Examination: Information Theory

MODEL ANSWER and MARKING SCHEME 2003

First Examiner: Professor L F Turner

Paper Code: E4.40 / S020

Second Marker: Dr J A Barria

Question

1

The transition matrix is as follows

$$\begin{array}{cc} & \text{Following Digit} \\ & \begin{array}{cc} 0 & 1 \end{array} \\ \text{Preceding Digit} \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} \cdot 9 & \cdot 1 \\ \cdot 7 & \cdot 3 \end{bmatrix} \end{array}$$

Given that the source starts in the 0 state the following sequence of events occurs leading to the state probabilities

1st Digit Generated $\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \cdot 9 & \cdot 1 \\ \cdot 7 & \cdot 3 \end{bmatrix} \Rightarrow \begin{bmatrix} \cdot 9 & \cdot 1 \end{bmatrix}$ state after 1st

2nd Digit Generated $\begin{bmatrix} \cdot 9 & \cdot 1 \end{bmatrix} \begin{bmatrix} \cdot 9 & \cdot 1 \\ \cdot 7 & \cdot 3 \end{bmatrix} \Rightarrow \begin{bmatrix} \cdot 88 & \cdot 12 \end{bmatrix}$ state after 2nd

3rd Digit Generated $\begin{bmatrix} \cdot 88 & \cdot 12 \end{bmatrix} \begin{bmatrix} \cdot 9 & \cdot 1 \\ \cdot 7 & \cdot 3 \end{bmatrix} \Rightarrow \begin{bmatrix} \cdot 876 & \cdot 124 \end{bmatrix}$ state after 3rd.

4th Digit Generated $\begin{bmatrix} \cdot 876 & \cdot 124 \end{bmatrix} \begin{bmatrix} \cdot 9 & \cdot 1 \\ \cdot 7 & \cdot 3 \end{bmatrix} \Rightarrow \begin{bmatrix} \cdot 8752 & \cdot 1248 \end{bmatrix}$ state after 4th

5th Digit Generated $\begin{bmatrix} \cdot 8752 & \cdot 1248 \end{bmatrix} \begin{bmatrix} \cdot 9 & \cdot 1 \\ \cdot 7 & \cdot 3 \end{bmatrix} \Rightarrow \begin{bmatrix} \cdot 87504 & \cdot 12496 \end{bmatrix}$

This agrees with the steady-state probabilities which from markov theory are $\frac{\cdot 7}{\cdot 8} = \cdot 850$ and $\frac{\cdot 1}{\cdot 8} = \cdot 125$

Now the entropy of the source when in state S_K is given by

$$H(X/S_K) = -\sum_i P(x_i/S_K)\log_2 P(x_i/S_K)$$

and the average entropy

$$= \sum_K P(S_K)\,H(X/S_K) = -\sum P(S_K)\sum P(x_i/S_K)\log_2 P(x_i/S_K)$$

We have two states $S_0 (=0)$; $S_1 (=1)$ and the conditional entropies are

$$-P(x_0/S_0)\log_2 P(x_0/S_0) - P(x_1/S_0)\log_2 P(x_1/S_0) \quad (1)$$

and $$-P(x_0/S_1)\log_2 P(x_0/S_1) - P(x_1/S_1)\log_2 P(x_1/S_1) \quad (2)$$

These are determined solely by the transition probabilities, so we have

$$H(X/S_0) = -(\cdot 9\log_2 \cdot 9 + \cdot 1\log_2 \cdot 1) = 0\cdot 4684$$

$$H(X/S_1) = -(\cdot 7\log_2 \cdot 7 + \cdot 3\log_2 \cdot 3) = 0\cdot 8812.$$

The entropy as the source generates its symbols are thus the above, weighted by the appropriate $P(S_0)$ and $P(S_1)$ as calculated on previous page.

We thus get $H_1 = 1 \times \cdot 4684 + 0 \times \cdot 8812 = 0\cdot 4684$ bits/symbol

$H_2 = \cdot 88 \times \cdot 4684 + \cdot 12 \times \cdot 8812 = 0\cdot 5179$ ..

$H_3 = \cdot 876 \times \cdot 4684 + \cdot 124 \times \cdot 8812 = 0 \cdot 5196$

$H_4 = \cdot 8752 \times \cdot 4684 + \cdot 1248 \times \cdot 8812 = 0 \cdot 52$

$H_5 = \cdot 875 \times \cdot 4684 + \cdot 125 \times \cdot 8812 = 0 \cdot 52$

// $H_5 = \cdot 52$ bits/binary digit is the steady-state // entropy.

16 marks

If source is considered as memoryless then the probability of generating 0 and 1 will be $P(0) = \cdot 875$ and $P(1) = \cdot 125$

Hence the zero-order entropy is

$$H(X)_{\text{memoryless}} = -\cdot 875 \log_2 \cdot 875 - \cdot 125 \log_2 \cdot 125$$

$$= 0 \cdot 545 \text{ bits/binary digit}$$

3 marks

If we encode using a Shannon/Fano code and we encode pairs of digits we get

	Prob	Code
X_0 (00)	49/64	1
X_1 (01)	7/64	01
X_2 (10)	7/64	001
X_3 (11)	1/64	000

3 marks

So average codeword length is $\frac{49}{64} \times 1 + \frac{14}{64} + \frac{21}{64} + \frac{3}{64} = \frac{87}{64}$

Hence average/digit $= 0 \cdot 68$ binary digits/digit

Conclusions :

(1) Memoryless entropy close to actual entropy

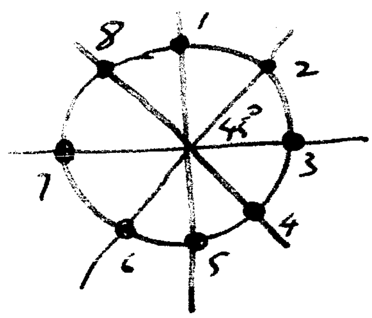
(2) If encode with s/c then need to encode more than 2 digits at a time to approach memoryless entropy

(3) If encode long enough block then will approach true entropy

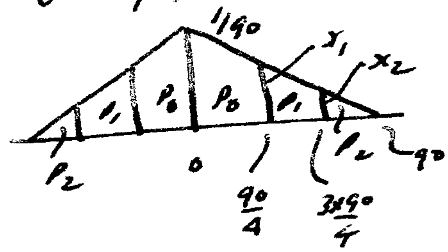
(3 marks)

Total mark (25)

2



The jitter pdf is



The equation is $P(\theta) = -\frac{1}{8100}\theta + \frac{1}{90}$ for pdf to right of zero.

So we have $x_1 = \frac{3}{4\times 90}$; $x_2 = \frac{1}{4\times 90}$

Hence $P_0 = 7/32$, $P_1 = 8/32$; $P_2 = 1/32$

Now $2P_0$ is the probability that jitter will not be sufficient to cause symbol error, P_1 is the probability that symbol will be decoded as an adjacent symbol and P_2 is probability that symbol will be decoded as a symbol two symbols away so we have a channel transition matrix as follows

$$
\begin{array}{c|cccccccc}
 & y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 \\
\hline
x_1 & 2P_0 & P_1 & P_2 & 0 & 0 & 0 & P_2 & P_1 \\
x_2 & P_1 & 2P_0 & P_1 & P_2 & 0 & 0 & 0 & P_2 \\
x_3 & P_2 & P_1 & 2P_0 & P_1 & P_2 & 0 & 0 & 0 \\
x_4 & 0 & P_2 & P_1 & 2P_0 & P_1 & P_2 & 0 & 0 \\
x_5 & 0 & 0 & P_2 & P_1 & 2P_0 & P_1 & P_2 & 0 \\
x_6 & 0 & 0 & 0 & P_2 & P_1 & 2P_0 & P_1 & P_2 \\
x_7 & P_2 & 0 & 0 & 0 & P_2 & P_1 & 2P_0 & P_1 \\
x_8 & P_1 & P_2 & 0 & 0 & 0 & P_2 & P_1 & 2P_0
\end{array}
$$

So; channel is doubly symmetric and hence capacity is

$$C = \log_2 K + \sum_{i=1}^{K} q_i \log_2 q_i ,$$

where K is number of output symbols and q_i's are elements of row of matrix.

Hence we obtain

$$C = \log_2 8 + 2P_0 \log_2 2P_0 + 2P_1 \log_2 P_1 + 2P_2 \log_2 P_2$$

bits / transmitted symbol

On substituting $P_0 = 7/32$, $P_1 = 4/32$ and $P_2 = 1/32$ we arrive at

$$C = 3 - .522 - 1 - .3125$$

$$= 1.165 \text{ bits / transmitted symbol}$$

20 marks

To achieve channel capacity in this case the input symbols $x_1, \ldots, x_8$ have to be used equally often AND we have to encode three data digits (binary digits) at a time

So we first have to encode the source into binary digits, and then use three of these at time

Similar

	P.A	
x_1	$(1/2)$	1
x_2	$(1/4)$	01
x_3	$(1/8)$	001
x_4	$(1/8)$	001

So we get $P(0) = P(1) = 1/2$

and we then simply take three binary digits at a time to get the x_i's for transmission. They will have equal probability

Total mark = 25

3

Part 1 $I = H(X) - H(X/Y)$

This can be interpreted as the difference between the a priori and a posteriori uncertainties. We are $H(X)$ uncertain about the input to the channel before receiving Y, and are $H(X/Y)$ uncertain about the input after having received Y.
$H(X/Y)$ could be interpreted as the amount of information still needed in order to be certain as to the input

$I = H(X) - H(X/YZ)$

$= H(X) - H(X/Y) + \{H(X/Y) - H(X/YZ)\}$ (*)

This can be related to the situation in which we send X and receive Y followed by Z.
The sequence pair in (*) can be interpreted in terms of successive reductions in our uncertainty about the input
i.e. $H(X) - H(X/Y)$ is the reduction in uncertainty about X once we have received Y. We remain $H(X/Y)$ uncertain about X

Limitations. Although we can argue in terms of the reduction in uncertainty, and we can argue through the capacity theorem that $I = H(X) - H(X/Y)$ (or its maximum) can be shown to be the rate at which information can be communicated error free, we cannot argue anything about certainty of information reception through an argument relating to the reduction in uncertainty. To achieve error-free communication we have to employ channel coding.

(8 marks)

Part 2 Consider $I = H(X) - H(X/Y)$,

where $H(X) = -\sum_i P(x_i) \log P(x_i)$

$$H(X/Y) = -\sum_i \sum_j P(x_i y_j) \log P(x_i/y_j)$$

$$\therefore I = \sum_i \sum_j P(x_i y_j) \log P(x_i/y_j)/P(x_i)$$

$$= \sum \sum P(x_i y_j) \log \frac{P(x_i y_j)}{P(y_j) \cdot P(x_i)}$$

$$= \sum \sum P(x_i/y_j) P(y_j) \log P(y_j/x_i)/P(y_j)$$

$$= \sum \sum P(y_j) P(x_i/y_j) \log 1/P(y_j) - \sum \sum P(y_j) P(x_i/y_j) \cdot \log 1/P(x_i/y_j)$$

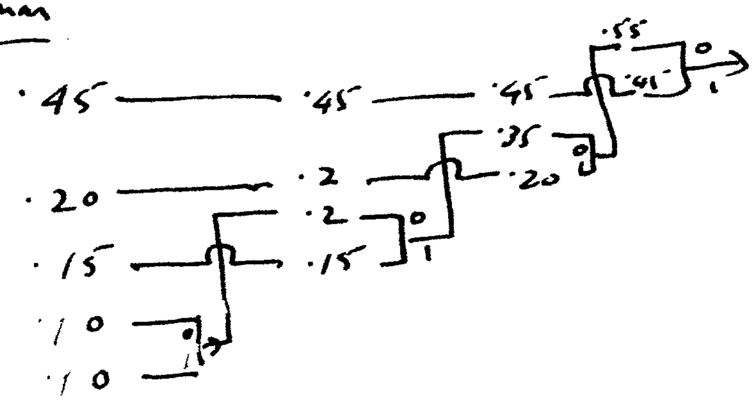
$= H(Y) - H(Y/X)$

5 marks

Part 3 Consider by way of example
(will accept any sensible example)

	Prob	S/F
x_1	·45	0
x_2	·20	1 0
x_3	·15	1 1 0
x_4	·10	1 1 1 0
x_5	·10	1 1 1 1

Huffman



Which yields the Huffman code

$x_1 \rightarrow 1$
$x_2 \rightarrow 01$
$x_3 \rightarrow 001$
$x_4 \rightarrow 0000$
$x_5 \rightarrow 0001$

5 marks

Both are non-prefix codes (instantaneous). The codewords in this case are of the same length and hence the codes are of the same efficiency. The codes have different code words but that does not matter.

The Huffman technique is the optimum, although more often than not there is no difference in performance between the S/F and H technique.

The most general and easiest way of implementing is to use ROM look-ups for both encoding and decoding.

With this technique the inputs x_i are used to address the ROM and the codewords are stored in the ROM locations. At the receiver the code words are used as addresses and the symbols x_i are located in the corresponding address locations.

Part 4

Let the source symbols be $x_1, \ldots, x_K$ and let the associated probabilities be $P_1, \ldots, P_K$ in which $P_1 \geq P_2 > P_3 \ldots \geq P_{K-1} \geq P_K$

Let the corresponding codewords be of length $n_1, \ldots, n_K$ respectively.

Now suppose some $n_i \neq n_K$ is longer than n_K.

It is immediately clear that by interchanging n_i and n_K

(1) no change in code structure occurs so it remains instantaneous and hence uniquely decodable

(2) the average codeword length is $\sum_i P_i n_i$ and that this will be reduced by the change, so the code becomes more efficient i.e. closer to optimum.

Thus the code can always be re-arranged so that the longer codewords are associated with the smaller source symbol probabilities and hence after re-arrangement the two least-probable code words will be the longest.

Now consider n_{K-1} and n_K (the re-arranged longest) (7 marks)

If n_K is longer than n_{K-1} then on account of the prefix property of the instantaneous code the extra digit in n_K can be dropped.

Total 25 marks

4

Part 1 The statement in itself does not constitute a restriction since we can send as many data bits/pulse as we like by increasing the size of the symbol set. But if we are power constrained then the symbols get closer together and hence if noise is present then the error probability increase

(3 marks)

Part 2

$$C = \max_{P(x)} I = \max_{P(x)} \left[H(X) - H(X/Y) \right] \text{ or } \max_{P(x)} \left[H(Y) - H(Y/X) \right]$$

I will accept a solution in which I is expressed as a function of $P(x_1)$, $P(x_2)$ — the input probabilities — and the function I is maximized by obtaining

$$\frac{\partial I}{\partial P(x_1)} = 0 \, ; \quad \frac{\partial I}{\partial P(x_2)} = 0$$ with the

solution being obtained for $P(x_1)$, $P(x_2)$ which satisfies the above subject to the constraint $P(x_1) + P(x_2) = 1$

Will also accept the following alternative.

$I = I(Y) - H(Y/X)$

with channel matrix $\begin{array}{c|cc} & y_1 & y_2 \\ \hline x_1 & 1-p & p \\ x_2 & p & 1-p \end{array}$

Now $H(Y/X) = -\sum_{j=1}^{2} P(x_j)\left[p\log p + (1-p)\log(1-p)\right]$

$= -(p\log p + (1-p)\log(1-p))$

and $H(Y) = -\{P(y_1)\log P(y_1) + P(y_2)\log P(y_2)\}$

But $P(y_1) = P(x_1)P(y_1/x_1) + P(x_2)P(y_1/x_2)$

$= P(x_1)(1-p) + P(x_2)\cdot p$

$P(y_2) = P(x_1)\cdot p + (1-p)P(x_2)$

If $P(x_1) = P(x_2)$ then $P(y_1) = P(y_2)$

hence $H(Y) = 1$ and

$I = 1 + p\log p + (1-p)\log(1-p)$

13 marks

13

Part 3. In order to transmit at capacity it is necessary that the input binary digits to the channel be used with equal probability

Now the source is

		Prob
x_1	$\to 0$	$1/3$
x_2	$\to 1$	$1/3$
x_3	$\to$	$1/3$

Consider the following coding in which the x_i's are mapped into binary digits

Case I

x_1	00
x_2	10
x_3	11

$\therefore P(0) = P(1) = 1/2$

$\therefore$ works perfect

But Entropy of source $= \log_2 3 = 1.59$ bits/symbol

We are using 2 bits/symbol $\therefore$ wasteful

Case II

Using S/F code

x_1	0
x_2	10
x_3	11

$P(0) = 0.4$

$P(1) = 0.6$

$\therefore$ Poor match to channel – we require $P(0) = P(1) = \frac{1}{2}$

Case III S/F applied to extended symbol set

Prob		SF code
1/9	$x_1 x_1$ —	000
"	$x_1 x_2$ —	001
"	$x_1 x_3$ —	010
"	$x_2 x_1$ —	011
"	$x_2 x_2$ —	100
"	$x_2 x_3$ —	101
"	$x_3 x_1$ —	110
	$x_3 x_2$ —	1110
1/9	$x_3 x_3$ —	1111

$\therefore P(0) = 0.45$
$P(1) = 0.55$ } getting closer to 1/2

Average codeword length $= \frac{29}{9}$ / pair

$= \frac{29}{18} = 1.61$ / symbol — close to entropy

(9 marks)

If we repeat for coding of longer blocks of x_i's then $P(0)$ and $P(1)$ approach 1/2 and $\bar{n} \to H(x)$

The final need is for proper channel coder that is inserted between source coder output and channel to achieve reliable communication. Care needed to ensure that channel coder gives $P(0) = P(1) \doteq 1/2$ at its output.

(Total mark = 25)

5

Part 1 The sequence $S_i = x_{i_1}, x_{i_2}, \ldots, x_{i_N}$, where each x_{i_j} is chosen from the set of symbols $x_1, x_2, \ldots x_n$.

Thus
$$\tilde{H}(S_i) = -\sum_{i=1}^{n^N} P(S_i) \log P(S_i)$$

$$= -\sum_{i_1=1}^{n} \cdots \sum_{i_N=1}^{n} P(x_{i_1}, \ldots, x_{i_N}) \log P(x_{i_1}, \ldots, x_{i_N}) \quad (1)$$

But
$$P(x_{i_1}, \ldots, x_{i_N}) = P(x_{i_1}, \ldots, x_{i_M}) \cdot P\left(\frac{x_{i_{M+1}}}{x_{i_1}, \ldots, x_{i_M}}\right)$$

Hence on substituting into (1) we obtain

$$\tilde{H}(S_i) = -\sum_{i_1=1}^{n} \cdots \sum_{i_N=1}^{n} P(x_{i_1}, \ldots, x_{i_N}) \left[\log P(x_{i_1}, \ldots, x_{i_M}) + \log P\left(\frac{x_{i_{M+1}}}{x_{i_1}, \ldots, x_{i_M}}\right) + \cdots + \log P\left(\frac{x_{i_N}}{x_{i_{N-M}}, \ldots, x_{i_{N-1}}}\right)\right]$$

But
$$-\sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} P(x_{i_1}, \ldots x_{i_N}) \log P\left(\frac{x_{i_j}}{x_{i_{j-M}}, \ldots, x_{i_{j-1}}}\right) = H(X)$$

$$j = M+1, \ldots, N$$

where $H(X)$ is the true entropy of the memory source.

Thus we have

$$\tilde{H}(\sigma) = (N-M)H(x) - \sum_{i_1=1}^{n} \cdots \sum_{i_M=1}^{n} P(x_{i_1}, \ldots, x_{i_M}) \log P(x_{i_1} \ldots x_{i_M})$$

$$= (N-M)\, H(x) - \sum_{i_1=1}^{n} \cdots \sum_{i_M=1}^{n} P(x_{i_1}, \ldots, x_{i_M}) \log P(x_{i_1} \ldots x_{i_M})$$

But since $-\sum p_i \log p_i \geq 0$ we have the result $\tilde{H} = (N-M)\, H(x) + \delta$ (2)

$\delta \geq 0$

and for $N > M$; δ is independent of N.

(15 marks)

Part 2 Proof of Shannon's noiseless coding theorem for memory source

Take a block S_i of source symbols, which have probability $P(S_i)$

We know that can select code word of length L_i which satisfies the Kraft/McMillan equality provided

$$\log 1/P(S_i) \leq L_i < \log 1/P(S_i) + 1$$ (3)

Hence we have

$$\sum_{i=1}^{n^N} P(S_i)\log\frac{1}{P(S_i)} \leq \sum_{i=1}^{n^N} L_i P(S_i) < \sum_{i=1}^{n^N} P(S_i)\log 1/P(S_i) + \sum_{i=1}^{n^N} P(S_i)$$

Now from (1) this can be written

$$(N-M)H(X)+\delta \leq N_{ave} < (N-M)H(X)+\delta+1$$

on dividing by N, the block length, we get

$$H(X) \leq \frac{N_{ave}}{N} = n_{ave} < H(X)+\frac{1}{N}\rightarrow 0$$

10 marks

as $N\rightarrow\infty$.

$\therefore\ n_{ave} = H(X)$ = entropy of source

Implementation is generally impossible in practice since code book size grows exponentially with N.

e.g. 8-bit pcm $(x_1,\dots,x_{256})$

So if attempt to encode 3 samples at a time codebook $=(256)^3$!!

Total mark = 25

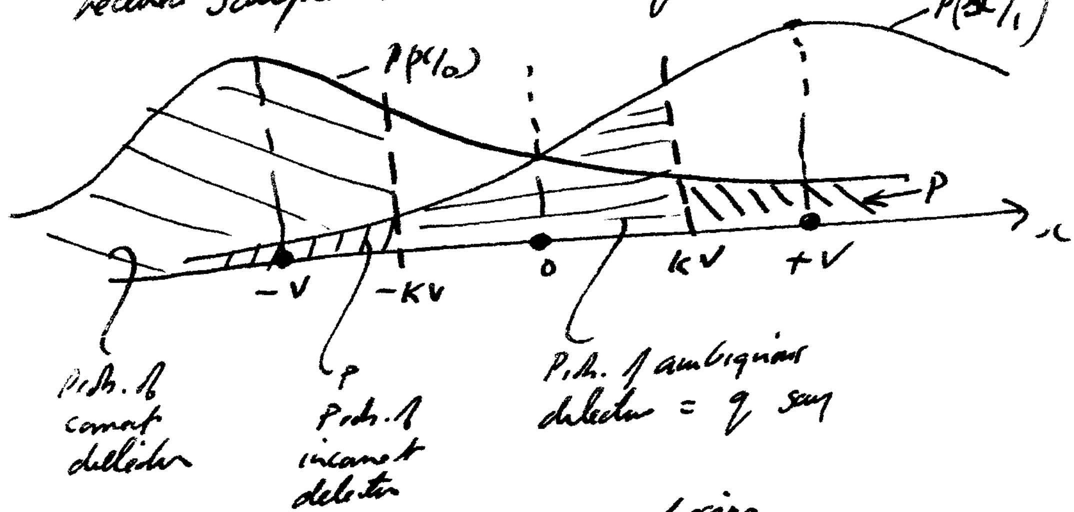
Examination: Information Theory

MODEL ANSWER and MARKING SCHEME

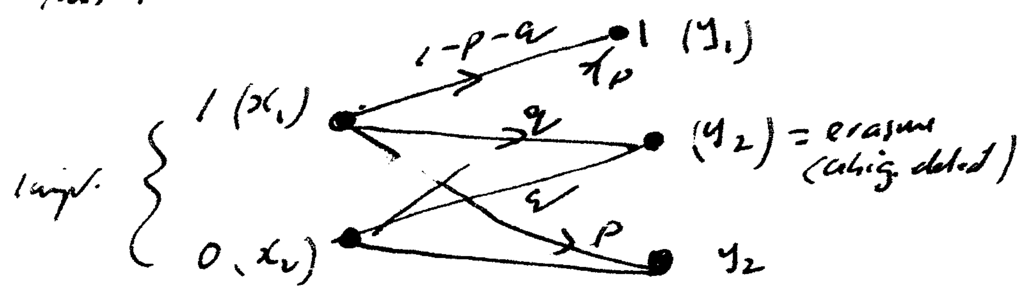
First Examiner: Professor L F Turner | Paper Code: E4.40 / S020

Second Marker: Dr J A Barria | Question |

Assume that cable attenuation is neglected. In this case the conditional pdf's relating to the received sample values are as follows:



Thus the channel transition ~~matrix~~ diagram is



and the channel transition matrix is

$$
\begin{array}{c} \\ x_1 \\ x_2 \end{array}
\begin{array}{c}
\begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\
\begin{bmatrix} 1-q-p & p & q \\ p & (1-p-q) & q \end{bmatrix}
\end{array}
$$

Now the channel matrix is uniform from the input, but not doubly uniform

$$\therefore I = H(Y) + (1-q-p)\log(1-q-p) + q\log q + p\log p$$

expect proof of this

Although channel not uniform from output it is easy to show in this case that $H(Y)$ is maximized when $P(x_1) = P(x_2) = 1/2$

expect this to be shown

With these probabilities

$$P(y_1) = P(x_1)(1-p-q) + P(x_2)\ldots = \frac{1-q}{2}$$

$$P(y_2) = P(x_1)q + P(x_2)q = q$$

$$P(y_3) = P(x_1)p + P(x_2)(1-p-q) = \frac{1-q}{2}$$

15 marks

Hence the capacity is

$$C = (1-q)\left[1-\log_2(1-q)\right] + (1-p-q)\log_2(1-p-q) + p\log_2 p \quad \text{bits/input}$$

— (1)

Examination: Information Theory

MODEL ANSWER and MARKING SCHEME

First Examiner: Professor L F Turner | Paper Code: E4.40 / S020

Second Marker: Dr J A Barria | Question | Page 23 of 23

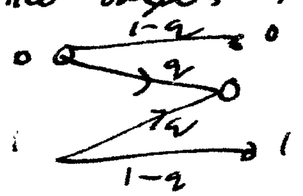
If $q = 0$ then C reduces to

$1 + p\log p + (1-p)\log(1-p)$ which is the capacity of the B.S.C.

The capacity in (1) is changed from that of the BSC since we are saying if signals are somewhat uncertain we then indicate this – i.e. we give outputs that are more certain than with BSC. (but most uncertain signals are indicated)

10 marks

If K becomes large, then $P \to 0$, i.e. we are attempting to rule out errors, at the expense of an increasing number of ambiguous outputs. The correct outputs are reduced in number, but only the most certain are given as 'positive' outputs. The transition diagram becomes



and the capacity as can be seen from (1) becomes $C = (1-q)$

Total mark = 25