ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 26 | Information Theory and Coding
Graded Homework, due Monday, Dec. 16, 2024 | Dec. 02, 2024

You are allowed (even encouraged) to discuss the problems on the homework with your colleagues. However, your solutions should be in your own words. If you collaborated on your solution, write down the name of your collaborators and your sources; no points will be deducted. But similarities in solutions beyond the listed collaborations will be considered as malpractice.

Notation:

$\exp_2(a) := 2^a$.

for a positive integer m, $[m] := \{1, \dots, m\}$.

PROBLEM 1. Suppose $\mathrm{scr}\colon \mathcal{A} \times \mathcal{B} \to \mathbb{R}$. We will call scr to be a 'score function'. Suppose $a(1), \dots, a(m)$ are elements of $\mathcal{A}$. For this collection, the *maximum score decoder* is the function $\mathrm{dec}_{\mathrm{scr}}\colon \mathcal{B} \to \{0, 1, \dots, m\}$ given by

$$\mathrm{dec}_{\mathrm{scr}}(b) = \arg\max_{i \in [m]} \mathrm{scr}(a(i), b),$$

if there is a unique maximizer; if not we set $\mathrm{dec}_{\mathrm{scr}}(b) = 0$ (i.e., in the case of a tie the decoder says "I can't decide").

(a) Suppose two score functions scr and scr′ are such that $\mathrm{scr}(a, b) - \mathrm{scr}'(a, b)$ is only a function of b. What can you say about $\mathrm{dec}_{\mathrm{scr}}$ and $\mathrm{dec}_{\mathrm{scr}'}$?

(b) Suppose two score functions scr and scr′ are such that $\mathrm{scr}(a, b) = \lambda\, \mathrm{scr}'(a, b)$ for a given constant $\lambda > 0$. What can you say about $\mathrm{dec}_{\mathrm{scr}}$ and $\mathrm{dec}_{\mathrm{scr}'}$?

For $t \in \mathbb{R}$, consider the *threshold decoder* $\mathrm{dec}_{t,\mathrm{scr}}$ that operates as follows: given b, form the list $L_t = \{i : \mathrm{scr}(a(i), b) \geq t\}$. If L_t consists of a single element i_0, we set $\mathrm{dec}_{t,\mathrm{scr}}(b) = i_0$. Otherwise $\mathrm{dec}_{t,\mathrm{scr}}(b) = 0$.

(c) Show that if $\mathrm{dec}_{t,\mathrm{scr}}(b) \neq 0$, then $\mathrm{dec}_{\mathrm{scr}}(b) = \mathrm{dec}_{t,\mathrm{scr}}(b)$.
Moral: If the threshold decoder makes the correct decision, so does the maximum score decoder.

PROBLEM 2. Let p_{XY} be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. Suppose the triple $(\tilde{X}, X, Y)$ has a joint distribution given by $p_{\tilde{X}XY}(\tilde{x}, x, y) = p_X(\tilde{x}) p_{XY}(x, y)$, i.e., (X, Y) is drawn according to p_{XY} and $\tilde{X}$ is independent of the pair (X, Y) but with the same marginal distribution as X. Suppose $\big((\tilde{X}_i, X_i, Y_i) : i = 1, 2, \dots\big)$ is a collection of i.i.d. random triples with distribution $p_{\tilde{X}XY}$. Given $s\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, let $\mathrm{scr}(x^n, y^n) = \sum_{i=1}^n s(x_i, y_i)$.

(a) Let $t_0 = \mathbb{E}\left[s(X, Y)\right] = \sum_{x,y} p_{XY}(x, y) s(x, y)$. Show that for any $t < t_0$,

$$\lim_{n \to \infty} \Pr(\mathrm{scr}(X^n, Y^n) < nt) = 0.$$

(b) Show that

$$\Pr(\mathrm{scr}(\tilde{X}^n, Y^n) \geq nt) \leq \exp_2[-n(t-\alpha)],$$

where $\alpha = \log_2 \mathbb{E}\left[\exp_2\left(s(\tilde{X}, Y)\right)\right] = \log_2 \sum_{x,y} p_X(x) p_Y(y) \exp_2(s(x,y))$.

Hint: $\mathbb{1}\{Z \geq z\} \leq \exp_2(Z)\exp_2(-z)$.

Given a channel $p_{Y|X}$, a probability distribution p_X, a blocklength n, and a rate R, set $m = \lceil 2^{nR} \rceil$, and construct a random encoder enc: $[m] \to \mathcal{X}^n$ by setting $\mathrm{enc}(i) = (E_{i1}, \dots, E_{in})$, where $(E_{ij} : i \in [m]\, j \in [n])$ is a collection of i.i.d. random variables with distribution p_X. Let the decoder be the threshold decoder $\mathrm{dec}_{nt,\mathrm{scr}}$ using the score function scr as defined above and threshold nt.

Let W be the transmitted message (uniformly chosen in $[m]$), and let $\hat{W}$ be the threshold decoder's decision.

(c) With $\tilde{X}^n$, X^n and Y^n as in the first paragraph, show that

$$\begin{aligned}\Pr(\hat{W} \neq W) &\leq \Pr(\mathrm{scr}(X^n, Y^n) < nt) + (m-1)\Pr(\mathrm{scr}(\tilde{X}^n, Y^n) \geq nt) \\ &\leq \Pr(\mathrm{scr}(X^n, Y^n) < nt) + 2^{nR}\Pr(\mathrm{scr}(\tilde{X}^n, Y^n) \geq nt).\end{aligned}$$

(d) With t_0 and α as in (a) and (b), show that for $R < t_0 - \alpha$, and $\epsilon > 0$, there is a choice of t such that for n large enough, $\Pr(\hat{W} \neq W) < \epsilon$.

Hint: Since $R < t_0 - \alpha$, there is a t strictly in between $R + \alpha$ and t_0. Now use (a), (b), and (c).

(e) Again, with t_0 and α as in (a) and (b), show that for $R < t_0 - \alpha$ and $\epsilon > 0$, for large enough n, there is an encoder enc with $m = \lceil 2^{nR} \rceil$ codewords, and a maximum score decoder $\mathrm{dec}_{\mathrm{scr}}$ using the score function scr with average error probability at most ϵ.

(f) Evaluate t_0 and α for the choice $s(x,y) := \log_2 \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}$. Use (e) together with 1(a) to conclude that for all rates $R < I(X;Y)$, there is an encoder of rate at least R with the decoder using the maximum likelihood rule $\mathrm{dec}(y^n) = \arg\max_{i\in[m]} p_{Y^n|X^n}\big(y^n \mid \mathrm{enc}(i)\big)$, which achieves an average error probability that can be made arbitrarily small.

(g) Suppose now the decoder uses the maximum likelihood rule adapted to the channel q instead of the true channel $p_{Y|X}$. Show that all rates up to

$$\max_{r_Y,\lambda} \left[\sum_{x,y} p_X(x) p_{Y|X}(y|x) \log_2 \frac{q(y|x)^\lambda}{r_Y(y)} - \log_2 \sum_{x,y} p_X(x) p_Y(y) \frac{q(y|x)^\lambda}{r_Y(y)} \right]$$

can be achieved with this (mismatched Maximum Likelihood) decoder. (The maximization is over all probability distributions r_Y and $\lambda > 0$.)

Hint: Use 1(b) and 1(c).

(h) Suppose our channel $p_{Y|X}$ is the a binary input binary output channel $p(1|0) = \delta_0$ and $p(0|1) = \delta_1$, with $\delta_i \leq 1/2$. Suppose the decoder is using the ML rule adapted to the BSC(ϵ) with $\epsilon < 1/2$. Show that all rates up to $1 - h_2\left(\frac{\delta_0+\delta_1}{2}\right)$ can be achieved by a suitable choice of p_X at the encoder.

Hint: Choose p_X, r_Y to be uniform distributions in (g) and find the maximizing λ.