ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 25 Information Theory and Coding
Graded Homework, due Friday December 10 Nov. 30, 2021

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 cheating.

PROBLEM 1. We are given a channel $p_{Y|X}$ and a distribution p_X on the input. Let $p_Y(y) = \sum_x p_X(x) p_{Y|X}(y|x)$ be the distribution of Y. Let q_Y be an arbitrary probability distribution on $\mathcal{Y}$.

(a) Show that $\sum_x p_X(x) D(p_{Y|X=x} \| q_Y) \geq I(X;Y)$ $(*)$.

(b) Let $F(p_X, q_Y, p_{Y|X})$ denote the left hand side of $(*)$. Show that for any q_Y

$$\max_{p_X} F(p_X, q_Y, p_{Y|X}) \geq C(p_{Y|X})$$

where $C(p_{Y|X})$ is the capacity of the channel.

(c) Show that for any q_Y, $\max_x D(p_{Y|X=x} \| q_Y) \geq C(p_{Y|X})$.

(d) Show that $\min_{q_Y} \max_x D(p_{Y|X=x} \| q_Y) = C(p_{Y|X})$.

[Hint: let p_X maximize $I(X;Y)$ and consider $q_Y = p_Y$, the distribution of Y when input has distribution p_X.]

PROBLEM 2. Suppose an information source $Y_1, Y_2, \ldots$ has a distribution that is parametrized by a parameter x in $\mathcal{X}$. We adopt the notations $p_{Y^n|X=x}$ and $H(Y^n|X=x)$ to denote the distribution and entropy of Y^n when the parameter has the value x.

We aim to design a source code $c_n : \mathcal{Y}^n \to \{0,1\}^*$ for the source Y^n.

(a) Suppose that the value x of the parameter is globally known, in particular, we can design our source code c_n with the knowledge of x. Show that for any uniquely decodable (u.d.) c_n, $E\big[\text{length}\big(c_n(Y^n)\big)\big] \geq H(Y^n|X=x)$.

For the rest of the problem the value x of the parameter is unknown. A reasonable goal in this case is to evaluate a source code c_n by

$$\text{regret}(c_n) := \max_x \tfrac{1}{n}\Big[E\big[\text{length}\big(c_n(Y^n)\big)\big|X=x\big] - H(Y^n|X=x)\Big].$$

In words, for each value of the parameter x we find the additional number of bits per letter c_n produces compared to an ideal code designed with the knowledge of x; regret(c_n) is the worst case (over x) of this penalty.

(b) Show that for any u.d. c_n there is a distribution q_{Y^n} on $\mathcal{Y}^n$ such that

$$\text{regret}(c_n) \geq \max_x \tfrac{1}{n} D(p_{Y^n|X=x} \| q_{Y^n}).$$

(c) Show that for any probability distribution q_{Y^n} on $\mathcal{Y}^n$ there is a prefix-free code c_n such that

$$\text{regret}(c_n) \leq \max_x \tfrac{1}{n} D(p_{Y^n|X=x} \| q_{Y^n}) + \tfrac{1}{n}.$$

(d) Let $R_n = \min\{\text{regret}(c_n) : c_n \text{ u.d.}\}$. Let $D_n = \min_{q_{Y^n}} \max_x \frac{1}{n} D(p_{Y^n|X=x} \| q_{Y^n})$. What is $\lim_{n\to\infty}(R_n - D_n)$?

(e) Let C_n denote the capacity of the channel with input X and output Y^n. What is $\lim_{n\to\infty}\big(R_n - \frac{1}{n}C_n\big)$?

(f) Suppose $|\mathcal{X}|$ is finite. What is $\lim_{n\to\infty} R_n$?

(g) Suppose $\mathcal{X} = [0,1]$, $\mathcal{Y} = \{0,1\}$ and when $X = x$, $Y_1, Y_2, \ldots$ are i.i.d. with $\Pr(Y_i = 1) = x$. Let $T = \sum_{i=1}^n Y_i$. Show that

  (i) $X - T - Y^n$

  (ii) $I(X; Y^n) = I(X; T)$.

(h) In the setting of (g) show that $R_n \leq \big(1 + \log(1+n)\big)/n$. [Hint: How many values can T take?]