The Solutions to Exam 2018

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

students did well in Q1

1.

a)

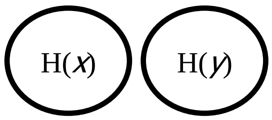
i) $H(X) = H(Y) = 1$ bit [2E]

ii) $H(X \mid Y) = H(Y \mid X) = 1$ bit [2E]

iii) $H(X, Y) = H(X) + H(Y \mid X) = 2$ bits [2E]

iv) $I(X; Y) = H(X) - H(X \mid Y) = 0$ [2E]

v) X and Y are independent. [2E]



b)

$I(X_1; Y_1) = 0$ [2B]

$I(X_2; Y_2) = 0$ [2B]

$I(X_{1:2}; Y_{1:2}) = 2H(X_1) = 2(-\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4}) = 1.62$ bits. [3B]

c) $D(\mathbf{p}||\mathbf{q}) = \sum p_i \log\frac{p_i}{q_i} = p\log\frac{p}{r} + (1-p)\log\frac{1-p}{1-r}$

$= \frac{1}{4}\log\frac{1}{2} + \frac{3}{4}\log\frac{3}{2} = 0.19$ [4E]

$D(\mathbf{q}||\mathbf{p}) = \sum q_i \log\frac{q_i}{p_i} = r\log\frac{r}{p} + (1-r)\log\frac{1-r}{1-p}$

$= \frac{1}{2}\log 2 + \frac{1}{2}\log\frac{2}{3} = 0.21$ [4E]

A small number of students didn't know how to calculate D(p||q).
A common mistake was to think D(p||q) = D(q||p)

2.

a)

i) (1) definition of $I(X;\hat{X})$. [1 each, B]

(2) Gaussian entropy; Shift doesn't change entropy.

(3) Conditioning reduces entropy.

(4) Given variance, Gaussian has max entropy.

(5) $Var(X-\hat{X}) \le D$.

(6) Algebra and $I(X;\hat{X}) \geq 0$

ii) (7) $\hat{X} + Z = X$

(8) $h(Z|\hat{X}) = h(Z) = \frac{1}{2}\log 2\pi e D$

(9) The lower bound is achievable $\Rightarrow$ $\geq$ becomes $=$.

(10) definition of rate-distortion.

b)

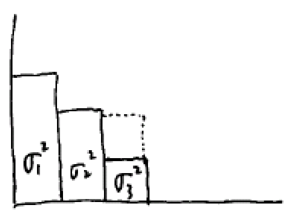
[3 A]

i) Single channel is when

$$3P \le \sigma_2^2 - \sigma_3^2$$

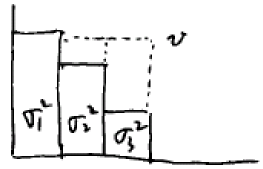
Capacity

$$C = \frac{1}{2}\log\left(1 + \frac{3P}{\sigma_3^2}\right)$$



In general students did well in Q2.
But some students got it confused with reverse waterfilling in source coding

ii) A pair of channel is when [3A]



$$\sigma_2^2 - \sigma_3^2 < 3P \le \sigma_1^2 - \sigma_2^2 + \sigma_1^2 - \sigma_3^2$$
$$= 2\sigma_1^2 - \sigma_2^2 - \sigma_3^2$$

$$3P = v - \sigma_2^2 + v - \sigma_3^2 \Rightarrow v = \frac{3P + \sigma_2^2 + \sigma_3^2}{2}$$

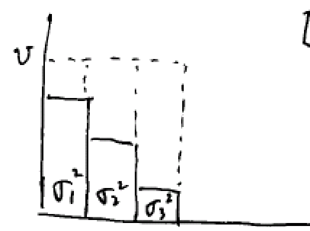
$$P_2 = v - \sigma_2^2 = \frac{3P - \sigma_2^2 + \sigma_3^2}{2}$$

$$P_3 = v - \sigma_3^2 = \frac{3P + \sigma_2^2 - \sigma_3^2}{2}$$

$$C = \frac{1}{2}\log\left(1 + \frac{P_2}{\sigma_2^2}\right) + \frac{1}{2}\log\left(1 + \frac{P_3}{\sigma_3^2}\right)$$
$$= \frac{1}{2}\log\left(1 + \frac{3P - \sigma_2^2 + \sigma_3^2}{2\sigma_2^2}\right) + \frac{1}{2}\log\left(1 + \frac{3P + \sigma_2^2 - \sigma_3^2}{2\sigma_3^2}\right)$$

iii) Three channels is when [4A]

$$3P > 2\sigma_1^2 - \sigma_2^2 - \sigma_3^2$$



$$3P = v - \sigma_1^2 + v - \sigma_2^2 + v - \sigma_3^2$$

$$\Rightarrow v = \frac{3P + \sigma_1^2 + \sigma_2^2 + \sigma_3^2}{3} = P + \frac{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}{3}$$

$$P_1 = v - \sigma_1^2 = P + \frac{\sigma_2^2 + \sigma_3^2 - 2\sigma_1^2}{3}$$

$$P_2 = v - \sigma_2^2 = P + \frac{\sigma_1^2 + \sigma_3^2 - 2\sigma_2^2}{3}$$

$$P_3 = v - \sigma_3^2 = P + \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_3^2}{3}$$

$$C = \frac{1}{2}\log\left(1 + \frac{P_1}{\sigma_1^2}\right) + \frac{1}{2}\log\left(1 + \frac{P_2}{\sigma_2^2}\right) + \frac{1}{2}\log\left(1 + \frac{P_3}{\sigma_3^2}\right)$$

A common mistake is to take it for granted that the second term equals h_phi(x). Please note that this isn't true in general, even if they have the same mean and K. You're required to give the steps of derivation.

c) Write the Gaussian pdf as

$$\varphi(\mathbf{x}) = \left|2\pi\mathbf{K}\right|^{-1/2} \exp\left(-\tfrac{1}{2}\mathbf{x}^T\mathbf{K}^{-1}\mathbf{x}\right)$$

Then

$$D(f\,\|\,\varphi) = -h_f(\mathbf{x}) - E_f \log\varphi(\mathbf{x}) \qquad [1A]$$

where

$$-E_f \log\varphi(\mathbf{x}) = -(\log e)\,E_f\left(-\tfrac{1}{2}\ln\left(\left|2\pi\mathbf{K}\right|\right) - \tfrac{1}{2}\mathbf{x}^T\mathbf{K}^{-1}\mathbf{x}\right)$$

$$= \tfrac{1}{2}(\log e)\left(\ln\left(\left|2\pi\mathbf{K}\right|\right) + \mathrm{tr}\left(E_f\,\mathbf{x}\mathbf{x}^T\mathbf{K}^{-1}\right)\right) \qquad [3A]$$

$$= \tfrac{1}{2}(\log e)\left(\ln\left(\left|2\pi\mathbf{K}\right|\right) + \mathrm{tr}(\mathbf{I})\right)$$

$$= \tfrac{1}{2}\log\left(\left|2\pi e\mathbf{K}\right|\right) = h_\varphi(\mathbf{x})$$

Finally

$$D(f\,\|\,\varphi) = h_\varphi(\mathbf{x}) - h_f(\mathbf{x}) \qquad [1A]$$

Although part c) looks tricky, it actually follows from the definition of KL divergence

3.

a) [1B each]

(1) definition of conditional entropy

(2) row entropies $H(y \mid x = x)$ are identical

(3) algebra

(4) definition of mutual information

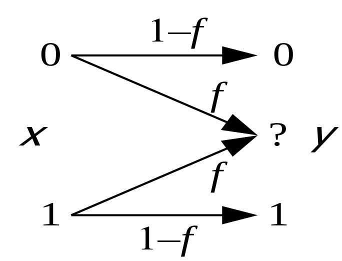
(5) from (3)

(6) uniform bound on entropy H(y)

(7) upper bound (6) is achievable with uniform input distribution

b) Average score of Part b) is low, probably because it's new theory

(i) The transition matrix of a BEC



can be rearranged into

$$\begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & ? \end{array} \\ \left(\begin{array}{cc|c} 1-f & 0 & f \\ 0 & 1-f & f \end{array}\right) \end{array}$$

Partition this matrix into two blocks, and note that both blocks are symmetric. So this is a generally symmetric channel. [3T]

However, BEC is not a weakly symmetric channel, because the column sums are not identical in general (unless f = 1/3). [1A]

(ii)

The transition matrix of the first channel is given by

$$\begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.7 \end{pmatrix}$$

which can be rearranged into [1T]

$$\left(\begin{array}{cc|c} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.7 & 0.2 \end{array}\right)$$

Again, with the above partition, both blocks are symmetric. So this channel is generally symmetric. We can still use the formula

$$I(x;y) = H(y) - H(y \mid x) = H(y) - H(\mathbf{Q}_{1,:})$$

A common mistake is to use formula C = log 3 - H(Q_1,:). Please note that this formula only holds for symmetric and weakly symmetric channels.

to calculate capacity, with uniform input distribution (because it achieves capacity). The difference here is that the output distribution is not uniform anymore. Thus, [2T]

$$C = H(0.4, 0.4, 0.2) - H(0.7, 0.2, 0.1) = 1.52 - 1.16 = 0.36 \text{ bits}$$

The transition matrix of the second channel is given by

$$\begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.1 & 0.7 \end{pmatrix}$$

There is no way to do a similar partition, so it is not generally symmetric. [1T]

c) Part c) is hard; very few students worked it out

(i) Recall the definition of mutual information

$$I(X;Y) = \sum_y P(y|x)Q(x)\log\frac{P(y|x)}{\sum_x Q(x)P(y|x)} \qquad (*)$$

Finding the capacity can be reformulated as the following optimization problem

$$\max_{Q(x)} I(X;Y) \qquad \text{subject to} \quad \sum_x Q(x) = 1$$

Using the method of Lagrange multiplier, we form the objective function

$$J = I(X;Y) + \lambda \sum_x Q(x)$$

Taking partial derivative with respect to Q(x), we obtain [3T]

$$\frac{\partial J}{\partial Q(x)} = I(x;Y) - \log e - \lambda$$

Therefore, letting this partial derivative be 0, we have

$$I(x;Y) = constant = C$$

Substituting this back to Eq. (*), we have [2T]

$$\max_{Q(x)} I(X;Y) = C$$

(ii)

For a generally symmetric channel, if the input distribution is uniform, then

$$I(x;Y) = \sum_y P(y|x)\log\frac{P(y|x)}{\sum_x \frac{1}{|X|}P(y|x)}$$

Note that within each symmetric block of a partition, $\sum_x \frac{1}{|X|}P(y|x)$ is identical for all y's, because its columns are permutations of each other. [2T]

This implies that if we form a matrix of entries

$$P(y|x)\log\frac{P(y|x)}{\sum_x \frac{1}{|X|}P(y|x)}$$

its rows will be permutations of each other. Thus $I(x;Y)$ is a constant for all x's. Therefore, the condition in (i) is satisfied, and accordingly the uniform distribution achieves capacity. [3T]

4.

a)

i) Capacity region [5B]

$$R_1 < C\left(\frac{P_1}{N}\right)$$

$$R_2 < C\left(\frac{P_2}{N}\right)$$

$$R_1 + R_2 < C\left(\frac{P_1+P_2}{N}\right)$$

At the corner point, the decoder decodes one user first, treating the other user as noise. Thus, it achieves rate $R_1 = C\left(\frac{P_1}{P_2+N}\right)$. After that, the decoder subtracts off user 1, meaning user 2 is only subject to noise. Thus, it can achieve rate $R_2 = C\left(\frac{P_2}{N}\right)$. This strategy is called successive interference cancellation or "Onion peeling".

ii) [5E]

$$C\left(\frac{P_1}{N}\right) + C\left(\frac{P_2}{P_1+N}\right)$$

$$= \frac{1}{2}\log\left(1+\frac{P_1}{N}\right) + \frac{1}{2}\log\left(1+\frac{P_2}{P_1+N}\right)$$

$$= \frac{1}{2}\log\left(\frac{P_1+N}{N}\cdot\frac{P_1+P_2+N}{P_1+N}\right)$$

$$= \frac{1}{2}\log\left(\frac{P_1+P_2+N}{N}\right)$$

$$= \frac{1}{2}\log\left(1+\frac{P_1+P_2}{N}\right)$$

$$= C\left(\frac{P_1+P_2}{N}\right)$$

iii)

$$d = \lim_{P\to\infty}\frac{C\left(\frac{mP}{N}\right)}{C\left(\frac{P}{N}\right)} = \lim_{P\to\infty}\frac{\frac{1}{2}\log\left(1+\frac{mP}{N}\right)}{\frac{1}{2}\log\left(1+\frac{P}{N}\right)}$$

$$= \lim_{P\to\infty}\frac{\log\left(\frac{mP}{N}\right)}{\log\left(\frac{P}{N}\right)} = \lim_{P\to\infty}\frac{\log(m)+\log\left(\frac{P}{N}\right)}{\log\left(\frac{P}{N}\right)} = 1$$

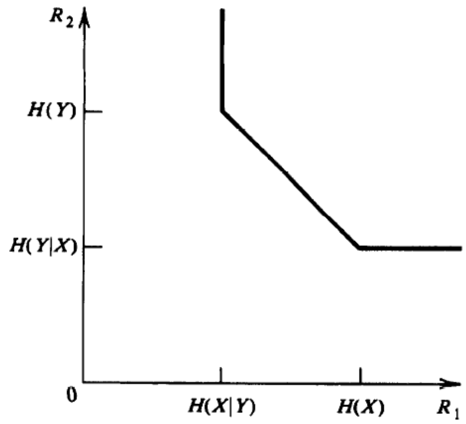
The DoF per user is 1/m, which tends to zero as m increases. [5T]

A small number of students forgot L'Hospital's rule, or couldn't apply it correctly.

b)

The Slepian-Wolf region is given by

$$R_1 \geq H(X \mid Y)$$
$$R_2 \geq H(Y \mid X)$$
$$R_1 + R_2 \geq H(X, Y)$$



[4E]

In this question,

$H(x, y) = -½\log½ - ¼\log¼ - 0\log 0 - ¼\log¼ = 1.5$ bits [2E]

$H(y \mid x) = -½\log \frac{2}{3} - ¼\log \frac{1}{3} - 0\log 0 - ¼\log 1 = 0.689$ bits [2E]

$H(x \mid y) = 0.5$ bits [2E]