Solution for Moed A 2014

1) **True or False**

a) False. For example:

$Y\backslash X$	0	1
0	0.25	0.25
1	0.5	0

For this distribution we can calculate $H_b(0.25) = H(X) < H(X|Y=0) = 1$.

b) True. For the converse, if multiplying by a would increase the capacity then we would use it in the point to point channel. To achieve, we divide by a and apply the decoding procedure as in the point to point channel.

c) True. This can be done as in the previous question. Another approach is by noting that the function $(\cdot)^3$ is a bijective function and therefore by having Y we indeed can recover $\tilde{Y}$ and:

$$\begin{aligned}
I(X;\tilde{Y}) &= H(\tilde{Y}) - H(\tilde{Y}|X) \\
&= H(\tilde{Y},Y) - H(Y,\tilde{Y}|X) \\
&= H(Y) - H(Y|X) \\
&= I(X;Y)
\end{aligned}$$

and thus $\max_{p(x)} I(X;Y) = \max_{p(x)} I(X;\tilde{Y}) = \frac{1}{2}\log(1+SNR)$.

2) **Cascaded Additive modulo-2 with dependent noise**

a) The capacity of the first channel is straightforward and equal to $C_1 = 1 - H_b(\epsilon)$.
To calculate the capacity C_2 between X and Y_2 we note that the noise between X and Y_2 is $N_1 + N_2$. Now, we have a simple point to point channel with this noise. The distribution of the noise is $N_1 + N_2 \sim Bern(\epsilon\bar{\alpha_1} + \bar{\epsilon}\alpha_2)$. Therefore, the capacity is:

$$C_2 = 1 - H_b(\epsilon\bar{\alpha_1} + \bar{\epsilon}\alpha_2)$$

b) All relations can occur, here are some examples:

i) For $C_1 = C_2$, substitute $\epsilon = 0, \alpha_2 = 0$. It then follows that $C_1 = C_2 = 1$. We are trying to produce two clean channels. The first channel is by $\epsilon = 0$ and therefore $Y_1 = X$. Now when N_1 is equal to zero we only need to make sure that N_2 does not change C_2. This is done by setting $\alpha_2 = 0$. (α_1 can take any value since $N_2 \neq 1$)

ii) For $C_1 < C_2$, substitute $\epsilon = \frac{1}{3}, \alpha_1 = 1, \alpha_2 = 0$. It then follows that $C_1 = 1 - H_b(\frac{1}{3})$ and $C_2 = 1$. The idea is to correct the noise that N_1 induces. We make sure that $N_1 + N_2 = 0$ and therefore $C_2 = 1$. ϵ can take any value different from zero.

iii) For $C_1 > C_2$, substitute $\epsilon = \frac{1}{3}, \alpha_1 = \frac{1}{2}, \alpha_2 = \frac{1}{2}$. It then follows that $C_1 = 1 - H_b(\frac{1}{3})$and $C_2 = 0$. This case is trivial, there are many examples to achieve this.

c) We begin with calculating the capacity of the third user, $C_3 = \max_{p(x)} I(X; Y_1, Y_2)$:

$$
\begin{aligned}
I(X; Y_1, Y_2) &= H(Y_2, Y_1) - H(Y_2, Y_1|X) && (1)\\
&= H(Y_2, Y_1) - H(Y_2|Y_1, X) - H(Y_1|X) && (2)\\
&= H(Y_2, Y_1) - H(Y_2|Y_1, X, N_1) - H(N_1) && (3)\\
&= H(Y_2, Y_1) - H(X + N_1 + N_2|Y_1, X, N_1) - H(N_1) && (4)\\
&= H(Y_2, Y_1) - H(N_2|Y_1, X, N_1) - H(N_1) && (5)\\
&= H(Y_2, Y_1) - H(N_2|N_1) - H(N_1) && (6)
\end{aligned}
$$

Symmetry in $H(Y_2, Y_1)$ applies that the maximum is attained for $p(x = 0) = 0.5$. Then it follows immediately that $H(Y_1) = 1$, and $H(Y_2|Y_1) = 0.5H(Y_2|Y_1 = 0) + 0.5H(Y_2|Y_1 = 1)$. We try to minimize C_1 by setting $\epsilon = 0.5$ and therefore $C_1 = 0$.
We calculate the term $H(Y_2|Y_1)$ for $\epsilon = 0.5$:

$$H(Y_2|Y_1) = 0.5H_b(0.5(\alpha_1 + \alpha_2)) + 0.5H_b(0.5(\bar{\alpha_1} + \bar{\alpha_2}).$$

Note that at this point C_1 is fixed to zero and the only parameters left are α_1 and α_2. Let's write the terms:

$$
\begin{aligned}
&C_1 = 1 - H_b(0.5) = 0 && (7)\\
&C_2 = 1 - H_b(0.5(\bar{\alpha_1} + \alpha_2)) && (8)\\
&C_3 = 1 + 0.5H_b(0.5(\alpha_1 + \alpha_2)) + 0.5H_b(0.5(\bar{\alpha_1} + \bar{\alpha_2})) - 0.5(H_b(\alpha_1) + H_b(\alpha_2)) - H_b(0.5) && (9)
\end{aligned}
$$

By choosing $\alpha_1 = 0$ and $\alpha_2 = 0.5$, we have that $C_3 > C_2$. There are many more examples.

3) **Network Coding for Broadcast Channel**

a) The minimum transmissions to node i is $k - |\mathcal{M}_i|$. The minimum number of transmission required for the whole system is $N_{trans} = \max\{k - |\mathcal{M}_1|, k - |\mathcal{M}_2|\}$.

b) To achieve this bound we first transmit all messages in the set $\mathcal{M}_1^c \bigcap \mathcal{M}_2^c$ directly to the receivers. Then, we transmit linear combinations of the missing parts. For example, assume that receiver 1 has M_1 and receiver 2 has M_2, then the transmitter will send a linear combination of (M_1, M_2). One should verify that by having all combinations, both receivers can recover the set of the messages $\mathcal{M}$. It is sufficient to work in bitwise XOR over the field $\mathbb{F}_2$, since $\mathcal{M}_i$ is known at the transmitter for each i.

c) From the same arguments, $N_{trans} = \max_i\{k - |\mathcal{M}_i|\}$.

d) The answer is the same as in subsection b only now the algorithm is preformed over τ number of sets.

4) **Conditional Information Divergence**

a) True or False. All arguments are true.

i)

$$\begin{aligned}
D\big(P_{A,B}\big|\big|Q_{A,B}\big) &= \sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} P_{A,B}(a,b)\log\left(\frac{P_{A,B}(a,b)}{Q_{A,B}(a,b)}\right) \\
&= \sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} P_{A,B}(a,b)\log\left(\frac{P_{B|A}(b|a)P_A(a)}{Q_{B|A}(b|a)Q_A(a)}\right) \\
&= \sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} P_{A,B}(a,b)\left[\log\left(\frac{P_{B|A}(b|a)}{Q_{B|A}(b|a)}\right)+\log\left(\frac{P_A(a)}{Q_A(a)}\right)\right] \\
&= D\big(P_{B|A}\big|\big|Q_{B|A}|P_A\big) + \sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} P_{A,B}(a,b)\log\left(\frac{P_A(a)}{Q_A(a)}\right) \\
&= D\big(P_{B|A}\big|\big|Q_{B|A}|P_A\big) + \sum_{a\in\mathcal{A}} P_A(a)\log\left(\frac{P_A(a)}{Q_A(a)}\right)\sum_{b\in\mathcal{B}} P_{B|A}(b|a) \\
&= D\big(P_{B|A}\big|\big|Q_{B|A}|P_A\big) + \sum_{a\in\mathcal{A}} P_A(a)\log\left(\frac{P_A(a)}{Q_A(a)}\right) \\
&= D\big(P_{B|A}\big|\big|Q_{B|A}|P_A\big) + D\big(P_A\big|\big|Q_A\big)
\end{aligned}$$

ii)

$$\begin{aligned}
D\big(P_{A,B}\big|\big|P_AP_B\big) &= \sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} P_{A,B}(a,b)\log\left(\frac{P_{A,B}(a,b)}{P_A(a)P_B(b)}\right) \\
&= \sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} P_{A,B}(a,b)\log\left(\frac{P_{B|A}(b|a)P_A(a)}{P_A(a)P_B(b)}\right) \\
&= \sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} P_{A,B}(a,b)\log\left(\frac{P_{B|A}(b|a)}{P_B(b)}\right) \\
&= D\big(P_{B|A}\big|\big|P_B|P_A\big)
\end{aligned}$$

iii) This follows from the known relation $I(X;Y) = D\big(P_{A,B}\big|\big|P_AP_B\big)\big)$.

iv)

$$\begin{aligned}
D\big(P_{A|B}\big|\big|Q_{A|B}|P_B\big) &= \sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} P_{A,B}(a,b)\log\left(\frac{P_{A|B}(a,b)}{Q_{A|B}(a,b)}\right) \\
&= \sum_{b\in\mathcal{B}} P_B(b)\sum_{a\in\mathcal{A}} P_{A|B}(a|b)\log\left(\frac{P_{A|B}(a,b)}{Q_{A|B}(a,b)}\right) \\
&= \sum_{b\in\mathcal{B}} P_B(b) D\big(P_{A|B=b}\big|\big|Q_{A|B=b}\big)
\end{aligned}$$

For the last part we use Shannon-Fano code which is known to be optimal for a dyadic distribution.

a) Using Shannon-Fano code, and multiplexing by Z, we have

$$\begin{aligned} L &= \sum_{z \in \mathcal{Z}} p(z) \sum_{x \in \mathcal{X}} p(x|z) \log_2 \frac{1}{p(x|z)} \\ &= \sum_{(x,z) \in \mathcal{X} \times \mathcal{Z}} p(x,z) \log_2 \frac{1}{p(x|z)} \\ &= H(X|Z) \end{aligned}$$

b) Since we encode with $Q_{X|Z}$, the average length is:

$$\begin{aligned} L &= \sum_{z \in \mathcal{Z}} p(z) \sum_{x \in \mathcal{X}} p(x|z) \log_2 \frac{1}{q(x|z)} \\ &= \sum_{(x,z) \in \mathcal{X} \times \mathcal{Z}} p(x,z) \log_2 \frac{1}{q(x|z)} \\ &= \sum_{(x,z) \in \mathcal{X} \times \mathcal{Z}} p(x,z) \log_2 \frac{p(x|z)}{q(x|z)} - \sum_{(x,z) \in \mathcal{X} \times \mathcal{Z}} p(x,z) \log_2 p(x|z) \\ &= D\big(P_{X|Z} \big|\big| Q_{X|Z} \big| P_Z\big) + H(X|Z) \end{aligned}$$

c) The answer is the same as in the previous question since we still encode with $Q_{X|Z}$.