End of Proof of the Kraft Inequality

Now suppose codeword lengths $\ell_i$ are given with $\sum_{i=1}^{m} D^{-\ell_i} \leq 1$. Without loss of generality, assume $\ell_1 \leq \cdots \leq \ell_m$. Then $\sum_{i=1}^{m} D^{\ell_m-\ell_i} \leq D^{\ell_m}$. Consider a $D$-ary tree of potential codewords of maximum length $\ell_m$. Pick any codeword of length $\ell_1$. Then all potential codewords attached to the right are not available to assign to other source letters. However,

$$D^{\ell_m} - D^{\ell_m-\ell_1} \geq \sum_{i=2}^{m} D^{\ell_m-\ell_i} \geq 1.$$

Hence, there are codewords of length $\ell_{\text{max}}$ available. Pick one and trace a path back to the root of the tree. This shows that there is a codeword of length $\ell_2$ available. Continuing in this way, we obtain codewords $d_1, \ldots, d_m$ that satisfy the prefix condition and $\ell(d_i) = \ell_i$. □

Properties of Uniquely Decodable Codes

**Theorem 3.9** (Brockway McMillan, 1956). Let $|X| < \infty$. If $C : X \rightarrow D^*$ is uniquely decodable, then the Kraft inequality $\sum_{x \in X} D^{-\ell(C(x))} \leq 1$ holds.

**Proof.** (Karush, 1961). Let $n$ be an arbitrary positive integer, and consider the identity

$$\left( \sum_{x \in X} D^{-\ell(C(x))} \right)^n = \prod_{k=1}^{n} \left( \sum_{x_k \in X} D^{-\ell(C(x_k))} \right) = \sum_{x_1 \in X} \sum_{x_2 \in X} \cdots \sum_{x_n \in X} D^{-[\ell(C(x_1)) + \cdots + \ell(C(x_n))]}.$$

Since $\ell(C(x_1)) + \cdots + \ell(C(x_n)) = \ell(C^*(x))$, where $x = x_1 \cdots x_n$, we can write

$$\left( \sum_{x \in X} D^{-\ell(C(x))} \right)^n = \sum_{x \in X^n} D^{-\ell(C^*(x))}.$$

For $i \geq 1$, let $B_i := \{x \in X^n : \ell(C^*(x)) = i\}$. Put $I := \{i : B_i \neq \emptyset\}$. Note that if $\ell_{\text{max}} := \max_{x \in X} \ell(C(x))$, then $i \in I \Rightarrow i \leq n\ell_{\text{max}}$ since $\ell(C^*(x)) = \sum_{k=1}^{n} \ell(C(x_k)) \leq \ell_{\text{max}}$. Therefore, the Kraft inequality holds.
\[ \sum_{k=1}^{n} \ell_{\text{max}} = n \ell_{\text{max}}. \]

Next, observe that the \( B_i \) are disjoint. Hence, \( \sum_{x \in X^n} D^{-\ell(C^*(x))} = \sum_{i \in I} \sum_{x \in B_i} D^{-i}. \)

Now for \( x \in B_i \), \( C^*(x) \) is a unique element of \( D^i \) because the codes is uniquely decodable. Hence, \( |B_i| \leq D^i \), and we can write

\[ \sum_{i \in I} \sum_{x \in B_i} D^{-i} = \sum_{i \in I} D^{-i} |B_i| \leq \sum_{i \in I} 1 = |I| \leq n \ell_{\text{max}}. \]

So

\[ \left( \sum_{x \in X} D^{-\ell(C(x))} \right)^n \leq n \ell_{\text{max}}, \]

or

\[ \sum_{x \in X} D^{-\ell(C(x))} \leq (n \ell_{\text{max}})^{1/n} \rightarrow 1, \quad \text{as } n \rightarrow \infty. \]

\[ \square \]

**Remark.** McMillan’s Theorem shows that a uniquely decodable code satisfies the Kraft inequality. Hence, by Kraft’s Theorem, we can always find a prefix code with the same code lengths as the uniquely decodable code. This prefix code has the same expected length as the uniquely decodable code, but the prefix code incurs less delay and is easier to decode.

**Definition 3.10.** An independent, identically distributed (i.i.d.) sequence of \( X \) valued random variables \( \{X_n\}_{n=1}^{\infty} \) is called a discrete memoryless source (DMS).

**Definition 3.11.** If \( p \) is a probability mass function (pmf) on \( X \), the entropy of \( p \) is

\[ H(p) := \sum_{x \in X} p(x) \log \frac{1}{p(x)} = -\sum_{x \in X} p(x) \log p(x) \geq 0, \]

where \( 0 \log 0 := 0 \), and base of log can be any real number greater than one. If the base is 2, the entropy is measured in **bits**; if the base is \( e \), the entropy is measured in **nats**.

**Definition 3.12.** If \( X \) is an \( X \) valued RV with pmf \( p \), we sometimes write \( H(X) \) in place of \( H(p) \). If \( \{X_n\}_{n=1}^{\infty} \) is a DMS with common pmf \( p(x) := P(X_n = x) \) for all \( n \geq 1 \), then we call \( H(X) := H(X_1) = H(p) \) the entropy of the DMS.