Theorem: Kraft-McMillan Inequality

The two following theorems give necessary and sufficient conditions for the existence of a prefix-free code or a uniquely decodable code for a given set of codeword lengths.

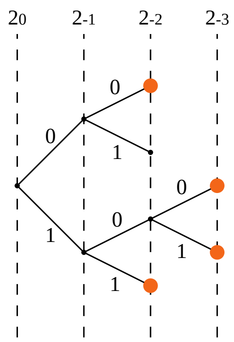
Theorem: Kraft's inequality

Let $\{\ell_1, \ell_2, \ldots, \ell_m\}$ be a set of m positive natural numbers. There exists a prefix-free code with image $\mathcal{C} = \{c_1, \ldots, c_m\}$ and codeword lengths $\ell(c_i) = \ell_i$, if and only if

$$\sum_{i=1}^{m} 2^{-\ell_i} \leq 1.$$

$\Rightarrow$

For the forward direction, suppose we have a prefix-free code C with image $\mathcal{C} = \{c_1, \ldots, c_m\}$ and codeword lengths $\ell(c_i) = \ell_i$. View this code as a tree, with codewords only on the leaves (but not necessarily all the leaves), and assign a weight of 2^{-d} to every node in the tree at depth d (including the leaves):



Note that the weight of each node is exactly the sum of the weight of its direct children, and thereby that the weight of the root is exactly the weight of all of the leaves. Since every codeword c_i resides on a leaf of depth ℓ_i (but not all leaves are necessarily occupied), the weight of the root is *at least* the sum of all the codeword weights:

$$\sum_{i=1}^{m} 2^{-\ell_i} \leq 2^0 = 1.$$

$\Leftarrow$

Shopping at the codeword supermarket as described in the following video: 03MacKayShoppingCodewords.MP4

A stronger version of Kraft's inequality holds as well, this time for uniquely decodable codes:

Theorem: McMillan inequality

For a uniquely decodable code with image $\mathcal{C} = \{c_1, \ldots, c_m\}$ and codeword lengths $\ell_i := \ell(c_i)$, it holds that

$$\sum_{i=1}^{m} 2^{-\ell_i} \leq 1.$$

Proof

Let $\mathcal{C}$ be a uniquely decodable code as in the theorem statement. We can write

$$S := \sum_{c \in \mathcal{C}} \frac{1}{2^{\ell(c)}} = \sum_{\ell=L_{\min}}^{L_{\max}} \frac{n_\ell}{2^\ell}$$

where $L_{\min} = \min_{c \in \mathcal{C}} \ell(c)$, $L_{\max} = \max_{c \in \mathcal{C}} \ell(c)$, and $n_\ell = |\{c \in \mathcal{C} \mid \ell(c) = \ell\}|$. Furthermore, for any $k \in \mathbb{N}$, consider the kth power of S,

$$S^k = \sum_{c_1, \ldots, c_k \in \mathcal{C}^k} \frac{1}{2^{\ell(c_1) + \ldots + \ell(c_k)}} = \sum_{\ell=kL_{\min}}^{kL_{\max}} \frac{n_\ell^{(k)}}{2^\ell}$$

where $n_\ell^{(k)}$ is defined as $n_\ell^{(k)} = \left|\left\{(c_1, \ldots, c_k) \in \mathcal{C}^k \mid \sum_i \ell(c_i) = \ell(c_1|\ldots|c_k) = \ell\right\}\right|$. Note that

$$n_\ell^{(k)} = \sum_{x \in \{0,1\}^\ell} \left|\{(c_1, \ldots, c_k) \in \mathcal{C}^k \mid c_1|\ldots|c_k = x\}\right| \leq \sum_{x \in \{0,1\}^\ell} 1 = 2^\ell$$

where the inequality follows from the unique decodability of $\mathcal{C}$. Thus, we can conclude that

$$S^k \leq (L_{\max} - L_{\min}) \cdot k$$

for all $k \in \mathbb{N}$, so S^k grows at most linearly in k, from which follows that $S \leq 1$ (for if not, S^k would grow exponentially in k).

Kraft's and McMillan's inequality together lead to the conclusion that the lengths of an optimal prefix-free code and an optimal uniquely decodable code coincide:

Corollary

Let P_X be a source. For every uniquely decodable code C, there exists a prefix-free code C' such that $\ell_C(P_X) = \ell_{C'}(P_X)$. Hence,

$$\ell_{\min}^{\text{p.f.}}(P_X) = \ell_{\min}^{\text{u.d.}}(P_X).$$

From now on, we will just write $\ell_{\min}(P_X)$ to denote either of these measures for average length. A code C for which $\ell_C(P_X) = \ell_{\min}(P_X)$ is called **optimal** for the source P_X.