Prefix-free code (instantaneous code)

code is prefix-free if no codeword is a prefix of any other

binary code tree: all codewords must be leaves for prefix-freedom.

PF code is full if no new codewords can be added without destroying PF property.

How do we know if we can construct a PF code for a given source with a particular codeword length?

Kraft Inequality (1949)

Let $\mathcal{X} = \{x_1, \dots, x_M\}$ and let $\ell(x_1), \dots, \ell(x_M)$ be codeword lengths.

Then $\sum_{j=1}^{M} 2^{-\ell(x_j)} \le 1$ if and only if PF code exists.

PF code is full when holds with equality.

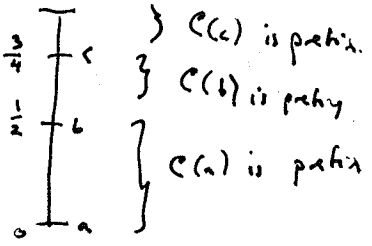
proof:

① model the space of codewords as unit interval $[0,1)$, since any codeword can be represented as a binary expansion.

$$0.b_1 b_2 \cdots = b_1 2^{-1} + b_2 2^{-2} + \dots$$

$$= \tfrac{1}{2} b_1 + \tfrac{1}{4} b_2 + \dots$$

$a \to 0$	0
$b \to 10$	$\frac{1}{2}$
$c \to 11$	$\frac{3}{4}$



② so a given codeword $C(x_j) = c_j$ is prefix for all codewords in an interval of size $2^{-\ell_j}$ starting at c_j.

③ prefix-free property $\Leftrightarrow$ intervals don't overlap.

→ check that PF satisfies $\sum_{j=1}^{m} 2^{-l_j} \leq 1$

Can obtain PF code by reading off interval boundaries.

---

optimizing PF codes

minimize $\sum_{j=1}^{m} p_j l_j$ s.t. $\sum_{j=1}^{m} 2^{-l_j} \leq 1$ over choices $l_1, \ldots, l_m$.

to get lower bound, ignore integer constraints (so Kraft inequality holds with equality)

using Lagrange multipliers, get

$$l_j = -\log p_j$$

$$\bar{L} \geq \sum_j p_j l_j = -\sum_j p_j \log p_j = H(X).$$

to get upper bound, for original integer program, round real-valued optimal lengths up.

$$l_j = \lceil -\log p_j \rceil \quad \text{which is Shannon-Fano-Elias code.}$$
$$< 1 - \log p_j$$

$$\bar{L} = \sum_j p_j l_j < \sum_j p_j (1 - \log p_j) = 1 - \sum p_j \log p_j = H(X) + 1$$

So $H(X) \leq \bar{L} < H(X) + 1$

Asymptotic Equipartition Property (AEP), a restatement of the law of large numbers

- Let $-\frac{1}{n} \log p(\vec{x}) = -\frac{1}{n} \sum_{k=1}^{n} \log p(x_k)$ be the empirical entropy of a particular source sequence. Thought to be generated iid according to r.v. X with entropy $H(X)$.

- The typical set is defined as set of sequences whose empirical entropy is close to the true entropy. $A_\varepsilon^{(n)}$ is $\vec{x} \in \mathcal{X}^n$ s.t.

$$\left| -\frac{1}{n} \log p(\vec{x}) - H(X) \right| \leq \varepsilon \quad \text{or}$$

$$2^{-n(H(X)+\varepsilon)} \leq p(\vec{x}) \leq 2^{-n(H(X)-\varepsilon)}$$

AEP: ① if $\vec{x} \in A_\varepsilon^{(n)}$ then $2^{-n(H(X)+\varepsilon)} \leq p(\vec{x}) \leq 2^{-n(H(X)-\varepsilon)}$

② for sufficiently large n, $\Pr\left[A_\varepsilon^{(n)}\right] > 1-\varepsilon$

③ for sufficiently large n,

$$(1-\varepsilon)\, 2^{n(H(X)-\varepsilon)} \leq \left|A_\varepsilon^{(n)}\right| \leq 2^{n(H(X)+\varepsilon)}$$

- Total number of typical sequences $\approx 2^{nH(X)}$ which is generally much less than all possible sequences, if $H(X) < \log|\mathcal{X}|$.

- when n large, can essentially think of sequence X as being drawn uniformly from typical set

Back to block coding

Let $X_1, X_2, \ldots, X_n$ be iid r.v. from $p(x)$.

want to find short descriptions for such sequences.

Achievable scheme:

① divide sequences in $\mathcal{X}^n$ into $A_\varepsilon^{(n)}$ and its complement, order two subsets in some fixed way.

② represent typical sequences by indexing with binary codewords of length no more than $n(H+\varepsilon)+1$, since less than $2^{n(H+\varepsilon)}$ elements in $A_\varepsilon^{(n)}$

→ prefix with a 0, so total length $\leq n(H+\varepsilon)+2$.

③ represent each sequence not in $A_\varepsilon^{(n)}$ using no more than $n \log|\mathcal{X}| + 1$ bits, and prefix with 1.

Performance of achievable scheme

let $\ell(x^n)$ be length of codeword for x^n, suppose n sufficiently large for $\Pr[A_\varepsilon^{(n)}] \geq 1-\varepsilon$.

$$E[\ell(X^n)] = \sum_{x^n} p(x^n)\ell(x^n) = \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n)\ell(x^n) + \sum_{x^n \notin A_\varepsilon^{(n)}} p(x^n)\ell(x^n)$$

$$\leq \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n)\left(n(H(X)+\varepsilon)+2\right) + \sum_{x^n \notin A_\varepsilon^{(n)}} p(x^n)\left(n \log|\mathcal{X}|+2\right)$$

$$= \Pr[A_\varepsilon^{(n)}]\left(n(H+\varepsilon)+2\right) + \left(1-\Pr[A_\varepsilon^{(n)}]\right)\left(n \log|\mathcal{X}|+2\right)$$

$$\leq n(H+\varepsilon) + \varepsilon n(\log|\mathcal{X}|)+2$$

letting $\varepsilon' = \varepsilon + \varepsilon \log|\mathcal{X}| + \frac{2}{n}$

$= n(H+\varepsilon')$

So achievable scheme has $E\left[\frac{1}{n}\ell(X^n)\right] = H(X)+\varepsilon$ for large n.

Also converse: if we use block code with blocklength n and coding rate less than $H(X) - \zeta$ where $\zeta > 0$ (doesn't change with n) then error probability goes to 1 as $n \to \infty$.

Proof idea: there isn't a smaller set than the typical set that contains most of the probability mass, measured in exponential sense. ($a_n \doteq b_n$ means $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$)