

# Fundamentals of Information Theory

## Homework 4

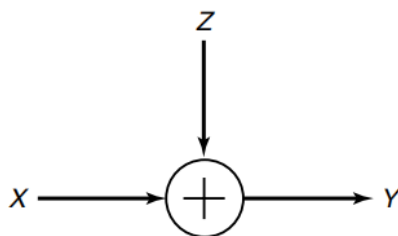
Name: Zhou Shouchen

Student ID: 2021533042

Due 23:59 (CST), Nov. 17, 2024

# Problem 1

7.2 Additive noise channel. Find the channel capacity of the following discrete memoryless channel:



where  $\Pr\{Z = 0\} = \Pr\{Z = a\} = \frac{1}{2}$ . The alphabet for  $x$  is  $\mathbf{X} = \{0, 1\}$ . Assume that  $Z$  is independent of  $X$ . Observe that the channel capacity depends on the value of  $a$ .

## Solution

We have  $\mathcal{X} = \{0, 1\}, \mathcal{Z} = \{0, 1\}, \mathcal{Y} = \{0, 1, a, a + 1\}, Y = X + Z$ .

So we need to discuss the value of  $a$ .

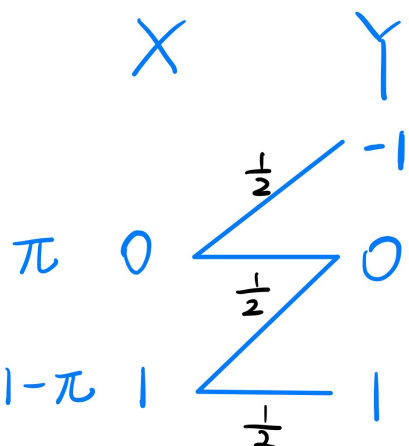
<1> When  $a = 0$ , then  $P(Z = 0) = P(Z = a) = \frac{1}{2}$ , which is not a suitable probability distribution. If we correct the probability into  $P(Z = 0) = 1$ , then we have

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= H(X) \quad (\text{since } Y = X, \text{ so when given } Y, X \text{ is deterministic}) \\ &\leq \log |\mathcal{X}| \\ &= 1 \end{aligned}$$

When  $p(x) = \left(\frac{1}{2}, \frac{1}{2}\right)$ , the inequality takes the equality.

So the channel capacity is  $C = \max_{p(x)} I(X; Y) = 1$  bit.

<2> When  $a = -1$ , then  $\mathcal{Y} = \{0, 1, -1\}$ , the DMC is as follows:



If we define  $p(X = 0) = \pi, p(X = 1) = 1 - \pi$ , then

$$P(Y = 0) = P(Y = 0|X = 0)P(X = 0) + P(Y = 0|X = 1)P(X = 1) = \frac{1}{2}\pi + \frac{1}{2}(1 - \pi) = \frac{1}{2}$$

Also, we can get that  $P(X = 0|Y = 0) = \frac{P(Y = 0|X = 0)P(X = 0)}{P(Y = 0)} = \pi$ ,  $P(X = 1|Y = 0) = 1 - \pi$ .

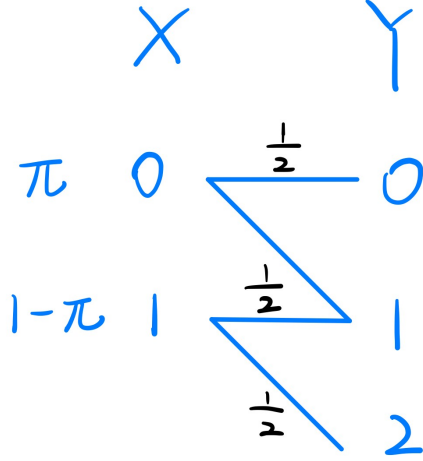
So we have:

$$\begin{aligned}
I(X; Y) &= H(X) - H(X|Y) \\
&= H(X) - \sum_y H(X|Y = y)P(Y = y) \\
&= H(X) - H(X|Y = 0)p(Y = 0) \quad (\text{since when } Y = \pm 1, X \text{ is deterministic, } H(X|Y = -1) = H(X|Y = 1) = 0) \\
&= H(\pi, 1 - \pi) - \frac{1}{2}H(\pi, 1 - \pi) \\
&= \frac{1}{2}H(\pi, 1 - \pi) \\
&\leq \frac{1}{2}\log 2 = \frac{1}{2}
\end{aligned}$$

If  $\pi = \frac{1}{2}$ , i.e.  $p(x) = \left(\frac{1}{2}, \frac{1}{2}\right)$ , the inequality takes the equality.

So the channel capacity is  $C = \max_{p(x)} I(X; Y) = \frac{1}{2}$  bit.

<3> When  $a = 1$ , then  $\mathcal{Y} = \{0, 1, 2\}$ , the DMC is as follows:



Which is quite similar to the case when  $a = -1$ .

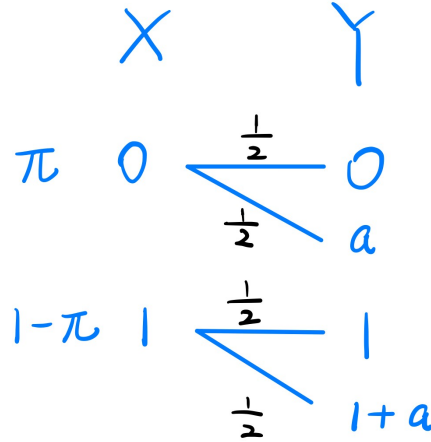
i.e. With the quite similar process, we can get that

$$\begin{aligned}
P(Y = 1) &= \frac{1}{2} \\
P(X = 0|Y = 1) &= \pi \\
I(X; Y) &= H(X) - H(X|Y) \\
&= H(X) - \sum_y H(X|Y = y)P(Y = y) \\
&= H(X) - H(X|Y = 1)p(Y = 1) \quad (\text{When } Y = 0, 2, X \text{ is deterministic, } H(X|Y = 0) = H(X|Y = 2) = 0) \\
&= H(\pi, 1 - \pi) - \frac{1}{2}H(\pi, 1 - \pi) \\
&= \frac{1}{2}H(\pi, 1 - \pi) \leq \frac{1}{2}\log 2 = \frac{1}{2}
\end{aligned}$$

If  $\pi = \frac{1}{2}$ , i.e.  $p(x) = \left(\frac{1}{2}, \frac{1}{2}\right)$ , the inequality takes the equality.

So the channel capacity is  $C = \max_{p(x)} I(X; Y) = \frac{1}{2}$  bit.

<4> When  $a \neq 0, a \neq \pm 1$ , then the channel is as follows:



Which is actually a noisy channel with non-overlapping outputs.

$$\begin{aligned}
 I(X; Y) &= H(X) - H(X|Y) \\
 &= H(X) \quad (\text{When given } Y, X \text{ is deterministic}) \\
 &\leq \log |\mathcal{X}| = 1
 \end{aligned}$$

So the channel capacity is  $C = \max_{p(x)} I(X; Y) = 1$  bit.

So above all, the channel capacity of the following discrete memoryless channel is as follows:

$$C = \begin{cases} 1 & , \text{ if } a = 0 , \text{ and the probability is corrected into } P(X = 0) = 1 \\ \frac{1}{2} & , \text{ if } a = \pm 1 \\ 1 & , \text{ otherwise} \end{cases}$$

### Problem 2

7.3 Channels with memory have higher capacity. Consider a binary symmetric channel with  $Y_i = X_i \oplus Z_i$ , where  $\oplus$  is mod 2 addition, and  $X_i, Y_i \in \{0, 1\}$ . Suppose that  $\{Z_i\}$  has constant marginal probabilities  $\Pr\{Z_i = 1\} = p = 1 - \Pr\{Z_i = 0\}$ , but that  $Z_1, Z_2, \dots, Z_n$  are not necessarily independent. Assume that  $Z^n$  is independent of the input  $X^n$ . Let  $C = 1 - H(p, 1 - p)$ . Show that

$$\max_{p(x_1, x_2, \dots, x_n)} I(X_1, X_2, \dots, X_n; Y_1, Y_2, \dots, Y_n) \geq nC$$

### Solution

Since  $Y_i = X_i \oplus Z_i \Rightarrow Z_i = X_i \oplus Y_i$ , so  $p(X^n|Y^n) = p(Z^n|Y^n)$ . Then we have

$$\begin{aligned} I(X^n; Y^n) &= H(X^n) - H(X^n|Y^n) \\ &= H(X^n) - H(Z^n|Y^n) \quad (p(X^n|Y^n) = p(Z^n|Y^n)) \\ &\geq H(X^n) - H(Z^n) \quad (\text{Conditioning reduces entropy}) \\ &= H(X^n) - \sum_{i=1}^n H(Z_i|Z_1, \dots, Z_{i-1}) \\ &\geq H(X^n) - \sum_{i=1}^n H(Z_i) \quad (\text{Conditioning reduces entropy}) \\ &= H(X^n) - nH(p, 1 - p) \end{aligned}$$

And since  $|\mathcal{X}| = 2$ , so we have

$$H(X^n) = \sum_{i=1}^n H(X_i|X_1, \dots, X_{i-1}) \leq \sum_{i=1}^n H(X_i) \leq \sum_{i=1}^n \log |\mathcal{X}| = n$$

When  $X_i$  are independent, and  $P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$ , the inequality holds the equality. So

$$\begin{aligned} \max_{p(x_1, x_2, \dots, x_n)} I(X_1, X_2, \dots, X_n; Y_1, Y_2, \dots, Y_n) &\geq \max_{p(X^n)} (H(X^n) - nH(p, 1 - p)) \\ &= n - nH(p, 1 - p) \\ &= n(1 - H(p, 1 - p)) \\ &= nC \end{aligned}$$

So above all, we have proved that when  $X_i \stackrel{i.i.d.}{\sim} \text{Bern}\left(\frac{1}{2}\right)$ , the ‘max’ could be taken, and

$$\max_{p(x_1, x_2, \dots, x_n)} I(X_1, X_2, \dots, X_n; Y_1, Y_2, \dots, Y_n) \geq nC$$

### Problem 3

7.4 Channel capacity. Consider the discrete memoryless channel  $Y = X + Z(\text{mod } 11)$ , where

$$Z = \begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

and  $X \in \{0, 1, \dots, 10\}$ . Assume that  $Z$  is independent of  $X$ .

- (a) Find the capacity.
- (b) What is the maximizing  $p^*(x)$  ?

### Solution

(a) Since  $Y = X + Z(\text{mod } 11)$ , i.e.  $Z = Y - X + 11(\text{mod } 11)$ . And  $X \perp Z$ , so

$$\begin{aligned} H(Y|X) &= H(Z|X) = H(Z) = H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = \log 3 \\ I(X; Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \log 3 \\ &\leq \log 11 - \log 3 \end{aligned}$$

When  $Y \sim DUnif(11)$ , the inequality takes equality. And we can prove that  $Y$  could get such distribution with certain  $p^*(x)$  in (b).

So the channel capacity is  $C = \max_{p(x)} I(X; Y) = \log \frac{11}{3}$  bits.

(b) Then we prove that we can construct  $p^*(x)$ , s.t.  $\forall y = 0, \dots, 10, P(Y = y) = \frac{1}{11}$ .

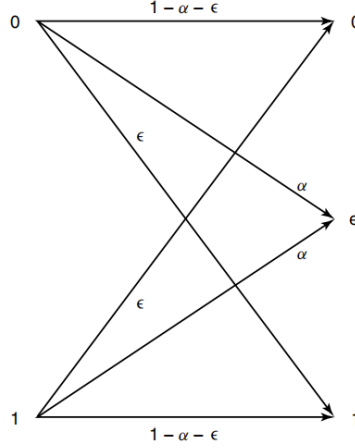
We define  $p_x = P(X = x) = \frac{1}{11}$ . Then we have

$$\begin{aligned} P(Y = y) &= \sum_x P(X = x)P(Y = y|X = x) \\ &= \frac{1}{3}p_{(y+8)\text{mod } 11} + \frac{1}{3}p_{(y+9)\text{mod } 11} + \frac{1}{3}p_{(y+10)\text{mod } 11} \\ &= 3 * \frac{1}{3} * \frac{1}{11} \\ &= \frac{1}{11} \end{aligned}$$

So above all, we have proved that  $p^*(x) = \frac{1}{11}, \forall x = 0, \dots, 10$  is the maximizing  $p(x)$ .

#### Problem 4

7.13 Erasures and errors in a binary channel. Consider a channel with binary inputs that has both erasures and errors. Let the probability of error be  $\epsilon$  and the probability of erasure be  $\alpha$ , so the channel is follows:



- (a) Find the capacity of this channel.
- (b) Specialize to the case of the binary symmetric channel ( $\alpha = 0$ ).
- (c) Specialize to the case of the binary erasure channel ( $\epsilon = 0$ ).

#### Solution

- (a) Let  $p(X = 0) = \pi, p(X = 1) = 1 - \pi$ , then we have

$$\begin{aligned}
 p(Y = 0) &= \sum_x p(Y = 0|X = x)p(X = x) = \pi(1 - \alpha - 2\epsilon) + \epsilon \\
 p(Y = 1) &= \sum_x p(Y = 1|X = x)p(X = x) = \pi(2\epsilon + \alpha - 1) + 1 - \alpha - \epsilon \\
 P(Y = e) &= \alpha
 \end{aligned}$$

So we have

$$\begin{aligned}
 I(X; Y) &= H(Y) - H(Y|X) \\
 &= H(Y) - \sum_x p(X = x)H(Y|X = x) \\
 &= H(\alpha, \pi(1 - \alpha - 2\epsilon) + \epsilon, \pi(2\epsilon + \alpha - 1) + 1 - \alpha - \epsilon) - H(1 - \alpha - \epsilon, \alpha, \epsilon) \\
 &= -[\pi(2\epsilon + \alpha - 1) + 1 - \alpha - \epsilon] \log[\pi(2\epsilon + \alpha - 1) + 1 - \alpha - \epsilon] - [\pi(1 - \alpha - 2\epsilon) + \epsilon] \log[\pi(1 - \alpha - 2\epsilon) + \epsilon] \\
 &\quad + \epsilon \log \epsilon + (1 - \alpha - \epsilon) \log(1 - \alpha - \epsilon) \\
 &\triangleq f(\pi) \\
 f'(\pi) &= -(2\epsilon + \alpha - 1) \log[\pi(2\epsilon + \alpha - 1) + 1 - \alpha - \epsilon] - (1 - \alpha - 2\epsilon) \log[\pi(1 - \alpha - 2\epsilon) + \epsilon] \\
 f''(\pi) &= -(2\epsilon + \alpha - 1)^2 \log_e 2 \left[ \frac{1}{\pi(2\epsilon + \alpha - 1) + 1 - \alpha - \epsilon} + \frac{1}{\pi(1 - \alpha - 2\epsilon) + \epsilon} \right]
 \end{aligned}$$

Then we need to discuss the sign of  $2\epsilon + \alpha - 1$  to determine the sign of  $f''(\pi)$ . Since  $\pi \in [0, 1]$ , so:

- 1. If  $2\epsilon + \alpha - 1 = 0$ , then  $f''(\pi) = 0$ .

2. If  $2\epsilon + \alpha - 1 > 0$ , then

$$[\pi(2\epsilon + \alpha - 1) + 1 - \alpha - \epsilon]_{\min} = [\pi(2\epsilon + \alpha - 1) + 1 - \alpha - \epsilon] \Big|_{\pi=0} = 1 - \alpha - \epsilon \geq 0$$

$$[\pi(1 - \alpha - 2\epsilon) + \epsilon]_{\min} = [\pi(1 - \alpha - 2\epsilon) + \epsilon] \Big|_{\pi=1} = 1 - \alpha - \epsilon \geq 0$$

So  $f''(\pi) \leq 0$ .

3. If  $2\epsilon + \alpha - 1 < 0$ , then

$$[\pi(2\epsilon + \alpha - 1) + 1 - \alpha - \epsilon]_{\min} = [\pi(2\epsilon + \alpha - 1) + 1 - \alpha - \epsilon] \Big|_{\pi=1} = \epsilon \geq 0$$

$$[\pi(1 - \alpha - 2\epsilon) + \epsilon]_{\min} = [\pi(1 - \alpha - 2\epsilon) + \epsilon] \Big|_{\pi=0} = \epsilon \geq 0$$

So  $f''(\pi) \leq 0$ .

So above all, we have proved that  $f''(\pi) \leq 0$  always holds, so  $f(\pi)$  is concave.

In order to get that maximum value of  $f(\pi)$ , we have

$$f'(\pi^*) = 0 \Rightarrow \pi^* = \frac{1}{2}$$

Then we have

$$C = \max_{p(x)} I(X; Y) = f(\pi^*) = H\left(\alpha, \frac{1-\alpha}{2}, \frac{1-\alpha}{2}\right) - H(1-\alpha-\epsilon, \alpha, \epsilon)$$

And since we have

$$H\left(\alpha, \frac{1-\alpha}{2}, \frac{1-\alpha}{2}\right) = -\alpha \log \alpha - 2 \cdot \frac{1-\alpha}{2} \log \frac{1-\alpha}{2} = -\alpha \log \alpha - (1-\alpha) \log(1-\alpha) + 1 - \alpha = H(\alpha, 1-\alpha) + 1 - \alpha$$

So above all, when  $p(x) = \left(\frac{1}{2}, \frac{1}{2}\right)$ , the inequality takes the equality, i.e.

$$C = H(\alpha, 1-\alpha) - H(1-\alpha-\epsilon, \alpha, \epsilon) + 1 - \alpha$$

(b) When  $\alpha = 0$ ,  $C = 1 - H(\epsilon, 1-\epsilon)$ .

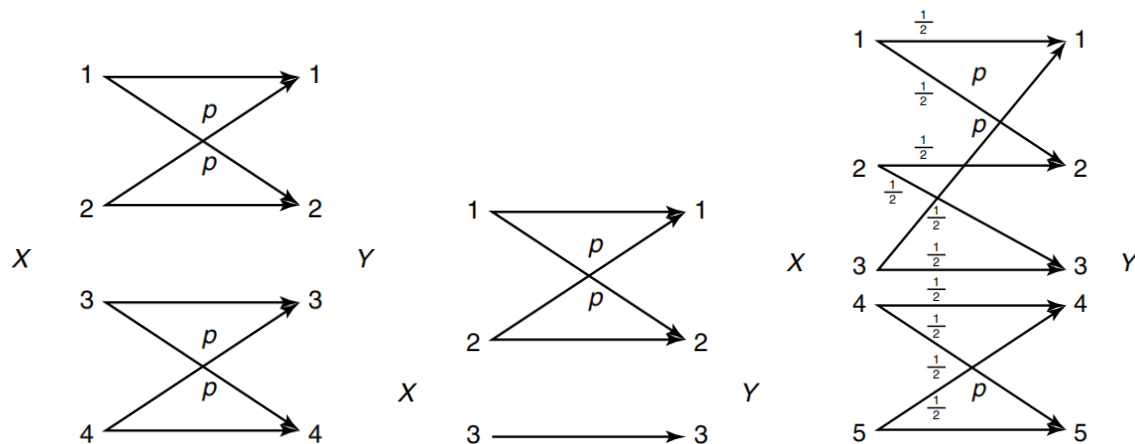
(c) When  $\epsilon = 0$ ,  $C = 1 - \alpha$



# Problem 5

7.34 Capacity. Find the capacity of

- (a) Two parallel BSCs:
- (b) BSC and a single symbol:
- (c) BSC and a ternary channel:



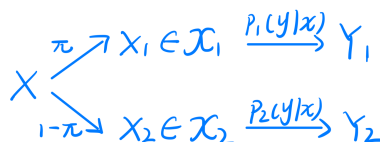
- (d) Ternary channel:

$$p(y|x) = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

# Solution

Since for (a), (b), (c): the channels are all combined with two parallel channels, so we can introduce two Lemmas.

Lemma1: If the channel is combined with two parallel channels with probability  $\pi$  to the first channel, and  $1 - \pi$  to the second channel, which is as followed.



Suppose the channel capacity of the total channel is  $C$ , and the capacity of the two channels are  $C_1$  and  $C_2$  respectively. Then we have

$$2^C = 2^{C_1} + 2^{C_2}$$

Proof:

Let  $\Theta(X)$  be the auxiliary variable, indicating that  $X$  is send to the  $\Theta(X)$ 's channel. Then we have

$$\Theta(X) = \begin{cases} 1 & , X \in \mathcal{X}_1 \\ 2 & , X \in \mathcal{X}_2 \end{cases}$$

$$C_1 = \max_{p_1(x)} I(X; Y_1)$$

$$C_2 = \max_{p_2(x)} I(X; Y_2)$$

And we also have that  $p(\Theta = 1) = p(X \in \mathcal{X}_1) = \pi, p(\Theta = 2) = p(X \in \mathcal{X}_2) = 1 - \pi$ . And since when given  $Y$  or given  $X$ , we could know  $\Theta$ , so we have  $I(X; \Theta|Y) = H(\Theta|Y) - H(\Theta|X, Y) = 0$ . i.e.

$$\begin{aligned} I(X; Y, \Theta) &= I(X; \Theta) + I(X; Y|\Theta) \\ &= I(X; Y) + I(X; \Theta|Y) = I(X; Y) \end{aligned}$$

Then we have

$$\begin{aligned} I(X; Y) &= I(X; \Theta) + I(X; Y|\Theta) \\ &= H(\Theta) - H(\Theta|X) + I(X; Y|\Theta) \\ &= H(\Theta) + I(X; Y|\Theta) \quad (\Theta \text{ is deterministic when } X \text{ is given}) \\ &= H(\pi, 1 - \pi) + I(X; Y|\Theta) \end{aligned}$$

As for  $I(X; Y|\Theta)$ , we have

$$\begin{aligned} I(X; Y|\Theta) &= H(X|\Theta) - H(X|Y, \Theta) \\ &= \sum_{\theta} p(\Theta = \theta) H(X|\Theta = \theta) - \sum_{\theta} p(\Theta = \theta) H(X|Y, \Theta = \theta) \\ &= \sum_{\theta} p(\Theta = \theta) [H(X|\Theta = \theta) - H(X|Y, \Theta = \theta)] \\ &= \sum_{\theta} p(\Theta = \theta) I(X; Y|\Theta = \theta) \\ &= \pi I(X_1; Y_1) + (1 - \pi) I(X_2; Y_2) \\ &\leq \pi C_1 + (1 - \pi) C_2 \quad (C_1 \text{ and } C_2 \text{ are the capacity of the two channels}) \end{aligned}$$

So we have

$$I(X; Y) \leq H(\pi, 1 - \pi) \pi C_1 + (1 - \pi) C_2 = -\pi \log \pi - (1 - \pi) \log(1 - \pi) + \pi C_1 + (1 - \pi) C_2 \triangleq f(\pi)$$

Since we want to maximize  $f(\pi)$ , so it is obviously that  $\pi = 0$  or  $\pi = 1$  is not the optimal solution. And from the property of capacity, we have  $C_1, C_2 \geq 0$ , so we have

$$\begin{aligned} f'(\pi) &= \log(1 - \pi) - \log \pi + C_1 - C_2 \\ f''(\pi) &= -\log_e 2 \left( \frac{1}{1 - \pi} + \frac{1}{\pi} \right) < 0 \end{aligned}$$

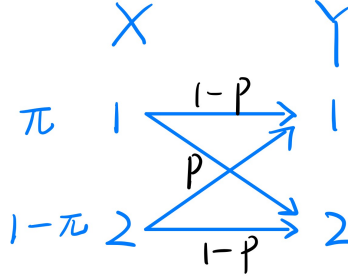
So  $f(\pi)$  is a concave function, so the optimal solution is

$$f'(\pi^*) = 0 \Rightarrow \pi^* = \frac{2^{C_1}}{2^{C_1} + 2^{C_2}}$$

So we have

$$\begin{aligned} C &= \max_{p(x)} I(X; Y) = \max_{\pi} f(\pi) = f(\pi^*) \\ &= \log(2^{C_1} + 2^{C_2}) \end{aligned}$$

So we have proved that  $2^C = 2^{C_1} + 2^{C_2}$ . □



Lemma2: For a BSC. The probability of error is  $p$ , then the capacity is  $1 - H(p, 1 - p)$ .

$$\begin{aligned}
 I(X; Y) &= H(Y) - H(Y|X) \\
 &= H(Y) - \sum_x p(x) H(Y|X = x) \\
 &= H(Y) - H(p, 1 - p) \\
 &\leq 1 - H(p, 1 - p)
 \end{aligned}$$

And we have when  $P(X = 1) = P(X = 2) = \frac{1}{2}$ , then  $P(Y = 0) = \sum_x P(X = x)P(Y = 1|X = x) = \frac{1}{2}$ , and so does  $P(Y = 1)$ .

So the inequality takes the equality, so we have when  $p(x) = \left(\frac{1}{2}, \frac{1}{2}\right)$

$$C = \max_{p(x)} I(X; Y) = 1 - H(p, 1 - p)$$

So with the two Lemmas, we could solve the problem easier.

(a) The two channels are paralleled BSC, so we have  $C_1 = C_2 = 1 - H(p, 1 - p)$ , so we have

$$C = \log \left( 2^{1-H(p, 1-p)} + 2^{1-H(p, 1-p)} \right) = 2 - H(p, 1 - p)$$

Where  $\pi^* = \frac{2^{C_1}}{2^{C_1} + 2^{C_2}} = \frac{1}{2}$ .

When  $p(x) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ ,  $I(X; Y)$  could take the maximum value.

(b) The first channel is a BSC, so  $C_1 = 1 - H(p, 1 - p)$ . And for the second channel,  $Y_2 = X_2 = 3$ , so

$$I(X_2; Y_2) = H(X_2) - H(X_2|Y_2) = H(X_2) = 0$$

i.e.  $C_2 = 0$ . So we have

$$C = \log \left( 2^{1-H(p, 1-p)} + 2^0 \right) = \log \left( 2^{1-H(p, 1-p)} + 1 \right)$$

Where  $\pi^* = \frac{2^{C_1}}{2^{C_1} + 2^{C_2}} = \frac{2}{2 + 2^{H(p, 1-p)}}$ .

When  $p(x) = \left( \frac{1}{2 + 2^{H(p, 1-p)}}, \frac{1}{2 + 2^{H(p, 1-p)}}, \frac{2^{H(p, 1-p)}}{2 + 2^{H(p, 1-p)}} \right)$ ,  $I(X; Y)$  could take the maximum value.

(c) The second channel is a BSC, and  $p = \frac{1}{2}$ , so  $C_2 = 1 - H(p, 1 - p) = 0$ .  
And for the first channel:

$$\begin{aligned}
I(X; Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \sum_x p(x) H(Y|X = x) \\
&= H(Y) - \sum_x p(x) H\left(\frac{1}{2}, \frac{1}{2}\right) \\
&= H(Y) - 1 \\
&\leq \log 3 - 1
\end{aligned}$$

When  $p(X = 1) = p(X = 2) = p(X = 3) = \frac{1}{3}$ , we have

$$P(Y = 1) = \sum_x P(X = x) P(Y = 1|X = x) = \frac{1}{3}$$

Similarly,  $P(Y = 2) = P(Y = 3) = \frac{1}{3}$ . So the inequality takes the equality, so we have  $C_1 = \log 3 - 1$ .  
So we have

$$C = \log(2^0 + 2^{\log 3 - 1}) = \log 5 - 1$$

Where  $\pi^* = \frac{2^{C_1}}{2^{C_1} + 2^{C_2}} = \frac{3}{5}$ .

When  $p(x) = \left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$ ,  $I(X; Y)$  could take the maximum value.

(d)

$$\begin{aligned}
I(X; Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \sum_x p(x) H(Y|X = x) \\
&= H(Y) - H\left(\frac{1}{3}, \frac{2}{3}\right) \\
&\leq \log 3 - H\left(\frac{1}{3}, \frac{2}{3}\right)
\end{aligned}$$

When  $p(X = 1) = p(X = 2) = p(X = 3) = \frac{1}{3}$ , we have

$$P(Y = 1) = \sum_x P(X = x) P(Y = 1|X = x) = \frac{1}{3}$$

Similarly,  $P(Y = 2) = P(Y = 3) = \frac{1}{3}$ . So the inequality takes the equality.

So we have  $C = \log 3 - H\left(\frac{1}{3}, \frac{2}{3}\right)$ .

And we have

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

So above all, we have  $C = \frac{2}{3}$ .

### Problem 6

7.37 Joint typicality. Let  $(X_i, Y_i, Z_i)$  be i.i.d. according to  $p(x, y, z)$ . We will say that  $(x^n, y^n, z^n)$  is jointly typical [written  $(x^n, y^n, z^n) \in A_\epsilon^{(n)}$ ] if

- $p(x^n) \in 2^{-n(H(X) \pm \epsilon)}$ .
- $p(y^n) \in 2^{-n(H(Y) \pm \epsilon)}$ .
- $p(z^n) \in 2^{-n(H(Z) \pm \epsilon)}$ .
- $p(x^n, y^n) \in 2^{-n(H(X, Y) \pm \epsilon)}$ .
- $p(x^n, z^n) \in 2^{-n(H(X, Z) \pm \epsilon)}$ .
- $p(y^n, z^n) \in 2^{-n(H(Y, Z) \pm \epsilon)}$ .
- $p(x^n, y^n, z^n) \in 2^{-n(H(X, Y, Z) \pm \epsilon)}$ .

Now suppose that  $(\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n)$  is drawn according to  $p(x^n)p(y^n)p(z^n)$ . Thus,  $\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n$  have the same marginals as  $p(x^n, y^n, z^n)$  but are independent. Find (bounds on)  $\Pr \left\{ (\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n) \in A_\epsilon^{(n)} \right\}$  in terms of the entropies  $H(X), H(Y), H(Z), H(X, Y), H(X, Z), H(Y, Z)$ , and  $H(X, Y, Z)$ .

### Solution

From the definition of jointly typicality, we have in probability:

$$\begin{aligned} -\frac{1}{n} \log p(x^n) &\rightarrow H(X), -\frac{1}{n} \log p(y^n) \rightarrow H(Y), -\frac{1}{n} \log p(z^n) \rightarrow H(Z), -\frac{1}{n} \log p(x^n, y^n) \rightarrow H(X, Y) \\ -\frac{1}{n} \log p(x^n, z^n) &\rightarrow H(X, Z), -\frac{1}{n} \log p(y^n, z^n) \rightarrow H(Y, Z), -\frac{1}{n} \log p(x^n, y^n, z^n) \rightarrow H(X, Y, Z) \end{aligned}$$

i.e.  $\exists N_1, \dots, N_7 \in \mathbb{N}$ , s.t.

$$\begin{aligned} \Pr \left[ \left| -\frac{1}{n} \log p(x^n) - H(X) \right| \geq \epsilon \right] &< \delta = \frac{\epsilon}{7} \quad \text{for } n \geq N_1 \\ \Pr \left[ \left| -\frac{1}{n} \log p(y^n) - H(Y) \right| \geq \epsilon \right] &< \delta = \frac{\epsilon}{7} \quad \text{for } n \geq N_2 \\ \Pr \left[ \left| -\frac{1}{n} \log p(z^n) - H(Z) \right| \geq \epsilon \right] &< \delta = \frac{\epsilon}{7} \quad \text{for } n \geq N_3 \\ \Pr \left[ \left| -\frac{1}{n} \log p(x^n, y^n) - H(X, Y) \right| \geq \epsilon \right] &< \delta = \frac{\epsilon}{7} \quad \text{for } n \geq N_4 \\ \Pr \left[ \left| -\frac{1}{n} \log p(x^n, z^n) - H(X, Z) \right| \geq \epsilon \right] &< \delta = \frac{\epsilon}{7} \quad \text{for } n \geq N_5 \\ \Pr \left[ \left| -\frac{1}{n} \log p(y^n, z^n) - H(Y, Z) \right| \geq \epsilon \right] &< \delta = \frac{\epsilon}{7} \quad \text{for } n \geq N_6 \\ \Pr \left[ \left| -\frac{1}{n} \log p(x^n, y^n, z^n) - H(X, Y, Z) \right| \geq \epsilon \right] &< \delta = \frac{\epsilon}{7} \quad \text{for } n \geq N_7 \end{aligned}$$

Then let  $N = \max\{N_1, \dots, N_7\}$ , we have  $\forall n \geq N$ :

$$\begin{aligned} p(x^n \notin A_\epsilon^{(n)}(p_X)) &< \frac{\epsilon}{7}, \quad p(y^n \notin A_\epsilon^{(n)}(p_Y)) < \frac{\epsilon}{7}, \quad p(z^n \notin A_\epsilon^{(n)}(p_Z)) < \frac{\epsilon}{7}, \quad p((x^n, y^n) \notin A_\epsilon^{(n)}(p_{X,Y})) < \frac{\epsilon}{7} \\ p((x^n, z^n) \notin A_\epsilon^{(n)}(p_{X,Z})) &< \frac{\epsilon}{7}, \quad p((y^n, z^n) \notin A_\epsilon^{(n)}(p_{Y,Z})) < \frac{\epsilon}{7}, \quad p((x^n, y^n, z^n) \notin A_\epsilon^{(n)}(p_{X,Y,Z})) < \frac{\epsilon}{7} \end{aligned}$$

Then we have

$$\begin{aligned}
p \left[ (x^n, y^n, z^n) \in A_\epsilon^{(n)}(P_{X,Y,Z}) \right] &= 1 - p \left[ (x^n, y^n, z^n) \notin A_\epsilon^{(n)}(P_{X,Y,Z}) \right] \\
&\geq 1 - p \left( x^n \notin A_\epsilon^{(n)}(P_X) \right) - p \left( y^n \notin A_\epsilon^{(n)}(P_Y) \right) - p \left( z^n \notin A_\epsilon^{(n)}(P_Z) \right) \\
&\quad - p \left( (x^n, y^n) \notin A_\epsilon^{(n)}(P_{X,Y}) \right) - p \left( (x^n, z^n) \notin A_\epsilon^{(n)}(P_{X,Z}) \right) - p \left( (y^n, z^n) \notin A_\epsilon^{(n)}(P_{Y,Z}) \right) \\
&\quad - p \left( (x^n, y^n, z^n) \notin A_\epsilon^{(n)}(P_{X,Y,Z}) \right) \text{ (against any of the requirements of jointly typical set)} \\
&\geq 1 - \frac{\epsilon}{7} \times 7 = 1 - \epsilon \rightarrow 1 \quad \text{as } n \rightarrow \infty
\end{aligned}$$

With  $p \left[ (x^n, y^n, z^n) \in A_\epsilon^{(n)}(P_{X,Y,Z}) \right] \geq 1 - \epsilon$ , we can get the lower bound of  $\left| A_\epsilon^{(n)}(P_{X,Y,Z}) \right|$ :

$$\begin{aligned}
1 - \epsilon &\leq p \left[ (x^n, y^n, z^n) \in A_\epsilon^{(n)}(P_{X,Y,Z}) \right] \\
&= \sum_{(x^n, y^n, z^n) \in A_\epsilon^{(n)}(P_{X,Y,Z})} p(x^n, y^n, z^n) \\
&\leq \left| A_\epsilon^{(n)}(P_{X,Y,Z}) \right| 2^{-n(H(X,Y,Z) - \epsilon)}
\end{aligned}$$

So the lower bound of  $\left| A_\epsilon^{(n)}(P_{X,Y,Z}) \right|$  is

$$(1 - \epsilon) 2^{n(H(X,Y,Z) - \epsilon)} \leq \left| A_\epsilon^{(n)}(P_{X,Y,Z}) \right|$$

As for the Upper bound of  $\left| A_\epsilon^{(n)}(P_{X,Y,Z}) \right|$ , we have

$$\begin{aligned}
1 &= \sum_{(x^n, y^n, z^n)} p(x^n, y^n, z^n) \\
&\geq \sum_{(x^n, y^n, z^n) \in A_\epsilon^{(n)}(P_{X,Y,Z})} p(x^n, y^n, z^n) \\
&\geq \left| A_\epsilon^{(n)}(P_{X,Y,Z}) \right| 2^{-n(H(X,Y,Z) + \epsilon)}
\end{aligned}$$

So the upper bound of  $\left| A_\epsilon^{(n)}(P_{X,Y,Z}) \right|$  is

$$2^{n(H(X,Y,Z) + \epsilon)} \geq \left| A_\epsilon^{(n)}(P_{X,Y,Z}) \right|$$

Therefore, we have

$$2^{n(H(X,Y,Z) - \epsilon)} \leq \left| A_\epsilon^{(n)}(P_{X,Y,Z}) \right| \leq 2^{n(H(X,Y,Z) + \epsilon)}$$

Then for the bound of  $\Pr \left\{ \left( \tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n \right) \in A_\epsilon^{(n)}(P_{X,Y,Z}) \right\}$ , we have

$$\begin{aligned}
\Pr \left\{ \left( \tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n \right) \in A_\epsilon^{(n)}(P_{X,Y,Z}) \right\} &= \sum_{(x^n, y^n, z^n) \in A_\epsilon^{(n)}(P_{X,Y,Z})} p(x^n) p(y^n) p(z^n) = \left| A_\epsilon^{(n)}(P_{X,Y,Z}) \right| p(x^n) p(y^n) p(z^n) \\
&\in \left[ (1 - \epsilon) 2^{n(H(X,Y,Z) - \epsilon)} \cdot 2^{-n(H(X) + \epsilon)} \cdot 2^{-n(H(Y) + \epsilon)} \cdot 2^{-n(H(Z) + \epsilon)}, \right. \\
&\quad \left. 2^{n(H(X,Y,Z) + \epsilon)} \cdot 2^{-n(H(X) - \epsilon)} \cdot 2^{-n(H(Y) - \epsilon)} \cdot 2^{-n(H(Z) - \epsilon)} \right] \\
&= \left[ (1 - \epsilon) 2^{n(H(X,Y,Z) - H(X) - H(Y) - H(Z) - 4\epsilon)}, 2^{n(H(X,Y,Z) - H(X) - H(Y) - H(Z) + 4\epsilon)} \right]
\end{aligned}$$

So above all, the bound of  $\Pr \left\{ \left( \tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n \right) \in A_\epsilon^{(n)}(P_{X,Y,Z}) \right\}$  is

$$(1 - \epsilon) 2^{n(H(X,Y,Z) - H(X) - H(Y) - H(Z) - 4\epsilon)} \leq \Pr \left\{ \left( \tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n \right) \in A_\epsilon^{(n)}(P_{X,Y,Z}) \right\} \leq 2^{n(H(X,Y,Z) - H(X) - H(Y) - H(Z) + 4\epsilon)}$$