The Solutions

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

1.

a)

i) $I(y;x) = H(x) + H(y) - H(x,y) = E \log \dfrac{p(x,y)}{p(x)p(y)} = \sum p(x,y) \log \dfrac{p(x,y)}{p(x)p(y)}$

Mutual information is the average amount of information that you get about x from observing the value of y. In particular, in communications, mutual information is the amount of information transmitted through a noisy channel. [3B]

ii) Max $I(y;x)$ = channel capacity, achieved by a capacity-achieving input distribution.
Min $I(y;x)$ = 0, achieved by a trivial input distribution with 0 entropy. [3B]

iii)

[3A]

$$\mathbf{p} = \left[1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6 \right] \Rightarrow H(\mathbf{p}) = 2.585$$

$$\mathbf{q} = \left[1/10 \quad 1/10 \quad 1/10 \quad 1/10 \quad 1/10 \quad 1/2 \right] \Rightarrow H(\mathbf{q}) = 2.161$$

$$D(\mathbf{p} \,\|\, \mathbf{q}) = E_{\mathbf{p}}(-\log q_x) - H(\mathbf{p}) = 2.935 - 2.585 = 0.35$$

$$D(\mathbf{q} \,\|\, \mathbf{p}) = E_{\mathbf{q}}(-\log p_x) - H(\mathbf{q}) = 2.585 - 2.161 = 0.424$$

iv) It is NOT a distance, since it is not symmetric. [3A]

v) $I(y;x) = \sum p(x,y) \log \dfrac{p(x,y)}{p(x)p(y)} = D(p(x,y) \,\|\, p(x)p(y))$

[3B]

b)

i) To find the stationary distribution, we write [3A]

$$T^T \begin{pmatrix} a \\ 1-a \end{pmatrix} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \begin{pmatrix} a \\ 1-a \end{pmatrix} = \begin{pmatrix} a \\ 1-a \end{pmatrix}$$

$$\Rightarrow \quad (1-p)a + p(1-a) = (1-2p)a + p = a$$

$$\Rightarrow \quad a = 1/2$$

which is a uniform distribution. [2A]

It follows that the entropy rate is given by
$H(x) = aH(p) + (1-a)H(p) = H(p)$ [3A]

ii) Obviously, $H(x)$ is maximized when p = ½. [2A]

2.

a) 1100001101010000001101 01

location	parsing	encoding
0000		
0001	1	(0000,1)
0010	10	(0001,0)
0011	0	(0000,0)
0100	00	(0011,0)
0101	11	(0001,1)
0110	01	(0011,1)
0111	010	(0110,0)
1000	000	(0100,0)
1001	011	(0110,1)
1010	0101	(0111,1)
1011		
1100		
1101		
1110		
1111		

[10E, 1 each step]

b)

Obviously, both x and y are uniform, with entropy H(x) = H(y) = 1.

i) $H(x) = 1$

The probability of a particular sequence x is given by

$$p(x) = \left(\tfrac{1}{2}\right)^m \left(\tfrac{1}{2}\right)^{n-m} = \left(\tfrac{1}{2}\right)^n \qquad m\text{: the number of ones}$$

Thus,

$$-\frac{1}{n}\log p(x) = -\frac{1}{n}\log\left(\tfrac{1}{2}\right)^n = 1 = H(X)$$

Therefore, all 2^n sequences are in the typical set.

[3A]

ii) $H(Y)=1$

Similarly, all 2^n sequences y are in the typical set

[2A]

iii) $H(X,Y) = H(X) + H(Y|X) = 1 + H(0.2) = 1.72$

From the joint distribution, we deduce

$$p(\boldsymbol{x},\boldsymbol{y}) = 0.4^{n-k}0.1^k = 2^{-n}0.8^{n-k}0.2^k$$

where k is the number of positions where they differ. [3A]

We also have

$$-\frac{1}{n}\log p(\boldsymbol{x},\boldsymbol{y}) = 1 - \frac{1}{n}\log(0.8^{n-k}0.2^k)$$

Joint typicality requires

$$|-\frac{1}{n}\log p(\boldsymbol{x},\boldsymbol{y}) - H(X,Y)| = \left|-\frac{1}{n}\log(0.8^{n-k}0.2^k) - 0.72\right| < \varepsilon = 0.1$$

From the following table, only the case k=2 is eligible. The size of the typical set is 45, while the probability is 0.302. [3A]

k	$\binom{n}{k}$	$0.8^{n-k}0.2^k$	$-\frac{1}{n}\log(0.8^{n-k}0.2^k)$	probability
0	1	0.107	0.322	0.107
1	10	0.0268	0.522	0.268
2	45	0.00671	0.722	0.302
3	120	0.00168	0.922	0.201
	…			

[4A]

3.

a)

(1) definition of mutual info [1B]
(2) conditional entropy = average row entropy $H(Y|X) = \sum_i H(Y|X = i)$ [1B]
(3) algebra [1B]
(4) $H(X) \leq 1$ [1B]
(5) definition of capacity, = is achieved by uniform input distribution [1B]

b) Erasure probability of W^-:

$1 - (1-f)^2 = 2f - f^2$ [2E]

$C(W^-) = (1-f)^2 = \left(1 - \frac{3}{4}\right)^2 = \frac{1}{16}$

[3E]

Erasure probability of W^+:

f^2 [2E]

$C(W^+) = 1 - f^2 = 1 - \left(\frac{3}{4}\right)^2 = \frac{7}{16}$

[3E]

c) Input (0 0 0 1 0 1 1 1) [1E]

First level (0 0 1 1 1 1 0 1) [3E]

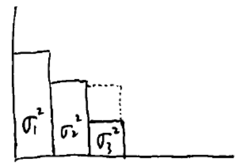
Second level (1 1 1 1 1 0 0 1) [3E]

Third level (0 1 1 0 1 0 0 1) 🡐 codeword [3E]

4. a)

i) Single channel is when

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

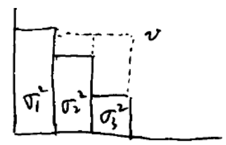


Capacity

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

[5A]

ii) A pair of channel is when



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

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

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

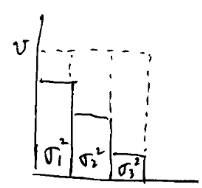
$$P_3 = \nu - \sigma_3^2 = \frac{3P + \sigma_2^2 - \sigma_3^2}{2}$$

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

[5A]

iii) Three channels is when

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



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

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

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

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

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

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

[5A]

b)

We can expand the mutual information

$$I(X; Y) = h(Y) - h(Y \mid X) = h(Y) - h(Z)$$

and $h(Z) = \log 2$, since Z is uniform over [-1, 1]. [1T]

The output Y is a sum of a discrete and a continuous random variable, and if the probabilities of X are p_{-1}, p_0, p_1 , then the output distribution of Y has piecewise constant distribution

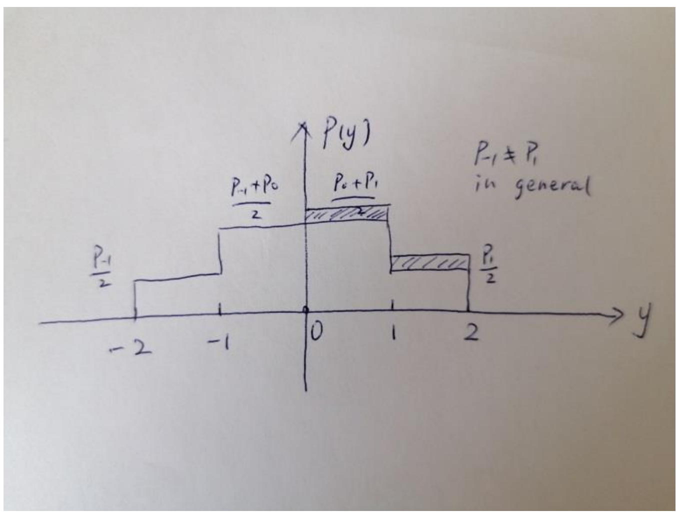
uniform with weight $\frac{p_{-1}}{2}$ for $-2 \le y \le -1$, [1T]

uniform with weight $\frac{p_{-1}+p_0}{2}$ for $-1 \le y \le 0$, [1T]

uniform with weight $\frac{p_1+p_0}{2}$ for $0 \le y \le 1$, [1T]

uniform with weight $\frac{p_1}{2}$ for $1 \le y \le 2$. [1T]

See the following sketch:



Given that Y ranges from -2 to 2, the maximum entropy that it can have is a uniform over this range. [2T]

This can be achieved if the distribution of X is (1/2, 0, 1/2). [1T]

Then $h(Y) = \log 4$ and the capacity of this channel is [2T]

$$C = \log 4 - \log 2 = \log 2 \text{ bits}.$$

