Telecom Paris Information Theory TP42

ASSIGNMENT 1

For Exercises 1-3 we use $\mathcal{C}$ to denote the code. The codeword symbols belong to $A = \{a, b\}$ and we use ε to denote the empty string.

Exercise 1 (Uniquely decodable and instantaneous codes)**.** For each of the following codes, determine if it is prefix-free. Which of these are uniquely decodable?

1. $\mathcal{C} = \{a, ba, bba, bbb\}$.
2. $\mathcal{C} = \{a, ab, abb, abbb\}$.
3. $\mathcal{C} = \{a, ab, ba\}$.
4. $\mathcal{C} = \{b, abb, abbba, bbba, baabb\}$.

Exercise 2 (Dangling suffixes)**.** For two sets E and D containing strings from alphabet A, define $E^{-1}D$ as the set of residual words obtained from D by removing some prefix that belongs to E. Formally,

$$E^{-1}D = \{y : xy \in D \text{ and } x \in E\}.$$

Calculate $\mathcal{C}^{-1}\mathcal{C}$ for the examples above.

Exercise 3 (Test for unique decodability)**.** Define the recursion

$$\begin{aligned} V_1 &= \mathcal{C}^{-1}\mathcal{C} \setminus \{\varepsilon\}, \\ V_{n+1} &= \mathcal{C}^{-1}V_n \cup V_n^{-1}\mathcal{C}, \ \ n \geq 1. \end{aligned}$$

Continue the recursion until $V_n \ni \varepsilon$; if not, until $V_n = V_m$ for some $m < n$.

1. For which of the above examples does the recursion terminate due to the first condition? Conclude that this happens if and only if the code is not uniquely decodable.
2. Does the above recursion terminate always? What is the complexity of the above algorithm in terms of the number of codewords and their lengths?

Exercise 4. (Uniquely decodable codes) Given an alphabet $\mathcal{X} = \{1, \ldots, m\}$ and a probability distribution $P = (p_1, \ldots, p_m)$ on $\mathcal{X}$, solve (using Lagrange multipliers) the following convex optimization problem:

$$\min_{\ell_1, \ldots, \ell_m \in \mathbb{R}} \sum_{i=1}^{m} p_i \ell_i \ \text{ subject to } \sum_{i=1}^{m} 2^{-\ell_i} \leq 1.$$

Conclude that for a uniquely decodable code, the minimum expected codeword length is greater than or equal to $H(P)$. Why is it *greater than or equal to* and not *equal to*?

Exercise 5 (Coin tosses and Kraft's inequality)**.** You are given a prefix-free code and a fair coin. Continue tossing the coin until you see a codeword. What is the probability that you will stop? What is the point of this experiment?

Exercise 6 (Entropy)**.** Let X and Y be the outcomes of a pair of dice thrown independently (hence each independently takes on values in $\{1, 2, 3, 4, 5, 6\}$ with equal probabilities). Let $Z = X + Y$ and let $Q = Z \mod 2$. Compute the following entropies: $H(X)$, $H(Y)$, $H(Z)$, $H(Q)$.

Exercise 7 (Entropy)**.** Let X be a random variable taking values in M points $a_1, \ldots, a_M$ and let $p_X(a_M) = \alpha$. Show that

$$H(X) = -\alpha \log \alpha - (1 - \alpha) \log(1 - \alpha) + (1 - \alpha)H(Y)$$

where Y is a random variable taking values in $M - 1$ points $a_1, \ldots, a_{M-1}$ with probabilities $P_Y(a_j) = P_X(a_j)/(1 - \alpha)$ for $1 \leq j \leq M - 1$. Show that

$$H(X) \leq -\alpha \log \alpha - (1 - \alpha) \log(1 - \alpha) + (1 - \alpha) \log(M - 1)$$

and determine the condition for equality.