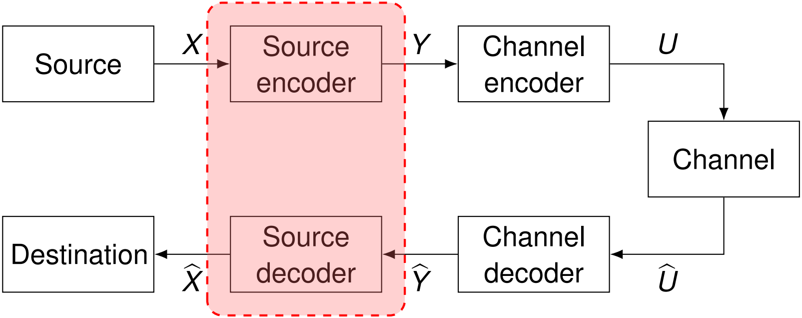
LUND
UNIVERSITY

Information Theory

Lecture 3
Optimal source coding and Kraft inequality

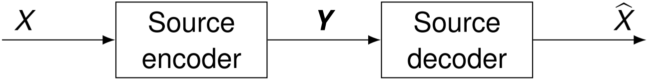
STEFAN HÖST

Communication system



Shannon's model for a communication system

Source coding



Definition

A source code is a mapping from the random variable $X \in \{x_1, \dots, x_k\}$ to a vector of variable length ℓ, $\boldsymbol{y} = (y_1 y_2 \dots y_\ell)$, where $y_i \in \mathbb{Z}_D$ are drawn from a D-ary alphabet.

The length of the codeword corresponding to x is denote ℓ_x.

The efficiency of a code is shown by the average codeword length,

$$L = E[\ell_x] = \sum_x p(x)\ell_x$$

Code classification

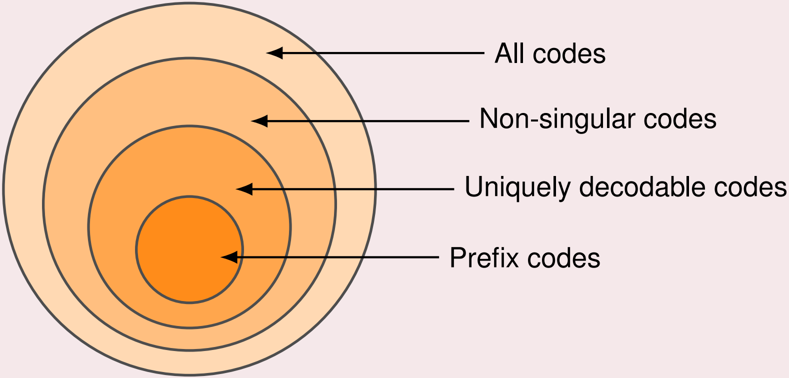
Definition

A code is said to be

- Non-singular if each source symbol is mapped to a distinct code vector.
- Uniquely decodable if each sequence of source symbols is mapped to a unique sequence of code symbols, i.e. the mapping from source sequences is non-singular.
- Prefix (or instantaneous) if no codeword is a prefix of another codeword.

Code classification

Illustration of code classifiation



LUND UNIVERSITY

Prefix codes

D-ary tree

- A *D*-ary tree is a tree where each node has either *D* or 0 branches.
- A full *D*-ary tree of depth *n* is a tree with depth *n* and D^n end nodes (leaves).

Code representation

A prefix code can be represented in a *D*-ary tree, with branches representing symbols. The depth of a path is the length of the codeword.

Optimal code

Definition

An optimal prefix code is a prefix code that minimizes the expected codeword length

$$L = E[\ell_x] = \sum_x p(x)\ell_x$$

over all prefix codes for the same source alphabet.

Lemma (Path length lemma)

In a tree representation of a prefix code, the average codeword length $L = E[\ell_x]$ equals the sum of probabilities for the inner nodes, incluing the root.

Kraft inequality

Theorem (Kraft inequality)

There exists a D-ary prefix code with codeword lengths $\ell_1, \ell_2, \dots, \ell_k$ if and only if

$$\sum_{i=1}^{k} D^{-\ell_i} \leq 1$$

Kraft inequality — Sketch of proof

Prefix code $\Rightarrow$ Ineqality

- Assume prefix code with $\max_x \ell_x = \ell_{\text{max}}$
- Represent the codewords in a full tree with $D^{\ell_{\text{max}}}$ leaves
- Remove subtree after codewords

$\Rightarrow \sum_x D^{\ell_x} \leq 1$

Kraft inequality — Sketch of proof

Ineqality $\Rightarrow$ Prefix code

- Assume $\sum_i D^{\ell_{x_i}} \leq 1$ and $\ell_{x_1} \leq \ell_{x_2} \leq \cdots \leq \ell_{x_k} = \ell_{\text{max}}$
- Start with full tree of depth $D^{\ell_{max}}$
- Put in shortest codewords and remove subtree after
- After $j < k$ steps there are unused nodes left at depth $D^{\ell_{\text{max}}}$

At $j = k - 1$ there are at least one node left
$\Rightarrow$ The code can be constructed

LUND UNIVERSITY

Kraft inequality

Theorem (McMillan inequality)

There exists a D-ary uniquely decodable code with codeword lengths $\ell_1, \ell_2, \ldots, \ell_k$ if and only if

$$\sum_{i=1}^{k} D^{-\ell_i} \leq 1$$

Uniquely decodable codes will never be better than prefix codes.

Bounds on L

Theorem

The expected codeword length $L = E[\ell_x]$ of a prefix code is lower bounded by the entropy of the source, i.e.

$$L \geq H_D(X) = \frac{H(X)}{\log D}$$

with equality if and only if the optimal codeword lengths $\ell_x = -\log_D p(x)$ is used.

Theorem

For every random variable X there exists a D-ary prefix code such that

$$L < H_D(X) + 1$$

Optimal codeword length

Theorem

Given a random variable X, the average codeword length for an optimal D-ary prefix code satisfies

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

with equality to the left if and only if $\ell_x = \log_D p(x)$.