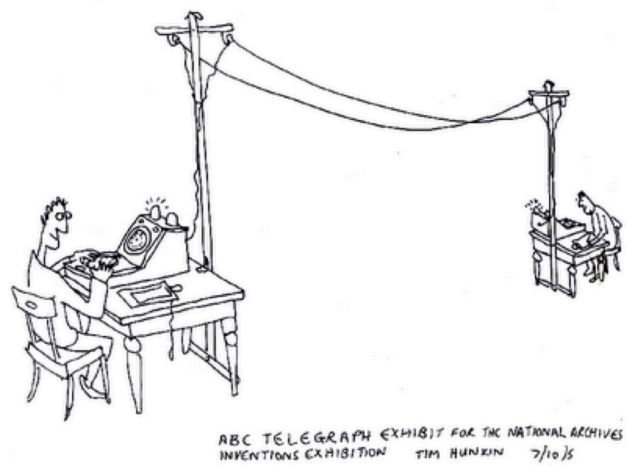
Lecture 7: Source Coding and Kraft Inequality

- Codes
- Kraft inequality and consequences

Horse Racing



Which horse won?

p_i	Code 1	Code 2
1/2	000	0
1/4	001	10
1/8	010	110
1/16	011	1110
1/64	100	111100
1/64	101	111101
1/64	110	111110
1/64	111	111111
El_i	3	2

$$H(X) = -\sum p_i \log p_i = 2\text{bits}$$

How to find the best code?

Codes

- Source code C for a random variable X is

$$C(x) : \mathcal{X} \to \mathcal{D}^*$$

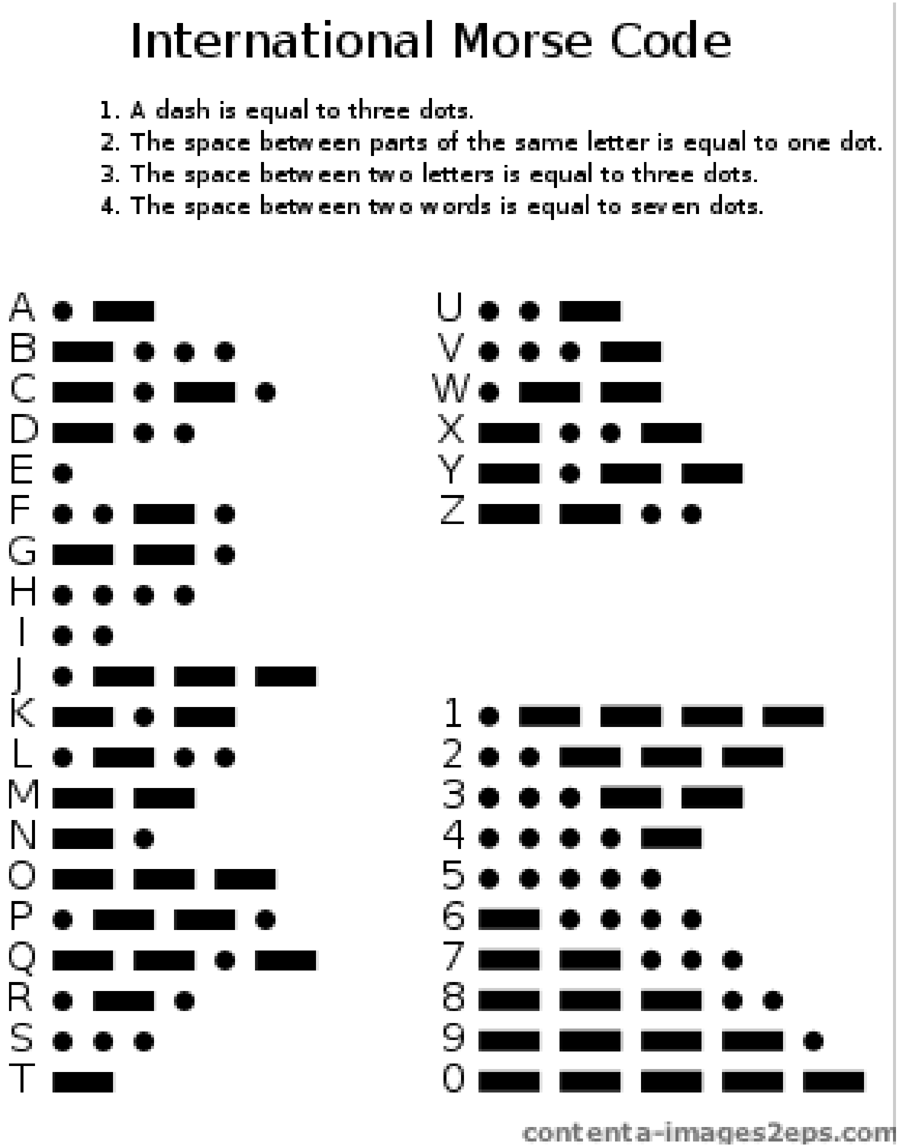
  $\mathcal{D}^*$: set of finite-length strings of symbol from D-ary alphabet $\mathcal{D}$

- Code length: $l(x)$

- Example: $C(\text{red}) = 00$, $C(\text{blue}) = 11$, $\mathcal{X} = \{\text{red, blue}\}$, $\mathcal{D} = \{0, 1\}$

Morse's code (1836)

- A code for English alphabet of four symbols
- Developed for electric telegraph system
- $\mathcal{D} = \{\text{dot, dash, letter space, word space}\}$
- Short sequences represent frequent letters
- Long sequences represent infrequent letter


International Morse Code
1. A dash is equal to three dots.
2. The space between parts of the same letter is equal to one dot.
3. The space between two letters is equal to three dots.
4. The space between two words is equal to seven dots.
contenta-images2eps.com

Source coding applications

- Magnetic recording: cassette, hardrive, USB...
- Speech compression
- Compact disk (CD)
- Image compression: JPEG

Still an active area of research:

- Solid state hard drive
- Sensor network: distributed source coding

What defines a good code

- Non-singular:
$$x \neq x' \Rightarrow C(x) \neq C(x')$$

- non-singular enough to describe a single RV X

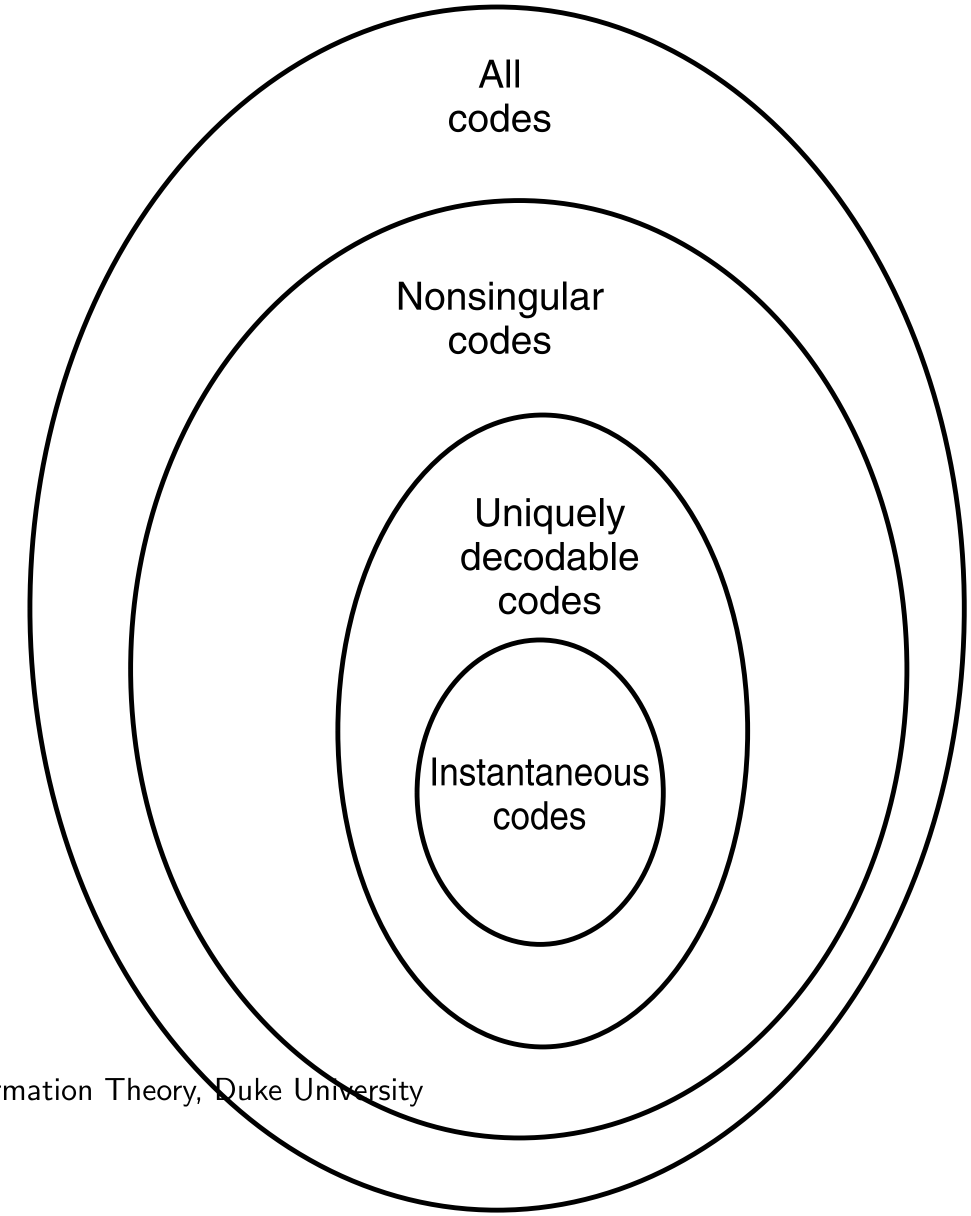
- When we send sequences of value of X, without "comma" can we still uniquely decode

- Uniquely decodable if extension of the code is nonsingular

$$C(x_1)C(x_2)\cdots C(x_n)$$

X	Singular	Nonsingular not uniquely decodable	Uniquely decoable	Prefix
1	0	0	10	0
2	0	010	00	10
3	0	01	11	110
4	0	10	110	111

- Uniquely decodable if only one possible source string producing it
- However, we have to look at entire string to determine
- Prefix code (instantaneous code): no codeword is a prefix of any other code


All
codes
Nonsingular
codes
Uniquely
decodable
codes
Instantaneous
codes

Expected code length

- Expected length $L(C)$ of a source code $C(x)$ for X with pdf $p(x)$

$$L(C) = \sum_{x \in \mathcal{X}} p(x) l(x)$$

- We wish to construct instantaneous codes of minimum expected length

Kraft inequality

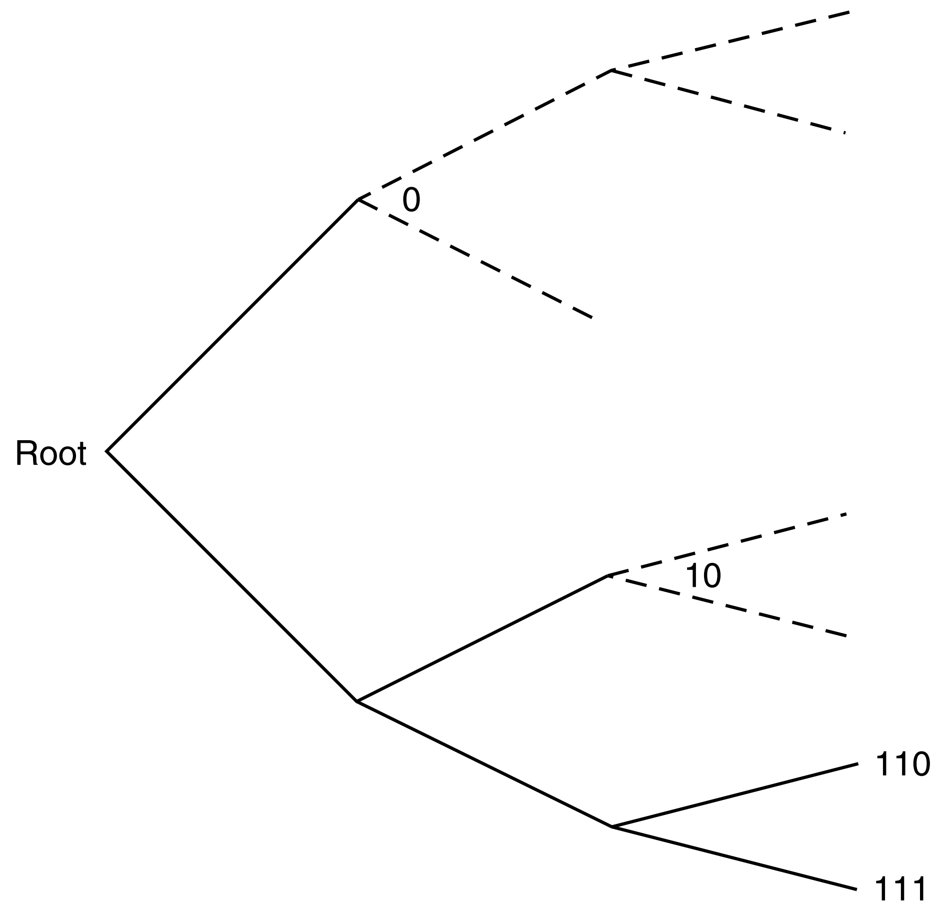
- By Kraft in 1949
- Coded over alphabet size D
- m codes with length $l_1, \ldots, l_m$
- The code length of all instantaneous code must satisfy Kraft inequality

$$\sum_{i=1}^{m} D^{-l_i} \leq 1$$

- Given $l_1, \ldots, l_m$ satisfy Kraft, can construct instantaneous code
- Can be extended to uniquely decodable code (McMillan inequality)

Proof of Kraft inequality

- Consider D-ary tree
- Each codeword is represented by a leaf node
- Path from the root traces out the symbol
- Prefix code: no codeword is an ancestor of any other codeword on the tree
- Each code eliminates its descendants as codewords


0
Root
10
110
111

- $l_{\max}$ be the length of longest codeword
- A codeword at level l_i has $D^{l_{\max}-l_i}$ descendants
- Descendant sets must be disjoint:

$$\sum D^{l_{\max}-l_i} \leq D^{l_{\max}}$$

$$\Rightarrow \sum D^{-l_i} \leq 1$$

- Converse: if $l_1, \ldots, l_{\max}$ satisfy Kraft inequality, can label first node at depth l_1, remove its descendants...
- Can extend to infinite prefix code $l_{\max} \to \infty$

Optimal expected code length

- One application of Kraft inequality
- Expected code length of D-ary is lower bounded by entropy:

$$L \geq H_D(X)$$

Proof:

$$\begin{aligned} L - H_D(X) &= \sum p_i l_i - \sum p_i \log_D \frac{1}{p_i} \\ &= D(p||r) + \log_D \frac{1}{c} \geq 0 \end{aligned}$$

$$r_i = D^{-l_i} / \sum_j D^{-l_j}, \quad c = \sum D^{-l_i} \leq 1$$

- Achieve minimum code length if
  - $c = 1$: Kraft inequality is equality
  - $r_i = p_i$: approximated pdf using D-ary alphabet is exact

- How to construct such an optimal code?

- Finding the D-adic distribution that is closet to distribution of X

- Construct the code by converse of Kraft inequality

Construction of optimal codes

- Finding the D-adic distribution that is closet to distribution of X is impractical because finding the closest D-adic distribution is not obvious

- Good suboptimal procedure
  - Shannon-Fano coding
  - Arithmetic coding

- Optimal procedure: Huffman coding

First step: finding optimal code length

- Solving optimization problem

$$\begin{aligned} \text{minimize}_{l_i} \quad & \sum_{i=1}^{m} p_i l_i \\ \text{subject to} \quad & \sum_{i=1}^{m} D^{-l_i} \le 1. \end{aligned}$$

- Solve using Lagrangian multiplier

$$J = \sum_{i=1}^{m} p_i l_i + \lambda(\sum_{i=1}^{m} D^{-l_i} - 1)$$

- Solution:

$$l_i^* = -\log_D p_i.$$

- Achieves the lower bound:

$$L^* = \sum p_i l_i^* = -\sum p_i \log_D p_i = H_D(X).$$

- Problem: $-\log_D p_i$ may not be an integer!

- Rounding up

$$l_i = \lceil -\log_D p_i \rceil.$$

  may not be optimal.

- Usable code constructions?

Summary

- Nonsingular $>$ Uniquely decodable $>$ Instantaneous codes
- Kraft inequality for Instantaneous code
- Entropy is lower bound on expected code length