Learner's Space 2025
Information Theory and Coding
Assignment 2

Suketu Patni

June 22, 2025

1. Find the capacity of the binary DMC with the transition probability model $\begin{pmatrix} f & 1-f \\ f & 1-f \end{pmatrix}$ where $f \in (0, 1)$. [3 marks]

2. Let $n \geq 2$ and let T be an $n \times n$ matrix such that all the entries of T are non-negative, and the sum of entries in each row and column of T is 1. Let $\vec{p} = (p_1, p_2, \ldots, p_n)$ be a probability vector. Show that $H(\vec{p}T) \geq H(\vec{p})$. [5 marks]

3. **Bounds on Expected Codelengths**
Let A (often called an **alphabet**) denote a finite set of symbols $\{a_1, a_2, \ldots, a_n\}$. We will use the notation A^N to denote the set of N-length strings, each of whose characters is a symbol from A. Let A^+ be the set of all strings of finite length composed of elements of A.

Definition 0.1. *A **binary symbol code** C is a mapping from A to $\{0, 1\}^+$. For $x \in A$, $c(x)$ denotes the codeword corresponding to x, and $l(x)$ denotes its length, with $l_i = l(a_i)$.*

Definition 0.2. *The **extended code** C^+ is a mapping from A^+ to $\{0, 1\}^+$ defined as*

$$c^+(x_1 \cdot x_2 \cdot \ldots \cdot x_m) = c(x_1) \cdot c(x_2) \cdot \ldots \cdot c(x_m)$$

where $x_i \in A$ and $\cdot$ represents concatenation of two strings/symbols.

Definition 0.3. *A code is called **instantaneous** if no codeword is a prefix of any other codeword, i.e., $\forall a_i,\ a_j \in A$, $\nexists s \in \{0, 1\}^+$ such that $c(a_i) \cdot s = c(a_j)$.*

For instance, for a set of 4 symbols, the binary symbol code with images $\{0, 10, 110, 111\}$ is instantaneous.

Definition 0.4. *Let X be a random variable with finite image $\mathcal{X}$ where $|\mathcal{X}| = |A|$. Map each value in the image of X to a corresponding symbol in A (bijectively), and let this mapping be φ. The **expected length** of a symbol code C over X is*

$$L(C, X) = \sum_{x \in A} P(\varphi(X) = x) l(x) = \sum_{i=1}^{|A|} p_i l_i$$

where p_i is defined as $P(\varphi(X) = a_i)$.

We want any symbol code to have 3 important properties: one, that every encoded string must have a **unique decoding**; two, that it must be **easy to decode**; and three, that it should achieve **as much compression as possible**.

- **Unique decoding**: $\forall s_1,\, s_2 \in A^+,\, c^+(s_1) = c^+(s_2) \Rightarrow s_1 = s_2$.
- **Easy to decode**: A code is easiest to decode if it is instantaneous, in which case it is possible to identify the end of a codeword as soon as some of its starting bits arrive.
- **Compression**: Empirically, p_i is chosen to denote the probability of the symbol a_i popping up whenever we compose a message written with symbols in A that we want to encode. For instance, if A were the English alphabet, we would expect $p_{\text{`e'}}$ to be the highest. We wish to minimize the expected length of a symbol code as much as possible, as this will mean we will have to send less symbols on average.

  Note that unique decodability does not imply instantaneity, but instantaneity implies unique decodability.

a. Prove that a uniquely decodable binary symbol code C with alphabet A satisfies

$$\sum_{i=1}^{|A|} 2^{-l_i} \leq 1$$

This famous inequality is known as the Kraft-McMillan inequality. To prove this, try writing the expression for the LHS of the inequality when raised to some power N and then use the fact that the code is uniquely decodable. Of course, this is just a hint, and other ways of proof/heuristic arguments are also accepted. [6 marks]

b. Prove that for a set of lengths $\{l_1, l_2, \ldots, l_n\}$ satisfying the Kraft-McMillan inequality, there exists an instantaneous code over an alphabet of n symbols having those lengths. A simple counting argument should do the trick. [6 marks]

Definition 0.5. *A **complete** code is a uniquely decodable code which satisfies equality in Kraft-McMillan's inequality.*

c. Now we come to the connection with information theory. Using Gibbs' inequality on a cleverly defined probability distribution, show that for a uniquely decodable binary symbol code C with alphabet A over X with equal support and image, $H(X) \leq L(C, X)$ with equality iff the code is complete and the codelength of any symbol $a \in A$ is the SIC of $\varphi^{-1}(a)$ (where φ is as defined in Definition 0.4). [5 marks]

d. Give an interpretation to $L(C, X) - H(X)$ in terms of the KL divergence. [1 mark]

e. Prove that there is an instantaneous code with the same setup as in the previous part which achieves $L(C, X) \leq H(X) + 1$. [4 marks]