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

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

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

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 ℓ is the total length of all the codewords and m is the number of codewords.

□

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*?

Solution. In order to find the optimal $\ell = (\ell_1, \ldots, \ell_m)$, we define the Lagrangian

$$L(\ell, \lambda) = \sum_{i=1}^{m} p_i \ell_i - \lambda \left(\sum_{i=1}^{m} 2^{-\ell_i} - 1 \right).$$

Taking derivative with respect to ℓ_i and λ and setting them equal to 0 yields

$$\frac{dL}{d\ell_i} = p_i - \lambda 2^{-\ell_i} \ln 2 = 0$$

and $\lambda = 1/\ln 2$ since $\sum_{i=1}^{m} p_i = 1$, whereby

$$\ell_i = -\log_2 p_i.$$

Thus, the minimum is $L^* = \sum_{i=1}^{m} p_i \ell_i = H(P)$. The minimum expected codeword length for a uniquely decodable code, $\overline{L}$ is obtained by solving the same optimization problem with an additional constraint, namely, $\ell_i \in \mathbb{N}$; therefore, the minimum can only be larger. Hence, $\overline{L} \geq L^* = H(P)$. □

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?

Solution. Let ℓ_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_is 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 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: $H(X)$, $H(Y)$, $H(Z)$, $H(Q)$.

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

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 $H(Q) = 1$.

□

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.

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

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

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 $1 \leq j \leq M-1$, whereby

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

□