ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 32 | Information Theory and Coding
Solutions to Graded Homework | Dec. 18, 2023

PROBLEM 1. Suppose (U_1, V) is a pair of random variables with distribution p_{UV}, and suppose $U_2, \dots, U_m$ are i.i.d. random variables with distribution p_U, independent of (U_1, V).

Let $\mathrm{score}(u, v) := p_{V|U}(v|u)$, and let $S_i = \mathrm{score}(U_i, V)$. For $i = 2, \dots, m$, let $B_i = \mathbb{1}\{S_i \geq S_1\}$, and let $L = \sum_{i=2}^{m} B_i$. Note that the event $\{L \geq 1\}$ includes the event $\{S_1 \text{ is not the highest score}\}$.

(a) Show that for any $r \geq 0$ and $i \geq 2$,

$$\mathbb{E}[B_i \mid U_1 = u_1, V = v] \leq \sum_u p_U(u) \left[\frac{p_{V|U}(v|u)}{p_{V|U}(v|u_1)}\right]^r.$$

Hint: For non-negative a, b, r, the inequality $\mathbb{1}\{a \geq b\} \leq (a/b)^r$ holds.

Solution: This follows by computing the expectation and using the hint, as

$$\begin{aligned}
\mathbb{E}[B_i \mid U_1 = u_1, V = v] &= \mathbb{E}[\mathbb{1}\{S_i \geq S_1\} \mid U_1 = u_1, V = v] \\
&= \mathbb{E}[\mathbb{1}\{\mathrm{score}(U_i, v) \geq \mathrm{score}(u_1, v)\}] \\
&= \sum_{u_i} p_U(u_i) \mathbb{1}\{p_{V|U}(v|u_i) \geq p_{V|U}(v|u_1)\} \\
&\leq \sum_{u_i} p_U(u_i) \frac{p_{V|U}(v|u_i)^r}{p_{V|U}(v|u_1)^r},
\end{aligned}$$

and we are done by changing the summation variable u_i to u.

(b) For $i \geq 2$, show that $\mathbb{E}[B_i] \leq \sum_v \left[\sum_u p_U(u) \sqrt{p_{V|U}(v|u)}\right]^2$.
Hint: Use (a) with a careful choice of r.

Solution: We write $\mathbb{E}[B_i]$ as the average of $\mathbb{E}[B_i \mid U_1 = u_1, V = v]$ as

$$\begin{aligned}
\mathbb{E}[B_i] &= \sum_{u_1, v} p_{UV}(u_1, v) \mathbb{E}[B_i \mid U_1 = u_1, V = v] \\
&\leq \sum_{u_1, v} p_{UV}(u_1, v) \sum_u p_U(u) \left[\frac{p_{V|U}(v|u)}{p_{V|U}(v|u_1)}\right]^r \\
&= \sum_{u_1, v} p_{V|U}(v|u_1) p_U(u_1) \sum_u p_U(u) \left[\frac{p_{V|U}(v|u)}{p_{V|U}(v|u_1)}\right]^r \\
&= \sum_v \left[\sum_{u_1} p_U(u_1) p_{V|U}(v|u_1)^{1-r}\right] \left[\sum_u p_U(u) p_{V|U}(v|u)^r\right].
\end{aligned}$$

Pick $r = \frac{1}{2}$, and we are done.

(c) Show that

$$\Pr(S_1 \text{ is not the highest score}) \leq (m-1) \sum_v \left[\sum_u p_U(u) \sqrt{p_{V|U}(v|u)}\right]^2.$$

Hint: $\Pr(L \geq 1) \leq \mathbb{E}[L]$.

Solution: This follows from

$$\begin{aligned}
\Pr(S_1 \text{ is not the highest score}) &\leq \Pr(L \geq 1) \\
&\leq \mathbb{E}[L] = \sum_{i=2}^{m} \mathbb{E}[B_i] \\
&\leq (m-1) \sum_v \Big[\sum_u p_U(u) \sqrt{p_{V|U}(v|u)} \Big]^2 .
\end{aligned}$$

Define $R_1(p_U, p_{V|U}) := -\log \sum_v \big[\sum_u p_U(u)\sqrt{p_{V|U}(v|u)}\big]^2$.

(d) With $p_{X^n}(x^n) = \prod_{i=1}^n p_X(x_i)$, and $p_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n p_{Y|X}(y_i|x_i)$, show that $R_1(p_{X^n}, p_{Y^n|X^n}) = nR_1(p_X, p_{Y|X})$.

Solution: Consider

$$\begin{aligned}
R_1(p_{X^n}, p_{Y^n|X^n}) &= -\log \sum_{y^n} \Big[\sum_{x^n} p_{X^n}(x^n) \sqrt{p_{Y^n|X^n}(y^n|x^n)} \Big]^2 \\
&= -\log \sum_{y^n} \Big[\sum_{x^n} \prod_{i=1}^n p_X(x_i) \sqrt{\prod_{i=1}^n p_{Y|X}(y_i|x_i)} \Big]^2 \\
&= -\log \sum_{y^n} \Big[\sum_{x^n} \prod_{i=1}^n p_X(x_i) \sqrt{p_{Y|X}(y_i|x_i)} \Big]^2 \\
&= -\log \sum_{y^n} \Big[\prod_{i=1}^n \sum_{x_i} p_X(x_i) \sqrt{p_{Y|X}(y_i|x_i)} \Big]^2 \\
&= -\log \sum_{y^n} \prod_{i=1}^n \Big[\sum_{x_i} p_X(x_i) \sqrt{p_{Y|X}(y_i|x_i)} \Big]^2 \\
&= -\log \prod_{i=1}^n \sum_{y_i} \Big[\sum_{x_i} p_X(x_i) \sqrt{p_{Y|X}(y_i|x_i)} \Big]^2 \\
&= -\sum_{i=1}^n \log \sum_{y_i} \Big[\sum_{x_i} p_X(x_i) \sqrt{p_{Y|X}(y_i|x_i)} \Big]^2 \\
&= -\sum_{i=1}^n \log \sum_{y} \Big[\sum_{x} p_X(x) \sqrt{p_{Y|X}(y|x)} \Big]^2 = nR_1(p_X, p_{Y|X})
\end{aligned}$$

(e) Given a channel $p_{Y|X}$ and input distribution p_X, show that for every $0 \leq R < R_1(p_X, p_{Y|X}) =: R_1$, and positive integer n, there is a code with $m = \lceil 2^{nR} \rceil$ codewords and with average probability of error $\bar{p}_e \leq 2^{-n(R_1 - R)}$.

Hint: Choose m codewords $X^n(1), \ldots, X^n(m)$, i.i.d. from distribution p_{X^n}. Make use of what you already showed in (d) and (c).

Solution: As suggested in the hint, choose $m = \lceil 2^{nR} \rceil$ codewords $X^n(1), \ldots, X^n(m)$, i.i.d. from distribution p_{X^n}. The encoder maps the message $j = 1, \ldots, m$ to the codeword $X^n(j)$. Upon receiving Y^n, the decoder computes $S_i = \text{score}(X^n(i), Y^n)$ for each $i = 1, \ldots, m$, and declares that $\hat{j} = \arg\max_{i=1,\ldots,m} S_i$ was sent (in case of ties, decide arbitrarily). Without loss of generality, assume that the message 1 was sent.

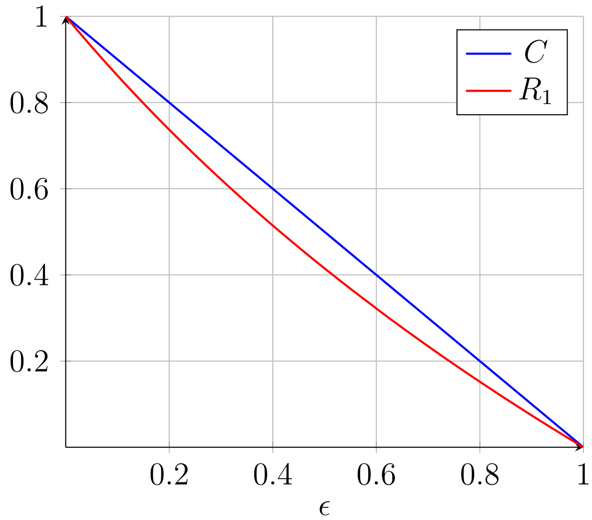
The average probability of error $\bar{p}_e$, averaged over the random codebook generation, is upper bounded by

$$\begin{aligned}\Pr(S_1 \text{ is not the highest score}) &\leq (m-1)\sum_{y^n}\left[\sum_{x^n} p_{X^n}(x^n)\sqrt{p_{Y^n|X^n}(y^n|x^n)}\right]^2 \\ &\leq 2^{nR}2^{-nR_1} = 2^{-n(R_1-R)}.\end{aligned}$$

Hence, since the average $\bar{p}_e$ over the choice of codewords is lesser than $2^{-n(R_1-R)}$, there exists a code with $\bar{p}_e \leq 2^{-n(R_1-R)}$ and $m = \lceil 2^{nR} \rceil$ codewords.

(f) With $p_{Y|X}$ being the Binary Erasure Channel and for p_X the uniform distribution on the input alphabet, compute and sketch R_1 (defined above) and C (the channel capacity) as a function of the erasure probability. Comment on the plots obtained.

Solution: When $p_{Y|X}$ is a BEC with erasure probability ϵ, by simply computing the above expression for R_1, we have $R_1 = -\log \frac{1+\epsilon}{2}$. The channel capacity for the BEC is given by $C = 1 - \epsilon$. Clearly, $R_1 \leq C$, with a strict inequality except for $\epsilon = 0$ or 1. In (e), we showed that for rates below R_1, we can achieve an exponential decay of $\bar{p}_e$. We now see that there are rates below capacity at which it is still, at this point in the problem, unclear whether or not we can achieve exponential decay (all we know thanks to the channel coding theorem is that the probability of error can be made to decay to zero for rates below the capacity; we do not know whether this decay is exponential in n).



(g) Continuing with the notation of (a)–(c), for any $r \geq 0$, and for any $0 \leq \rho \leq 1$, show that

$$\mathbb{E}[L^\rho \mid U_1 = u_1, V = v] \leq (m-1)^\rho \left(\sum_u p_U(u)\left[\frac{p_{V|U}(v|u)}{p_{V|U}(v|u_1)}\right]^r\right)^\rho.$$

Hint: Use of the bound you found in (a) to upper bound $\mathbb{E}[L \mid U_1 = u_1, V = v]$; note that $z \in [0,\infty) \mapsto z^\rho$ is concave, so, $\mathbb{E}[Z^\rho] \leq \mathbb{E}[Z]^\rho$.

Solution: Since $z \mapsto z^\rho$ is concave for $0 \le \rho \le 1$, by Jensen's inequality,

$$\begin{aligned}
\mathbb{E}[L^\rho \mid U_1 = u_1, V = v] \le \mathbb{E}[L \mid U_1 = u_1, V = v]^\rho &= \mathbb{E}\left[\sum_{i=2}^m B_i \,\middle|\, U_1 = u_1, V = v\right]^\rho \\
&= \left(\sum_{i=2}^m \mathbb{E}[B_i \mid U_1 = u_1, V = v]\right)^\rho \\
&\le \left((m-1)\sum_u p_U(u)\left[\frac{p_{V|U}(v|u)}{p_{V|U}(v|u_1)}\right]^r\right)^\rho \\
&= (m-1)^\rho\left(\sum_u p_U(u)\left[\frac{p_{V|U}(v|u)}{p_{V|U}(v|u_1)}\right]^r\right)^\rho.
\end{aligned}$$

(h) For any $0 \le \rho \le 1$, and $r \ge 0$, show that

$$\mathbb{E}[L^\rho] \le (m-1)^\rho \sum_v \Big[\sum_{u'} p_U(u')p_{V|U}(v|u')^{1-r\rho}\Big]\Big[\sum_u p_U(u)p_{V|U}(v|u)^r\Big]^\rho.$$

Hint: Use (g).

Solution: Simply computing, we have

$$\begin{aligned}
\mathbb{E}[L^\rho] &= \sum_{u_1,v} p_{UV}(u_1,v)\mathbb{E}[L^\rho \mid U_1 = u_1, V = v] \\
&\le \sum_{u_1,v} p_{UV}(u_1,v)(m-1)^\rho\left[\sum_u p_U(u)\left(\frac{p_{V|U}(v|u)}{p_{V|U}(v|u_1)}\right)^r\right]^\rho \\
&\le (m-1)^\rho \sum_{u_1,v} p_U(u_1)p_{V|U}(v|u_1)\left[\sum_u p_U(u)\left(\frac{p_{V|U}(v|u)}{p_{V|U}(v|u_1)}\right)^r\right]^\rho \\
&= (m-1)^\rho \sum_v \Big[\sum_{u_1} p_U(u_1)p_{V|U}(v|u_1)^{1-r\rho}\Big]\Big[\sum_u p_U(u)p_{V|U}(v|u)^r\Big]^\rho,
\end{aligned}$$

and we are done by changing the summation variable from u_1 to u'.

(i) For any $0 \le \rho \le 1$, show that

$$\mathbb{E}[L^\rho] \le (m-1)^\rho \sum_v \Big[\sum_u p_U(u)p_{V|U}(v|u)^{1/(1+\rho)}\Big]^{1+\rho}.$$

Hint: Examine (h) for the choice $r = 1/(1+\rho)$.

Solution: As suggested in the hint, substituting $r = 1/(1+\rho)$ into the result of part (h) gives the desired result.

For $0 < \rho \le 1$, define $R_\rho(p_U, p_{V|U}) := -\rho^{-1}\log \sum_v \Big[\sum_u p_U(u)p_{V|U}(v|u)^{1/(1+\rho)}\Big]^{1+\rho}$. (Observe that setting $\rho = 1$ recovers R_1.)

(j) Given a channel $p_{Y|X}$ and input distribution p_X, show that for every $0 \le R < R_\rho(p_X, p_{Y|X}) =: R_\rho$, positive integer n, there is a code with $m = \lceil 2^{nR} \rceil$ codewords and with average probability of error $\bar{p}_e \le 2^{-n\rho(R_\rho - R)}$.

Hint: Observe that $\Pr(L \ge 1) \le \mathbb{E}[L^\rho]$ and follow the reasoning in (d) and (e).

Solution: First observe that $R_\rho(p_{X^n}, p_{Y^n|X^n}) = nR_\rho(p_X, p_{Y|X})$. As in (e), again pick the $m = \lceil 2^{nR} \rceil$ codewords $X^n(1), \dots, X^n(m)$, i.i.d. from distribution p_{X^n}. The encoder maps the message $j = 1, \dots, m$ to the codeword $X^n(j)$. The decoder receives Y^n and computes $S_i = \text{score}(X^n(i), Y^n)$ for each $i = 1, \dots, m$, and declares that $\hat{j} = \arg\max_{i=1,\dots,m} S_i$ was sent (deciding arbitrarily in case of ties). Again, assuming w.l.o.g. that message 1 was sent, the average value of $\bar{p}_e$ (over the choice of codewords) is upper bounded by the probability that S_1 is not the highest, which is further upper bounded by

$$\begin{aligned}
\Pr(L \geq 1) &\leq \mathbb{E}[L^\rho] \\
&\leq (m-1)^\rho \sum_v \Big[\sum_u p_U(u) p_{V|U}(v|u)^{1/(1+\rho)} \Big]^{1+\rho} \\
&\leq 2^{nR\rho} 2^{-nR_\rho \rho} = 2^{-n\rho(R_\rho - R)}.
\end{aligned}$$

Hence, there is a code with $m = \lceil 2^{nR} \rceil$ codewords and $\bar{p}_e \leq 2^{-n\rho(R_\rho - R)}$.

(k) Show that $\lim_{\rho \to 0^+} R_\rho(p_U, p_{V|U}) = I(U;V)$. Conclude from this and (j) that for any channel $p_{Y|X}$ and $R < C(p_{Y|X})$ there is a number $\beta > 0$ such that, for every positive integer n there is a code for the channel with $m = \lceil 2^{nR} \rceil$ codewords and with error probability $\bar{p}_e \leq 2^{-n\beta}$.

Solution: We first show that $\lim_{\rho \to 0^+} R_\rho(p_U, p_{V|U}) = I(U;V)$. Starting from the left-hand side, observing that it is of the $\frac{0}{0}$-form, and applying L'Hôpital's rule, $\lim_{\rho \to 0^+} R_\rho(p_U, p_{V|U})$ equals

$$\lim_{\rho \to 0^+} -\frac{d}{d\rho} \log \sum_v \Big[\sum_u p_U(u) p_{V|U}(v|u)^{1/(1+\rho)} \Big]^{1+\rho}.$$

Define $F(\rho) := \sum_v \big[\sum_u p_U(u) p_{V|U}(v|u)^{1/(1+\rho)} \big]^{1+\rho}$, and note that $F(0) = 1$. Using the relation $\frac{d}{d\rho} \log F(\rho) = \frac{\log e}{F(\rho)} \frac{d}{d\rho} F(\rho)$, we see that $\lim_{\rho \to 0^+} R_\rho = -\log(e) \frac{dF(\rho)}{d\rho}\big|_{\rho=0}$. To evaluate $dF/d\rho$, observe that $F(\rho)$ is of the form $\sum_v f(v,\rho)^{1+\rho}$, with each $f(v,\rho)$ the sum $\sum_u p_U(u) p_{V|U}(v|u)^{1/(1+\rho)}$. Further observe that $f(v,0) = p_V(v)$.

Use the relations $d(f(\rho)^{1+\rho})/d\rho = f(\rho)^{1+\rho}[\ln f(\rho) + (1+\rho) f(\rho)^{-1} f'(\rho)]$ — at $\rho = 0$ this equals $f(0)[\ln f(0) + f'(0)]$ — and $d(z^{1/(1+\rho)})/d\rho = -(1+\rho)^{-2} z^{1/(1+\rho)} \ln z$ — at $\rho = 0$ this equals $-z \ln z$ — to find

$$\frac{dF}{d\rho}\Big|_{\rho=0} = \sum_v \Big[p_V(v) \ln p_V(v) - \sum_u p_U(u) p_{V|U}(v|u) \ln p(v|u) \Big].$$

We recognize the right hand side as $[-H(V) + H(V|U)]/\log(e)$. Consequently, $\lim_{\rho \to 0+} R_\rho(p_U, p_{V|U}) = H(V) - H(V|U) = I(U;V)$.

Since $R < C(p_{Y|X})$, there is a p_X for which $R < I(X;Y)$. Moreover, as $I(X;Y) = \lim_{\rho \to 0^+} R_\rho(p_X, p_{Y|X})$, there is a $\rho > 0$ for which $R < R_\rho(p_X, p_{Y|X})$. The claim now follows with $\beta = \rho(R_\rho(p_X, p_{Y|X}) - R)$ and (j).