Homework 4 Simple Solution

TA: Heng-Chien Liou*

1. (Mutual information) [8]

a) How much information does the length of a sequence give about the content of a sequence? In this problem, let us consider a sequence of i.i.d. Ber(1/2) random variables $X_1, X_2, \ldots$.

Let N_0 be a random variable taking value at 6 with probability 1/3 and at 12 with probability 2/3. Furthermore, N_0 is independent of the sequence $\{X_i \mid i = 1, 2, \ldots\}$. Also let N_1 denote the length of the sequence when the first "1" appears. Obviously N_1 is also a random variable.

Compute $\mathrm{I}(N_0; X_1, X_2, \ldots, X_{N_0})$ and $\mathrm{I}(N_1; X_1, X_2, \ldots, X_{N_1})$. [4]

b) Consider a sequence of n binary random variables $(X_1, X_2, \ldots, X_n)$. Each sequence with an even number of 1's has probability $2^{-(n-1)}$, and each sequence with an odd number of 1's has probability 0. Compute the following:

$$\mathrm{I}(X_1; X_2),\ \mathrm{I}(X_2; X_3|X_1),\ \mathrm{I}(X_3; X_4|X_1, X_2),\ \ldots,\ \mathrm{I}(X_{n-1}; X_n|X_1, X_2, \ldots, X_{n-2}). \quad [4]$$

Solution:

a) By definition:

$$\begin{aligned}\mathrm{I}(N_0; X_1, X_2, \cdots, X_{N_0}) &= \mathrm{H}(N_0) - \mathrm{H}(N_0|X_1, X_2, \cdots, X_{N_0}) = \mathrm{H}(N_0)\\ &= \frac{1}{3}\log 3 + \frac{2}{3}\log\left(\frac{3}{2}\right) = \log 3 - \frac{2}{3}.\end{aligned}$$

Similarly, $\mathrm{I}(N_1; X_1, X_2, \cdots, X_{N_1}) = \mathrm{H}(N_1)$. We only need to calculate $\mathrm{H}(N_1)$.
By problem 4 b) of HW1, $\mathrm{H}(N_1) = 2$.

b) We will first show that $X_1, X_2, \cdots, X_{n-1}$ are mutually independent. Note that for all $x_1, \cdots, x_{n-1} \in \{0,1\}^{n-1}$,

$$\begin{aligned}\mathsf{Pr}\{X_1 = x_1, \cdots, X_{n-1} = x_{k-1}\} &= \mathsf{Pr}\{X_1 = x_1, \cdots, X_{n-1} = x_{k-1}, X_n = 0\}\\ &\quad + \mathsf{Pr}\{X_1 = x_1, \cdots, X_{n-1} = x_{k-1}, X_n = 1\}\\ &= 2^{-(n-1)}\end{aligned}$$

since either $x_1 \oplus \cdots \oplus x_{n-1} \oplus 0 = 0$ or $x_1 \oplus \cdots \oplus x_{n-1} \oplus 1 = 0$.

*with contribution by Chen-Hao Hsiao and Wen-Shao Ho

Therefore, for $\mathcal{I} \subseteq \{1, 2, \cdots, n-1\}$, by calculating the marginal distribution, we have

$$\Pr\left\{\bigcap_{i\in\mathcal{I}} X_i = x_i\right\} = \prod_{i\in\mathcal{I}} \Pr\{X_i = x_i\} = 2^{-|\mathcal{I}|}.$$

By the mutual independence derived above, for $k = 1, 2 \cdots, n-2$,

$$\begin{aligned}\mathrm{I}(X_k; X_{k+1}|X_1, \cdots, X_{k-1}) &= \mathrm{H}(X_k|X_1, \cdots, X_{k-1}) - \mathrm{H}(X_k|X_1, \cdots, X_{k-1}, X_{k+1}) \\ &= \mathrm{H}(X_k) - \mathrm{H}(X_k) = 0.\end{aligned}$$

Also,

$$\begin{aligned}\mathrm{I}(X_{n-1}; X_n|X_1, \cdots, X_{n-2}) &= \mathrm{H}(X_{n-1}|X_1, \cdots, X_{n-2}) - \mathrm{H}(X_{n-1}|X_1, \cdots, X_{n-2}, X_n) \\ &= \mathrm{H}(X_{n-1}) - \mathrm{H}(X_1 \oplus \cdots \oplus X_{n-2} \oplus X_n|X_1, \cdots, X_{n-2}, X_n) \\ &= 1 - 0 = 1.\end{aligned}$$

We use $\mathrm{H}(f(X)|X) = 0$ in the second equation.

2. (Data processing) [12]

a) Let $X_1 - X_2 - X_3 - X_4$ form a Markov chain. Prove that

$$\mathrm{I}(X_1; X_3) + \mathrm{I}(X_2; X_4) \le \mathrm{I}(X_1; X_4) + \mathrm{I}(X_2; X_3). \qquad [6]$$

b) Let $X_1 - X_2 - (X_3, X_4)$ form a Markov chain. Prove that

$$\mathrm{I}(X_1; X_3) + \mathrm{I}(X_1; X_4) \le \mathrm{I}(X_1; X_2) + \mathrm{I}(X_3; X_4). \qquad [6]$$

Solution:

a)

$$\begin{aligned}\mathrm{I}(X_1; X_3) + \mathrm{I}(X_2; X_4) &= \mathrm{I}(X_1; X_3) + \mathrm{I}(X_1, X_2; X_4) - \mathrm{I}(X_1; X_4|X_2) && \text{chain rule} \\ &= \mathrm{I}(X_1; X_3) + \mathrm{I}(X_1, X_2; X_4) && \text{markov} \\ &= \mathrm{I}(X_1; X_3) + \mathrm{I}(X_1; X_4) + \mathrm{I}(X_2; X_4|X_1) && \text{chain rule} \\ &\le \mathrm{I}(X_1; X_3) + \mathrm{I}(X_1; X_4) + \mathrm{I}(X_2; X_3|X_1) && \text{data processing} \\ &= \mathrm{I}(X_1; X_4) + \mathrm{I}(X_3; X_1, X_2) && \text{chain rule} \\ &= \mathrm{I}(X_1; X_4) + \mathrm{I}(X_2; X_3) && \text{markov}\end{aligned}$$

$$\begin{aligned}\mathrm{I}(X_1; X_3) + \mathrm{I}(X_1; X_4) &= \mathrm{I}(X_1; X_3, X_4) - \mathrm{I}(X_1; X_4|X_3) + \mathrm{I}(X_1; X_4) && \text{chain rule} \\ &\le \mathrm{I}(X_1; X_2) - \mathrm{I}(X_1; X_4|X_3) + \mathrm{I}(X_1; X_4) && \text{data processing} \\ &= \mathrm{I}(X_1; X_2) + \mathrm{I}(X_1; X_4)\end{aligned}$$

$$
\begin{aligned}
&- (\mathrm{I}(X_1;X_4) + \mathrm{I}(X_3;X_4|X_1) - \mathrm{I}(X_3;X_4)) && \text{chain rule} \\
&\le \mathrm{I}(X_1;X_2) + \mathrm{I}(X_3;X_4) && \text{nonnegative}
\end{aligned}
$$

3. (Sum Channel) [16]

Consider l DMC's

$$\left\{ (\mathcal{X}^{(i)}, \mathsf{P}^{(i)}_{Y|X}, \mathcal{Y}^{(i)}) \,\middle|\, i = 1, 2, \ldots, l \right\},$$

where DMC $(\mathcal{X}^{(i)}, \mathsf{P}^{(i)}_{Y|X}, \mathcal{Y}^{(i)})$ has channel capacity $\mathrm{C}^{(i)}$, for $1 \le i \le l$. The channel input alphabets are disjoint, and so are the channel output alphabets, that is,

$$\mathcal{X}^{(i)} \cap \mathcal{X}^{(j)} = \mathcal{Y}^{(i)} \cap \mathcal{Y}^{(j)} = \emptyset, \ \forall\, i \ne j.$$

The **sum channel** $(\mathcal{X}^{\oplus}, \mathsf{P}^{\oplus}_{Y|X}, \mathcal{Y}^{\oplus})$ associated to these channels is defined as follows:

- Input alphabet is the union $\mathcal{X}^{\oplus} := \cup_{i=1}^{l} \mathcal{X}^{(i)}$ of the individual input alphabets.
- Output alphabet is the union $\mathcal{Y}^{\oplus} := \cup_{i=1}^{l} \mathcal{Y}^{(i)}$ of the respective output alphabets.
- At each time slot the transmitter chooses to use *one and only one* of the l channels to transmit a symbol, that is,

$$\mathsf{P}^{\oplus}(y|x) := \begin{cases} \mathsf{P}^{(i)}_{Y|X}(y|x), & \text{if } x \in \mathcal{X}^{(i)} \text{ and } y \in \mathcal{Y}^{(i)} \\ 0, & \text{otherwise} \end{cases}$$

a) Introduce a random variable I indicating which DMC is used in the sum channel, that is,

$$I = i \quad \text{if } X \in \mathcal{X}^{(i)},\ i = 1, 2, ..., l.$$

Show that for the sum channel $\mathsf{P}^{\oplus}_{Y|X}$, $\mathrm{I}(X;Y) = \mathrm{I}(X;Y|I) + \mathrm{H}(I)$. [4]

b) Find the capacity of the sum channel in terms of $\{\mathrm{C}^{(i)} \mid i = 1, 2, \ldots, l\}$. [6]

c) Find the optimal input probability distribution for the sum channel in terms of the optimal input probability distributions for the individual channels. [6]

Solution:

- Let $X^{\oplus}$ be the r.v. of imput sum channel
- Let $Y^{\oplus}$ be the r.v. of output sum channel

$$
\begin{aligned}
\mathrm{C}^{\oplus} &= \max_{\mathsf{P}^{\oplus}_X} \mathrm{I}\left(X^{\oplus}; Y^{\oplus}\right) \\
&= \max_{\mathsf{P}^{\oplus}_X} \mathrm{I}\left(X^{\oplus}; Y^{\oplus}, I\right) && \mathcal{X}^{(i)} \text{ disjoint} \\
&= \max_{\mathsf{P}^{\oplus}_X} \left\{ \mathrm{I}\left(X^{\oplus}; I\right) + \mathrm{I}\left(X^{\oplus}; Y^{\oplus} \middle| I\right) \right\} && \text{chain rule}
\end{aligned}
$$

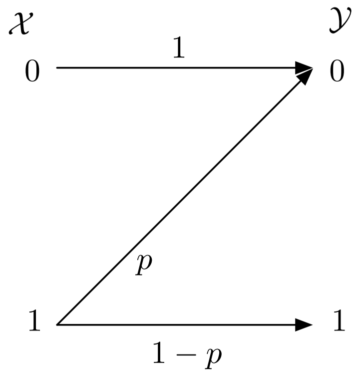
$$
\begin{aligned}
&= \max_{\mathsf{P}_X^{\oplus}} \left\{ \mathrm{H}(I) + \mathrm{I}\left(X^{\oplus}; Y^{\oplus} \middle| I\right) \right\} && I \text{ is deterministic of } X^{\oplus} \\
&= \max_{\mathsf{P}_X^{\oplus}} \sum_{i=1}^{l} \mathsf{P}_I(i) \left[\log \frac{1}{\mathsf{P}_I(i)} + \mathrm{I}\left(X^{\oplus}; Y^{\oplus} \middle| I = i\right) \right] \\
&= \max_{\mathsf{P}_I} \sum_{i=1}^{l} \mathsf{P}_I(i) \log \frac{2^{\mathrm{C}^{(i)}}}{\mathsf{P}_I(i)} \\
&= \log \sum_{i=1}^{l} 2^{\mathrm{C}^{(i)}} && \text{Jensen's inequality}
\end{aligned}
$$

The equality of the Jensen's inequality holds iff $\frac{P_I(i)}{2^{\mathrm{C}^{(i)}}} = \text{constant}, \forall i = 1, \ldots, l.$

Let $\mathsf{P}_X^{(i)}(x)$ be the optimal input probability distribution for the i-th channel. The optimal input probability distribution for the sum channel is

$$
\mathsf{P}_{X^{\oplus}}(x) = \sum_{i=1}^{l} \mathsf{Pr}\left(X^{(i)} = x \middle| I = i\right) P_I(i) = \frac{\sum_{i=1}^{l} 2^{\mathrm{C}^{(i)}} \mathsf{P}_X^{(i)}(x) \mathbb{1}\{I(x) = i\}}{\sum_{j=1}^{l} 2^{\mathrm{C}^{(j)}}}, \quad \forall x \in \mathcal{X}^{\oplus}.
$$

4. (Z channel) [14]



The Z channel (depicted above) is one of the simplest asymmetric channel with its channel law described as follows.

$$
\mathsf{P}_{Y|X} = \begin{bmatrix} 1 & 0 \\ p & 1-p \end{bmatrix}.
$$

In the following, let us assume $p = 1/2$.

a) Find the capacity of the Z channel and a capacity achieving input distribution P_X^*. Also find P_Y^*, the output distribution induced by the input distribution P_X^*. [8]

b) Is the capacity achieving input distribution of the Z channel unique? [2]

c) Recall that $\mathrm{C} = \mathrm{D}\left(\mathsf{P}_{Y|X} \middle\| \mathsf{P}_Y^* \middle| \mathsf{P}_X^*\right)$ and can be viewed as a weighted average of

$$
\left\{ \mathrm{D}\left(\mathsf{P}_{Y|X}(\cdot|a) \middle\| \mathsf{P}_Y^*(\cdot)\right) \,\middle|\, a \in \mathcal{X} \right\}.
$$

For the Z channel, derive $\mathrm{D}\left(\mathsf{P}_{Y|X}(\cdot|0) \middle\| \mathsf{P}_Y^*(\cdot)\right)$ and $\mathrm{D}\left(\mathsf{P}_{Y|X}(\cdot|1) \middle\| \mathsf{P}_Y^*(\cdot)\right)$. [4]

Solution:

a) Note that $\mathcal{X}$ is an alphabet of size 2, hence we can model any input distribution as $\mathsf{P}_X = \mathrm{Ber}(q), 0 \le q \le 1$. For given $X \sim \mathrm{Ber}(q)$, we then have $Y \sim \mathrm{Ber}\left((1-p)q\right)$.

$$\begin{aligned}
\mathrm{C}(q) &:= \mathrm{I}(X;Y) = \mathrm{H}(Y) - \mathrm{H}(Y|X) \\
&= \mathsf{h_b}\left((1-p)q\right) - \mathsf{P}_X(0)\mathrm{H}(Y|X=0) - \mathsf{P}_X(1)\mathrm{H}(Y|X=1) \\
&= \mathsf{h_b}\left((1-p)q\right) - q\mathsf{h_b}\left(p\right)
\end{aligned}$$

$\mathrm{C}(q)$ can be understood as a real function of single variable. The channel capacity then can be denoted as $\mathrm{C} = \max_{q\in[0,1]} \mathrm{C}(q)$, and can be approached by simple calculus:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}q}\mathrm{C}(q) &= \frac{\mathrm{d}}{\mathrm{d}q}\left[\mathsf{h_b}\left((1-p)q\right) - q\mathsf{h_b}\left(p\right)\right] \\
&= (1-p)\log\frac{1-(1-p)q}{(1-p)q} - \mathsf{h_b}\left(p\right) \\
\mathrm{C}(0) &= 0 \\
\mathrm{C}(1) &= \mathsf{h_b}\left(1-p\right) - \mathsf{h_b}\left(p\right)
\end{aligned}$$

by the extreme value theorem, if we restrict $p = 1/2$, it becomes clear that $q = 2/5$ is the unique maximizer. Hence

$$\mathrm{C} = \max_{\mathsf{P}_X} \mathrm{I}(X;Y) = \max_{q\in[0,1]} \mathrm{C}(q) = \mathrm{C}(2/5) = \mathsf{h_b}\left(1/5\right) - 2/5 = \log 5 - 2.$$

And $\mathsf{P}_X^* = \mathrm{Ber}(2/5)$ and $\mathsf{P}_Y^* = \mathrm{Ber}(1/5)$.

b) Yes, it is unique.

c)

$$\begin{aligned}
\mathrm{D}\left(\mathsf{P}_{Y|X}(\cdot|0)\middle\|\mathsf{P}_Y^*(\cdot)\right) &= 1\log\frac{1}{1-1/5} + 0\log\frac{0}{1/5} = \log 5 - 2 \\
\mathrm{D}\left(\mathsf{P}_{Y|X}(\cdot|1)\middle\|\mathsf{P}_Y^*(\cdot)\right) &= \frac{1}{2}\log\frac{1/2}{1-1/5} + \frac{1}{2}\log\frac{1/2}{1/5} = \log 5 - 2 \\
\mathrm{D}\left(\mathsf{P}_{Y|X}\middle\|\mathsf{P}_Y^*\middle|\mathsf{P}_X^*\right) &= \log 5 - 2
\end{aligned}$$