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 $\mathsf{x}, \mathbf{x}, \mathbf{X}$ 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) $\log x = \frac{\ln x}{\ln 2}$ 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 $H(\mathbf{p}) \le H(\mathbf{q})$. 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]

$$H(x, y, z) = H(x) + H(y \mid x) + H(z \mid x, y)$$
$$= H(x) + H(z \mid x) + H(y \mid x, z)$$

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

$$I(y;z) \overset{\text{(i)}}{=} H(z) - H(z\,|\,y)$$
$$\overset{\text{(ii)}}{=} H(p_z) - H(z\,|\,y=0)(1-p_y) - H(z\,|\,y=1)p_y$$
$$\overset{\text{(iii)}}{=} H(g+(1-2g)p_y) - H(g)$$

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]

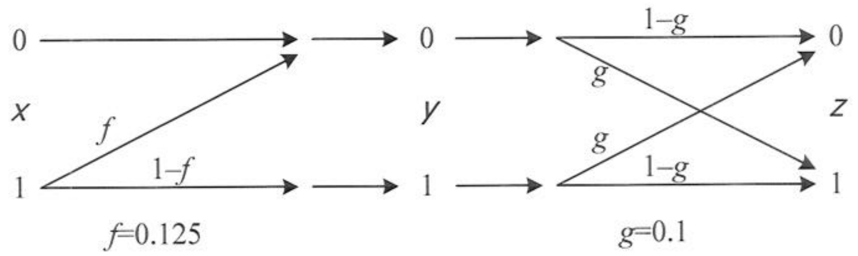

Figure 3.1

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

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

[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 $C \le \tfrac{1}{2}\log\left(\dfrac{P+N}{Q}\right)$

(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]

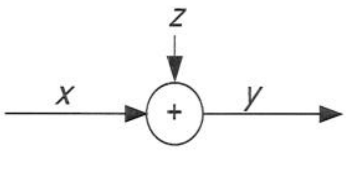

Figure 4.1

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:

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

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

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

(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]

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

(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]

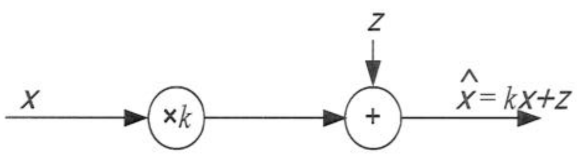

Figure 6.1

Master - June 2008

2008 E4.40/SO20 Solutions

Key to letters on mark scheme: B=Bookwork, C=New computed example, A=New analysis

1. (a) If $|\boldsymbol{X}|=n$ then $q_i = n^{-1}\ \forall i$. From the Information Inequality (given in question) [3B]

$$0 \le D(\mathbf{p}\,\|\,\mathbf{q}) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right) = -H(\mathbf{p}) + \sum_i p_i \log n \quad \Rightarrow \quad H(\mathbf{p}) \le \log n = H(\mathbf{q})$$

(b) (i) From the question,

$$H(x)+H(y\,|\,x)+H(z\,|\,x,y) = H(x)+H(z\,|\,x)+H(y\,|\,x,z).$$

However because $z = x + y$, $H(z\,|\,x,y) = H(y\,|\,x,z) = 0$. Also, because x and y are independent, $H(y\,|\,x) = H(y)$. Hence [5A]

$$H(x)+H(y) = H(x)+H(z\,|\,x) \le H(x)+H(z)$$

where the final inequality follows because conditioning reduces entropy. Subtracting $H(x)$ from both sides gives the required result.

(ii) If x is uniformly distributed and $y = 1 - x$, then $H(x) = H(y) = 1$ but $z \equiv 1$ so $H(z) = 0$. [3C]

(c) (i) $H(z) = \log 720 = 9.49$ bits. [1C]

(ii) The six wires R, …, W will form three pairs but you have no way of telling which pair is which. Thus there are 6 possible arrangements of the three pairs and two possible arrangements of the wires within each pair giving a total of $6 \times 2^3 = 48$ equally likely possibilities. Thus the entropy of z is $\log 48 = 5.58$ bits. [3C]

(iii) You now have three interconnection groups of sizes 1, 2 and 3. Because the sizes are all different, you can identify which is which and so your only uncertainty is the arrangement of wires within each group. Thus the number of possibilities is now $1! \times 2! \times 3! = 12$ which gives an entropy of $\log 12 = 3.58$ bits. [3C]

(iv) This final measurement uniquely identifies the wires since the members of any single group in the previous part are now in different groups. Thus the entropy is now 0. [2C]

2. (a) $\mathbf{p}_x = [0.65 \quad 0.35]$ [1C]

$$H(\mathcal{X}) = 0.65\times H(0.05/0.65) + 0.35\times H(0.05/0.35)$$
$$= 0.65\times H(0.077) + 0.35\times H(0.143)$$
$$= 0.65\times 0.391 + 0.35\times 0.592 = 0.461 \text{ bits}$$ [3C]

(b) Huffman Code for inputs [00, 11, 01, 10] with probs [0.6, 0.3, 0.05, 0.05] are [0, 10, 110, 111] giving an expected code length of 1.5 bits. [4C]

(c) The joint probability of (x_{i-1}, y_i) is

$$0.8\begin{pmatrix} .6 & .05 \\ .05 & .3 \end{pmatrix} + 0.2\begin{pmatrix} .05 & .6 \\ .3 & .05 \end{pmatrix} = \begin{pmatrix} .49 & .16 \\ .1 & .25 \end{pmatrix}$$ [3C]

and the joint probability of (y_{i-1}, y_i) is

$$0.8\begin{pmatrix} .49 & .16 \\ .1 & .25 \end{pmatrix} + 0.2\begin{pmatrix} .1 & .25 \\ .49 & .16 \end{pmatrix} = \begin{pmatrix} .412 & .178 \\ .178 & .232 \end{pmatrix}$$ [3C]

(d) We have

$$H(y_i \mid x_{i-1}) = 0.65\times H(0.16/0.65) + 0.35\times H(0.1/0.35)$$
$$= 0.65\times H(0.246) + 0.35\times H(0.286)$$
$$= 0.65\times 0.8051 + 0.35\times 0.8631 = 0.8254$$ [2C]

also, noting that $\mathbf{p}_y = [0.59 \quad 0.41]$,

$$H(y_i \mid y_{i-1}) = 0.59\times H(0.178/0.59) + 0.41\times H(0.178/0.41)$$
$$= 0.59\times H(0.302) + 0.41\times H(0.434)$$
$$= 0.59\times 0.8834 + 0.41\times 0.9875 = 0.926$$ [2C]

For a hidden markov process, we have $H(y_i \mid y_{i-1}, x_j) \le H(\mathcal{Y}) \le H(y_i \mid y_{i-1})$ for any $j \le i-1$ and, in particular for $j = i-1$. But since y_{i-1} depends only on x_{i-1}, we have $H(y_i \mid y_{i-1}, x_{i-1}) = H(y_i \mid x_{i-1})$. [2C]

3. (a) (i) This is the definition of mutual information. [1B]

(ii) We can decompose conditional entropy as a weighted sum of entropy conditional on specific values. [1B]

(iii) For both $y = 0$ and $y = 1$, z is a Bernoulli variable with probability vector $[1-g \quad g]$ and so its entropy is $H(g)$. [1B]

$H(z) = H(p_z)$ and, as given in the question preamble, $p_z = g + (1-2g)p_y$.

The value of $I(y;z)$ is maximized by making $g + (1-2g)p_y = 0.5$ which occurs when $p_y = ½$. The channel capacity is therefore $1 - H(g) = 1 - 0.469 = 0.531$ bits. [2B]

(b) In a similar way, we have

$$\begin{aligned} I(x;y) &= H(y) - H(y \mid x) \\ &= H(y) - H(y \mid x = 0)(1-p_x) - H(y \mid x = 1)p_x \\ &= H(y) - 0 - H(f)p_x \\ &= H\big(p_x(1-f)\big) - p_x H(f) \end{aligned}$$

[3A]

Setting the derivative with respect to p_x to zero gives

$$\begin{aligned} 0 = {}^{dI}\!/_{dp} &= (1-f)\log(p^{-1}(1-f)^{-1} - 1) - H(f) \\ &= 0.875\log(1.143p^{-1} - 1) - 0.5436 \end{aligned}$$

$$\Rightarrow \quad 1.143p^{-1} - 1 = 2^{0.5436/0.875} = 2^{0.6212} = 1.5382$$

$$\Rightarrow \quad p = 1.143/2.5382 = 0.4503$$ [2C]

$$\begin{aligned} \Rightarrow \quad I &= H(0.4503 \times 0.875) - 0.4503 \times 0.5436 \\ &= H(0.394) - 0.2447 = 0.9673 - 0.2447 = 0.7226 \text{ bits} \end{aligned}$$ [2C]

(c) The transition probabilities of the combined channel are

	z=0	z=1
x=0	$1-g=0.9$	$g=0.1$
x=1	$f+g-2fg=0.2$	1-f-g+2fg=0.8

Following the previous derivation, and noting that $p_z = 0.1+0.7p_x$, the channel capacity is [2C]

$C = H(p_z) - p_x H(0.2) - (1-p_x)H(0.1)$ [1A]

Setting the derivative w.r.t p_x to zero gives

$$0 = {dC}/{dp} = 0.7\log\left(p_z^{-1}-1\right) - 0.7219 + 0.469$$

$$\Rightarrow \quad p_z^{-1} - 1 = 2^{0.3613} = 1.2846 \quad \Rightarrow \quad p_z = 0.4337$$ [2C]

$$\Rightarrow \quad p = (0.4377-0.1)/0.7 = 0.4824$$

$$\Rightarrow \quad C = H(0.4377) - 0.4824\times 0.7219 - 0.5176\times 0.469$$

$$= 0.9888 - 0.3483 - 0.2427$$ [2C]

$$= 0.3978 \text{ bits}$$

(d) It would be possible to increase to the minimum capacity of the sub channels, namely 0.531 bits. [1A]

4. (a) (i) Definition of mutual information.

(ii) Definition of y.

(iii) Trnaslation invariance by a conditional constant.

(iv) z is independent of x.

(v) Gaussian bound for $h(y)$ in terms of its variance. Known $h(z)$ since Gaussian.

(vi) Algebra [6B]

We have equality in step (v) only if y is Gaussian. Since we know that z is Gaussian, this implies that x is also Gaussian.

(b) (i) Rearranging the definition of Q gives $h(z) = ½\log 2\pi e Q$ and the result follows if this is inserted into the previous part after step (iv). [2A]

(ii) From the previous part, $2^{2h(z)} = 2\pi e Q$ and, if we choose x to be Gaussian, $2^{2h(x)} = 2\pi e P$. Hence the power inequality becomes

$$2^{2h(y)} \geq 2\pi e(P+Q) \quad \Rightarrow \quad h(y) \geq ½\log\left(2\pi e(P+Q)\right)$$ [2A]

Hence, from part (a),

$$I(x;y) = h(y) - h(z)$$
$$\geq ½\log 2\pi e(P+Q) - ½\log 2\pi e Q = ½\log\frac{P+Q}{Q}$$ [2A]

Thus this is a lower bound on the capacity with Gaussian x so the capacity with arbitrary x must be at least this large. [2A]

(c) (i) For a uniform distribution $h(z) = \log 6 = 2.585$. This gives $Q = 36(2\pi e)^{-1} = 2.1078$. The variance of z is $N = 36/12 = 3$. Hence

$$1.8153 = ½\log\frac{P+Q}{Q} \leq C \leq ½\log\frac{P+N}{Q} = 1.8396$$ [3A]

(ii) $H(x) = \log 3 = 1.585$. In this case, we can achieve error-free decoding since the inputs are separated by twice the maximum noise amplitude. Hence $H(x \mid y) = 0$ and $I(x;y) = 1.585$ bits. [3A]

5. (a) (i) Total probability cannot exceed 1.

(ii) A finite sum cannot exceed the minimum summand multiplied by the number of summands.

(iii) The minimum value of $P(\mathbf{x},\mathbf{y})$ within the jointly typical set is given in its definition.

(iv) Algebra [4B]

(b) (i) $\max_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} P(\mathbf{x},\mathbf{z}) = \max_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} P(\mathbf{x})P(\mathbf{z}) \le \max_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} P(\mathbf{x}) \max_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} P(\mathbf{z}) \le 2^{-nH(X)+n\varepsilon}2^{-nH(Y)+n\varepsilon}$ [2B]

(ii) We can write

$$\sum_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} P(\mathbf{x},\mathbf{z}) \le \left|J_\varepsilon^{(n)}\right| \max_{\mathbf{x},\mathbf{z}\in J_\varepsilon^{(n)}} P(\mathbf{x},\mathbf{z}) \le 2^{nH(X,Y)+n\varepsilon}2^{-nH(X)-nH(Y)+2n\varepsilon}$$ [4B]

$$= 2^{nH(X,Y)-nH(X)-nH(Y)+3n\varepsilon}$$

(c) (i) $H(X) = -(7/11)\log(7/11) - (4/11)\log(4/11)$. Therefore $\mathbf{x}\in T_\mathbf{x}$ if and only if $P(\mathbf{x}) = 2^{-11H(X)} = (7/11)^7(4/11)^4$. This can only happen if exactly 4 of the x_i are equal to 1. The number of ways of choosing 4 out of 11 x_i is C_{11}^4 and so the result follows. [2A]

(ii) If, in addition, $\mathbf{x},\mathbf{y}\in J_0^{(11)}$, then we require that $y_i = 1$ for 2 out of the 7 i for which $x_i = 0$ and for 3 out of the 4 i for which $x_i = 0$. This gives the required expression. [2A]

(iii) $P\left(\mathbf{x},\mathbf{y}\in J_0^{(11)}\right) = P\left(\mathbf{x},\mathbf{y}\in J_0^{(11)} \mid \mathbf{x}\in T_\mathbf{x}\right)P\left(\mathbf{x}\in T_\mathbf{x}\right) = 0.1345\times 0.2438 = 0.0328$. [2A]

(iv) We now have

$$P(\mathbf{x},\mathbf{z}\in J_0^{(11)} \mid \mathbf{x}\in T_\mathbf{x}) = C_7^2(5/11)^2(6/11)^5 C_4^3(5/11)^3(6/11) = 0.0429$$

Which gives [4A]

$$P\left(\mathbf{x},\mathbf{z}\in J_0^{(11)}\right) = P\left(\mathbf{x},\mathbf{y}\in J_0^{(11)} \mid \mathbf{x}\in T_\mathbf{x}\right)P\left(\mathbf{x}\in T_\mathbf{x}\right) = 0.0429\times 0.2438 = 0.0105$$

6. (a) (i) Definition of mutual information

(ii) Translation invariance, since $\hat{X}$ is conditionally constant.

(iii) Conditioning reduces entropy. Equality if the quantization error is independent of $\hat{X}$.

(iv) Gaussian bound for entropy with a given variance. Equality if $(X-\hat{X})$ is Gaussian.

(v) $\mathrm{Var}(X-\hat{X}) \le E\left((X-\hat{X})^2\right) \le D$ and $\log()$ is monotonic. Equality if $(X-\hat{X})$ is zero-mean and distortion is maximum allowed. [6B]

(b) (i) Since X and Z are independent and zero mean, we have $E(XZ)=0$ and so can ignore the cross terms. We have,

$$E\left((X-\hat{X})^2\right)=E\left((X-kX-Z)^2\right)=(1-k)^2\sigma^2+kD$$
$$=(D\sigma^{-2})^2\sigma^2+(1-D\sigma^{-2})D=D^2\sigma^{-2}+D-D^2\sigma^{-2}=D$$
[2C]

(ii) We have

$$\mathrm{Var}(\hat{X})=k^2\,\mathrm{Var}(X)+\mathrm{Var}(Z)=k^2\sigma^2+kD$$
$$=k(k\sigma^2+D)=k\left((1-D\sigma^{-2})\sigma^2+D\right)$$
$$=k(\sigma^2-D+D)=k\sigma^2=\sigma^2-D$$
[2C]

(iii) We have

$$I(X;\hat{X})=h(\hat{X})-h(\hat{X}\,|\,X)$$
$$\overset{(i)}{\le} \tfrac{1}{2}\log 2\pi e(\sigma^2-D)-h(\hat{X}-X\,|\,X)$$
$$\overset{(ii)}{=} \tfrac{1}{2}\log 2\pi e(\sigma^2-D)-h(Z)$$
$$=\tfrac{1}{2}\log 2\pi e(k\sigma^2)-\tfrac{1}{2}\log 2\pi e(kD)$$
$$=\tfrac{1}{2}\log 2\pi e\,\sigma^2 D^{-1}$$
$$\Rightarrow\quad R(D)\overset{(iii)}{\le} \tfrac{1}{2}\log 2\pi e\,\sigma^2 D^{-1}$$
[5A]

(i) because of the Gaussian bound and translation invariance. (ii) because Z is independent of X. (iii) because $R(D)$ is the minimum and so cannot exceed any specific example.

(c) We have $\sigma^2=1/12$ and $h(X)=\log(1)=0$. Also $(X-\hat{X})\in\left[-\tfrac{1}{4},+\tfrac{1}{4}\right]$ and is uniformly distributed so $D=E\left((X-\hat{X})^2\right)=1/48$. [2C]

The lower bound from part (a) is $0-\tfrac{1}{2}\log 2\pi e/48=-\tfrac{1}{2}\log 0.3558=0.7454$ bits.

The upper bound from part (b) is $\tfrac{1}{2}\log(48/12)=2$ bits. [2C]

The actual bit rate is necessarily above the lower bound. It also happens to be below the upper bound although this need not necessarily be true for a block length of only 1. [1C]