CO 739 Information Theory and Applications
University of Waterloo, Winter 2024
Instructor: Ashwin Nayak

Assignment 2, Mar. 1, 2024
Due: Fri., Mar. 15, 2024

Question 1. Let XY be jointly distributed random variables on the sample space $\mathcal{X} \times \mathcal{Y}$. For $x \in \mathcal{X}$, let $p(x)$ denote the distribution of $Y|(X = x)$. Let q be an arbitrary distribution over $\mathcal{Y}$.

(a) Prove that
$$\mathrm{I}(X : Y) \quad \leq \quad \mathbb{E}_{x \leftarrow X}\, \mathrm{S}(p(x) \| q) \ .$$

(b) Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of jointly distributed random variables, distributed as q. Suppose that $X_1 X_2 \cdots X_n \sim (Y_1 Y_2 \cdots Y_n)|E$, for some event E with $\Pr(E) \geq 2^{-\delta n}$. Prove that
$$\mathrm{S}(X_1 X_2 \cdots X_n \,\|\, Y_1 Y_2 \cdots Y_n) \quad \leq \quad \delta n \ ,$$
and hence that there is an index i such that $X_i \sim q'$, where q' is a distribution close to q in ℓ_1 distance: $\|q' - q\|_1 \in \mathrm{O}(\sqrt{\delta}\,)$.

Question 2. (a) Let XY and UV be two pairs of jointly distributed random variables with $X, U \in \mathcal{X}$ and $Y, V \in \mathcal{Y}$. Prove that $\mathfrak{h}(XY, UV) \geq \mathfrak{h}(X, U)$, i.e., Hellinger distance is monotonic under taking marginals.

(b) Let $X \coloneqq X_1 X_2 \cdots X_n$ be a random variable, with X_i being mutually independent. Let M be jointly distributed with X, and $S \subseteq [n]$ be a random variable independent of XM such that for every $i \in [n]$, we have $\Pr(i \in S) \leq p$. Prove that
$$\mathrm{I}(X_S : M | S) \quad \leq \quad p\, \mathrm{I}(X : M) \ .$$

Question 3. Let $\epsilon \in (0, 1)$. Let d be an even positive integer, and $a \in \{\pm 1\}^d$ with exactly $d/2$ coordinates equal to 1. Let $p(a)$ be the distribution over $[d]$ given by $p(a)_i \coloneqq \frac{1}{d}(1 + a_i \epsilon)$ for $i \in [d]$.

(a) Let $X(a) \sim p(a)$. Show that $\mathrm{H}(X(a)) \geq \log d - c\epsilon^2$ for a universal constant c.

(b) Let $\delta \in [0, 1)$. Suppose there is an algorithm that learns a with probability at least $1 - \delta$ given n independent samples from $p(a)$.
(i) Describe a protocol by which a sender can encode a string a as above using only samples from $p(a)$ so that the receiver can identify a with probability at least $1 - \delta$. What is the length of the encoding?
(ii) Derive as large a lower bound on n as you can.

Question 4. For $\epsilon \in (0, 1/2)$, we say a subset of strings $C \subseteq \{0, 1\}^n$ is an ϵ-*cover* if every n-bit string x is within Hamming distance ϵn from some element in C.

(a) Prove that any ϵ-cover C has cardinality $|C| \geq 2^{n(1 - \mathrm{h}(\epsilon))}$.

(b) Prove that for large enough n, a uniformly random subset of $\{0, 1\}^n$ of size $n^3\, 2^{(1 - \mathrm{h}(\epsilon))n}$ is an ϵ-cover with probability at least $1 - 2^{-\Omega(n)}$. (You may use without proof the inequality $\binom{n}{\epsilon n} \geq 2^{n\, \mathrm{h}(\epsilon)}/n$, for n large enough.)