ECE 5630 - Solutions Homework Assignment 5

1) See Lecture 17.

2) a) Consider the transition kernel $P_{Z|XY}$ described by $Z = \mathbb{1}_{\{X=Y\}}$. Let P_Z and Q_Z be the transformation of P_{XY} and Q_{XY}, respectively, when pushed through $P_{Z|XY}$. We then have that $P_Z = \mathsf{Bern}(p)$ and $Q_Z = \mathsf{Bern}(q)$. Using the f-divergence DPI, we get

$$I(X;Y) = D_{\mathsf{KL}}(P_{XY}\|Q_{XY}) \geq D_{\mathsf{KL}}(P_Z\|Q_Z) = D_{\mathsf{KL}}\big(\mathsf{Bern}(p)\|\mathsf{Bern}(q)\big).$$

Then,

$$\begin{aligned} I(X;Y) &\geq (1-p)\log\left(\frac{1-p}{1-q}\right) + p\log\left(\frac{p}{q}\right) \\ &= (1-p)\log\left(\frac{1}{1-q}\right) + p\log\left(\frac{1}{q}\right) - H_b(p) \\ &\geq p\log\left(\frac{1}{q}\right) - H_b(p), \end{aligned}$$

where the last inequality follows from the fact that $(1-p)\log(1/(1-q)) \geq 0$.

b) We have $H(X) = \log|\mathcal{X}|$. Then, from part (a), we get

$$H(X|Y) \leq \log|\mathcal{X}| + H_b(p) - p\log\left(\frac{1}{q}\right).$$

Using the total probability theorem, we get

$$q = Q_{XY}(X=Y) = \sum_{y\in\mathcal{X}} Q_{XY}(X=Y, Y=y) = \sum_{y\in\mathcal{X}} P_X(X=y)P_Y(Y=y) = \frac{1}{|\mathcal{X}|}\sum_{y\in\mathcal{X}} P_Y(Y=y) = \frac{1}{|\mathcal{X}|}.$$

Then,

$$H(X|Y) \leq (1-p)\log|\mathcal{X}| + H_b(p) = P_{XY}(X\neq Y)\log|\mathcal{X}| + H_b\left(P_{XY}(X\neq Y)\right),$$

where the last equality follows from the fact that $H_b(p) = H_b(1-p)$.

3) a) The predictor with minimum probability of error is $\widehat{X} = 1$ for which $P_e = 1 - p_1$.

b) We have

$$\begin{aligned} H(P) &= -p_1\log(p_1) - \sum_{i=2}^{m} p_i\log(p_i) \\ &= -p_1\log(p_1) - P_e\log(P_e) - P_e\sum_{i=2}^{m}\frac{p_i}{P_e}\log\left(\frac{p_i}{P_e}\right) \\ &= H(P_e) - P_e\sum_{i=2}^{m}\frac{p_i}{P_e}\log\left(\frac{p_i}{P_e}\right). \end{aligned}$$

Note that $p_2/P_e, \dots, p_m/P_e$ is a valid PMF on an alphabet of size $m-1$. So,

$$H(P) \leq H(P_e) - P_e \log(m-1) \tag{1}$$

with equality attained by $p_i = (1-p_1)/(m-1)$ for $i = 2, \dots, m$. Note that from the condition $p_1 \geq p_2 \dots \geq p_m$, we get $p_1 \geq 1/m$. Thus, the predictor with the minimum probability of error is still $\widehat{X} = 1$ as $p_1 \geq (1-p_1)/(m-1)$, and the probability of error remains unchanged.

c) We have that $H(P_e) \leq 1$. Then, using Inequality (1), we get $H(X) \leq 1 + P_e \log(m-1)$. Thus,

$$P_e \geq \frac{H(X) - 1}{\log(m-1)}.$$

4) a) Maximum capacity is $C = \log(5)$ bits. Achieved by taking $\mathcal{Z} = \{10, 20, 30\}$ and $P_X = \mathsf{Unif}(\mathcal{X})$.

b) We have that $H(Y|X) = H(Z) = \log(3)$. Thus,

$$C = \max_{P_X}(H(Y) - H(Y|X)) = \max_{P_X} H(Y) - \log(3).$$

Given that $Z \sim \mathsf{Unif}(\mathcal{Z})$, we have $\max_{P_X} H(Y) = \max_{P_Y} H(Y) = \log|\mathcal{Y}|$. A choice of $\mathcal{Z}$ that minimizes $|\mathcal{Y}|$ is given by $\mathcal{Z} = \{0, 1, 2\}$ for which $\mathcal{Y} = \{0, 1, 2, 3, 4, 5\}$ and, thus, $C = \log(6) - \log(3) = 1$. The distribution P_X that attains the capacity has the PMF $p_X = (1/2, 0, 0, 1/2)$.

5) Sections (a) and (b), see solution to Problem 10 in Chapter 2 of T. M. Cover and J. A. Thomas “Elements of Information Theory”, 2nd Edition, Wiley, NY, US, 2003. For Sections (c) and (d), consider the following.

c) Let $B \sim \mathsf{Ber}(\alpha)$ and define $\Theta = B + 1$ as a random variable with alphabet $\{1, 2\}$ that indicats which of the two channels is used. Let the channel input be $X := (\Theta, X_\Theta)$. Since the output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are disjoint, Θ is a function of Y, i.e. $X \leftrightarrow Y \leftrightarrow \Theta$.

Therefore,

$$\begin{aligned}
I(X;Y) &= I(X;Y,\Theta) \\
&= I(X_\Theta,\Theta;Y,\Theta) \\
&= I(\Theta;Y,\Theta) + I(X_\Theta;Y,\Theta|\Theta) \\
&= I(\Theta;Y,\Theta) + I(X_\Theta;Y|\Theta) \\
&= H(\Theta) + \alpha I(X_\Theta;Y|\Theta=1) + (1-\alpha) I(X_\Theta;Y|\Theta=2) \\
&= H_b(\alpha) + \alpha I(X_1;Y_1) + (1-\alpha) I(X_2;Y_2).
\end{aligned}$$

Thus, it follows that

$$C = \sup_{\alpha\in[0,1]} \Big[H_b(\alpha) + \alpha C_1 + (1-\alpha) C_2\Big],$$

which is a strictly concave function on α. Hence, the maximum exists and by elementary calculus, one can easily show $C = \log_2(2^{C_1} + 2^{C_2})$, which is attained with $\alpha = 2^{C_1}/(2^{C_1} + 2^{C_2})$.

If one interprets $M = 2^C$ as the effective number of noise free symbols, then the above result follows in a rather intuitive manner: we have $M_1 = 2^{C_1}$ noise free symbols from channel 1, and $M_2 = 2^{C_2}$ noise free symbols from channel 2. Since at each step we get to choose which channel to use, we essentially have $M_1 + M_2 = 2^{C_1} + 2^{C_2}$ noise free symbols for the new channel. Therefore, the capacity of this channel is $C = \log_2(2^{C_1} + 2^{C_2})$.

d) From part (b) we get that the capacity is $\log(2^{1-H(p)} + 2^0)$.

6) a) Since $X \leftrightarrow Y \leftrightarrow \tilde{Y}$ forms a Markov chain, we can apply the data processing inequality. Hence for every input variable $X \sim P$, we have $I(X;Y) \geq I(X;\tilde{Y})$. Let $\tilde{X} \sim \tilde{P}$ be the capacity achieving input variable (and distribution) for the channel $P_{\tilde{Y}|X}$, i.e., $\max_{X\sim P} I(X;\tilde{Y}) = I(\tilde{X};\tilde{Y})$. Then

$$C = \max_{X\sim P} I(X;Y) \geq I(\tilde{X};Y) \geq I(\tilde{X};\tilde{Y}) = \max_{X\sim P} I(X;\tilde{Y}) = \tilde{C}.$$

Thus, processing the output does not increase capacity.

b) We have equality (no decrease in capacity) in the above sequence of inequalities only if we have equality in data processing inequality, i.e., for the distribution that maximizes $I(X;\tilde{Y})$, we have $X \leftrightarrow \tilde{Y} \leftrightarrow Y$ forming a Markov chain. In other words, $\tilde{Y} = g(Y)$ should be a sufficient statistic for X.

7) a) $C_1 = 1 - H_b(\lambda_1)$. See Lecture 10 for the derivation of the capacity of the binary symmetric channel.

b) $C_2 = 1 - H_b(\lambda_2)$.

c) One can observe that $Q_{Z|X}$ is also a BSC with transition probability $\lambda_1 * \lambda_2 = \lambda_1(1-\lambda_2) + (1-\lambda_1)\lambda_2$. Thus, $C_3 = 1 - H_b(\lambda_1 * \lambda_2)$. Moreover, we have $X - Y - Z$ form a Markov chain. Then, for all $P_X \in \mathcal{P}(\mathcal{X})$, we have

$$I(X;Z) \leq I(X;Y,Z) = I(X;Y).$$

Thus, $C_3 \leq C_1$. Similarly, we get $C_3 \leq C_2$.

d) In this case, $C_3 = \min\{C_1, C_2\}$ as only one of the channels would be the bottleneck between X and Z.

e) Since $X - Y - Z$, then $(X;Y,Z) = I(X;Y)$. If the receiver can view both Y and Z, the capacity is $\max_{P_X} I(X;Y,Z) = \max_{P_X} I(X;Y) = C_1$.

8) To find the capacity of the product channel $(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}_1 \times \mathcal{Y}_2, P_{Y_1,Y_2|X_1,X_2})$, we need to find the distribution $P_{X_1,X_2} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ that maximizes $I(X_1,X_2;Y_1,Y_2)$. Since the transition kernel factors as $P_{Y_1,Y_2|X_1,X_2} = P_{Y_1|X_1}P_{Y_2|X_2}$, the joint distribution will be

$$P_{X_1,X_2,Y_1,Y_2} = P_{X_1,X_2}P_{Y_1|X_1}P_{Y_2|X_2}.$$

The above structure implies that $Y_1 \leftrightarrow X_1 \leftrightarrow X_2 \leftrightarrow Y_2$ forms a Markov chain and

$$\begin{aligned}
I(X_1,X_2;Y_1,Y_2) &= H(Y_1,Y_2) - H(Y_1,Y_2|X_1,X_2) \\
&\overset{(a)}{=} H(Y_1,Y_2) - H(Y_1|X_1,X_2) - H(Y_2|X_1,X_2) \\
&\overset{(b)}{=} H(Y_1,Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\
&\overset{(c)}{\leq} H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\
&= I(X_1;Y_1) + I(X_2;Y_2),
\end{aligned}$$

where (a) and (b) follow from Markovity, while (c) is met with equality if X_1 and X_2, which makes Y_1 and Y_2 independent. Therefore

$$\begin{aligned} C &= \max_{P_{X_1,X_2}} I(X_1, X_2; Y_1, Y_2) \\ &\leq \max_{P_{X_1,X_2}} I(X_1;Y_1) + \max_{P_{X_1,X_2}} I(X_2;Y_2) \\ &= \max_{P_{X_1}} I(X_1;Y_1) + \max_{P_{X_2}} I(X_2;Y_2) \\ &= C_1 + C_2. \end{aligned}$$

with equality if and only if $P_{X_1,X_2} = P^\star_{X_1} \otimes P^\star_{X_2}$, where $P^\star_{X_1}$ and $P^\star_{X_2}$, respectively, the capacity achieving distributions for C_1 and C_2.

9) a) See Figure 1.

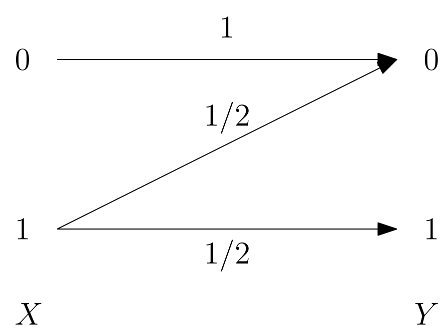

Fig. 1: Z-Channel

b) Let $P_X = \mathsf{Bern}(p)$. Then, $P_Y = \mathsf{Bern}(p/2)$ and $H(Y) = H_b(p/2)$. Also,

$$H(Y|X) = \mathbf{P}\left(X=0\right) H(Y|X=0) + \mathbf{P}\left(X=1\right) H(Y|X=1) = pH_b(1/2) = p.$$

Thus, we have $I(X;Y) = H(Y) - H(Y|X) = H_b(p/2) - p$. We then have that $I(X;Y)$ is concave in p. So we find the capacity of the channel, using the first-order optimality condition, i.e., $\mathsf{d}(H_b(p/2) - p)/\mathsf{d}p = 0$. That is, for $p^* := \operatorname{argmax}_{p \in [0,1]} I(X;Y)$, we have

$$\log\left(\frac{1 - p^*/2}{p^*/2}\right) = 2.$$

Thus, $p^* = 2/5$. Then, the capacity of the Z-channel is $C = \max_{P_X} I(X;Y) = H_b(1/5) - 2/5 = 0.322$, which is achieved by $P_X = \mathsf{Bern}(2/5)$.