ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 15 | Information Theory and Coding
Solutions to homework 7 | November 6, 2012

PROBLEM 1.

(a) The statistician calculates $\tilde{Y} = g(Y)$. Since $X \to Y \to \tilde{Y}$ forms a Markov chain, we can apply the data processing inequality. Hence for every distribution on X,

$$I(X;Y) \geq I(X;\tilde{Y}).$$

Let $\tilde{p}(x)$ be the distribution on x that maximizes $I(X;\tilde{Y})$. Then

$$C = \max_{p(x)} I(X;Y) \geq I(X;Y)_{p(x)=\tilde{p}(x)} \geq I(X;\tilde{Y})_{p(x)=\tilde{p}(x)} = \max_{p(x)} I(X;\tilde{Y}) = \tilde{C}.$$

Thus, the statistician is wrong and 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 $X \to \tilde{Y} \to Y$ forming a Markov chain, in other words if given $\tilde{Y}$, X and Y are independent.

PROBLEM 2.

$$Y = X + Z \qquad X \in \{0,1\}, \quad Z \in \{0,a\}$$

We have to distinguish various cases depending on the values of a.

$a = 0$ In this case, $Y = X$, and $\max I(X;Y) = \max H(X) = 1$. Hence the capacity is 1 bit per transmission.

$a \neq 0, \pm 1$ In this case, Y has four possible values $0, 1, a$ and $1+a$. Knowing Y, we know the X which was sent, and hence $H(X|Y) = 0$. Hence $\max I(X;Y) = \max H(X) = 1$, achieved for an uniform distribution on the input X.

$a = \pm 1$ In the case $a = 1$, Y has three possible output values, $0, 1$ and 2, and the channel is identical to the binary erasure channel discussed in class, with $\epsilon = 1/2$. As derived in class, the capacity of this channel is $1 - \epsilon = 1/2$ bit per transmission. The case of $a = -1$ is essentially the same and the capacity here is also $1/2$ bit per transmission.

PROBLEM 3. Since given X, one can determine Y from Z and vice versa, $H(Y|X) = H(Z|X) = H(Z) = \log 3$, regardless of the distribution of X. Hence the capacity of the channel is

$$\begin{aligned} C &= \max_{p_X} I(X;Y) \\ &= \max_{p_X} H(Y) - H(Y|X) \\ &= \log 11 - \log 3 \end{aligned}$$

which is attained when X has uniform distribution. The same result can also be seen by observing that this channel is symmetric.

Problem 4.

First we express $I(X;Y)$, the mutual information between the input and output of the Z-channel, as a function of $x = \Pr(X=1)$:

$$\begin{aligned} H(Y|X) &= x\,\mathcal{H}(\varepsilon) \\ H(Y) &= \mathcal{H}(\Pr(Y=1)) = \mathcal{H}((1-\varepsilon)x) \\ I(X;Y) &= H(Y) - H(Y|X) = \mathcal{H}((1-\varepsilon)x) - x\,\mathcal{H}(\varepsilon) \end{aligned} \tag{1}$$

We deduce that if $\varepsilon = 0$, the capacity equals 1 bit/symbol and is attained for $x = 1/2$. If $\varepsilon = 1$, then $I(X;Y) = 0$ for every $0 \le x \le 1$. Hence, the capacity is equal to zero and any value of x achieves it. From now on we assume $\varepsilon \neq 0, 1$.

Using elementary calculus, we have that

$$\frac{d}{dx} I(X;Y) = (1-\varepsilon)\log\left(\frac{1-(1-\varepsilon)x}{(1-\varepsilon)x}\right) - \mathcal{H}(\varepsilon)\ .$$

Imposing the condition $\frac{d}{dx}I(X;Y) = 0$ yields to the unique solution

$$x^*(\varepsilon) = \left((1-\varepsilon)(2^{\frac{\mathcal{H}(\varepsilon)}{1-\varepsilon}}+1)\right)^{-1}\ .$$

From (1) we have $I(X;Y) = 0$ for $x = 0$ and $x = 1$, and therefore the maximum of the mutual information is achieved for $x = x^*(\varepsilon)$. The capacity $C(\varepsilon)$ is given by

$$C(\varepsilon) = \mathcal{H}((1-\varepsilon)x^*(\varepsilon)) - x^*(\varepsilon)\,\mathcal{H}(\varepsilon) \text{ bits/symbol.}$$

Problem 5. Observe that with P_3 defined as in the problem, whateve distribution we choose for X, the random variables X, Y, Z form a Markov chain, i.e., given Y, the random variables X and Z are independent. The data processing theorem then yields:

$$\begin{aligned} I(X;Z) &\le I(X;Y) \le C_1 \\ I(X;Z) &\le I(Y;Z) \le C_2 \end{aligned}$$

and thus $I(X;Z) \le \min\{C_1, C_2\}$ for any distribution on X. We then conclude that $C_3 = \max_{p_X} I(X;Z) \le \min\{C_1, C_2\}$.

Problem 6.

1. As we are allowed to use only one of the channels then the input alphabet is simply $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$ and similarly $\mathcal{Y} = \mathcal{Y}_1 \cup \mathcal{Y}_2$. It is also easy to see that the transition distribution of the channel is as given in the problem statement.

2. Assume $q^\star$ is the capacity achieving distribution for the combined channel which assigns probabilities to $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$. Let $\alpha = \sum_{\mathcal{X}_1} q^\star(x)$ and let

$$q_1^\star = \begin{cases} q^\star(x)/\alpha & x \in \mathcal{X}_1 \\ 0 & x \in \mathcal{X}_2 \end{cases}\ , \quad q_2^\star = \begin{cases} 0 & x \in \mathcal{X}_1 \\ q^\star(x)/(1-\alpha) & x \in \mathcal{X}_2. \end{cases}$$

It is easy to see that $q_1^\star$ and $q_2^\star$ are both valid probability distributions and $q^\star = \alpha q_1^\star + (1-\alpha) q_2^\star$. We also have

$$\begin{aligned} H(Y|X) &= \sum_{x \in X} q^\star(x) H(Y|x) \\ &= \alpha \sum_{x \in \mathcal{X}_1} q_1^\star(x) H(Y|x) + (1-\alpha) \sum_{x \in \mathcal{X}_2} q_2^\star(x) H(Y|x) \\ &= \alpha H^\star(Y_1|X_1) + (1-\alpha) H^\star(Y_2|X_2), \end{aligned}$$

where $H^\star(Y_i|X_i)$, $i = 1, 2$ is the conditional entropy of the channel i when the assigned input distribution is $q_i^\star$, $i = 1, 2$ respectively. Assume that the output distribution of the channels is $o_i^\star$, $i = 1, 2$ if we assign $q_i^\star$, $i = 1, 2$ as the input distribution. It is easy to check that as the input and the output alphabet of the channels are disjoint, $o_1^\star$ will be concentrated on $\mathcal{Y}_1$ and similarly $o_2^\star$ will be concentrated on $\mathcal{Y}_2$. If we group the output of the channel into two disjoint groups $\mathcal{Y}_1$ and $\mathcal{Y}_2$, by applying the grouping property of the entropy, it can be seen that

$$H(Y) = \alpha H^\star(Y_1) + (1-\alpha)H^\star(Y_2) + h_2(\alpha),$$

where $H^\star(Y_i)$, $i = 1, 2$ is the output entropy of the channel i when the corresponding input distribution is $q_i^\star$, $i = 1, 2$ and h_2 denotes the binary entropy function. Hence, we have that

$$\begin{aligned} I(X;Y) &= \alpha(H^\star(Y_1) - H^\star(Y_1|X_1)) \\ &\quad + (1-\alpha)(H^\star(Y_2) - H^\star(Y_2|X_2)) + h_(\alpha) \\ &= \alpha I_1^\star + (1-\alpha)I_2^\star + h_2(\alpha). \end{aligned}$$

For a given value of α, it is seen that both $I_1^\star$ and $I_2^\star$ are maximized provided that $q_1^\star = p_1^\star$ and $q_2^\star = p_2^\star$. In other words, the optimal distribution over $\mathcal{Y}$ has the property that if we restrict and renormalize it over $\mathcal{Y}_i$, $i = 1, 2$ we should get the corresponding capacity achieving distribution for channel i respectively. This implies that the probability distribution must have the form proposed in the problem statement.

3. By the expression we obtained in the previous part and by replacing $I_i^\star = C_i$, $i = 1, 2$ we obtain that the capacity as a function of α is as follows

$$C(\alpha) = \alpha C_1 + (1-\alpha)C_2 + h_2(\alpha).$$

Taking the derivative with respect to α we obtain that $\alpha_{\text{opt}} = \frac{2^{C_1}}{2^{C_1}+2^{C_2}}$. Replacing in the expression for $C(\alpha)$ we have that

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

4. Replacing $C_1 = C_2 = 0$ we get $C = 1$. Although it seems counter intuitive, but notice that here there is an indirect communication link from the input to the output because as we said $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are disjoint and the receiver by observing the output, indirectly notices which channel has been selected for communication at the input and actually this indirect link behaves like a noiseless channel which by its own can carry one bit of information from the sender to the receiver and that is the reason why we obtain $C = 1$.