)$.
 - (ii) Demonstrate that the independence criterion is necessary by specifying a joint distribution for X and Y for which $H(X) = H(Y) = 1$ but $H(Z) = 0$. [3]

- (c) A cable connecting two buildings contains 6 indistinguishable wires; in order to use the cable, you need to determine which wire connects to which. The wires are labelled A, B, C, D, E, F at one end and R, S, T, U, V, W at the other.

The random variable $Z \in \{1 : 720\}$ indicates which of the $6! = 720$ possible connection patterns is true. You propose to determine Z by connecting various combinations of the wires together at one end while a friend measures the connectivity between wires at the other.

 - (i) Give the value of $H(Z)$ if all of the $6!$ possible connection patterns have equal probability. [1]
 - (ii) You connect the wires in pairs $A=B$, $C=D$, $E=F$ and determine the connectivity between the six wires R, ..., W. If m_1 denotes the result of this measurement, determine the value of $H(Z | m_1)$. [3]
 - (iii) m_2 denotes the result of measuring the connectivity between R, ..., W if you connect $A=B$ and $C=D=E$ instead of the pairwise connection pattern given in part (ii). Determine the value of $H(Z | m_2)$. [3]
 - (iv) You now connect $A=C$ and $B=D=F$ and measure the connectivity between R, ..., W. If m_3 denotes the result of this measurement, determine the value of $H(Z | m_2, m_3)$. [2]

2. The pixels of a binary-valued image are transmitted as a stream of bits, x_i . The bitstream is modelled as a stationary Markov process with the joint probability, $P(x_{i-1}, x_i)$ as follows:

		x_i	
		0	1
x_{i-1}	0	0.6	0.05
	1	0.05	0.3

The following values of $H(p)$ may be helpful in this question:

p	0.0769	0.1429	0.2462	0.2857	0.3017	0.4341
$H(p)$	0.3912	0.5917	0.8051	0.8631	0.8834	0.9875

- Determine the probability mass vector for x_i and the entropy rate, $H(\mathbf{X})$, of the process. [4]
- A Huffman encoder is used to encode pairs of bits, (x_{i-1}, x_i) . Design the encoder and determine the expected number of encoded bits per pixel-pair. [4]
- In a noisy version of the image, y_i , each pixel is corrupted independently by being inverted with probability 0.2. Determine the joint probability functions $P(x_{i-1}, y_i)$ and $P(y_{i-1}, y_i)$. [6]
- Calculate $H(y_i | x_{i-1})$ and $H(y_i | y_{i-1})$ and explain why the entropy rate of the Hidden Markov process $\{y_i\}$ must lie between these two values. [6]

3. Figure 3.1 shows two communications channels connected in series. The first connects the Bernoulli random variables X and Y while the second connects Y and Z . The probabilities that X , Y and Z equal 1 are p_x , $p_y = (1-f)p_x$ and $p_z = g + (1-2g)p_y$ respectively. The error probabilities are $f = 0.125$ and $g = 0.1$ as shown.

The following values of $H(p)$ may be helpful in this question:

p	0.1	0.2	0.394	0.4377
$H(p)$	0.469	0.7219	0.9673	0.9888

- (a) Considering first the binary symmetric channel linking Y and Z , justify each step of the following derivation

$$\begin{aligned}
 I(Y; Z) &\stackrel{(i)}{=} H(Z) - H(Z|Y) \\
 &\stackrel{(ii)}{=} H(p_z) - H(Z|Y=0)(1-p_y) - H(Z|Y=1)p_y \\
 &\stackrel{(iii)}{=} H(g + (1-2g)p_y) - H(g)
 \end{aligned}$$

Determine (as a numerical value) the value of p_y that maximizes this expression and hence the capacity of the channel. [5]

- (b) For the channel linking X and Y , derive an expression for $I(X; Y)$ in terms of f and p_x . Hence find the capacity of the channel and the value of p_x that attains it. [7]

You may assume without proof that $\frac{dH(p)}{dp} = \log(p^{-1} - 1)$.

- (c) Calculate the transition probabilities of the combined channel linking X to Z . Determine the capacity of this channel and the value of p_x that attains it. [7]
- (d) By how much could the capacity of the combined channel be increased if it was possible to recode Y before transmission through the binary symmetric channel. [1]

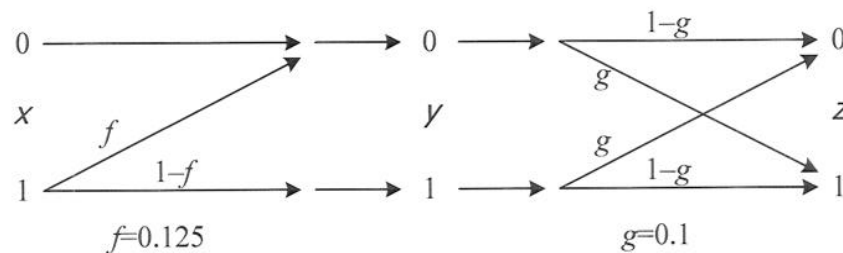


Figure 3.1

4. In the discrete-time channel of *Figure 4.1*, X and Y are continuous random variables and the zero-mean additive noise Z is identically distributed for each use of the channel and is independent of X . The variance of X is P and the variance of Z is N .
- (a) If Z is Gaussian, justify each step of the following

$$\begin{aligned}
 I(X; Y) &\stackrel{(i)}{=} h(Y) - h(Y|X) \stackrel{(ii)}{=} h(Y) - h(X+Z|X) \\
 &\stackrel{(iii)}{=} h(Y) - h(Z|X) \stackrel{(iv)}{=} h(Y) - h(Z) \\
 &\stackrel{(v)}{\leq} \frac{1}{2} \log(2\pi e(P+N)) - \frac{1}{2} \log(2\pi eN) \\
 &\stackrel{(vi)}{=} \frac{1}{2} \log\left(\frac{P+N}{N}\right)
 \end{aligned}
 \tag{6}$$

Hence give the channel capacity, C , and the distribution of X that attains it.

- (b) If, now, Z is non-Gaussian and we define the noise entropy power, Q , by

$$Q = (2\pi e)^{-1} 2^{2h(Z)}, \tag{2}$$

- (i) show that the channel capacity satisfies $C \leq \frac{1}{2} \log\left(\frac{P+N}{Q}\right)$
- (ii) using the “power inequality”, $2^{2h(Y)} \geq 2^{2h(X)} + 2^{2h(Z)}$, which you may assume without proof, derive a lower bound on C in terms of P and Q . [6]
- (c) Suppose now that $P = 24$ and that Z is uniformly distributed in the range -3 to $+3$.
- (i) Evaluate the capacity bounds from parts (b)(i) and (b)(ii). [3]
- (ii) Determine $I(X; Y)$ if X takes the values -6 , 0 and $+6$ with equal probability. [3]

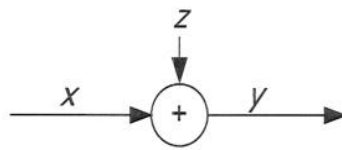


Figure 4.1

5. \mathbf{x} and \mathbf{y} are discrete-valued random vectors of length n where each pair (x_i, y_i) is drawn independently from the joint probability mass function $p_{xy}(x, y)$. The jointly typical set, $J_\varepsilon^{(n)}$, is the set of vector pairs satisfying the following conditions:

$$J_\varepsilon^{(n)} = \left\{ \mathbf{x}, \mathbf{y} : \begin{aligned} & \left| -n^{-1} \log(p_x(\mathbf{x})) - H(X) \right| \leq \varepsilon, \\ & \left| -n^{-1} \log(p_y(\mathbf{y})) - H(Y) \right| \leq \varepsilon, \\ & \left| -n^{-1} \log(p_{xy}(\mathbf{x}, \mathbf{y})) - H(X, Y) \right| \leq \varepsilon \end{aligned} \right\}$$

where $p_x(x)$ and $p_y(y)$ are the probability mass functions of x_i and y_i respectively.

The probability $p_x(\mathbf{x}) = \prod_{i=1}^n p_x(x_i)$ and similarly for $p_y(\mathbf{y})$ and $p_{xy}(\mathbf{x}, \mathbf{y})$.

- (a) Justify each of steps (i) to (iv) in the following derivation of an upper bound for $|J_\varepsilon^{(n)}|$, the size of $J_\varepsilon^{(n)}$:

$$1 \geq \sum_{\mathbf{x}, \mathbf{y} \in J_\varepsilon^{(n)}} p_{xy}(\mathbf{x}, \mathbf{y}) \stackrel{(i)}{\geq} |J_\varepsilon^{(n)}| \min_{\mathbf{x}, \mathbf{y} \in J_\varepsilon^{(n)}} p_{xy}(\mathbf{x}, \mathbf{y}) \stackrel{(iii)}{\geq} |J_\varepsilon^{(n)}| 2^{-nH(X, Y) - n\varepsilon} \stackrel{(iv)}{\Rightarrow} |J_\varepsilon^{(n)}| \leq 2^{nH(X, Y) + n\varepsilon} \quad [4]$$

- (b) \mathbf{z} is a discrete random vector, independent of \mathbf{x} , whose elements are drawn independently from the same probability mass function as y_i , i.e. $p_{xz}(x, z) = p_x(x)p_y(z)$.

$$(i) \quad \text{Show that } \max_{\mathbf{x}, \mathbf{z} \in J_\varepsilon^{(n)}} p_{xz}(\mathbf{x}, \mathbf{z}) \leq 2^{-nH(X) + n\varepsilon} 2^{-nH(Y) + n\varepsilon} \quad [2]$$

$$(ii) \quad \text{Hence derive an upper bound on } P(\mathbf{x}, \mathbf{z} \in J_\varepsilon^{(n)}) \quad [4]$$

- (c) Now suppose that $n = 11$ and $\varepsilon = 0$ and that $p_{xy}(x, y)$ is given by

	$y=0$	$y=1$
$x=0$	5/11	2/11
$x=1$	1/11	3/11

We define the typical set $T_{\mathbf{x}} = \{\mathbf{x} : -n^{-1} \log p_x(\mathbf{x}) = H(X)\}$.

- (i) Show that $\mathbf{x} \in T_{\mathbf{x}}$ if and only if exactly 4 of the x_i equal 1.

$$\text{Hence show that the probability of this is } P(\mathbf{x} \in T_{\mathbf{x}}) = C_{11}^4 (4/11)^4 (7/11)^7 \quad [2]$$

where $C_n^k = n!/(k!(n-k)!)$ denotes a binomial coefficient.

$$(ii) \quad \text{Explain why } P(\mathbf{x}, \mathbf{y} \in J_0^{(11)} | \mathbf{x} \in T_{\mathbf{x}}) = C_7^2 (2/7)^2 (5/7)^5 C_4^3 (3/4)^3 (1/4) \quad [2]$$

$$(iii) \quad \text{Hence determine the value of } P(\mathbf{x}, \mathbf{y} \in J_0^{(11)}) \quad [2]$$

$$(iv) \quad \text{If } \mathbf{z} \text{ is a random vector, independent of } \mathbf{x}, \text{ whose elements are independent Bernoulli variables with } P(z_i = 1) = 5/11, \text{ calculate } P(\mathbf{x}, \mathbf{z} \in J_0^{(11)}) \quad [4]$$

6. The continuous random variable X has zero mean and variance σ^2 . We define the information rate-distortion function for X to be $R(D) = \min I(X; \hat{X})$ where the minimum is taken over all conditional distributions $p(\hat{X} | X)$ for which $E((X - \hat{X})^2) \leq D$. You may assume without proof that $h(X) \leq h(Y) = \frac{1}{2} \log(2\pi\sigma^2)$ where Y is Gaussian with variance σ^2 .

- (a) Carefully justify each step in the following bound and given the conditions for equality in steps (iii) to (v): [6]

$$\begin{aligned}
 I(X; \hat{X}) &\stackrel{(i)}{=} h(X) - h(X | \hat{X}) \\
 &\stackrel{(ii)}{=} h(X) - h(X - \hat{X} | \hat{X}) \\
 &\stackrel{(iii)}{\geq} h(X) - h(X - \hat{X}) \\
 &\stackrel{(iv)}{\geq} h(X) - \frac{1}{2} \log(2\pi \text{Var}(X - \hat{X})) \\
 &\stackrel{(v)}{\geq} h(X) - \frac{1}{2} \log(2\pi D)
 \end{aligned}$$

- (b) In the diagram of Figure 6.1, Z is independent of X and is zero-mean Gaussian with variance kD where $k = 1 - D\sigma^{-2}$ for $D \leq \sigma^2$.

- (i) Show that $E((X - \hat{X})^2) = D$. [2]

- (ii) Show that $\text{Var}(\hat{X}) = \sigma^2 - D$. [2]

- (iii) By expanding $I(X; \hat{X})$ as $h(\hat{X}) - h(\hat{X} | X)$, show that $R(D) \leq \frac{1}{2} \log(\sigma^2 D^{-1})$. [5]

- (c) If X is uniformly distributed in the interval $(-\frac{1}{2}, +\frac{1}{2})$ and is encoded with 1-bit per sample as $\hat{X} \in \{-\frac{1}{4}, +\frac{1}{4}\}$, determine the distortion, $D = E((X - \hat{X})^2)$, together with the bounds defined in parts (a) and (b). Comment on the relationship between the actual bit-rate and the bounds. [5]

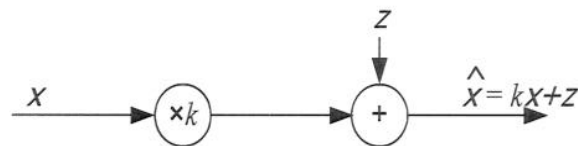


Figure 6.1