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 $\mathsf{x}, \mathbf{\mathsf{x}}, \mathbf{\mathsf{X}}$ denote a random scalar, vector and matrix respectively. The alphabet of a discrete random scalar, x, is denoted by $\boldsymbol{\mathcal{X}}$ and its size by $|\boldsymbol{\mathcal{X}}|$.

(b) $\mathsf{x}_{1:n}$ denotes the sequence $\mathsf{x}_1, \mathsf{x}_2, \cdots, \mathsf{x}_n$.

(c) The normal distribution function is denoted by:
$N(x;\mu,\sigma^2) = \left(2\pi\sigma^2\right)^{-½} \exp(-½(\mathrm{x} - \mu)^2\sigma^{-2})$

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

(e) $\log x = \dfrac{\ln x}{\ln 2}$ 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 $H(\mathbf{p}) \le H(\mathbf{q})$. You may assume without proof that $D(\mathbf{p} \| \mathbf{q}) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right) \ge 0$. [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]

$$\begin{aligned} H(x,y,z) &= H(x) + H(y \mid x) + H(z \mid x,y) \\ &= H(x) + H(z \mid x) + H(y \mid x,z) \end{aligned}$$

Show that if x and y are independent, $H(z) \ge H(y)$.

(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 $H(z \mid m_1)$. [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 $H(z \mid m_2)$. [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 $H(z \mid m_2, m_3)$. [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	x_i = 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

(a) Determine the probability mass vector for x_i and the entropy rate, $H(\boldsymbol{X})$, of the process. [4]

(b) 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]

(c) 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]

(d) Calculate $H(y_i \mid x_{i-1})$ and $H(y_i \mid y_{i-1})$ 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

$$\begin{aligned} I(Y;Z) &\overset{\text{(i)}}{=} H(Z) - H(Z \mid Y) \\ &\overset{\text{(ii)}}{=} H(p_Z) - H(Z \mid Y=0)(1-p_Y) - H(Z \mid Y=1)p_Y \\ &\overset{\text{(iii)}}{=} H(g + (1-2g)p_Y) - H(g) \end{aligned}$$

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 $\dfrac{dH(p)}{dp} = \log\left(p^{-1} - 1\right)$.

(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]

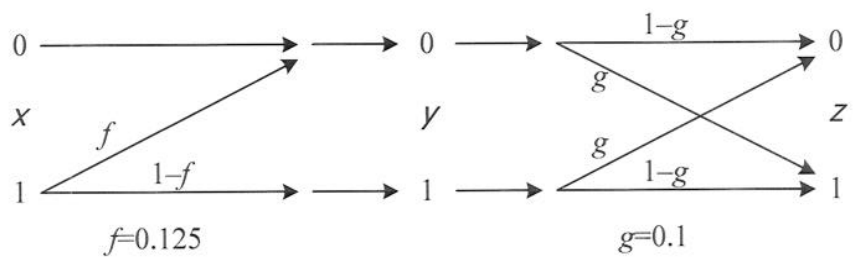

Figure 3.1

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

$$\begin{aligned} I(x;y) &\overset{\text{(i)}}{=} h(y) - h(y \mid x) \overset{\text{(ii)}}{=} h(y) - h(x+z \mid x) \\ &\overset{\text{(iii)}}{=} h(y) - h(z \mid x) \overset{\text{(iv)}}{=} h(y) - h(z) \\ &\overset{\text{(v)}}{\le} \tfrac{1}{2}\log\big(2\pi e(P+N)\big) - \tfrac{1}{2}\log(2\pi e N) \\ &\overset{\text{(vi)}}{=} \tfrac{1}{2}\log\left(\frac{P+N}{N}\right) \end{aligned}$$

[6]

Hence give the channel capacity, C, and the distribution of x that attains it. [2]

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

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

(i) show that the channel capacity satisfies $C \le \tfrac{1}{2}\log\left(\frac{P+N}{Q}\right)$

(ii) using the "power inequality", $2^{2h(y)} \ge 2^{2h(x)} + 2^{2h(z)}$, 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]

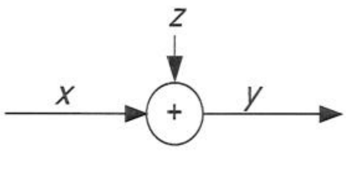

Figure 4.1

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:

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

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 $\left|J_\varepsilon^{(n)}\right|$, the size of $J_\varepsilon^{(n)}$: [4]

$$1 \overset{(i)}{\ge} \sum_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p_{xy}(\mathbf{x},\mathbf{y}) \overset{(ii)}{\ge} \left|J_\varepsilon^{(n)}\right| \min_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p_{xy}(\mathbf{x},\mathbf{y}) \overset{(iii)}{\ge} \left|J_\varepsilon^{(n)}\right| 2^{-nH(x,y)-n\varepsilon} \overset{(iv)}{\Rightarrow} \left|J_\varepsilon^{(n)}\right| \le 2^{nH(x,y)+n\varepsilon}$$

(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) Show that $\max_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} p_{xz}(\mathbf{x},\mathbf{z}) \le 2^{-nH(x)+n\varepsilon} 2^{-nH(y)+n\varepsilon}$ [2]

(ii) Hence derive an upper bound on $P\left(\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}\right)$. [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 $T_{\mathbf{x}} = \left\{\mathbf{x} : -n^{-1}\log p_x(\mathbf{x}) = H(x)\right\}$.

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

Hence show that the probability of this is $P(\mathbf{x}\in T_{\mathbf{x}}) = C_{11}^4 (4/11)^4 (7/11)^7$ where $C_n^k = n!/(k!(n-k)!)$ denotes a binomial coefficient. [2]

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

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

(iv) If $\mathbf{z}$ is a random vector, independent of $\mathbf{x}$, whose elements are independent Bernoulli variables with $P(z_i = 1) = 5/11$, calculate $P\left(\mathbf{x},\mathbf{z}\in J_0^{(11)}\right)$. [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 $p(\hat{x} \mid x)$ for which $E\left((x-\hat{x})^2\right) \le D$. You may assume without proof that $h(x) \le h(y) = ½\log\left(2\pi e\sigma^2\right)$ 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]

$$\begin{aligned} I(x;\hat{x}) &\overset{\text{(i)}}{=} h(x) - h(x \mid \hat{x}) \\ &\overset{\text{(ii)}}{=} h(x) - h(x-\hat{x} \mid \hat{x}) \\ &\overset{\text{(iii)}}{\ge} h(x) - h(x-\hat{x}) \\ &\overset{\text{(iv)}}{\ge} h(x) - ½\log\left(2\pi e \operatorname{Var}(x-\hat{x})\right) \\ &\overset{\text{(v)}}{\ge} h(x) - ½\log(2\pi e D) \end{aligned} .$$

(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 $D \le \sigma^2$.

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

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

(iii) By expanding $I(x;\hat{x})$ as $h(\hat{x}) - h(\hat{x} \mid x)$, show that $R(D) \le ½\log\left(\sigma^2 D^{-1}\right)$. [5]

(c) If x is uniformly distributed in the interval $(-½,+½)$ and is encoded with 1-bit per sample as $\hat{x} \in \{-¼,+¼\}$, determine the distortion, $D = E\left((x-\hat{x})^2\right)$, together with the bounds defined in parts (a) and (b). Comment on the relationship between the actual bit-rate and the bounds. [5]

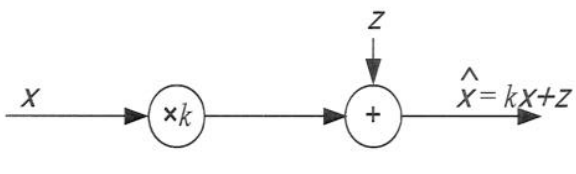

Figure 6.1