Information Theory II

Spring Semester 2017

Prof. Dr. A. Lapidoth / Prof. Dr. S. Moser

Signal and Information Processing Laboratory

Institut für Signal- und Informationsverarbeitung

Model Answers to Exercise 8 of April 13, 2017

http://www.isi.ee.ethz.ch/teaching/courses/it2.html

Problem 1 — *Capacity Region of Broadcast Channels Depends Only on Conditional Marginals*

Because $\phi_1(Y_1^n) \neq M_1$ implies $(\phi_1(Y_1^n), \phi_2(Y_2^n)) \neq (M_1, M_2)$, we have $P_\mathrm{e} \geq P_1$. By the same argument, $P_\mathrm{e} \geq P_2$ holds, and therefore

$$P_\mathrm{e} \geq \max\{P_1, P_2\}.$$

On the other hand,

$$\begin{aligned}
P_\mathrm{e} &= \Pr\bigl[(\phi_1(Y_1^n), \phi_2(Y_2^n)) \neq (M_1, M_2)\bigr] \\
&= \Pr\bigl[\phi_1(Y_1^n) \neq M_1 \ \text{ or } \ \phi_2(Y_2^n) \neq M_2\bigr] \\
&\leq \Pr\bigl[\phi_1(Y_1^n) \neq M_1\bigr] + \Pr\bigl[\phi_2(Y_2^n) \neq M_2\bigr] \\
&= P_1 + P_2 \\
&\leq 2 \cdot \max\{P_1, P_2\}.
\end{aligned}$$

This shows that P_e tends to zero if and only if both P_1 and P_2 tend to zero. Since P_1 and P_2 depend only on the conditional marginals $P_{Y_1|X}$ and $P_{Y_2|X}$, it follows that the capacity region depends only on $P_{Y_1|X}$ and $P_{Y_2|X}$.

Problem 2 — *Capacity Points*

a) From the definition of achievable rate pairs we see that R_1^* is the capacity of the single-user channel from X to Y_1. The same argument holds for R_2^*, so

$$\mathsf{R}_1^* = \max_{P_X} I(X; Y_1), \tag{1}$$

$$\mathsf{R}_2^* = \max_{P_X} I(X; Y_2). \tag{2}$$

b) Let X be distributed according to a distribution that achieves the maximum in (2). Then,

$$\mathsf{R}_2^* = I(X; Y_2) \overset{\text{(i)}}{\leq} I(X; Y_1) \leq \mathsf{R}_1^*,$$

where (i) follows from the data processing inequality.

Problem 3 ***Broadcast BEC***

Note first that the channel is stochastically degraded. Indeed, if we construct a channel $W(Y_2|Y_1)$ from Y_1 to Y_2 such that $W(0|0) = W(1|1) = \frac{1-\alpha_2}{1-\alpha_1}$ and otherwise the output is always ?, then we obtain a physically degraded broadcast channel with the same conditional marginals as our channel. (The channel $W(Y_2|Y_1)$ is well-defined because of the assumption $\alpha_1 < \alpha_2$.) Therefore, we know that the capacity region of our channel is given by the union of all rate pairs $(\mathsf{R}_1, \mathsf{R}_2)$ satisfying

$$\begin{aligned}\mathsf{R}_1 &\leq I(X;Y_1|U)\\ \mathsf{R}_2 &\leq I(U;Y_2)\end{aligned}$$

for some distribution of the form $P_U P_{X|U} P_{Y_1Y_2|X}$. We can rewrite the bound on R_1 as

$$\begin{aligned}\mathsf{R}_1 &\leq I(X;Y_1|U)\\ &= H(X|U) - H(X|U,Y_1)\\ &= H(X|U) - \sum_{y_1} P_{Y_1}(y_1)H(X|U,Y_1=y_1)\\ &\overset{\text{(i)}}{=} H(X|U) - P_{Y_1}(?)H(X|U,Y_1=?)\\ &\overset{\text{(ii)}}{=} H(X|U) - \alpha_1 H(X|U,Y_1=?)\\ &\overset{\text{(iii)}}{=} H(X|U) - \alpha_1 H(X|U)\\ &= (1-\alpha_1)H(X|U),\end{aligned}$$

where (i) holds because $Y_1 \in \{0,1\}$ implies $X = Y_1$, so the conditional entropy is zero in these cases; (ii) holds because $P_{Y_1}(?) = \alpha_1$ is true for all distributions $P_U P_{X|U}$ (the probability of an erasure does not depend on whether the channel input was a zero or a one); and (iii) holds because erasures are independent of the channel input.
(Note that $I(X;Y_1|U) = \sum_u P_U(u) I(X;Y_1|U=u)$, where $I(X;Y_1|U=u)$ is the mutual information between the input and output of a binary erasure channel with input distribution $P_{X|U=u}(x)$. One could also perform the above steps for the BEC to obtain $I(X;Y_1|U=u) = (1-\alpha_1)H(X|U=u)$ and use that $\sum_u P_U(u)(1-\alpha_1)H(X|U=u) = (1-\alpha_1)H(X|U)$.)
We upperbound R_2 as follows:

$$\begin{aligned}\mathsf{R}_2 &\leq I(U;Y_2)\\ &\overset{\text{(i)}}{=} I(U,X;Y_2) - I(X;Y_2|U)\\ &\overset{\text{(ii)}}{=} I(X;Y_2) + I(U;Y_2|X) - I(X;Y_2|U)\\ &\overset{\text{(iii)}}{=} I(X;Y_2) - I(X;Y_2|U)\\ &\overset{\text{(iv)}}{=} I(X;Y_2) - (1-\alpha_2)H(X|U)\\ &\overset{\text{(v)}}{\leq} (1-\alpha_2) - (1-\alpha_2)H(X|U)\\ &= (1-\alpha_2)(1-H(X|U)),\end{aligned}$$

where (i) and (ii) follow from the chain rule for mutual information; (iii) follows from Markovity; (iv) follows from a similar calculation as done above for R_1; and (v) follows because the mutual information between the input and the output of a channel is bounded by its capacity. Now substitute $H(X|U)$ by λ. Because $H(X|U)$ is between zero and one bit, we see that every achievable rate pair has to be in the union over all $\lambda \in [0,1]$ of the regions defined by

$$\mathsf{R}_1 \leq \lambda(1-\alpha_1), \tag{3}$$

$$\mathsf{R}_2 \leq (1-\lambda)(1-\alpha_2). \tag{4}$$

We next show that all rates satisfying (3) and (4) with strict inequality are achievable. This can be done by a simple time-sharing strategy for every block length n. Use λn channel uses to transmit data to the first receiver and use the remaining channel uses to transmit data to the second receiver. Since (3) and (4) are satisfied with strict inequality and since the capacity of a BEC with erasure probability α is $1-\alpha$ bits per channel use, the error probability of this scheme can be made arbitrarily small as the block length tends to infinity.

Problem 4 — *Converse for Deterministic Broadcast Channel*

For a given $(\mathsf{R}_1, \mathsf{R}_2, n)$ code, let M_1 and M_2 be independent and uniformly distributed over the message sets. Define the PMF P_X as

$$P_X(x) \triangleq \sum_{i=1}^{n} \frac{1}{n} P_{X_i}(x) \tag{5}$$

and observe that

$$\begin{aligned} I(M_1;\hat{M}_1) &\leq I(X^n;Y_1^n) \\ &= H(Y_1^n) - H(Y_1^n|X^n) \\ &\overset{\text{(i)}}{=} H(Y_1^n) \\ &= \sum_{i=1}^{n} H(Y_{1,i}|Y_1^{i-1}) \\ &\leq \sum_{i=1}^{n} H(Y_{1,i}) \\ &= n\sum_{i=1}^{n} \frac{1}{n} H(P_{Y_{1,i}}) \\ &\overset{\text{(ii)}}{\leq} nH\left(\sum_{i=1}^{n} \frac{1}{n} P_{Y_{1,i}}\right) \\ &\overset{\text{(iii)}}{=} nH(Y_1), \end{aligned} \tag{6}$$

where (i) holds because the broadcast channel is deterministic; (ii) follows from the concavity of the entropy; and (iii) holds because

$$\begin{aligned} \sum_{i=1}^{n} \frac{1}{n} P_{Y_{1,i}}(y_1) &= \sum_{i=1}^{n} \frac{1}{n} \sum_{x} P_{X_i Y_{1,i}}(x, y_1) \\ &= \sum_{i=1}^{n} \frac{1}{n} \sum_{x} P_{X_i}(x) P_{Y_1|X}(y_1|x) \\ &= \sum_{x} \sum_{i=1}^{n} \frac{1}{n} P_{X_i}(x) P_{Y_1|X}(y_1|x) \\ &\overset{\text{(iv)}}{=} \sum_{x} P_X(x) P_{Y_1|X}(y_1|x) \\ &= P_{Y_1}(y_1), \end{aligned}$$

where (iv) follows from (5) and P_{Y_1} is the distribution induced by P_X and the channel law. Similarly, we obtain

$$I(M_2;\hat{M}_2) \leq nH(Y_2), \tag{7}$$

$$I(M_1, M_2;\hat{M}_1, \hat{M}_2) \leq nH(Y_1, Y_2). \tag{8}$$

Since M_1 and M_2 are independent and uniformly distributed over the message sets, we have

$$\begin{aligned}\mathsf{R}_1 &= \frac{1}{n}H(M_1)\\ &= \frac{1}{n}I(M_1;\hat{M}_1) + \frac{1}{n}H(M_1|\hat{M}_1)\\ &\overset{\text{(i)}}{\leq} H(Y_1) + \delta_n,\end{aligned}$$

where (i) follows from (6) and from Fano's inequality; and similar statements follow from (7) and (8). As the block length n tends to infinity, δ_n has to tend to zero, and we obtain that every achievable rate pair $(\mathsf{R}_1, \mathsf{R}_2)$ has to be in the union over all P_X of the regions

$$\begin{aligned}\mathsf{R}_1 &\leq H(Y_1),\\ \mathsf{R}_2 &\leq H(Y_2),\\ \mathsf{R}_1 + \mathsf{R}_2 &\leq H(Y_1, Y_2)\end{aligned}$$

with joint PMF $P_X P_{Y_1,Y_2|X}$.