Information Theory ECE 231A | UCLA Fall quarter 2022-23
Prof. Suhas Diggavi | Handout # 4, Monday, 26th September 2022

---

# Homework Set #1

Due 11th October 2022, before 11:59pm.

Submit your solutions to Gradescope with Entry Code:**57DN5B**

---

## Problem 1 (Slackness in Kraft Inequality)

An instantaneous code has word lengths $l_1, l_2, \cdots, l_m$, which satisfy the *strict* inequality

$$\sum_{i=1}^{m} D^{-l_i} < 1.$$

The code alphabet is $\mathcal{D} = \{0, 1, 2, \cdots D-1\}$. Show that there exist arbitrarily large sequence of code symbols which cannot be decoded into sequence of codewords.

## Problem 2 (How many fingers has a Martian? )

Let

$$S = \begin{pmatrix} S_1, \cdots, S_m \\ p_1, \cdots, p_m \end{pmatrix}.$$

The $S_i$'s are encoded into strings from a $D$-symbol output alphabet in a uniquely decodable manner. If $m = 6$ and the codeword lengths are $(l_1, l_2, \cdots, l_6) = (1, 1, 2, 3, 2, 3)$, find a good lower bound on $D$. You may wish to explain the title of the problem.

## Problem 3 (Value of questions)

Let $X \sim p(x)$, where $X \in \{1, \ldots, m\}$ and $S \subseteq \{1, \ldots, m\}$. We are allowed to ask a question of the type "Is $X$ in $S$?". Denote the binary answer to this question by $Y$, *i.e.*,

$$Y = \begin{cases} 1, & \text{if } X \in S \\ 0, & \text{if } X \notin S \end{cases}$$

(a) Suppose $\mathbb{P}[X \in S] = q$, find the decrease in uncertainty in $X$ after knowing $Y$, *i.e.,* what is $H(X) - H(X|Y)$?

(b) Assume $m = 5$ and $X$ has the pmf $\left\{\frac{2}{20}, \frac{3}{20}, \frac{3}{20}, \frac{5}{20}, \frac{7}{20}\right\}$. Suppose you could ask only one question in the above form. What question would you ask to obtain the most information about $X$?

(c) If you are allowed to ask multiple questions, what strategy would you use to minimize the expected number of questions for the pmf in the previous part?

## Problem 4 (Fix-free codes)

A code is said to be "fix-free" if it is both a prefix-free code and a suffix-free code. Recall that $s$ is a prefix of $t$ if $t$ is of the form $t = sv$, the concatenation of $s$ and $v$ for some string $v$. Similarly we say $s$ is a *suffix* of $t$ if $t = vs$. For example, the suffixes of "banana" are "a", "na", "ana", "nana, "anana" and "banana". Let $l_1, \dots, l_k$ be $k$ integers satisfying $l_1 \leq \dots \leq l_k$. In this problem we will try to prove that if $\sum_{i=1}^{k} 2^{-l_i} \leq \frac{1}{2}$ then there exists a binary fix-free code with codeword lengths $l_1, \dots, l_k$.

A code $\mathcal{C}$ is said to be a *fix-free code* if and only if no codeword is the prefix or the suffix of any other codeword. Consider the following algorithm:

- Initialize $A_i = \{0,1\}^{l_i}$ as the set of available codewords of length $l_i$ for every $1 \leq i \leq k$.

**for** $i = 1 \dots k$ **do**
    **if** $A_i \neq \emptyset$ **then**
        - Pick $\mathcal{C}(i) \in A_i$.
        **for** $j = i+1 \dots k$ **do**
            - $(*)$ Remove from $A_j$ all the words which start with $\mathcal{C}(i)$.
            - $(**)$ Remove from $A_j$ all the words which end with $\mathcal{C}(i)$.
        **end**
    **else**
        - Algorithm failure.
    **end**
**end**

- Return $\mathcal{C} = \{\mathcal{C}(i) : \ 1 \leq i \leq k\}$.

(a) For every $1 \leq i \leq k$ and every $i < j \leq k$, show that the number of words in $A_j$ that start with $\mathcal{C}(i)$ is $2^{l_j - l_i}$, and that the number of words in $A_j$ that end with $\mathcal{C}(i)$ is $2^{l_j - l_i}$.

(b) Show that the number of words that are removed from $A_j$ in $(*)$ and $(**)$ is at most $2^{l_j - l_i + 1}$.

(c) Show that if $\sum_{i=1}^{k} 2^{-l_i} \leq \frac{1}{2}$, then the returned code $\mathcal{C}$ is fix-free and that the lengths of its codewords are $l_1, \dots, l_k$.

*Hint:* Show that if $\sum_{i=1}^{k} 2^{-l_i} \leq \frac{1}{2}$, then the algorithm will not fail.

(d) Let $U$ be a random variable taking values in an alphabet $\mathcal{U}$. Show that there exists a fix-free code $\mathcal{C} : \mathcal{U} \longrightarrow \{0,1\}^*$ such that $H(U) \leq \mathbb{E}[\text{length}(\mathcal{C}(U))] \leq H(U) + 2$.

## Problem 5 (Entropy of disjoint mixture)

Let $X_1$ and $X_2$ be discrete random variables drawn according to the probability mass functions $\mathbf{p}(.)$ and $\mathbf{q}(.)$, respectively, i.e., $\mathbf{p}(i) = \mathbb{P}[X_1 = i]$ and $\mathbf{q}(j) = \mathbb{P}[X_2 = j]$. The random variable $X_1$ takes an odd value from the alphabet $\mathcal{X}_1 = \{1, 3, 5, \dots, m-1\}$, and the random variable $X_2$ takes an even value from the alphabet $\mathcal{X}_2 = \{2, 4, 6, \dots, m\}$, where $m$ is an even integer. Let

$$X = \begin{cases} X_1 & \text{w.p. } 1-\gamma \\ X_2 & \text{w.p. } \gamma \end{cases} \tag{1}$$

Let $Y = f(X)$ be a function of $X$ defined as follows:

$$Y = \begin{cases} 1 & \text{if } X \text{ is odd} \\ 2 & \text{if } X \text{ is even} \end{cases} \tag{2}$$

(a) Find the relation between $H(X)$ and $H(X, Y)$.

(b) Find $H(X)$ in terms of $H(X_1)$, $H(X_2)$, and $\gamma$.

(c) If $H(X_1) > H(X_2)$, find the value of $\gamma$ that will maximize $H(X)$.

## Problem 6 (Source Coding)

Consider a random variables $X$ that takes five values $\{A, B, C, D, E, F\}$ with probabilities $\mathbf{p} = \{\frac{4}{16}, \frac{4}{16}, \frac{3}{16}, \frac{2}{16}, \frac{2}{16}, \frac{1}{16}\}$

(a) Compute the entropy of the random variable $X$.

(b) Construct a binary Huffman code of the random variable $X$? What is the expected length of this code?