Prefix or instantaneous codes

are source codes where no codeword is a prefix of another.

Example: \( S_A = \{ A, B, C, D, E \} \), \( TA = \{ 0, 1 \} \),

\[
\begin{align*}
A & \mapsto 01 \\
B & \mapsto 110 \\
C & \mapsto 111 \\
D & \mapsto 10 \\
E & \mapsto 00
\end{align*}
\]

Prefix codes are uniquely decodable.
Read the letters of the encoded message until they make a codeword, then write that letter down and start again.

1110111010000010
CABDEED

Proof of Kraft inequality

The picture below shows four “layers” of the word tree for \( TA = \{ 0, 1, 2 \} \). The word with no letters is denoted by \( \emptyset \).

There is a path from \( w_1 \) to \( w_2 \) if & only if \( w_1 \) is a prefix of \( w_2 \).
Now highlight the vertices corresponding to codewords.

e.g. \( C(A) = 0 \), \( C(B) = 10 \), \( C(C) = 110 \), \( C(D) = 111 \), \( C(E) = 112 \), \( C(F) = 12 \), \( C(G) = 2 \). No codeword is “above” another.
Kraft proof continued

Let \( L \) be the longest codeword length. Draw \( L + 1 \) layers of the tree (thus including all codewords).

There are \( D^L \) “leaves” (vertices on the bottom layer).

A codeword \( C(x) \) of length \( \ell(x) \) has \( D^{L-\ell(x)} \) leaves below it.

So \( \sum_x D^{L-\ell(x)} \leq D^L \). Dividing by \( D^L \) gives \( \sum_x D^{-\ell(x)} \leq 1 \).

Kraft converse

Now suppose we have integers \( n(x) \) satisfying \( \sum_{x \in \mathcal{S}} D^{-n(x)} \leq 1 \).

Write \( \mathcal{S} = \{ x_1, x_2, \ldots, x_k \} \) with \( n(x_1) \leq n(x_2) \leq \cdots \leq n(x_k) \).

Let \( L = n(x_1) \), and draw the \( L + 1 \) layer tree.

Note that \( \sum_{i=1}^{k} D^{L-n(x_i)} \leq D^L = \text{number of leaves} \).

Choose \( C(x_1) \) to be any word of length \( n(x_1) \), and erase tree below \( C(x_1) \). This removes \( D^{L-n(x_1)} \) leaves.

Then choose \( C(x_2) \) to be any word of length \( n(x_2) \), & erase tree below \( C(x_2) \), removing \( D^{L-n(x_2)} \) leaves. Continue in this way.

After \( m \)th step \( \sum_{i=1}^{m} D^{L-n(x_i)} \) leaves have been removed.

If \( m < k \) then \( \sum_{i=1}^{m} D^{L-n(x_i)} < D^L \). So there are leaves left.

Kraft converse example

Example: Suppose \( D = 3 \) and the \( n(x_i) \) are 1, 1, 2, 2, 3, 3, 3.

Since \( L = n(x_7) = 3 \), we draw the 27-leaf tree.

Choose a word of length 1 as a codeword & erase tree below it.

There are leaves left; so we can continue.

Choose another word of length 1 as a codeword & erase below.

There are leaves left; we can choose a word of length 2.

And another. Now the length 3 words.

Kraft converse concluded

Since we are taking the \( n(x_i) \) in increasing order, the next word to choose never has to be higher than anything already chosen. So it won’t be a prefix of an earlier one.

We have erased everything that has an earlier word as a prefix.

So as long as there are leaves left we can choose the next word and still have a prefix code.

It is given that \( \sum_{i=1}^{k} D^{-n(x_i)} \leq 1 \). So \( \sum_{i=1}^{k} D^{L-n(x_i)} \leq D^L \).

And \( \sum_{i=1}^{m} D^{L-n(x_i)} < D^L \) for \( m < k \).

The left hand side is the number of leaves removed after \( m \) steps; the right hand side the total number of leaves.

So while \( m < k \) there are leaves left & we can choose \( C(x_{m+1}) \).

So \( C(x_1), C(x_2), \ldots, C(x_k) \) can all be chosen.
Entropy and encoding

Suppose that our TA has $D$ letters, and we are encoding messages using a prefix code $C$. A message is a sequence of values of a random variable $X$. For each $x \in SA$ let $p(x)$ be the probability that a random letter of a random message is $x$. Define the entropy to the base $D$ of $X$ by

$$H_D(X) = -\sum_x p(x) \log_D p(x) = \frac{1}{\log_2 D} H(X).$$

We shall show that we need on average at least $H_D(X)$ letters of TA to encode each letter of SA.

That is, the expected codeword length must be $H_D(X)$ at least.

Shannon-Fano codes

If $Q = \sum_x D^{-\ell(x)}$ then $L(C) \geq H_D(X) - \log_D Q$. If $C$ is a prefix code, Kraft gives $Q \leq 1$. So $\log_D Q \leq 0$, and $L(C) \geq H_D(X)$.

Equality occurs iff $Q = 1$ and $D(p||q) = 0$ (in the notation of the previous proof).

But $D(p||q) = 0$ iff $p(x) = q(x) = D^{-\ell(x)}$, given $Q = 1$.

So to construct a code with $L(C) = H_D(X)$ we need to make $\ell(x) = -\log_D p(x)$ for all $x$. This satisfies the Kraft Inequality: $\sum_x D^{-\ell(x)} = \sum_x p(x) = 1$.

But $\ell(x)$ has to be an integer, and of course $-\log_D p(x)$ will not usually be an integer.

But we can choose $\ell(x)$ to be the least integer greater than or equal to $-\log_D p(x)$. Then $\sum_x D^{-\ell(x)} \leq 1$, as we need.

Codes satisfying this are called Shannon-Fano codes.

A lower bound on the expected length

The following result holds for all source codes $C$.

**Theorem:** Let $L(C) = \sum_x p(x) \ell(x)$, the expected length of $C$. Then $L(C) \geq H_D(X) - \log_D (\sum_x D^{-\ell(x)})$.

**Proof:** Let $Q = \sum_x D^{-\ell(x)}$. For each $x \in SA$ let $q(x) = D^{-\ell(x)}/Q$.

Then $\sum_x q(x) = 1$; so $q$ is a probability distribution on $SA$, and

$$0 \leq D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = \sum_x p(x) \log p(x) - \sum_x p(x) \log q(x) = -H(X) - \sum_x p(x) \log D^{-\ell(x)} + \sum_x p(x) \log Q$$

$$= -H(X) + \sum_x p(x) \ell(x) \log D + \log Q$$

Dividing by $\log D$ gives the result.

Expected length of Shannon-Fano codes

**Theorem:** A Shannon-Fano code from a source of entropy $H_D(X)$ to a $D$-letter target alphabet satisfies

$$H_D(X) \leq L(C) < H_D(X) + 1.$$

**Proof:** By the definition of Shannon-Fano codes we have

$$-\log_D p(x) \leq \ell(x) < -\log_D p(x) + 1.$$

Taking expectations gives the result.

However, it is not true that Shannon-Fano codes always achieve the smallest possible value of $L(C)$.

Suppose $D = 2$ and $SA$ has two letters with probabilities $\frac{1}{4}$, $\frac{3}{4}$. Since $-\log_2 \frac{1}{4} = 2$ and $-\log_2 \frac{3}{4} \approx 0.4$, Shannon-Fano gives codeword lengths of 2 and 1, and $L(C) = \frac{1}{4} \times 2 + \frac{3}{4} \times 1 = \frac{5}{4}$.

But we can get $L(C) = 1$ by assigning the two one-letter words of TA to the two letters of SA.