The Solutions to Exam 2017

B—bookwork, E—new example, T—new theory

Students did well in Question 1

1.

a) We have the following outcomes, each with probability ¼:

$x_1\ x_2$	00	01	10	11
$y = \max(x_1, x_2)$	0	1	1	1

i) Thus $P(y=0) = ¼$, $P(y=1) = ¾$,

$H(y) = -¼\log(1/4) - ¾\log(¾) = ½ + 0.31 = 0.81$ [3E]

ii) We have the joint distribution

$x_1\ \ y$	0	1
0	¼	¼
1	0	½

$H(y|x_1) = ½\, H(½) + ½\, H(1) = ½$ [3E]

$I(x_1\,;\,y) = H(y) - H(y|\,x_1) = 0.81 - ½ = 0.31$

iii) $I(x_{1:2}\,;\,y) = H(y) - H(y|\,x_{1:2})$

$= 0.81 - 0$ $\quad y$ is a function of x_1 and x_2

$= 0.81$ [3E]

b) Recall

A small number of students didn't remember relative entropy

$$D(\mathbf{p}\,||\,\mathbf{q}) = \sum_i p_i \log_2\left(\frac{p_i}{q_i}\right) \geq 0$$

$D(\mathbf{p}||\mathbf{q}) = \frac{1}{2}\log\frac{3}{2} + \frac{1}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{3}{4} = \log 3 - 1.5 = 0.085.$ [3E]

$D(\mathbf{q}||\mathbf{p}) = \frac{1}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{4}{3} + \frac{1}{3}\log\frac{4}{3} = \frac{5}{3} - \log 3 = 0.082$ [3E]

c) Recall Fano's inequality $H(x|y) \leq P(x\neq y)\log M + H(P(x\neq y))$. [1B]

$I(x\,;\,y) = H(x) - H(x\,|\,y)$ definition [1B]

$\geq \log M - [P(x\neq y)\log M - H(P(x\neq y))]$ Fano [2E]

$\geq \log M - [(1 - P(x = y))\log M + H(P(x\neq y))]$ algebra [2E]

$\geq P(x = y)\log M - H(P(x\neq y))$ algebra [2E]

$= P(x = y)\log M - H(P(x = y))$ because $H(P(x\neq y)) = H(P(x = y))$ [2E]

2.

Part a is bookwork

a) [1B each]

(1) total probability of the jointly typical set

(2) $p(\mathbf{x},\mathbf{y}) \le \max_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p(\mathbf{x},\mathbf{y})$

(3) $\max_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p(\mathbf{x},\mathbf{y}) \le 2^{-n(H(X,Y)-\varepsilon)}$, from definition of the jointly typical set

(4) algebra

(5) total probability of the jointly typical set <= 1

(6) $p(\mathbf{x},\mathbf{y}) \ge \min_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p(\mathbf{x},\mathbf{y})$

(7) $\min_{\mathbf{x},\mathbf{y}\in J_\varepsilon^{(n)}} p(\mathbf{x},\mathbf{y}) = 2^{-n(H(X,Y)+\varepsilon)}$, from definition of the jointly typical set

(8) algebra

Part b is a bit hard: students should truly understand the concept of typical sequences.

b)

i) Note that the marginal distribution of X is

$$P(X=0)=4/7, \qquad P(X=1)=3/7. \qquad [2E]$$

Because $\varepsilon = 0$, $\mathbf{x} \in T_\mathbf{x}$ if and only if exactly 4 of the X_i's are equal to 0, and the other 3 of the X_i's are equal to 1. Hence

$$P(\mathbf{x} \in T_\mathbf{x}) = \binom{7}{4}\left(\frac{4}{7}\right)^4\left(\frac{3}{7}\right)^3 = 0.294\,. \qquad [3E]$$

ii) If, in addition, $(\mathbf{x},\mathbf{y}) \in J_0^{(7)}$, we require that y_i =0 for 3 out of 4 index i's for which $x_i = 0$, and y_i =0 for 1 out of 3 index i's for which $x_i = 1$.
Thus

$$P(\mathbf{x},\mathbf{y} \in J_0^{(7)} \mid \mathbf{x} \in T_\mathbf{x}) = \binom{4}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1 \times \binom{3}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2 = 0.188\,. \qquad [3E]$$

iii) Hence we determine the value

$$\begin{aligned}
P(\mathbf{x},\mathbf{y} \in J_0^{(7)}) &= P(\mathbf{x},\mathbf{y} \in J_0^{(7)} \mid \mathbf{x} \in T_\mathbf{x})P(\mathbf{x} \in T_\mathbf{x}) + P(\mathbf{x},\mathbf{y} \in J_0^{(7)} \mid \mathbf{x} \notin T_\mathbf{x})P(\mathbf{x} \notin T_\mathbf{x}) \\
&= P(\mathbf{x},\mathbf{y} \in J_0^{(7)} \mid \mathbf{x} \in T_\mathbf{x})P(\mathbf{x} \in T_\mathbf{x}) + 0 \\
&= 0.188 \times 0.294 \\
&= 0.055
\end{aligned}$$

[3E]

iv) We also require that z_i =0 for 3 out of 4 index i's for which $x_i = 0$, and z_i =0 for 1 out of 3 index i's for which $x_i = 1$. But, since $\mathbf{z}$ is independent of $\mathbf{x}$,

$$P(\mathbf{x},\mathbf{z} \in J_0^{(7)} \mid \mathbf{x} \in T_\mathbf{x}) = \binom{4}{3}\left(\frac{4}{7}\right)^3\left(\frac{3}{7}\right)^1 \times \binom{3}{1}\left(\frac{4}{7}\right)^1\left(\frac{3}{7}\right)^2 = 0.101 \qquad [3E]$$

Thus

Students didn't do well in iv)

$$
\begin{aligned}
P(\mathbf{x},\mathbf{z} \in J_0^{(7)}) &= P(\mathbf{x},\mathbf{y} \in J_0^{(7)} \mid \mathbf{x} \in T_{\mathbf{x}})P(\mathbf{x} \in T_{\mathbf{x}}) \\
&= 0.101 \times 0.294 \\
&= 0.030
\end{aligned}
$$

[3E]

3. Bookwork

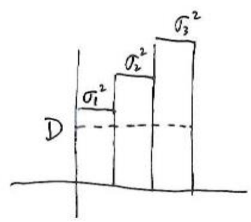
a)

(1) definition of mutual info [1B]
(2) chain rule [1B]
(3) x_i is a function of $y_{1:i-1}$ and w [1B]
(4) channel is memoryless [1B]
(5) indep. bound, or chain rule + conditioning reduces entropy [1B]
(6) definition of mutual info [1B]
(7) mutual info. <= capacity [1B]
(8) Fano's inequality [1B]
(9) algebra [1B]
(10) taking limit and $P_e^{(n)} \to 0$ [1B]

In general, students did well, but some got confused with water marking over parallel Gaussian channels, which lost marks. Note reverse water marking is source coding, not channel coding

b)

i) In this case, all 3 sources are encoded. [5E]



$0 < D < \sigma_1^2$

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

$$R(D) = \frac{1}{2}\log\frac{1}{0.5} + \frac{1}{2}\log\frac{2}{0.5} + \frac{1}{2}\log\frac{4}{0.5} = 3$$

ii) In this case, 2 sources are encoded, at distortion 1. [5E]

$$R(D) = \frac{1}{2}\log\frac{2}{1} + \frac{1}{2}\log\frac{4}{1} = \frac{3}{2}$$

iii) In this case, only 1source is encoded, at distortion 3 [5E]

$$R(D) = \frac{1}{2}\log\frac{4}{3} = 0.21$$

Note: D is average distortion, so distortion is 3 for the third source.

Again, solution is not unique. As long as students demonstrate good understanding of Markov chains, marks will be given.

4.

a)

i) Denote by N_1 the variance of Z_1, N'_2 the variance of Z'_2. We are supposed to verify

$$f(y_1, y_2|x) = f(y_1|x)f(y_2|y_1)$$

Definition of Markov chain [1E]

This is so because

$$f(y_1, y_2|x) = f(y_1, y_1 + z'_2|x)$$

From channel model [1E]

$$= f(y_1|x)f(y_1 + z'_2|x, y_1)$$

Chain rule [1E]

$$= f(y_1|x)f(y_1 + z'_2|y_1)$$

Z'_2 is independent of x [1E]

$$= f(y_1|x)f(y_2|y_1)$$

Obvious [1E]

ii) Part ii) is bookwork

Encoding: The sender uses the first codebook with power αP at rate R_1, and the second codebook with power $(1-\alpha)P$ at rate R_2, sends the sum of two codewords. Both codebooks are i.i.d. Gaussian. [3B]

Decoding: Bad receiver Y_2 treats Y_1 as noise, yielding a rate [2B]

$$R_2 \le C\left(\frac{(1-\alpha)P}{\alpha P + N_2}\right)$$

Good receiver Y_1 first decodes the second message X_2. It is able to do so because its channel is better: [2B]

$$R_2 \le C\left(\frac{(1-\alpha)P}{\alpha P + N_2}\right) \le C\left(\frac{(1-\alpha)P}{\alpha P + N_1}\right)$$

Then it subtracts out X_2, and decodes his own message. Since the channel is clean now, the rate is given by [3B]

$$R_1 \le C\left(\frac{\alpha P}{N_1}\right)$$

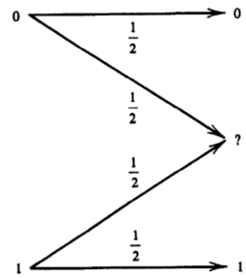
b) Part b is hard; few students got the correct answer

Setting $X_2 = 0$, we can send at a rate of 1 bit per transmission from sender 1. Similarly, setting $X_1 = 0$, we can send at a rate $R_2 = 1$. This gives us two extreme points of the capacity region. [2T]

Can we do better? Let us assume that $R_1 = 1$, so that the codewords of X_1 must include all possible binary sequences; X_1 would look like a Bernoulli(1/2) process. This acts like

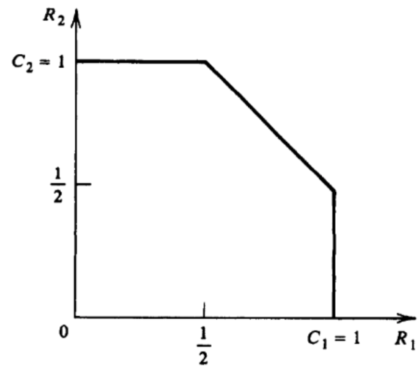
The key point is to realize this binary erasure channel

noise for the transmission from X_2. For X_2, the channel looks like a binary erasure channel in the following figure, whose capacity is 1/2 bit per transmission. [3T]



Hence when sending at maximum rate 1 for sender 1, we can send an additional 1/2 bit from sender 2, and vice versa. We can verify that these rates are the best that can be achieved. [3T]

The capacity region for this multi-access channel is illustrated as follows.



[2T]