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 $\varepsilon$ 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. $\mathcal{C} = \{b, abb, abbba, bbba, baabb\}$.

*Solution.* 1. Prefix-free 2. Uniquely decodable, not prefix-free 3. Not uniquely decodable 4. Not uniquely decodable. □

**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.

*Solution.* 1. $\{\epsilon\}$ 2. $\{\epsilon, b, bb, bbb\}$ 3. $\{\epsilon, b\}$ 4. $\{\epsilon, bba, aabb, ba\}$. □

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

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

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?

*Solution.* 1. Examples $2, 3$, and $4$.

2. The above recursion terminates always since there are only a finite number of dangling suffixes for a given code. The time complexity of the algorithm is $O(\ell m)$ where $\ell$ is the total length of all the codewords and $m$ is the number of codewords.

□

**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.

*Solution.* By considering encoding of sequences of equal lengths, the definition given in the exercise is implied by the definition given in class, namely, a code is *uniquely decodable* if for any sequence of message symbols

$$u = u_1, \ldots, u_m$$

and

$$v = v_1, \ldots, v_k,$$

the condition

$$f(u) \overset{\text{def}}{=} f(u_1)f(u_2)\ldots f(u_m) = f(v_1)f(v_2)\ldots f(v_k) \overset{\text{def}}{=} f(v)$$

implies that $u = v$, that is $m = k$ and $u_i = v_i$ for $1 \leq i \leq m$. Conversely, assume that the definition in the question holds. We want to show that the condition $f(u) = f(v)$ implies that $u = v$. By considering the encoding of the concatenation of $u$ and $v$ we get

$$f(uv) = f(u)f(v) = f(v)f(u) = f(vu)$$

Without loss of generality suppose that $m \leq k$. Then the above equality implies that

$$f(u_1)f(u_2)\ldots f(u_m) = f(v_1)f(v_2)\ldots f(v_m)$$

and

$$f(v_{m+1}), \ldots, f(v_k) = \emptyset,$$

which is possible only if $u = v$. □

**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 $L_2$?

*Solution.* Since all instantaneous codes are uniquely decodable, we have $L_2 \leq L_1$. Suppose $L_2$ is attained by a uniquely decodable code with certain codeword lengths. Then, by the Kraft-McMillan inequality, there exists a prefix-free code with the same codeword lengths. Since $L$ depends only on the codeword lengths, this implies that $L_1 \leq L_2$. □

**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 $i$th codeword, and $n$ is the number of codewords.

*Solution.* It can be checked that all the statements are equivalent to the following. In the tree representation of the code, there is no leaf node without a sibling. □

**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?

*Solution.* Let $\ell_i$ be the length of the $i$-th codeword and let $A_i$ be the event that we see the $i$-th codeword. Then, probability that we will stop is

$$P\left(\cup_i A_i\right) = \sum_i P(A_i) = \sum_i 2^{-\ell_i},$$

where the first identity follows since $A_i$s are disjoint (for a prefix-free code). Now, $\sum_i 2^{-\ell_i} \leq 1$ since it equals the probability of an event, which proves the Kraft's inequality.
*Remark* – This proof illustrates the powerful technique of *probabilistic method*[1] which is a recurring theme in information theory. □

**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: $H(X)$, $H(Y)$, $H(Z)$, $H(Q)$.

*Solution.* $X$ and $Y$ are uniform random variables over $\{1, 2, 3, 4, 5, 6\}$, so

$$H(X) = H(Y) = \log_2(6).$$

The probability distribution of $Z$ is

$$Z = \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{pmatrix}$$

So, $H(Z) = 3.27$. The probability distribution of $Q$ is

$$Q = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

an so $H(Q) = 1$.

□

[1] *Noga Alon and Joel H. Spencer. The Probabilistic Method. John Wiley & Sons, 3rd edition, 2008. Exer. 1.8, p. 12.*

**Exercise 9** (Entropy)**.** Let $X$ be a random variable taking values in $M$ points $a_1, \dots, a_M$ 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 $a_1, \dots, a_{M-1}$ with probabilities $P_Y(a_j) = P_X(a_j)/(1-\alpha)$ for $1 \le j \le M-1$. Show that

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

and determine the condition for equality.

*Solution.*

$$\begin{aligned} H(X) &= -\sum_{j=1}^{M} p_X(a_j) \log p_X(a_j) \\ &= -\alpha \log \alpha - \sum_{j=1}^{M-1} p_X(a_j) \log p_X(a_j) \\ &= -\alpha \log \alpha - (1-\alpha) \sum_{j=1}^{M-1} \frac{p_X(a_j)}{1-\alpha} \log \left( \frac{p_X(a_j)}{(1-\alpha)} (1-\alpha) \right) \\ &= -\alpha \log \alpha - (1-\alpha) \sum_{j=1}^{M-1} p_Y(a_j) \Big\{ \log p_Y(a_j) + \log(1-\alpha) \Big\} \\ &= -\alpha \log \alpha - (1-\alpha)\log(1-\alpha) + (1-\alpha)H(Y) \end{aligned}$$

where for the final equality we used $\sum_{j=1}^{M-1} p_Y(a_j) = 1-\alpha$.

To prove that

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

it suffices to observe that $Y$ takes at most $M-1$ values, hence its entropy is at most $\log(M-1)$. Equality is achieved when the distribution of $Y$ is uniform over the $M-1$ points; that is, when $p_Y(a_j) = 1/(M-1)$ for $1 \le j \le M-1$, whereby

$$p_X = \left( \frac{1-\alpha}{M-1}, \frac{1-\alpha}{M-1}, \dots, \frac{1-\alpha}{M-1}, \alpha \right).$$

□