Telecom Paris | ACCQ202, Information Theory

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

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

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 (Alternative definition of unique decodability)**.** An $f : \mathcal{X} \to \mathcal{Y}$ code is called uniquely decodable if for any messages $u = u_1 \cdots u_k$ and $v = v_1 \cdots v_k$ (where $u_1, v_1, \cdots, u_k, v_k \in \mathcal{X}$) with

$$f(u_1)f(u_2)\cdots f(u_k) = f(v_1)f(v_2)\cdots f(v_k),$$

we have $u_i = v_i$ for all i. That is, as opposed to the definition given in class, we require that the codes of any pair of messages with the same length are equal. Prove that the two definitions are equivalent.

Exercise 5 (Uniquely decodable and instantaneous codes)**.** Let $L = \sum_{i=1}^{n} p_i l_i^2$ be the expected value of the square of the word lengths associated with an encoding of the random variable X. Let $L_1 = \min L$ over all instantaneous codes; and let $L_2 = \min L$ over all uniquely decodable codes. What inequality relationship exists between L_1 and L_2?

Exercise 6 (Equality in Kraft's inequality)**.** An f prefix code is called full if it loses its prefix property by adding any new codeword to it. A string x is called undecodable if it is impossible to construct a sequence of codeword symbols such that x is a prefix of their concatenation. Show that the following three statements are equivalent.

a. f is full,

b. there is no undecodable string with respect to f,

c. $\sum_{i=1}^{n} s^{-l_i} = 1$, where s is the cardinality of the code alphabet, l_i is the codeword length of the ith codeword, and n is the number of codewords.

Exercise 7 (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 8 (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 9 (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.