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 13 of June 1, 2017

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

Problem 1 ***Optimal Guessing***

a) We have $P_X(0) = 0.3$ and $P_X(1) = 0.7$. From class we know that it is optimal to guess in decreasing order of probabilities, so guessing first $X = 1$ and then $X = 0$ is optimal. The corresponding guessing function is $\mathsf{G}^*(1) = 1$ and $\mathsf{G}^*(0) = 2$. Therefore,

$$\mathsf{E}[\mathsf{G}^*(X)^\rho] = \sum_x P_X(x)\mathsf{G}^*(x)^\rho = \frac{7}{10} + \frac{3}{10}\cdot 2^\rho.$$

b) We have

$$\begin{aligned}\mathsf{E}[\mathsf{G}^*(X|Y)^\rho] &= \sum_y P_Y(y)\,\mathsf{E}[\mathsf{G}^*(X|Y=y)^\rho] \\ &= \sum_y P_Y(y)\sum_x P_{X|Y=y}(x)\,\mathsf{E}[\mathsf{G}^*(x|Y=y)^\rho] \\ &\overset{\text{(i)}}{=} \sum_y P_Y(y)\Big\{P_{X|Y=y}(1)\cdot 1^\rho + P_{X|Y=y}(0)\cdot 2^\rho\Big\} \\ &= P_X(1)\cdot 1^\rho + P_X(0)\cdot 2^\rho \\ &= \frac{7}{10} + \frac{3}{10}\cdot 2^\rho,\end{aligned}$$

where (i) holds because $P_{X|Y=y}(1) > P_{X|Y=y}(0)$ for both $y = 0$ and $y = 1$, so the optimal guessing order is independent of the side information. Observe that $\mathsf{E}[\mathsf{G}^*(X|Y)^\rho] = \mathsf{E}[\mathsf{G}^*(X)^\rho]$ holds for all $\rho > 0$ even though X and Y are not independent.

c) We have

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log\mathsf{E}[\mathsf{G}^*(X^n)^\rho] &= \rho H_{\frac{1}{1+\rho}}(X) \\ &= (1+\rho)\log\sum_x P_X(x)^{\frac{1}{1+\rho}} \\ &= (1+\rho)\log\left(\left(\frac{3}{10}\right)^{\frac{1}{1+\rho}} + \left(\frac{7}{10}\right)^{\frac{1}{1+\rho}}\right).\end{aligned}$$

Problem 2 ***Rényi Entropy***

a) We first treat the case $\alpha \in (0,1)$: observe that for all $\alpha \in (0,1)$ and $p \in [0,1]$, $p^\alpha \geq p$ holds (with equality if and only if $p = 0$ or $p = 1$). Therefore,

$$\sum_x P_X(x)^\alpha \geq \sum_x P_X(x) = 1, \tag{1}$$

and because $\frac{1}{1-\alpha} > 0$,

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_x P_X(x)^\alpha \geq \frac{1}{1-\alpha} \log 1 = 0.$$

Equality holds if and only if equality holds in (1), which is the case if and only if $P_X(x) \in \{0,1\}$ for all x, i.e., if and only if X is deterministic.

In the case $\alpha > 1$, observe that $p^\alpha \leq p$ holds for all $\alpha > 1$ and $p \in [0,1]$ (with equality if and only if $p = 0$ or $p = 1$), so

$$\sum_x P_X(x)^\alpha \leq \sum_x P_X(x) = 1, \tag{2}$$

and because $\frac{1}{1-\alpha} < 0$,

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_x P_X(x)^\alpha \geq \frac{1}{1-\alpha} \log 1 = 0.$$

Equality holds if and only if equality holds in (2), which is the case if and only if $P_X(x) \in \{0,1\}$ for all x, i.e., if and only if X is deterministic.

b) Without loss of generality, we use the natural logarithm:

$$\begin{aligned}
\lim_{\alpha\to 1} H_\alpha(X) &= \lim_{\alpha\to 1} \frac{1}{1-\alpha} \ln \sum_x P_X(x)^\alpha \\
&= \lim_{\alpha\to 1} \frac{\ln \sum_{x\in\mathrm{supp}(P_X)} P_X(x)^\alpha}{1-\alpha} \\
&\overset{\text{(i)}}{=} \lim_{\alpha\to 1} \frac{\left\{\sum_{x\in\mathrm{supp}(P_X)} P_X(x)^\alpha\right\}^{-1} \cdot \sum_{x\in\mathrm{supp}(P_X)} P_X(x)^\alpha \ln P_X(x)}{-1} \\
&= -\sum_{x\in\mathrm{supp}(P_X)} P_X(x) \ln P_X(x) \\
&= H(X),
\end{aligned}$$

where (i) follows from L'Hôpital's rule because both the numerator and the denominator tend to zero as α tends to one.

Problem 3 ***Rényi Divergence***

a) We first consider the case when an x exists with $P(x) > 0$ and $Q(x) = 0$. Then, $D(P\|Q) = \infty$, $D_\alpha(P\|Q) = \infty$ for $\alpha > 1$, and

$$\lim_{\alpha\uparrow 1} \sum_x P(x)^\alpha Q(x)^{1-\alpha} = \lim_{\alpha\uparrow 1} \sum_{x\in\mathrm{supp}(P)\cap\mathrm{supp}(Q)} P(x)^\alpha Q(x)^{1-\alpha} = \sum_{x\in\mathrm{supp}(P)\cap\mathrm{supp}(Q)} P(x) < 1,$$

so

$$\lim_{\alpha\uparrow 1} D_\alpha(P\|Q) = \lim_{\alpha\uparrow 1} \frac{1}{\alpha-1}\log\sum_x P(x)^\alpha Q(x)^{1-\alpha} = \infty.$$

In the other case, we have $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$ and, using $\mathcal{S} = \mathrm{supp}(P) \cap \mathrm{supp}(Q) = \mathrm{supp}(P)$,

$$\begin{aligned}
\lim_{\alpha\to 1} D_\alpha(P\|Q) &= \lim_{\alpha\to 1}\frac{1}{\alpha-1}\ln\sum_x P(x)^\alpha Q(x)^{1-\alpha} \\
&= \lim_{\alpha\to 1}\frac{\ln\sum_{x\in\mathcal{S}} P(x)^\alpha Q(x)^{1-\alpha}}{\alpha-1} \\
&\overset{\text{(i)}}{=} \lim_{\alpha\to 1}\frac{\left\{\sum_{x\in\mathcal{S}} P(x)^\alpha Q(x)^{1-\alpha}\right\}^{-1}\cdot\sum_{x\in\mathcal{S}} P(x)^\alpha Q(x)^{1-\alpha}\ln\frac{P(x)}{Q(x)}}{1} \\
&= \sum_{x\in\mathcal{S}} P(x)\ln\frac{P(x)}{Q(x)} \\
&= D(P\|Q),
\end{aligned}$$

where (i) follows from L'Hôpital's rule because both the numerator and the denominator tend to zero as α tends to one.

b) Since $U(x) = \frac{1}{|\mathcal{X}|}$ for all x, we have

$$\begin{aligned}
D_\alpha(P\|U) &= \frac{1}{\alpha-1}\log\sum_x P(x)^\alpha\left(\frac{1}{|\mathcal{X}|}\right)^{1-\alpha} \\
&= \frac{1}{\alpha-1}\log|\mathcal{X}|^{\alpha-1} - \frac{1}{1-\alpha}\log\sum_x P(x)^\alpha \\
&= \log|\mathcal{X}| - H_\alpha(P).
\end{aligned}$$

Problem 4 — *Guessing with Chosen Side Information*

a) We have

$$\begin{aligned}
\mathsf{E}[\mathsf{G}^*(X^n|f_n(X^n))^\rho] &\overset{\text{(i)}}{\geq} \frac{1}{|\mathcal{Y}|^\rho}\,\mathsf{E}[\mathsf{G}^*(X^n)^\rho] \\
&\overset{\text{(ii)}}{\geq} \frac{1}{|\mathcal{Y}|^\rho}\,\frac{1}{1+\rho}\,\frac{1}{(n+1)^{(1+\rho)|\mathcal{X}|}}\,2^{n\max_{Q\in\mathcal{P}_n(\mathcal{X})}\{\rho H(Q)-D(Q\|P)\}} \\
&= 2^{-n\rho\mathsf{R}}\,\frac{1}{1+\rho}\,\frac{1}{(n+1)^{(1+\rho)|\mathcal{X}|}}\,2^{n\max_{Q\in\mathcal{P}_n(\mathcal{X})}\{\rho H(Q)-D(Q\|P)\}} \\
&= 2^{-n\rho\mathsf{R}}\,\frac{1}{1+\rho}\,\frac{1}{(n+1)^{(1+\rho)|\mathcal{X}|}}\,2^{n\rho[H_{1/(1+\rho)}(X)-\delta_n]} \\
&= \frac{1}{1+\rho}\,\frac{1}{(n+1)^{(1+\rho)|\mathcal{X}|}}\,2^{n\rho[H_{1/(1+\rho)}(X)-\delta_n-\mathsf{R}]},
\end{aligned}$$

where (i) and (ii) follow from the lecture notes; and δ_n is defined as

$$\begin{aligned}
\delta_n &\triangleq \frac{1}{\rho}\left[\max_{Q\in\mathcal{P}(\mathcal{X})}\{\rho H(Q)-D(Q\|P)\} - \max_{Q\in\mathcal{P}_n(\mathcal{X})}\{\rho H(Q)-D(Q\|P)\}\right] \\
&= H_{\frac{1}{1+\rho}}(X) - \frac{1}{\rho}\cdot\max_{Q\in\mathcal{P}_n(\mathcal{X})}\{\rho H(Q)-D(Q\|P)\}.
\end{aligned}$$

Because $\mathsf{R} < H_{1/(1+\rho)}(X)$ and because δ_n tends to zero as n tends to infinity, there exist $\gamma > 0$ and n_0 such that

$$H_{\frac{1}{1+\rho}}(X) - \delta_n - \mathsf{R} \geq \gamma$$

holds for all $n \geq n_0$. Consequently, for all such n,

$$\mathsf{E}[\mathsf{G}^*(X^n|f_n(X^n))^\rho] \geq \frac{1}{1+\rho} \frac{1}{(n+1)^{(1+\rho)|\mathcal{X}|}} 2^{n\rho\gamma}, \tag{3}$$

and since the RHS of (3) tends to infinity as n tends to infinity, the LHS of (3) must also tend to infinity as n tends to infinity.

b) We partition every type class into $k_n = \left\lceil \frac{2^{n\mathsf{R}}}{2(n+1)^{|\mathcal{X}|}} \right\rceil$ subsets. Because R is positive, there exists a n_0 such that

$$2^{n\mathsf{R}} \geq 2(n+1)^{|\mathcal{X}|} \tag{4}$$

holds for all $n \geq n_0$. Since we are only interested in the large-n asymptotics, we assume from now on that n is such that (4) holds. Then, the total number of sets can be bounded as

$$\begin{aligned} |\mathcal{P}_n(\mathcal{X})| \cdot k_n &\leq (n+1)^{|\mathcal{X}|} \cdot k_n \\ &\leq (n+1)^{|\mathcal{X}|} \cdot \left(\frac{2^{n\mathsf{R}}}{2(n+1)^{|\mathcal{X}|}} + 1 \right) \\ &= \frac{1}{2} \cdot 2^{n\mathsf{R}} + \frac{1}{2} \cdot 2(n+1)^{|\mathcal{X}|} \\ &\overset{\text{(i)}}{\leq} \frac{1}{2} \cdot 2^{n\mathsf{R}} + \frac{1}{2} \cdot 2^{n\mathsf{R}} \\ &= 2^{n\mathsf{R}}, \end{aligned}$$

where (i) follows from (4). Since there are at most $2^{n\mathsf{R}}$ such sets, we can use f_n to encode the set to which a sequence x^n belongs. For a given sequence x^n of type Q, the decoder therefore learns to which subset of which type class the sequence belongs and the number of guesses is upper bounded by $\left\lceil \frac{|\mathcal{T}^n(Q)|}{k_n} \right\rceil$. Consequently,

$$\begin{aligned} \mathsf{E}[\mathsf{G}^*(X^n|f_n(X^n))^\rho] &= \sum_{Q \in \mathcal{P}_n(\mathcal{X})} \sum_{x^n \in \mathcal{T}^n(Q)} P_X^n(x^n) \cdot \mathsf{G}^*(x^n|f_n(x^n))^\rho \\ &\leq \sum_{Q \in \mathcal{P}_n(\mathcal{X})} \sum_{x^n \in \mathcal{T}^n(Q)} P_X^n(x^n) \cdot \left\lceil \frac{|\mathcal{T}^n(Q)|}{k_n} \right\rceil^\rho \\ &= \sum_{Q \in \mathcal{P}_n(\mathcal{X})} P_X^n(\mathcal{T}^n(Q)) \cdot \left\lceil \frac{|\mathcal{T}^n(Q)|}{k_n} \right\rceil^\rho \\ &\overset{\text{(i)}}{\leq} \sum_{Q \in \mathcal{P}_n(\mathcal{X})} P_X^n(\mathcal{T}^n(Q)) \cdot \left\{ 1 + 2^\rho \left(\frac{|\mathcal{T}^n(Q)|}{k_n} \right)^\rho \right\} \\ &= 1 + 2^\rho \sum_{Q \in \mathcal{P}_n(\mathcal{X})} P_X^n(\mathcal{T}^n(Q)) \cdot |\mathcal{T}^n(Q)|^\rho \cdot k_n^{-\rho} \\ &\overset{\text{(ii)}}{\leq} 1 + 2^\rho \sum_{Q \in \mathcal{P}_n(\mathcal{X})} 2^{-nD(Q\|P_X)} \cdot 2^{n\rho H(Q)} \cdot 2^{-n\rho\mathsf{R}} \, 2^\rho (n+1)^{\rho|\mathcal{X}|} \\ &= 1 + 4^\rho (n+1)^{\rho|\mathcal{X}|} \cdot 2^{-n\rho\mathsf{R}} \sum_{Q \in \mathcal{P}_n(\mathcal{X})} 2^{n(\rho H(Q) - D(Q\|P_X))} \end{aligned}$$

$$\leq 1 + 4^{\rho}(n+1)^{\rho|\mathcal{X}|} \cdot 2^{-n\rho\mathsf{R}} \cdot (n+1)^{|\mathcal{X}|} \, 2^{n \max_{Q \in \mathcal{P}(\mathcal{X})} \{\rho H(Q) - D(Q \| P_X)\}}$$
$$= 1 + 4^{\rho}(n+1)^{(1+\rho)|\mathcal{X}|} \cdot 2^{n\rho[H_{1/(1+\rho)}(X) - \mathsf{R}]}, \qquad (5)$$

where (i) follows from the inequality

$$\lceil \xi \rceil^{\rho} \leq 1 + 2^{\rho}\xi^{\rho}, \qquad (6)$$

which holds for all $\rho > 0$ and $\xi \geq 0$; and (ii) holds because $k_n \geq \frac{2^{n\mathsf{R}}}{2(n+1)^{|\mathcal{X}|}}$. (It is easy to see that (6) is valid: for $\xi \in [0,1]$, it holds trivially, and for $\xi \geq 1$, we have $\xi + 1 \leq 2\xi$ and therefore $\lceil \xi \rceil^{\rho} \leq (\xi+1)^{\rho} \leq (2\xi)^{\rho} \leq 1 + 2^{\rho}\xi^{\rho}$.)

Because $\mathsf{R} > H_{1/(1+\rho)}(X)$, the RHS of (5) tends to one as n tends to infinity. Since the LHS of (5) is lower bounded by one, it follows from the sandwich theorem that the LHS of (5) also tends to one as n tends to infinity.