École Polytechnique Fédérale de Lausanne
School of Computer and Communication Sciences

Handout 21 Information Theory and Coding
Solutions to Homework 8 Nov. 16, 2015

Problem 1.

$$Y_i = X_i \oplus Z_i,$$

where

$$Z_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

and Z_i are not necessarily independent.

$$\begin{aligned} I(X_1,\dots,X_n;Y_1,\dots,Y_n) &= H(X_1,\dots,X_n) - H(X_1,\dots,X_n|Y_1,\dots,Y_n) \\ &= H(X_1,\dots,X_n) - H(Z_1,\dots,Z_n|Y_1,\dots,Y_n) \\ &\geq H(X_1,\dots,X_n) - H(Z_1,\dots,Z_n) \\ &\geq H(X_1,\dots,X_n) - \sum H(Z_i) \\ &= H(X_1,\dots,X_n) - nH(p) \\ &= n - nH(p), \end{aligned}$$

if $X_1,\dots,X_n$ are chosen i.i.d. $\sim$ Bern(1/2). The capacity of the channel with memory over n uses of the channel is

$$\begin{aligned} nC^{(n)} &= \max_{p(x_1,\dots,x_n)} I(X_1,\dots,X_n;Y_1,\dots,Y_n) \\ &\geq I(X_1,\dots,X_n;Y_1,\dots,Y_n)_{p(x_1,\dots,x_n)=\text{Bern}(1/2)} \\ &\geq n(1-H(p)) \\ &= nC. \end{aligned}$$

Hence channels with memory have higher capacity. The intuitive explanation for this result is that the correlation between the noise decreases the effective noise; one could use the information from the past samples of the noise to combat the present noise.

Problem 2. To find the capacity of the product channel, we must find the distribution $p(x_1,x_2)$ on the input alphabet $\mathcal{X}_1 \times \mathcal{X}_2$ that maximizes $I(X_1,X_2;Y_1,Y_2)$. Since the joint distribution

$$p(x_1,x_2,y_1,y_2) = p(x_1,x_2)p(y_1|x_1)p(y_2|x_2),$$

$Y_1 \to X_1 \to X_2 \to Y_2$ forms a Markov chain and therefore

$$\begin{aligned} I(X_1,X_2;Y_1,Y_2) &= H(Y_1,Y_2) - H(Y_1,Y_2|X_1,X_2) && (1) \\ &= H(Y_1,Y_2) - H(Y_1|X_1,X_2) - H(Y_2|X_1,X_2) && (2) \\ &= H(Y_1,Y_2) - H(Y_1|X_1) - H(Y_2|X_2) && (3) \\ &\leq H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) && (4) \\ &= I(X_1;Y_1) + I(X_2;Y_2), && (5) \end{aligned}$$

where (2) and (3) follow from Markovity, and we have equality in (4) if Y_1 and Y_2 are independent. Equality occurs when X_1 and X_2 are independent. Hence

$$\begin{aligned} C &= \max_{p(x_1,x_2)} I(X_1, X_2; Y_1, Y_2) \\ &\leq \max_{p(x_1,x_2)} I(X_1; Y_1) + \max_{p(x_1,x_2)} I(X_2; Y_2) \\ &= \max_{p(x_1)} I(X_1; Y_1) + \max_{p(x_2)} I(X_2; Y_2) \\ &= C_1 + C_2. \end{aligned}$$

with equality iff $p(x_1, x_2) = p^*(x_1)p^*(x_2)$ and $p^*(x_1)$ and $p^*(x_2)$ are the distributions for which $C_1 = I(X_1; Y_2)$ and $C_2 = I(X_2; Y_2)$ respectively.

Problem 3.

(a)

$$\begin{aligned} I(X;Y) = I(X_k, K; Y_k, K) &= I(K; Y_k, K) + I(X_k; Y_K, K|K) = H(K) + I(X_k; Y_k|K) \\ &= h_2(\alpha) + \mathbb{P}_K[1].I(X_k; Y_k|K=1) + \mathbb{P}_K[2]I(X_k; Y_k|K=2) \\ &= h_2(\alpha) + \alpha.I(X_1; Y_1) + (1-\alpha)I(X_2; Y_2) \end{aligned}$$

(b) The distribution of X is determined by α and by the distributions of X_1 and X_2. It is clear from the expression in (a) that for any given α, $I(X;Y)$ is maximized when $I(X_1; Y_1)$ and $I(X_2; Y_2)$ are maximized, i.e., when the distribution of X_1 (resp. X_2) achieves the capacity of P_1 (resp. P_2). We conclude that the value of α in the capacity achieving distribution is the one that maximizes the function $f(\alpha) = h_2(\alpha) + \alpha C_1 + (1-\alpha)C_2$. The derivative of f is:

$$f'(\alpha) = -\log_2(\alpha) - \frac{1}{\ln 2} + \log_2(1-\alpha) + \frac{1}{\ln 2} + C_1 - C_2 = C_1 - C_2 + \log_2 \frac{1-\alpha}{\alpha}.$$

We have $f'(\alpha) = 0$ (resp. $f'(\alpha) > 0$, $f'(\alpha) < 0$) if $\alpha = \alpha^*$ (resp. $\alpha < \alpha^*$, $\alpha > \alpha^*$), where $\alpha^* = \frac{2^{C_1}}{2^{C_1} + 2^{C_2}}$. This means that $f(\alpha)$ is maximized at $\alpha = \alpha^*$. Therefore, the capacity achieving distribution is such that $\alpha = \frac{2^{C_1}}{2^{C_1} + 2^{C_2}}$ and X_1 (resp. X_2) achieves the capacity of the channel P_1 (resp. P_2).

(c) From (b), we have:

$$\begin{aligned} C &= -\frac{2^{C_1}}{2^{C_1} + 2^{C_2}} \log_2 \frac{2^{C_1}}{2^{C_1} + 2^{C_2}} - \frac{2^{C_2}}{2^{C_1} + 2^{C_2}} \log_2 \frac{2^{C_2}}{2^{C_1} + 2^{C_2}} + \frac{2^{C_1}C_1}{2^{C_1} + 2^{C_2}} + \frac{2^{C_2}C_2}{2^{C_1} + 2^{C_2}} \\ &= -\frac{2^{C_1}}{2^{C_1} + 2^{C_2}} C_1 + \frac{2^{C_1}}{2^{C_1} + 2^{C_2}} \log_2(2^{C_1} + 2^{C_2}) - \frac{2^{C_2}}{2^{C_1} + 2^{C_2}} C_2 \\ &\quad + \frac{2^{C_2}}{2^{C_1} + 2^{C_2}} \log_2(2^{C_1} + 2^{C_2}) + \frac{2^{C_1}C_1}{2^{C_1} + 2^{C_2}} + \frac{2^{C_2}C_2}{2^{C_1} + 2^{C_2}} \\ &= \log_2(2^{C_1} + 2^{C_2}). \end{aligned}$$

Therefore, $2^C = 2^{C_1} + 2^{C_2}$.

PROBLEM 4. The assertion is clearly true with $n = 1$. To complete the proof by induction we need to show that the cascade of a BSC with parameter $q = \frac{1}{2}(1 - (1 - 2p)^n)$ with a BSC with parameter p is equivalent to a BSC with parameter $\frac{1}{2}(1 - (1 - 2p)^{n+1})$. To do so, observe that for a cascade of a BSC with parameter q and a BSC with parameter p, when a bit is sent, the opposite bit will be received if exactly one of the channels makes a flip, and this happens with probability $(1 - q)p + (1 - p)q$. Thus, the cascade is equivalent to a BSC with this parameter. For $q = \frac{1}{2}(1 - (1 - 2p)^n)$,

$$(1-q)p + (1-p)q = \frac{1}{2}(1 + (1-2p)^n)p + \frac{1}{2}(1 - (1-2p)^n)(1-p) = \frac{1}{2}(1 - (1-2p)^{n+1}),$$

and the assertion is proved.

Alternate proof: the cascade makes flips the incoming bit if an odd number of the elements of the cascade flip. Thus the cascade is equivalent to a BSC with parameter

$$a = \sum_{k:k \text{ odd}} \binom{n}{k} p^k (1-p)^{n-k}.$$

Let $b = \sum_{k:k \text{ even}} \binom{n}{k} p^k (1-p)^{n-k}$. Observe that

$$a + b = \sum_k \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1,$$

and

$$-a + b = \sum_k \binom{n}{k} (-p)^k (1-p)^{n-k} = (-p + 1 - p)^n = (1-2p)^n.$$

Subtracting the two equalities and dividing by two, we get $a = \frac{1}{2}(1 + (1 - 2p)^n)$.

PROBLEM 5. Let $P'_{X,Y}(x, y) = P_{Y|X}(y|x)Q'(x)$, $P'_Y(y) = \sum_{x\in\mathcal{X}} P'_{X,Y}(x, y)$ and $P_Y(y) = \sum_{x\in\mathcal{X}} P_{Y|X}(y|x)Q(x)$. We then have for any Q'

$$\begin{aligned}
&\sum_{x\in\mathcal{X}} Q'(x) \sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) \log\left(\frac{P_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}} P_{Y|X}(y|x')Q(x')}\right) - I(Q') \\
&= E_{P'_{X,Y}} \log \frac{P_{Y|X}}{P_Y} - I(Q') \\
&= E_{P'_{X,Y}} \left(\log \frac{P_{Y|X}}{P_Y} - \log \frac{P'_{X,Y}}{Q'_X P'_Y}\right) \\
&= E_{P'_{X,Y}} \log \frac{P'_Y}{P_Y} \\
&= E_{P'_Y} \log \frac{P'_Y}{P_Y} \\
&= D(P'_Y || P_Y) \geq 0
\end{aligned}$$

with equality if and only if $Q' = Q$. To prove (b), notice in the upper bound of part (a), that the inner summation is a function of x and that the outer summation is an average of this function with respect to the distribution $Q'(x)$. The average of a function is upper bounded by the maximum value that the function takes, and hence (b) follows.