ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 15 | Information Theory and Coding
Midterm exam | Oct. 29, 2024

3 problems, 43 points, 180 minutes.
1 sheet (2 pages) of notes allowed.

Good Luck!

PLEASE WRITE YOUR NAME ON EACH SHEET OF YOUR ANSWERS.

PLEASE WRITE THE SOLUTION OF EACH PROBLEM ON A SEPARATE SHEET.

(All logarithms are taken to the base 2.)

PROBLEM 1. (16 points)

Suppose $f : [0, \infty) \to \mathbb{R} \cup \{\pm\infty\}$ is a decreasing, convex function, and p and q are probability distributions on an alphabet $\mathcal{U}$ (i.e., $p(u) \geq 0$ and $\sum_{u \in \mathcal{U}} p(u) = 1$, similarly for q). Define

$$K_f(p, q) = \sum_{u \,:\, p(u) > 0} p(u)\, f\left(\frac{q(u)}{p(u)}\right).$$

(a) (2 pts) Show that $K_f(p, q) \geq f(1)$, and equality happens if $q = p$.
Hint: Make sure to use convexity.

Suppose U is a random variable with distribution p. A "prediction" about U is a probability distribution q on $\mathcal{U}$ — basically saying "I believe we will see the value u with probability $q(u)$". A prediction q is assigned a score via $\text{score}(q) = \mathbb{E}\left[\frac{1}{q(U)}\right] = \sum_u \frac{p(u)}{q(u)}$.

(b) (3 pts) Let $p_{1/2}(u) = \frac{p(u)^{1/2}}{A}$, where $A = \sum_u p(u)^{1/2}$ to ensure that $p_{1/2}$ is a probability distribution. Show that for any probability distribution q, $\text{score}(q) \geq A^2$, with equality if $q = p_{1/2}$.
Hint: First show that with $f(x) = 1/x$, $\text{score}(q) = A^2 K_f(p_{1/2}, q)$.

(c) (3 pts) Suppose $c : \mathcal{U} \to \{0, 1\}^*$ is a uniquely decodable code. Show that $\mathbb{E}[2^{\text{length}(c(U))}] \geq A^2$.

Fix $\alpha > 0$.

(d) (3 pts) Replace the score function above with $\text{score}_\alpha(q) = \mathbb{E}[q(U)^{-\alpha}] = \sum_u \frac{p(u)}{q(u)^\alpha}$. Show that for any q, $\text{score}_\alpha(q) \geq (A_{1/1+\alpha})^{1+\alpha}$, with equality if $q = p_{1/1+\alpha}$, where we define $p_s(u) = p(u)^s / A_s$ where $A_s = \sum_u p(u)^s$.
Hint: Choose f appropriately and express $\text{score}_\alpha(q)$ in terms of $K_f(p_s, q)$ for some s.

(e) (2 pts) Show that for any uniquely decodable code $c : \mathcal{U} \to \{0, 1\}^*$,

$$\mathbb{E}[2^{\alpha\, \text{length}(c(U))}] \geq (A_{1/1+\alpha})^{1+\alpha}.$$

(f) (3 pts) Show that there exists a prefix-free code $c : \mathcal{U} \to \{0, 1\}^*$ such that

$$\mathbb{E}[2^{\alpha\, \text{length}(c(U))}] \leq 2^\alpha (A_{1/1+\alpha})^{1+\alpha}.$$

Problem 2. (12 points)

For this problem, we define the following notation different from that used in the lectures. Fix a natural number n. Let $(X_1, \ldots, X_n)$ be a vector of binary random variables, with each X_i taking values in $\{0,1\}$. For $i, j = 1, \ldots, n$, let $X_i^j = (X_i, \ldots, X_j)$ if $i \leq j$ and empty if $i > j$. Let $X_{\neq i}$ denote the vector X_1^n without the i^{th} element, i.e., $X_{\neq i} = (X_1^{i-1}, X_{i+1}^n)$. Also let $X_{(\bar{i})}$ denote the vector X_1^n with its i^{th} element flipped, i.e., $X_{(\bar{i})} = (X_1^{i-1}, 1 - X_i, X_{i+1}^n)$.

(a) (3 pts) Show that $\sum_{i=1}^n H(X_i \mid X_{\neq i}) \leq H(X_1^n)$.

Let A be a subset of $\{0,1\}^n$, i.e., A consists of binary vectors of length n. Denote by $E(A)$ the set of pairs of vectors in A that differ at *exactly* one position, i.e.,

$$\begin{aligned} E(A) &= \{(x_1^n, \tilde{x}_1^n) \in A \times A \text{ such that } \tilde{x}_i \neq x_i \text{ for exactly one } i\} \\ &= \{(x_1^n, \tilde{x}_1^n) \in A \times A \text{ such that } \tilde{x}_1^n = x_{(\bar{i})} \text{ for some } i\}. \end{aligned}$$

Let $(X_1, \ldots, X_n)$ be randomly and uniformly chosen from A.

(b) (3 pts) Fix $x_1^n \in A$. Compute $H(X_i \mid X_{\neq i} = x_{\neq i})$.
Hint: Consider two cases: $x_{(\bar{i})} \in A$ and $x_{(\bar{i})} \notin A$.

(c) (3 pts) Show that $H(X_i \mid X_{\neq i}) = \frac{1}{|A|} \sum_{x_1^n \in A} \mathbb{1}\{x_{(\bar{i})} \in A\}$.

(d) (3 pts) Show that $\sum_{i=1}^n H(X_i \mid X_{\neq i}) = \frac{|E(A)|}{|A|}$ and conclude that $|E(A)| \leq |A| \log |A|$.
Hint: Use (a).

Problem 3. (15 points)

(a) (2 pts) Suppose p is a probability distribution on $\mathcal{U}$. Show that for any probability distribution q on $\mathcal{U}$, $\max_{u \in \mathcal{U}} \log \frac{p(u)}{q(u)} \geq 0$. Additionally, show that $\min_q \max_{u \in \mathcal{U}} \log \frac{p(u)}{q(u)} = 0$, where the minimization is over all probability distributions q on $\mathcal{U}$.

(b) (2 pts) Show that $\min_q \max_{u \in \mathcal{U}} \log \frac{f(u)}{q(u)} = \log K$, where $K = \sum_u f(u)$ for a nonnegative function f.

Hint: Use (a).

Suppose from now on that for every θ in some parameter set Θ, we have a probability distribution p_θ on $\mathcal{U}$.

(c) (2 pts) Show that $\min_q \max_{u \in \mathcal{U}, \theta \in \Theta} \log \frac{p_\theta(u)}{q(u)} = S$, where $S = \log \sum_{u \in \mathcal{U}} \max_{\theta \in \Theta} p_\theta(u)$.

Hint: Use (b).

(d) (3 pts) Suppose we do not know the probability distribution of a random variable U, except that the distribution is one of the p_θ above. Show that there is a prefix-free code $c : \mathcal{U} \to \{0,1\}^*$ such that, for every $\theta \in \Theta$ and every $u \in \mathcal{U}$, $\operatorname{length} c(u) \leq \log \frac{1}{p_\theta(u)} + S + 1$, where S is as in part (c) above.

Suppose we know that $U_1, U_2, \ldots,$ are i.i.d. Bernoulli(θ) random variables, but we do not know the value of $\theta \in [0,1]$. For $u^n \in \{0,1\}^n$, define $p_\theta(u^n) = \theta^{k(u^n)}(1-\theta)^{n-k(u^n)}$, where $k(u^n)$ is the number of 1's in the sequence $(u_1, \ldots, u_n)$. With this definition, $\Pr(U^n = u^n) = p_\theta(u^n)$.

(e) (3 pts) Show that for any u^n, we have $\max_{\theta \in [0,1]} p_\theta(u^n) = \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k}$ with $k = k(u^n)$, and conclude that

$$\sum_{u^n \in \{0,1\}^n} \max_{\theta \in [0,1]} p_\theta(u^n) = \sum_{i=0}^{n} \binom{n}{i} \left(\frac{i}{n}\right)^i \left(1 - \frac{i}{n}\right)^{n-i}.$$

Hint: Differentiate $\log p_\theta(u^n)$ with respect to θ.

(f) (3 pts) Show that for each n, there is a prefix-free code $c_n : \{0,1\}^n \to \{0,1\}^*$ such that, for every $\theta \in [0,1]$ and every $u^n \in \{0,1\}^n$,

$$\operatorname{length} c_n(u^n) \leq \log \frac{1}{p_\theta(u^n)} + \log(1+n) + 1.$$

Hint: Use (d) and (e).