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 11 of May 11, 2017

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

Problem 1 ***Method of Types***

a) To compute $|\mathcal{P}_n(\mathcal{X})|$, we make use of the following bijection between the set $\mathcal{P}_n(\mathcal{X})$ and the set of all sequences consisting of exactly n stars and $|\mathcal{X}|-1$ separators: the number of stars at the beginning of a sequence corresponds to the number of occurrences of the first symbol from $\mathcal{X}$, the number of stars between the first and the second separator corresponds to the number of occurrences of the second symbol from $\mathcal{X}$, and so on. For example, the type $P_{\mathbf{x}} = \left(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}\right) \in \mathcal{P}_5(\mathcal{X})$ corresponds to the sequence "*|**|**". The number of all sequences consisting of exactly n stars and $|\mathcal{X}|-1$ separators, and therefore $|\mathcal{P}_n(\mathcal{X})|$, is

$$\binom{n+|\mathcal{X}|-1}{|\mathcal{X}|-1}$$

because the sequences have length $n+|\mathcal{X}|-1$ and separators have to appear at $|\mathcal{X}|-1$ positions.

Here, $|\mathcal{P}_5| = \binom{7}{2} = 21$. This is much less than the upper bound

$$|\mathcal{P}_5| \le (n+1)^{|\mathcal{X}|} = 6^3 = 216.$$

The improved version of the upper bound is already closer:

$$|\mathcal{P}_5| \le (n+1)^{|\mathcal{X}|-1} = 6^2 = 36.$$

b) The type of $\mathbf{x} = (b,c,c,a,b)$ is $P_{\mathbf{x}}(a) = \frac{1}{5}$, $P_{\mathbf{x}}(b) = \frac{2}{5}$, $P_{\mathbf{x}}(c) = \frac{2}{5}$, i.e.,

$$P_{\mathbf{x}} = \left(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}\right).$$

c) The set of sequences having one particular type is called *type class*. In our case the size of the type class of $P_{\mathbf{x}}$ is

$$\left|\mathcal{T}^5(P_{\mathbf{x}})\right| = \frac{5!}{1!2!2!} = 30.$$

In order to compare this to the bounds derived in class we need to compute

$$\mathsf{H}(P_{\mathbf{x}}) = -\frac{1}{5}\log\frac{1}{5} - 2\cdot\frac{2}{5}\log\frac{2}{5} = \log 5 - \frac{4}{5}\log 2.$$

Hence,

$$\left|\mathcal{T}^5(P_{\mathbf{x}})\right| \le e^{n\,\mathsf{H}(P_{\mathbf{x}})} = e^{5\log 5 - 4\log 2} = \frac{5^5}{2^4} = \frac{3125}{16} \approx 195;$$

and

$$\left|\mathcal{T}^5(P_{\mathbf{x}})\right| \geq \frac{1}{(n+1)^{|\mathcal{X}|}}\, e^{n\,\mathsf{H}(P_{\mathbf{x}})} = \frac{1}{216}\cdot\frac{3125}{16} \approx 0.90.$$

The improved lower bound is

$$\left|\mathcal{T}^5(P_{\mathbf{x}})\right| \geq \frac{1}{(n+1)^{|\mathcal{X}|-1}}\, e^{n\,\mathsf{H}(P_{\mathbf{x}})} = \frac{1}{36}\cdot\frac{3125}{16} \approx 5.4.$$

These bounds are poor for small n. The lower bound can be improved further by using the exact number of types, $|\mathcal{P}_5| = 21$:

$$\left|\mathcal{T}^5(P_{\mathbf{x}})\right| \geq \frac{1}{|\mathcal{P}_5|}\, e^{n\,\mathsf{H}(P_{\mathbf{x}})} = \frac{1}{21}\cdot\frac{3125}{16} \approx 9.3.$$

d) We already computed $\mathsf{H}(P_{\mathbf{x}}) = \log 5 - \frac{4}{5}\log 2$. Next we compute

$$D(P_{\mathbf{x}}\|Q) = \frac{1}{5}\log\frac{1/5}{2/3} + 2\cdot\frac{2}{5}\log\frac{2/5}{1/6} = \frac{1}{5}\log\frac{2^7\cdot 3^5}{5^5}.$$

Then

$$\begin{aligned} Q^5(b,c,c,a,b) &= e^{-n(\mathsf{H}(P_{\mathbf{x}})+D(P_{\mathbf{x}}\|Q))} = e^{-5\log 5+4\log 2-\log 2^7-\log 3^5+\log 5^5} \\ &= 2^{-3}\cdot 3^{-5} = \frac{1}{1944} \approx 5.14\cdot 10^{-4}. \end{aligned}$$

Obviously, this could also have been computed directly as

$$Q^5(b,c,c,a,b) = \left(\frac{2}{3}\right)^1\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2 = \frac{1}{2^3\cdot 3^5} = \frac{1}{1944}.$$

e) Note that

$$\Pr\bigl[P_{\mathbf{X}} = P_{(b,c,c,a,b)}\bigr] = Q^5\bigl(\mathcal{T}^5(P_{(b,c,c,a,b)})\bigr).$$

We know from class that for $\mathbf{x} = (b,c,c,a,b)$

$$Q^5\bigl(\mathcal{T}^5(P_{\mathbf{x}})\bigr) \leq e^{-nD(P_{\mathbf{x}}\|Q)} = \frac{5^5}{2^7\cdot 3^5} = \frac{3125}{31104} \approx 0.100 = 10.0\cdot 10^{-2},$$

and

$$Q^5\bigl(\mathcal{T}^5(P_{\mathbf{x}})\bigr) \geq \frac{1}{(n+1)^{|\mathcal{X}|}}\cdot e^{-nD(P_{\mathbf{x}}\|Q)} = \frac{1}{216}\cdot\frac{3125}{31104} \approx 4.65\cdot 10^{-4}.$$

This value can be improved as follows:

$$Q^5\bigl(\mathcal{T}^5(P_{\mathbf{x}})\bigr) \geq \frac{1}{(n+1)^{|\mathcal{X}|-1}}\cdot e^{-nD(P_{\mathbf{x}}\|Q)} = \frac{1}{36}\cdot\frac{3125}{31104} \approx 0.279\cdot 10^{-2}.$$

Since all sequences in a type class have the same probability, the exact value of $Q^5\bigl(\mathcal{T}^5(P_{\mathbf{x}})\bigr)$ can be computed as

$$Q^5\bigl(\mathcal{T}^5(P_{\mathbf{x}})\bigr) = \left|\mathcal{T}^5(P_{\mathbf{x}})\right| e^{-n\,\mathsf{H}(P_{\mathbf{x}})-nD(P_{\mathbf{x}}\|Q)} = 30\cdot\frac{1}{1944} = \frac{5}{324} \approx 1.54\cdot 10^{-2}.$$

a) We first treat some special cases. Note that $D(P\|Q)$ is finite if and only if $\operatorname{supp}(P) \subseteq \operatorname{supp}(Q)$.

i) If $\mathsf{R}_n \leq H(Q)$, then $\min_{P: H(P) \geq \mathsf{R}_n} D(P\|Q) = D(Q\|Q) = 0$, so $P^* = Q$.

ii) If $\mathsf{R}_n = \log \operatorname{supp}(Q)$, then P^* is the uniform distribution over the support of Q, since this is the only distribution satisfying $\operatorname{supp}(P) \subseteq \operatorname{supp}(Q)$ and $H(P) \geq \mathsf{R}_n$.

iii) If $\mathsf{R}_n > \log \operatorname{supp}(Q)$, then no P exists such that $D(P\|Q)$ is finite (and indeed, inspecting the proof for the universal source coding scheme shows that in this case, the probability of error is zero).

From now on, we assume $H(Q) < \mathsf{R}_n < \log \operatorname{supp}(Q)$. We use Lagrange multipliers to minimize $D(P\|Q)$ over P subject to $\sum_x P(x) = 1$ and $H(P) = r$. Let

$$\mathrm{J}(P) = \sum_x P(x) \ln \frac{P(x)}{Q(x)} + \lambda_0 \left(\sum_x P(x) - 1 \right) + \lambda_1 \left(-\sum_x P(x) \ln P(x) - r \right).$$

Computing the partial derivative with respect to $P(\hat{x})$, we get

$$\frac{\partial \mathrm{J}(P)}{\partial P(\hat{x})} = \ln \frac{P(\hat{x})}{Q(\hat{x})} + 1 + \lambda_0 - \lambda_1 \ln P(\hat{x}) - \lambda_1.$$

Setting this partial derivative to zero for every $\hat{x}$, we obtain

$$P^*(x) = \left(Q(x)\, e^{\lambda_1 - \lambda_0 - 1} \right)^{\frac{1}{1-\lambda_1}},$$

which is, after substituting $\alpha = \frac{1}{1-\lambda_1}$, equivalent to

$$P^*(x) = \frac{Q(x)^\alpha}{\sum_{x'} Q(x')^\alpha}.$$

Having determined the "form" of P^*, it remains to show how to choose α and to justify that P^* indeed achieves the minimum. For the interested reader, we now show a way to solve these issues. Rewrite P^* as

$$P^*(x) = \frac{Q(x)^{\frac{1}{1+\rho}}}{\sum_{x'} Q(x')^{\frac{1}{1+\rho}}} \tag{1}$$

and let $\rho > 0$ be such that $H(P^*) = \mathsf{R}_n$. (The existence of such a ρ follows from the intermediate value theorem because $H(Q) < \mathsf{R}_n < \log \operatorname{supp}(Q)$, because $H(P^*)|_{\rho=0} = H(Q)$, because $\lim_{\rho\to\infty} H(P^*) = \log \operatorname{supp}(Q)$, and because $H(P^*)$ is continuous in ρ.) Then, for every PMF P satisfying $H(P) \geq \mathsf{R}_n$,

$$\begin{aligned}
D(P\|Q) &= \sum_x P(x) \log \frac{P(x)}{Q(x)} \\
&= \sum_x P(x) \log \frac{P(x)^{1+\rho} \cdot P^*(x)^{1+\rho}}{P^*(x)^{1+\rho} \cdot P(x)^\rho \cdot Q(x)} \\
&= (1+\rho) D(P\|P^*) + \rho H(P) + \sum_x P(x) \log \frac{P^*(x)^{1+\rho}}{Q(x)} \\
&\overset{\text{(i)}}{=} (1+\rho) D(P\|P^*) + \rho H(P) + \log \gamma
\end{aligned}$$

$$
\begin{aligned}
&= (1+\rho)D(P\|P^*) + \rho H(P) + \sum_x P^*(x)\log\frac{P^*(x)^{1+\rho}}{Q(x)} \\
&= (1+\rho)D(P\|P^*) + \rho H(P) + \sum_x P^*(x)\log\frac{P^*(x)}{P^*(x)^{-\rho}\cdot Q(x)} \\
&= (1+\rho)D(P\|P^*) + \rho H(P) - \rho H(P^*) + D(P^*\|Q) \\
&= (1+\rho)D(P\|P^*) + \rho\bigl(H(P) - H(P^*)\bigr) + D(P^*\|Q) \\
&\overset{\text{(ii)}}{\geq} D(P^*\|Q),
\end{aligned}
$$

where (i) holds because $\gamma = \frac{P^*(x)^{1+\rho}}{Q(x)} = \bigl\{\sum_{x'} Q(x')^{\frac{1}{1+\rho}}\bigr\}^{-(1+\rho)}$ does not depend on x; and (ii) holds because $\rho > 0$ and $H(P) \geq \mathsf{R}_n = H(P^*)$. Furthermore, equality in (ii) holds if and only if $P = P^*$. We conclude that in the case $H(Q) < \mathsf{R}_n < \log\operatorname{supp}(Q)$, P^* given by (1) is the unique minimizer of $D(P\|Q)$ subject to $H(P) \geq \mathsf{R}_n$.

b) To achieve $P_{\mathrm{e}}^{(n)} \to 0$, $H(Q) < \mathsf{R}$ is sufficient (and we know that even if the encoder and the decoder know Q, the probability of error tends to one if $H(Q) > \mathsf{R}$). If $\mathsf{R} \geq 1$, then $P_{\mathrm{e}}^{(n)} = 0$ is achieved for any binary distribution Q. If $\mathsf{R} \in (0,1)$, then there exists a unique $p^* \in \bigl(0, \frac{1}{2}\bigr)$ such that $H_{\mathrm{b}}(p^*) = \mathsf{R}$ because $H_{\mathrm{b}}(p)$ is strictly increasing in p for $p \in \bigl[0, \frac{1}{2}\bigr]$. Hence, $P_{\mathrm{e}}^{(n)} \to 0$ is achieved if $Q = (q, 1-q)$ satisfies $q < p^*$ or $q > 1 - p^*$.

Problem 3 — *Large Deviations*

Using Sanov's theorem we know that the exponent we are looking for is

$$
\inf_{f:\mathsf{E}_f[X^2]\geq\alpha^2} D\bigl(f\|\mathcal{N}(0,\sigma^2)\bigr). \tag{2}
$$

If $\alpha^2 \leq \sigma^2$, then (2) is zero, which is achieved by choosing $f = \mathcal{N}\bigl(0,\sigma^2\bigr)$. From now on, assume that $\alpha^2 > \sigma^2$. For a fixed f satisfying $\mathsf{E}_f\bigl[X^2\bigr] \geq \alpha^2$, define $\beta^2 \triangleq \mathsf{E}_f\bigl[X^2\bigr] \geq \alpha^2 \geq \sigma^2$ and observe that

$$
\begin{aligned}
D\bigl(f\|\mathcal{N}(0,\sigma^2)\bigr) &= \int_{-\infty}^{\infty} f(x)\ln\frac{f(x)}{\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{x^2}{2\sigma^2}}}\,\mathrm{d}x \\
&= -h(f) + \int_{-\infty}^{\infty} f(x)\ln\Bigl(\sqrt{2\pi\sigma^2}\,e^{\frac{x^2}{2\sigma^2}}\Bigr)\,\mathrm{d}x \\
&= -h(f) + \ln\sqrt{2\pi\sigma^2} + \frac{1}{2\sigma^2}\int_{-\infty}^{\infty} f(x)x^2\,\mathrm{d}x \\
&= -h(f) + \frac{1}{2}\Bigl(\ln 2\pi\sigma^2 + \frac{\beta^2}{\sigma^2}\Bigr) \\
&\overset{\text{(i)}}{\geq} -\frac{1}{2}\ln 2\pi e\beta^2 + \frac{1}{2}\Bigl(\ln 2\pi\sigma^2 + \frac{\beta^2}{\sigma^2}\Bigr) \\
&= \frac{1}{2}\Bigl(\frac{\beta^2}{\sigma^2} - 1 - \ln\frac{\beta^2}{\sigma^2}\Bigr) \\
&\overset{\text{(ii)}}{\geq} \frac{1}{2}\Bigl(\frac{\alpha^2}{\sigma^2} - 1 - \ln\frac{\alpha^2}{\sigma^2}\Bigr)\text{ nats,}
\end{aligned} \tag{3}
$$

where (i) holds because Gaussians maximize differential entropy for a fixed second moment; and (ii) holds because the expression in parentheses is monotonically increasing in β^2 for $\beta^2 \geq \sigma^2$ (this can be verified by taking the derivative with respect to β^2). Since (i) and (ii) hold with equality for the choice $f = \mathcal{N}\bigl(0,\alpha^2\bigr)$, the minimum in (2) in the case $\alpha^2 > \sigma^2$ is given by the RHS of (3).