Solutions: Homework Set # 7

Problem 1

(a) Let us write the mutual information $I(X^n;Y^n)$ as follows

$$\begin{aligned} I(X^n;Y^n) &= I(X_1, X_2^n; Y^{n-1}, Y_n) \\ &= I(\underbrace{X_1}_{A}, \underbrace{Y^{n-1}}_{B}; \underbrace{Y^{n-1}}_{B}, \underbrace{Y_n}_{C}). \end{aligned}$$

We can write

$$\begin{aligned} I(A,B;B,C) &\overset{(1)}{=} I(A;BC) + I(B;BC|A) \\ &= I(A;BC) + H(B|A) + \underbrace{H(B|ABC)}_{0}, \end{aligned}$$

where (1) follows from chain rule. So we have

$$\begin{aligned} I(X^n;Y^n) &= I(X_1;Y^n) + H(Y^{n-1}|X_1) \\ &\overset{(1)}{=} \sum_{i=1}^{n} I(Y_i;X_1|Y^{i-1}) + \sum_{i=1}^{n} H(Y_i|\underbrace{X_1, Y^{i-1}}_{X^i}) \\ &\overset{(2)}{=} I(X_1;Y_1) + \sum_{i=1}^{n} H(Y_i|X_i) \\ &\overset{(3)}{=} I(X_1;Y_1) + nh_2(p), \end{aligned}$$

where
(1) follows by applying the chain rule for mutual information and conditional entropy;
(2) follows from the fact that $I(Y_i;X_1|Y^{i-1}) = 0$ for $1 < i \le n$, because we can write

$$\begin{aligned} I(Y_i;X_1|Y^{i-1}) &= H(Y_i|Y^{i-1}) - H(Y_i|X_1, Y^{i-1}) \\ &= H(Y_i|X_2^i) - H(Y_i|X^i) \\ &= H(Y_i|X_i) - H(Y_i|X_i) \\ &= 0; \end{aligned}$$

(3) follows from the fact that we have a binary symmetric channel so we have $H(Y_i|X_i) = h_2(p)$ where $h_2(p) \triangleq -p\log_2 p - (1-p)\log_2 1 - p$.

So we can write

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} I(X^n;Y^n) &= \lim_{n\to\infty} \frac{1}{n} \left[I(X_1;Y_1) + nh_2(p)\right] \\ &= h_2(p), \end{aligned}$$

because $I(X_1;Y_1) \le H(X_1) \le 1$.

(b) We know that the capacity of the binary symmetric channel is $1 - h_2(p)$ so there exist some value of p such that $h_2(p) > 1 - h_2(p)$ or equivalently $h_2(p) > 1/2$.

Note that $\lim_{n\to\infty} \frac{1}{n} I(X^n; Y^n)$ is not the rate at which this scheme can convey information from source to destination. To find the communication rate for this scheme we should find the mutual information between transmitted message W and received sequence of bits Y^n namely $\lim_{n\to\infty} \frac{1}{n} I(W; Y^n)$. This is in fact the quantity that we calculate in part (c). So part (a) does not imply that the capacity with feedback increases because the quantity that we calculate in (a) is not capacity.

(c) Note that we have the following Markov chain

$$W \leftrightarrow X_1 \leftrightarrow Y_1 \leftrightarrow \cdots \leftrightarrow Y_n,$$

or more compactly

$$W \leftrightarrow X_1 \leftrightarrow Y^n.$$

Using data processing inequality we can write

$$I(W; Y^n) \leq I(X_1; Y^n).$$

But we have calculated $I(X_1; Y^n)$ in part (a) where we obtained $I(X_1; Y^n) = I(X_1; Y_1)$. So we can write

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} I(W; Y^n) &\leq \lim_{n\to\infty} \frac{1}{n} I(X_1; Y^n) \\
&= \lim_{n\to\infty} \frac{1}{n} I(X_1; Y_1) \\
&\leq \lim_{n\to\infty} \frac{1}{n} H(X_1) \\
&\leq \lim_{n\to\infty} \frac{1}{n} \cdot 1 \\
&= 0.
\end{aligned}$$

So the actual capacity of this channel is 0.

Problem 2

(a) Let $p_{x_1 x_2}$ denote the probability of $X_1 = x_1$ and $X_2 = x_2$. then from definition:

$$\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= H(X_1, X_2) - H(X_1, X_2 | Y_1, Y_2) \\
&= H(X_1, X_2) \quad \text{since } X_1 \text{ and } X_2 \text{ can be known exactly from } Y_1 \text{ and } Y_2 \\
&= -p_{00} \log p_{00} - p_{01} \log p_{01} - p_{10} \log p_{10} - p_{11} \log p_{11}
\end{aligned}$$

(b)

$$\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= H(X_1, X_2) \\
&\leq \log_2(4)
\end{aligned}$$

where the equality holds for $p_{0,0} = p_{1,0} = p_{0,1} = p_{1,1} = \frac{1}{4}$.

(c) We assume in this part that $p_{0,0} = p_{1,0} = p_{0,1} = p_{1,1} = \frac{1}{4}$ and find $I(X_1; Y_1)$. To this end, let's first calculate $p_{Y_1}(y_1)$.

$$p_{Y_1}(0) = p_{Y_1,Y_2}(0,0) + p_{Y_1,Y_2}(0,1) = \frac{1}{2},$$

and

$$p_{Y_1}(1) = p_{Y_1,Y_2}(1,0) + p_{Y_1,Y_2}(1,1) = \frac{1}{2}.$$

Thus $H(Y_1) = 1$. Furthermore, $p_{Y_1|X_1}(y_1|x_1)$ is found to be:

$$\begin{aligned} p_{Y_1|X_1}(0|0) &= \frac{1}{2} \\ p_{Y_1|X_1}(1|0) &= \frac{1}{2} \\ p_{Y_1|X_1}(0|1) &= \frac{1}{2} \\ p_{Y_1|X_1}(1|1) &= \frac{1}{2}. \end{aligned}$$

Thus $H(Y_1|X_1) = \frac{1}{2} + \frac{1}{2} = 1$. So finally,

$$\begin{aligned} I(X_1; Y_1) &= H(Y_1) - H(Y_1|X_1) \\ &= 1 - 1 \\ &= 0. \end{aligned}$$

Thus, the distribution on the input sequence that achieves capacity does not necessarily maximize the mutual information between individual symbols and their corresponding outputs. What to note here is that although the knowledge of Y_1 and Y_2 together, leaves no uncertainty in X_1 and X_2, the knowledge of Y_1 alone give no information about X_1. Just to notice this yourself, if you have received only $Y_1 = 0$, can you decide if $X_1 = 0$ was sent or $X_2 = 0$? In fact, under the proposed uniform distribution of $p_{0,0} = p_{1,0} = p_{0,1} = p_{1,1} = \frac{1}{4}$, given Y_1 the events $X_1 = 0$ and $X_2 = 0$ would be equally likely.

Problem 3

Since we know the Fano's inequality for random variables (not the random sequences) we start by relating $H(X^k|Y_k)$ to $H(X_i|Y_i)$:

$$\begin{aligned} H(X^k|Y^k) &= \sum_{i=1}^{k} H(X_i|Y^k X_1^{i-1}) \\ &\leq \sum_{i=1}^{k} H(X_i|Y_i) \quad \text{since conditioning reduces entropy} \\ &\leq \sum_{i=1}^{k} p_i(e)\log(|\mathcal{X}|) + H(p_i(e)) \quad \text{Fano's inequality for } H(X_i|Y_i) \end{aligned}$$

where $p_i(e) = Pr(X_i \neq Y_i)$.
Let's write

$$p_i(e) = Pr(X_i \neq Y_i) = \mathbb{E}1_{X_i \neq Y_i}$$

(This is true by definition:

$$\mathbb{E}1_{X_i \neq Y_i} = Pr(X_i \neq Y_i) \times 1 + Pr(X_i = Y_i) \times 0 = Pr(X_i \neq Y_i) = p_i(e).)$$

So

$$\begin{aligned} H(X^k|Y^k) &= \sum_{i=1}^{k} p_i(e) \log(|\mathcal{X}|) + H(p_i(e)) \\ &= \sum_{i=1}^{k} \mathbb{E}1_{X_i \neq Y_i} \log(|\mathcal{X}|) + \sum_{i=1}^{k} H(\mathbb{E}1_{X_i \neq Y_i}) \end{aligned}$$

To find the expression of the question we show that

- $\sum_{i=1}^{k} \mathbb{E}1_{X_i \neq Y_i} = \mathbb{E} \sum_{i=1}^{k} 1_{X_i \neq Y_i} = \mathbb{E} d_H(X^k, Y^k)$ (By definition of Hamming distance)
- $\sum_{i=1}^{k} H(\mathbb{E}1_{X_i \neq Y_i}) \leq kH(\frac{1}{k}\mathbb{E} d_H(X^k, Y_k))$

$$\begin{aligned} \sum_{i=1}^{k} H(\mathbb{E}1_{X_i \neq Y_i}) &= k \sum_{i=1}^{k} \frac{1}{k} H(\mathbb{E}1_{X_i \neq Y_i}) \\ &\leq kH(\sum_{i=1}^{k} \frac{1}{k} \mathbb{E}1_{X_i \neq Y_i}) \\ &= kH(\frac{1}{k}\mathbb{E} d_H(X^k, Y^k)) \end{aligned}$$

Problem 4

(a) Let us assume

$$\max_{x \in \mathcal{X}} \frac{D(W_{Y|X}||P_Y)}{c(x)} = T.$$

This means that

$$\forall x \in \mathcal{X} \ : \quad D(W_{Y|X}||P_Y) \leq Tc(x).$$

Then we have

$$\frac{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||P_Y)}{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) c(x)} \leq \frac{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) Tc(x)}{\sum_{x \in \mathcal{X}} \tilde{P}_X(x) c(x)} = T.$$

(b) We have

$$
\begin{aligned}
&\sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||P_Y) - \sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||\tilde{P}_Y) \\
=&\sum_{x\in\mathcal{X}} \tilde{P}_X(x) \left[D(W_{Y|X}||P_Y) - D(W_{Y|X}||\tilde{P}_Y)\right] \\
=&\sum_{x\in\mathcal{X}} \tilde{P}_X(x) \left[\sum_{y\in\mathcal{Y}} W_{Y|X}(y|x) \left[\log \frac{W_{Y|X}(y|x)}{P_Y(y)} - \log \frac{W_{Y|X}(y|x)}{\tilde{P}_Y(y)}\right]\right] \\
=&\sum_{x\in\mathcal{X}} \tilde{P}_X(x) \sum_{y\in\mathcal{Y}} W_{Y|X}(y|x) \left[\log \frac{\tilde{P}_Y(y)}{P_Y(y)}\right] \\
=&\sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} \tilde{P}_X(x) W_{Y|X}(y|x) \left[\log \frac{\tilde{P}_Y(y)}{P_Y(y)}\right] \\
=&\sum_{y\in\mathcal{Y}} \log \frac{\tilde{P}_Y(y)}{P_Y(y)} \sum_{x\in\mathcal{X}} \tilde{P}_X(x) W_{Y|X}(y|x) \\
=&\sum_{y\in\mathcal{Y}} \tilde{P}_Y(y) \log \frac{\tilde{P}_Y(y)}{P_Y(y)} = D(\tilde{P}_Y||P_Y) \\
\geq& 0.
\end{aligned}
$$

The second part can be proved easily bu replacing the above inequality in the result of part (a).

(c) It is easy to see that

$$\max_{x\in\mathcal{X}} \frac{D(W_{Y|X}||P_Y^*)}{c(x)} \leq \lambda.$$

Using part (b) we have

$$\frac{\sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||\tilde{P}_Y)}{\sum_{x\in\mathcal{X}} \tilde{P}_X(x) c(x)} \leq \max_{x\in\mathcal{X}} \frac{D(W_{Y|X}||P_Y^*)}{c(x)} \leq \lambda.$$

Having the equality, we conclude that

$$\frac{\sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||\tilde{P}_Y)}{\sum_{x\in\mathcal{X}} \tilde{P}_X(x) c(x)} = \frac{\sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||P_Y^*)}{\sum_{x\in\mathcal{X}} \tilde{P}_X(x) c(x)} = \lambda.$$

Thus

$$\sum \tilde{P}_X(x) D(W_{Y|X}||\tilde{P}_Y) - \sum P_X^*(x) D(W_{Y|X}||P_Y^*) = -D(\tilde{P}_Y||P_Y^*) = 0,$$

so we deduce that $\tilde{P}_Y = P_Y^*$.

(d) We have

$$\frac{I(X;Y)}{\mathbb{E}[c(x)]} = \frac{\sum_{x\in\mathcal{X}} P_X(x) D(W_{Y|X}||P_Y)}{\sum_{x\in\mathcal{X}} P_X(x) c(x)} \leq \max_{x\in\mathcal{X}} \frac{D(W_{Y|X}||P_Y)}{c(x)}.$$

By part (c), the distribution that maximizes the capacity per unit cost, C_{cost}, is $P^*(x)$ characterized by

$$\frac{D(W_{Y|X}||P_Y^*)}{c(x)} \leq \lambda, \quad \forall x \in \mathcal{X},$$

and

$$\frac{D(W_{Y|X}||P_Y^*)}{c(x)} = \lambda, \quad \forall x: \; P_X^*(x) > 0,$$

where $P_Y^* = \sum_x P_X^*(x) W_{Y|X}(y|x)$.