

# IT 2014 Solutions

A — New Application

B — Bookwork

E — New Example

T — New Theory

1. a) Joint distribution

| x \ y | 0             | 1             |
|-------|---------------|---------------|
| 0     | 0             | $\frac{1}{3}$ |
| 1     | $\frac{1}{3}$ | $\frac{1}{3}$ |

$$p(X=0) = \frac{1}{3} \quad p(X=1) = \frac{2}{3}$$

$$p(Y=0) = \frac{1}{3} \quad p(Y=1) = \frac{2}{3}$$

$$H(X) = H(Y) = -\frac{1}{3} \log \frac{1}{3} - \frac{2}{3} \log \frac{2}{3} = 0.918 \quad [2]E$$

$$H(X|Y) = H(Y|X) = \frac{1}{3} H(0) + \frac{2}{3} H\left(\frac{1}{2}\right) = \frac{2}{3} = 0.667 \quad [2]E$$

average row entropy,  $H(p)$ : entropy function

$$H(X, Y) = H(X) + H(Y|X) = 0.918 + \frac{2}{3} = 1.58 \quad [1]E$$

$$I(X; Y) = H(X) - H(X|Y) = 0.918 - \frac{2}{3} = 0.251 \quad [1]E$$

b)  $I(X_1; Y_1) = I(X_1; X_2) = 0$  Since  $X_1$  and  $X_2$  are iid. [1]E

$$I(X_2; Y_2) = 0 \quad [1]E$$

$$I(X_{1:2}; Y_{1:2}) = I(X_{1:2}; X_{2:1}) = H(X_{1:2}) = 2 \quad [3]E$$

$$\begin{aligned} I(X_1, X_2 | Y_3) &= \frac{1}{2} I(X_1; X_2 | Y_3=0) + \frac{1}{2} I(X_1; X_2 | Y_3=1) \\ &= \frac{1}{2} I(X_1; X_1 | Y_3=0) + \frac{1}{2} I(X_1; X_1 \oplus 1 | Y_3=1) \quad [3]E \\ &= \frac{1}{2} H(X_1) + \frac{1}{2} H(X_1) \\ &= H(X_1) = 1 \end{aligned}$$

c) i) Let  $(\pi_0 \pi_1)$  be the stationary distribution

Then,

$$(\pi_0 \pi_1) \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} = (\pi_0 \pi_1)$$

We have

$$\pi_0(1-p) + \pi_1 q = \pi_0 \quad [2]A$$

$$\pi_0 p = \pi_1 q$$

$$\pi_0 = \frac{q}{p} \pi_1$$

Since  $\pi_0 + \pi_1 = 1$ , we have

$$\left(1 + \frac{q}{p}\right) \pi_1 = 1 \quad [2]A$$

Thus,

$$\pi_1 = \frac{p}{p+q} \quad \pi_0 = \frac{q}{p+q}$$

$$ii) H(X) = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}) \quad [2]A$$

$$= \pi_0 H(p) + \pi_1 H(q) \quad \text{average row entropy}$$

$$= \frac{q}{p+q} H(p) + \frac{p}{p+q} H(q) \quad (*) \quad [2]A$$

iii) We know  $H(X_n | X_{n-1}) \leq H(X_n) \leq 1$ , so

$$H(X) \leq 1$$

[3]A

Examining Eq. (\*), we find  $H(X) = 1$  if

$$p = q = 0.5.$$

2. a) (1) chain rule [1B]  
i) (2) chain rule in another way [1B]

(3)  $H(e|x, y) \geq 0$  entropy is non-negative [1B]

$H(e|y) \leq H(e)$  conditioning reduces entropy [1B]

(4) total probability theorem

(5)  $H(e) = H(p_e)$  [2B]

Given  $y$  and  $e=0$ ,  $X=y$ , so entropy = 0.

Given  $y$  and  $e=1$ ,  $X \neq y$  but can take any of the  $|X|-1$  values, so entropy  $\leq \log(|X|-1)$

(6) algebra [1B]

(7)  $H(p_e) \leq 1$  [1B]

ii) The optimum detection is given by

$$\hat{x} = \begin{cases} 1 & y=a \\ 2 & y=b \\ 3 & y=c \end{cases} \quad [2E]$$

Thus,  $P_e$  is equal to the sum of the off-diagonal elements, i.e.,

$$P_e = \frac{1}{2} \quad [1E]$$

To evaluate Fano's inequality, we need

$$\begin{aligned} H(X|Y) &= \frac{1}{3} H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) + \frac{1}{3} H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) + \frac{1}{3} H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) \\ &= H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) \\ &= \frac{1}{2} \log 2 + \frac{1}{4} \log 4 + \frac{1}{4} \log 4 \\ &= 1.5 \end{aligned} \quad [2E]$$

Hence,

$$\begin{aligned} P_e &\geq \frac{H(X|Y) - 1}{\log(|X| - 1)} \\ &= \frac{1.5 - 1}{\log 2} \\ &= 0.5 \end{aligned} \quad [2E]$$

which is exactly the same as the actual  $P_e$ .

b) By definition

$$R(D) = \min I(X, \hat{X})$$

Such that  $E[(X - \hat{X})^2] = D$  and  $p(x) = \int p(x, \hat{x}) d\hat{x}$ .

The second condition is obviously true from Fig 2.2.

[2T]

Now check the first condition:

$$\begin{aligned} E[(X - \hat{X})^2] &= E\left[\left(\frac{D}{\sigma^2} X - \frac{\sigma^2 - D}{\sigma^2} Z\right)^2\right] \\ &= \frac{D^2}{\sigma^4} \sigma^2 + \frac{(\sigma^2 - D)^2}{\sigma^4} \frac{D \sigma^2}{\sigma^2 - D^2} \\ &= \frac{D^2}{\sigma^2} + \frac{(\sigma^2 - D) D}{\sigma^2} \\ &= D. \end{aligned} \quad [2T]$$

The mutual information

$$\begin{aligned} I(X, \hat{X}) &= h(\hat{X}) - h(\hat{X}|X) \\ &= h(\hat{X}) - h\left(\frac{\sigma^2 - D}{\sigma^2} Z\right) \end{aligned} \quad [2T]$$

Note that  $\hat{X}$  has zero mean and variance

$$\begin{aligned} E[\hat{X}^2] &= \left(\frac{\sigma^2 - D}{\sigma^2}\right)^2 \left(\sigma^2 + \frac{\sigma^2 \cdot D}{\sigma^2 - D}\right) \\ &= \sigma^2 - D \end{aligned} \quad [2T]$$

$$\text{So } h(\hat{X}) \leq \frac{1}{2} \log [2\pi e (\sigma^2 - D)]$$

Gaussian has maximum entropy

Hence,

$$\begin{aligned} I(X; \hat{X}) &\leq \frac{1}{2} \log [2\pi e(\sigma^2 - D)] - h(Z) - \log \frac{\sigma^2 - D}{\sigma^2} \\ &= \frac{1}{2} \log [2\pi e(\sigma^2 - D)] - \frac{1}{2} \log \left[ 2\pi e \frac{D\sigma^2}{\sigma^2 - D} \right] - \frac{1}{2} \log \left( \frac{\sigma^2 - D}{\sigma^2} \right)^2 \\ &= \frac{1}{2} \log \frac{\sigma^2}{D} \quad [2T] \end{aligned}$$

Since  $I(X; \hat{X}) \leq \frac{1}{2} \log \frac{\sigma^2}{D}$  for this example, the minimum mutual information also  $\leq \frac{1}{2} \log \frac{\sigma^2}{D}$ . QED.

3

a) i) [1B] each

- (1) chain rule
- (2) conditioning reduces entropy
- (3) definition of mutual info
- (4)  $X_i$ 's are independent
- (5) from (2)
- (6) definition of mutual info
- (7) definition of mutual info
- (8) memoryless channel
- (9) chain rule
- (10) conditioning reduces entropy

ii) From (3)-(6), if the channel has memory, i.e.,  $Y_i$ 's are correlated for independent  $X_i$ 's, then

$$I(X_{1:n}, Y_{1:n}) \geq \sum I(X_i, Y_i) \quad [3A]$$

On the other hand, if the channel is memoryless, then

$$I(X_{1:n}, Y_{1:n}) \leq \sum I(X_i, Y_i) \quad [3A]$$

Therefore, a channel with memory has higher capacity.

b) i) weakly symmetric

$$C = \log 3 - H\left(\frac{1}{3}, \frac{1}{6}, \frac{1}{2}\right) \\ = 0.126$$

[3E]

ii) symmetric

$$C = \log 3 - H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) \\ = 0.085$$

[3E]

iii) symmetric

$$C = \log 4 - H\left(\frac{1}{2}, \frac{1}{2}\right) \\ = 1$$

[3E]

4. a) Under no interference, it is a Gaussian channel

$$C = \frac{1}{2} \log\left(1 + \frac{P}{N}\right) \quad (**) \quad [2B]$$

Under very strong interference,  $Y_1$  first decodes  $X_2$ , while treating  $X_1$  as noise. It then subtracts  $X_2$  and decodes  $X_1$ , which is a clean channel with the same capacity as (\*\*). [2B]

This is possible if the rate of  $X_2$  is less than [2A]

$$\frac{1}{2} \log\left(1 + \frac{a^2 P}{P+N}\right)$$

Thus, we have the same capacity if

$$\frac{1}{2} \log\left(1 + \frac{P}{N}\right) \leq \frac{1}{2} \log\left(1 + \frac{a^2 P}{P+N}\right) \quad [4A]$$

$$\frac{P}{N} \leq \frac{a^2 P}{P+N}$$

$$a^2 \leq \frac{P+N}{N} = 1 + \frac{P}{N}$$

b) From the joint distribution, we obtain the marginal distributions:

$$P_{U_1} = (\alpha + \beta, \frac{\gamma}{m-1}, \dots, \frac{\gamma}{m-1}) \quad [1E]$$

$$P_{U_2} = (\alpha + \gamma, \frac{\beta}{m-1}, \dots, \frac{\beta}{m-1}) \quad [1E]$$

Thus,

$$H(U_1) = -(\alpha + \beta) \log(\alpha + \beta) - (m-1) \frac{\gamma}{m-1} \log\left(\frac{\gamma}{m-1}\right)$$

$$= -(\alpha + \beta) \log(\alpha + \beta) - \gamma \log\left(\frac{\gamma}{m-1}\right) \quad [2E]$$

$$= H(\alpha + \beta, \gamma) + \gamma \log(m-1)$$



$$\text{Similarly, } H(U_2) = H(\alpha + r, \beta) + \beta \log(m-1) \quad [2E]$$

$$= -(\alpha + r) \log(\alpha + r) - \beta \log \frac{\beta}{m-1}$$

Also,

$$H(U_1, U_2) = -\alpha \log \alpha - (m-1) \frac{\beta}{m-1} \log \left( \frac{\beta}{m-1} \right) - (m-1) \frac{r}{m-1} \log \left( \frac{r}{m-1} \right) \quad [2E]$$

$$= -\alpha \log \alpha - \beta \log \frac{\beta}{m-1} - r \log \frac{r}{m-1}$$

$$= H(\alpha, \beta, r) + \beta \log(m-1) + r \log(m-1)$$

$$H(U_1|U_2) = H(U_1, U_2) - H(U_2) \quad [2E]$$

$$= H(\alpha, \beta, r) - H(\alpha + r, \beta) + r \log(m-1)$$

$$H(U_2|U_1) = H(U_1, U_2) - H(U_1) \quad [2E]$$

$$= H(\alpha, \beta, r) - H(\alpha + \beta, r) + \beta \log(m-1)$$

Hence, the slepian-wolf region is

$$R_1 \geq H(U_1|U_2) \quad [3P]$$

$$R_2 \geq H(U_2|U_1)$$

$$R_1 + R_2 \geq H(U_1, U_2)$$

[For this question, the expressions are not unique.]