Coding and Information Theory Overview
Chapter 1: Source Coding - B

Xuejun Liang

2022 Fall

Quick Review (1)

- The source alphabet of $S = \{s_1, s_2, \ldots, s_q\}$
- Probability distribution $P = (p_1, p_2, \ldots, p_q)$
- Code alphabet $T = \{t_1, \ldots, t_r\}$
- Code word: a sequence of symbols from T
- Encode source s = $X_1 X_2 X_3$ ..., where X_n = s_i
- Source code (simply code) C = $\{w_1, w_2, \ldots, w_q\}$
- The average word-length of C is

$$L(C) = \sum_{i=1}^{q} p_i l_i \,. \quad where\ l_i = |w_i|$$

Quick Review (2)

- A uniquely decodable code C
- Compute C_n and C_∞

  Algorithm to compute C1, C2, …, Cn_1, Cn

  C0 = C

  For each code-word cw in C

  For each code cw_1 in Cn_1

  If cw_1 is prefix of cw, then add cw/cw_1 in Cn

  If cw is prefix of cw_1, then add cw_1/cw in Cn
- The Sardinas-Patterson Theorem (Theorem 1.10)
  - A code C (finite) is uniquely decodable if and only if $C \cap C_\infty = \emptyset$

Example 1.12

- Let C be the ternary code {01, 1, 2, 210}.
  - C1 = {10}
  - C2 = {0}
  - C3 = {1}
  - Now $1 \in C \cap C_\infty$. So C is not uniquely decodable

– $10 \in$ C1 because **2.10 = 210,** with $2 \in$ C and $210 \in$ C

– $0 \in$ C2 because **1.0 = 10,** with $1 \in$ C and $10 \in$ C1

– $1 \in$ C3 because **0.1 = 01,** with $0 \in$ C2 and $01 \in$ C

Example 1.12 (Cont.)

- Let C be the ternary code {01, 1, 2, 210}.
  - C1 = {10} because **2.10 = 210,** with $2 \in C$ and $210 \in C$
  - C2 = {0} because **1.0 = 10,** with $1 \in C$ and $10 \in C1$
  - C3 = {1} because **0.1 = 01,** with $0 \in C2$ and $01 \in C$
  - Now $1 \in C \cap C_\infty$. So C is not uniquely decodable
- Can you find an example of non-unique decodability?

1.3 Instantaneous Codes

- Example 1.14
  - Consider the binary code C given by

$$s_1 \mapsto 0,\ s_2 \mapsto 01,\ s_3 \mapsto 11.$$

  - We have $\mathcal{C}_1 = \mathcal{C}_2 = \cdots = \{1\}$, so $\mathcal{C}_\infty = \{1\}$;
  - Thus $\mathcal{C} \cap \mathcal{C}_\infty = \emptyset$, so C is uniquely decodable
  - Consider a finite message t = 0111 . ...
  - We can not decode until we know how many 1's.
  - We say that C is not **instantaneous**.

Instantaneous Codes (cont.)

- Example 1.16
  - Consider the binary code D given by

    $$s_1 \mapsto 0,\ s_2 \mapsto 10,\ s_3 \mapsto 11,$$

  - the reverse of the code C in Example 1.14
  - this is uniquely decodable
  - It is also instantaneous
- Formal definition
  - A code C is instantaneous if, for each sequence of code-words $w_{i_1}w_{i_2}, \ldots w_{i_n}$, every code-sequence beginning $t = w_{i_1}w_{i_2}, \ldots w_{i_n} \ldots$ is decoded uniquely as $s = s_{i_1}s_{i_2}\ldots s_{i_n}\ldots$, no matter what the subsequent symbols in t are.

Prefix Code

- A code C is a prefix code if no code-word w_i is a prefix (initial segment) of any code-word w_j $(i \neq j)$; equivalently, $w_j \neq w_i w$ for any $w \in T^*$,
  - That is, $C_1 = \emptyset$
- **Theorem 1.17**
  - A code C is instantaneous if and only if it is a prefix code.

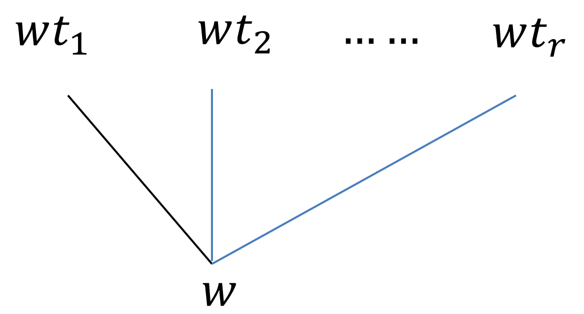
(⇒) If not prefix, then not instantaneous

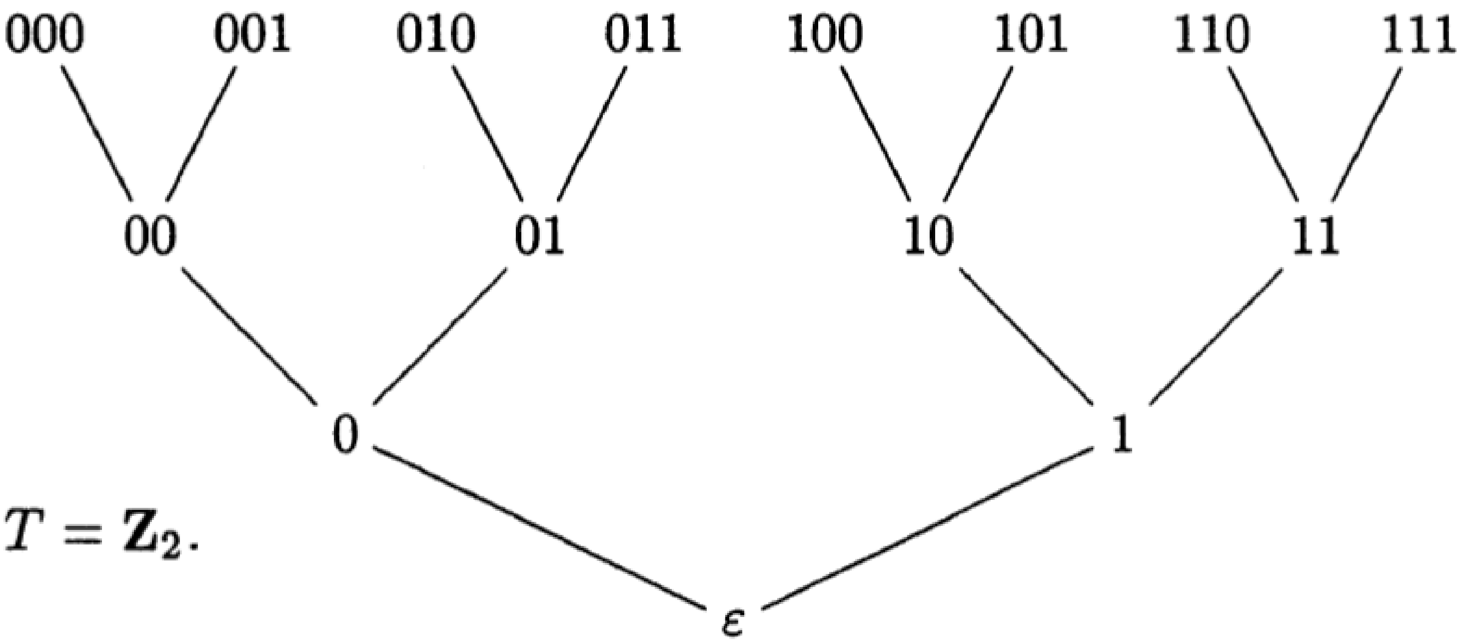
(⇐) If prefix, then instantaneous

- Note: A code C is instantaneous, if
  - $t = w_{i_1} w_{i_2}, \ldots w_{i_n}$ … is decoded uniquely as
  - $s = s_{i_1} s_{i_2} \ldots s_{i_n} \ldots,$

1.4 Constructing Instantaneous Codes

- $w \in T^*$
- $T = \{t_1, t_2, .., t_r\}$

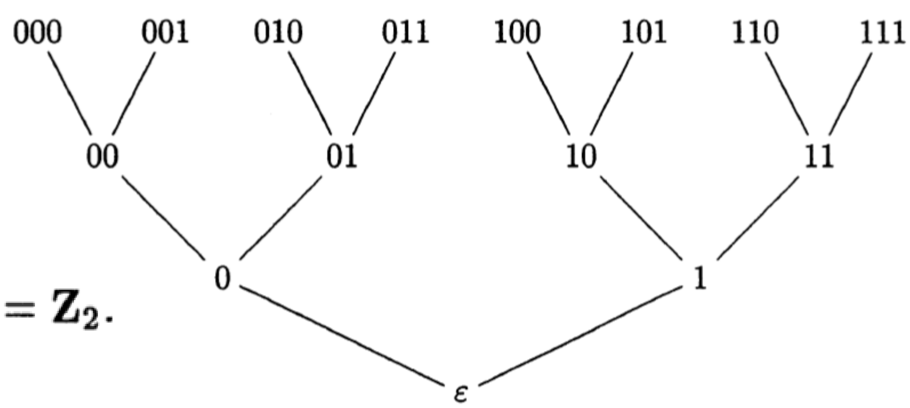




$T = \mathbf{Z}_2.$

Constructing Instantaneous Codes (Cont.)

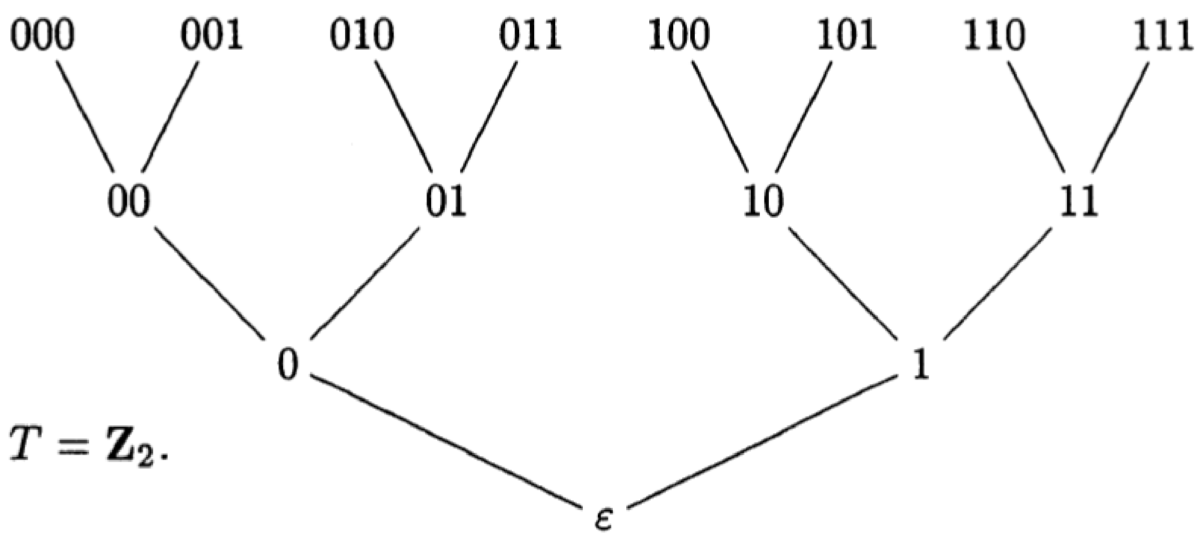
- A code C can be regarded as a finite set of vertices of the tree T^*.
- A word w_i is a prefix of w_j if and only if the vertex w_i is dominated by the vertex w_j
  - that is, there is an upward path in T^* from w_i to w_j
- C is instantaneous if and only if no vertex $w_i \in C$ is dominated by a vertex $w_j \in C$ $(i \neq j)$.



$T = \mathbf{Z}_2$.

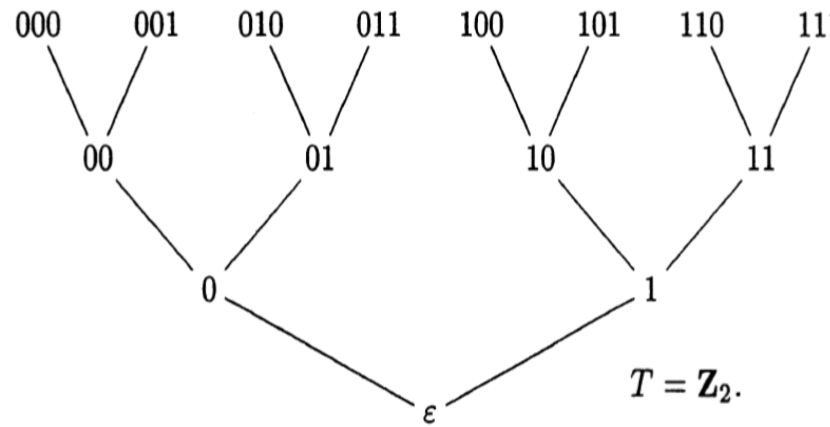
Example 1.18

- Let us find an instantaneous **binary** code C for a source S with five symbols $s_1, \dots, s_5$.



Example 1.19

- Is there an instantaneous **binary** code for this source S with word-lengths 1, 2, 3, 3, 4?
- No, Why?



- Is there an instantaneous **ternary** code for this source S with word-lengths 1, 2, 3, 3, 4?
- Yes. Why?

Example 1.19

- Is there an instantaneous **binary** code for this source S with word-lengths 1, 2, 3, 3, 4?
- No, Why?

- Is there an instantaneous **ternary** code for this source S with word-lengths 1, 2, 3, 3, 4?
- Yes. Why?