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 $G^*(1) = 1$ and $G^*(0) = 2$. Therefore,

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

b) We have

$$
E[G^*(X|Y)^\rho] = \sum_y P_Y(y) E[G^*(X|Y = y)^\rho]
$$

$$
= \sum_y P_Y(y) \sum_x P_{X|Y=y}(x) E[G^*(x|Y = y)^\rho]
$$

$$
= \sum_y P_Y(y) \left\{ P_{X|Y=y}(1) \cdot 1^\rho + P_{X|Y=y}(0) \cdot 2^\rho \right\}
$$

$$
= P_X(1) \cdot 1^\rho + P_X(0) \cdot 2^\rho
$$

$$
= \frac{7}{10} + \frac{3}{10} \cdot 2^\rho,
$$

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 $E[G^*(X|Y)^\rho] = E[G^*(X)^\rho]
holds for all $\rho > 0$ even though $X$ and $Y$ are not independent.

c) We have

$$
\lim_{n \to \infty} \frac{1}{n} \log E[G^*(X^n)^\rho] = \rho H_{\frac{1}{1+\rho}}(X)
$$

$$
= (1 + \rho) \log \sum_x P_X(x)^\frac{1}{1+\rho}
$$

$$
= (1 + \rho) \log \left( \left( \frac{3}{10} \right)^\frac{1}{1+\rho} + \left( \frac{7}{10} \right)^\frac{1}{1+\rho} \right).
$$
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,

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

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

$$\sum_x P_X(x)^\alpha \leq \sum_x P_X(x) = 1,$$

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:

$$\lim_{\alpha \to 1} H_\alpha(X) = \lim_{\alpha \to 1} \frac{1}{1-\alpha} \ln \sum_x P_X(x)^\alpha$$

$$= \lim_{\alpha \to 1} \frac{\ln \sum_{x \in \text{supp}(P_X)} P_X(x)^\alpha}{1-\alpha}$$

$$\overset{(i)}{=} \lim_{\alpha \to 1} \frac{\sum_{x \in \text{supp}(P_X)} P_X(x)^\alpha}{1-\alpha} \cdot \ln P_X(x)$$

$$= -\sum_{x \in \text{supp}(P_X)} P_X(x) \ln P_X(x)$$

$$= H(X),$$

where (i) follows from L’Hôpital’s rule because both the numerator and the denominator tend to zero as $\alpha$ 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

$$\lim_{\alpha \uparrow 1} \sum_x P(x)^\alpha Q(x)^{1-\alpha} = \lim_{\alpha \uparrow 1} \sum_{x \in \text{supp}(P) \cap \text{supp}(Q)} P(x)^\alpha Q(x)^{1-\alpha} = \sum_{x \in \text{supp}(P) \cap \text{supp}(Q)} P(x) < 1.$$
so
\[
\lim_{\alpha \to 1} D_\alpha(P\|Q) = \lim_{\alpha \to 1} \frac{1}{\alpha - 1} \log \sum_x P(x)^\alpha Q(x)^{1-\alpha} = \infty.
\]

In the other case, we have \(\text{supp}(P) \subseteq \text{supp}(Q)\) and, using \(S = \text{supp}(P) \cap \text{supp}(Q) = \text{supp}(P)\),
\[
\lim_{\alpha \to 1} D_\alpha(P\|Q) = \lim_{\alpha \to 1} \frac{1}{\alpha - 1} \ln \sum_x P(x)^\alpha Q(x)^{1-\alpha}
= \lim_{\alpha \to 1} \frac{\ln \sum_{x \in S} P(x)^\alpha Q(x)^{1-\alpha}}{\alpha - 1}
= \lim_{\alpha \to 1} \left\{ \left( \sum_{x \in S} P(x)^\alpha Q(x)^{1-\alpha} \right)^{-1} \cdot \sum_{x \in S} P(x)^\alpha Q(x)^{1-\alpha} \ln \frac{P(x)}{Q(x)} \right\} = \sum_{x \in S} P(x) \ln \frac{P(x)}{Q(x)} = D(P\|Q),
\]
where (i) follows from L'Hôpital's rule because both the numerator and the denominator tend to zero as \(\alpha\) tends to one.

b) Since \(U(x) = \frac{1}{|x|}\) for all \(x\), we have
\[
D_\alpha(P\|U) = \frac{1}{\alpha - 1} \log \sum_x P(x)^\alpha \left( \frac{1}{|x|} \right)^{1-\alpha}
= \frac{1}{\alpha - 1} \log |x|^{\alpha-1} - \frac{1}{1 - \alpha} \log \sum_x P(x)^\alpha
= \log |x| - H_\alpha(P).
\]

**Problem 4**

**Guessing with Chosen Side Information**

a) We have
\[
\mathbb{E}[G^*(X^n|f_n(X^n))] \geq \frac{1}{|\mathcal{Y}|^\rho} \mathbb{E}[G^*(X^n)^\rho]
\]
\[
\geq \frac{1}{|\mathcal{Y}|^\rho} \frac{1}{1 + \rho} \frac{1}{(n+1)^{(1+\rho)|X|}} 2^{n \max_{Q \in \mathcal{P}_n(X)} \{\rho H(Q) - D(Q\|P)\}}
= 2^{-n \rho \mathbb{E}} \frac{1}{1 + \rho} \frac{1}{(n+1)^{(1+\rho)|X|}} 2^{n \max_{Q \in \mathcal{P}_n(X)} \{\rho H(Q) - D(Q\|P)\}}
= 2^{-n \rho \mathbb{E}} \frac{1}{1 + \rho} \frac{1}{(n+1)^{(1+\rho)|X|}} 2^n \{H_{1/(1+\rho)}(X) - \delta_n\}
= \frac{1}{1 + \rho} \frac{1}{(n+1)^{(1+\rho)|X|}} 2^n \{H_{1/(1+\rho)}(X) - \delta_n\}
\]
where (i) and (ii) follow from the lecture notes; and \(\delta_n\) is defined as
\[
\delta_n = \frac{1}{\rho} \left[ \max_{Q \in \mathcal{P}_n(X)} \{\rho H(Q) - D(Q\|P)\} - \max_{Q \in \mathcal{P}_n(X)} \{\rho H(Q) - D(Q\|P)\}\right]
= H_{1/(1+\rho)}(X) - \frac{1}{\rho} \max_{Q \in \mathcal{P}_n(X)} \{\rho H(Q) - D(Q\|P)\}.
\]

© Amos Lapidoth and Stefan M. Moser, 2017 3
Because $R < H_{1/(1+\rho)}(X)$ and because $\delta_n$ tends to zero as $n$ tends to infinity, there exist $\gamma > 0$ and $n_0$ such that
\[
H_{1/(1+\rho)}(X) - \delta_n - R \geq \gamma
\]
holds for all $n \geq n_0$. Consequently, for all such $n$,
\[
E[G^*(X^n|f_n(X^n))]^\rho \geq \frac{1}{1 + \rho} \frac{1}{(n + 1)^{(1+\rho)|X|}} 2^{nR\gamma}, \tag{3}
\]
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 $k_n = \left\lceil \frac{2^{nR}}{2((n+1)^{|X|}} \right\rceil$ subsets. Because $R$ is positive, there exists a $n_0$ such that
\[
2^{nR} \geq 2(n + 1)^{|X|} \tag{4}
\]
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
\[
|\mathcal{P}_n(X)| \cdot k_n \leq (n + 1)^{|X|} \cdot k_n = (n + 1)^{|X|} \cdot \left( \frac{2^{nR}}{2(n + 1)^{|X|}} + 1 \right)
\leq \frac{1}{2} \cdot 2^{nR} + \frac{1}{2} \cdot 2(n + 1)^{|X|}
\leq \frac{1}{2} \cdot 2^{nR} + \frac{1}{2} \cdot 2^{nR}
= 2^{nR},
\]
where (i) follows from (4). Since there are at most $2^{nR}$ 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 $\left\lceil \frac{|T^n(Q)|}{k_n} \right\rceil$. Consequently,
\[
E[G^*(X^n|f_n(X^n))]^\rho = \sum_{Q \in \mathcal{P}_n(X)} \sum_{x^n \in T^n(Q)} P_X^n(x^n) \cdot G^*(x^n|f_n(x^n))^\rho
\leq \sum_{Q \in \mathcal{P}_n(X)} \sum_{x^n \in T^n(Q)} P_X^n(x^n) \cdot \left\lceil \frac{|T^n(Q)|}{k_n} \right\rceil^\rho
\leq \sum_{Q \in \mathcal{P}_n(X)} P_X^n(T^n(Q)) \cdot \left\lceil \frac{|T^n(Q)|}{k_n} \right\rceil^\rho (i)
\leq \sum_{Q \in \mathcal{P}_n(X)} P_X^n(T^n(Q)) \cdot \left\lceil \frac{|T^n(Q)|}{k_n} \right\rceil^\rho \cdot k_n^{-\rho}
\leq 1 + \rho \cdot \sum_{Q \in \mathcal{P}_n(X)} 2^{-nD(Q||P_X)} \cdot 2^{n\rho H(Q)} \cdot 2^{-n\rho R} 2^\rho (n + 1)^{\rho|X|}
\leq 1 + 4^\rho (n + 1)^{\rho|X|} \cdot 2^{-n\rho R} \sum_{Q \in \mathcal{P}_n(X)} 2^{n\rho H(Q) - D(Q||P_X))}$
\[ \leq 1 + 4^\rho (n + 1)^{\rho |X|} \cdot 2^{-n\rho R} \cdot (n + 1)^{|X|} 2^n \max_{Q \in \mathcal{F}(X)} \{ \rho H(Q) - D(Q \parallel P_X) \} \\
= 1 + 4^\rho (n + 1)^{(1 + \rho)|X|} \cdot 2^{n\rho} [H_{1/(1+\rho)}(X) - R], \] 

(5)

where (i) follows from the inequality

\[ [\xi]^{\rho} \leq 1 + 2^\rho \xi^\rho, \] 

(6)

which holds for all \( \rho > 0 \) and \( \xi \geq 0 \); and (ii) holds because \( k_n \geq \frac{2^{nR}}{(n+1)^{|X|}} \). (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 \( [\xi]^{\rho} \leq (\xi + 1)^{\rho} \leq (2\xi)^{\rho} \leq 1 + 2^\rho \xi^\rho \).

Because \( R > \frac{H_{1/(1+\rho)}(X)}{1} \), 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.