# Coding and Information Theory
# Overview
# Chapter 1: Source Coding

Xuejun Liang

2019 Fall

# Overview

- Information Theory and Coding Theory are two related aspects of the problem of how to transmit information efficiently and accurately from a source, through a channel, to a receiver.
- Based on Mathematics areas:
  - Probability Theory and Algebra
  - Combinatorics and Algebraic Geometry

# Important Problems

- How to compress information, in order to transmit it rapidly or store it economically
- How to detect and correct errors in information

# Information Theory vs. Coding Theory

- Information Theory uses probability distributions to quantify information (through the entropy function) , and to relate it to the average word-lengths of encodings of that information
  - In particular, Shannon's Fundamental Theorem Guarantees the existence of good error-correcting codes (ECCs)
- Coding Theory is to use mathematical techniques to construct ECCs, and to provide effective algorithms with which to use ECCs.

# Chapter 1: Source Coding

1.1 Definitions and Examples

1.2 Uniquely Decodable Codes

1.3 Instantaneous Codes

1.4 Constructing Instantaneous Codes

1.5 Kraft's Inequality

1.6 McMillan's Inequality

1.7 Comments on Kraft's and McMillan's Inequalities

# 1.1 Definitions and Examples

- A sequence $s = X_1 X_2 X_3 \ldots$ of symbols $X_n$, emitting comes from a source $S$
- The source alphabet of $S = \{s_1, s_2, \ldots, s_q\}$
- Consider $X_n$ as random variables and assume that
  - they are independent and
  - have the same probability distribution $p_i$.

$$\Pr(X_n = s_i) = p_i \quad \text{for } i = 1, \ldots, q.$$

$$p_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{q} p_i = 1$$

# Examples

- Example 1.1
  - $S$ is an unbiased die, $S$ = {1, . . . , 6} with $q$ = 6, $X_n$ is the outcome of the $n$-th throw, and $p_i = 1/6$.
- Example 1.2
  - $S$ is the weather at a particular place, with $X_n$ representing the weather on day $n$, $S$ = {good, moderate, bad}.

$$p_1 = 1/4,\ p_2 = 1/2,\ p_3 = 1/4.$$

- Example 1.3
  - $S$ is a book, $S$ consists of all the symbols used, $X_n$ is the $n$-th symbol in the book, and $p_i$ is the frequency of the $i$-th symbol in the source alphabet.

# Code alphabet, symbol, word

- Code alphabet $T = \{t_1, \ldots, t_r\}$ consisting of $r$ code-symbols $t_j$.
  - Depends on the technology of the channel
  - Call $r$ the radix (meaning "root“ or “base”)
  - Refer to the code as an $r$-ary code
  - When $r$ = 2, binary code, $T$ = $Z_2$ = {0, 1}
  - When $r$ = 3, ternary code, $T$ = $Z_3$ = {0, 1, 2}
- Code word: a sequence of symbols from $T$

# Encode and Example

- To encode $s$ = $X_1X_2X_3$ ..., we represent $X_n$ = $s_i$ by
  - $s_i$ → $w_i$ (its code word)
  - $s \rightarrow t$ (one by one)
  - we do not separate the code-words in $t$
- Example 1.4
  - If $S$ is an unbiased die, as in Example 1.1, take $T = Z_2$ and let $w_i$ be the binary representation of the source-symbol $s_i$ = $i$ ($i$ = 1, . . . ,6)
  - $s$ = 53214 → $t$ =10111101100
  - Could write $t$ = 101.11.10.1.100 for clearer exposition

# Define codes more precisely

- A word $w$ in $T$ is a finite sequence of symbols from $T$, its length $|w|$ is the number of symbols.
- The set of all words in $T$ is denoted by $T^*$, including empty word ε.
- The set of all non-empty words in $T$ is denoted by $T^+$

$$T^* = \bigcup_{n \geq 0} T^n \qquad \text{and} \qquad T^+ = \bigcup_{n > 0} T^n,$$

where $T^n = T \times \cdots \times T$

# Define codes more precisely (Cont.)

- A source code (simply a code) $C$ is a function $S \rightarrow T^+$

$$w_i = C(s_i) \in T^+, \qquad i = 1, 2, \ldots, q$$

- Regard $C$ as a finite set of words $w_1, w_2, \ldots, w_q$ in $T^+$.
- $C$ can be extended to a function $S^* \rightarrow T^*$

$$\mathbf{s} = s_{i_1} s_{i_2} \ldots s_{i_n} \mapsto \mathbf{t} = w_{i_1} w_{i_2} \ldots w_{i_n} \in T^*$$

- The image of this function is the set

$$C^* = \{w_{i_1} w_{i_2} \ldots w_{i_n} \in T^* \mid \text{each } w_{i_j} \in C,\ n \geq 0\}$$

- The average word-length of $C$ is

$$L(C) = \sum_{i=1}^{q} p_i l_i .$$

  - $where\ l_i = |w_i|$

# The aim is to construct codes $C$

a) there is easy and unambiguous decoding $t$ -> $s$ ,
b) the average word-length $L(C)$ is small.

- The rest of this chapter considers criterion (a) , and the next chapter considers (b).
- Example 1.5
  - The code $C$ in Example 1.4 has $l_1$ = 1, $l_2$ = $l_3$ = 2 and $l_4$ = $l_5$ = $l_6$ = 3, so

$$L(C) = \frac{1}{6}(1+2+2+3+3+3) = \frac{7}{3}.$$

# 1.2 Uniquely Decodable Codes

- A code $C$ is uniquely decodable (u.d. for short) if each $t \in T^*$ corresponds under $C$ to at most one $s \in S^*$;
  - in other words, the function $C : S^* \rightarrow T^*$ is one-to-one,
- Will always assume that the code-words $w_i$ in $C$ are distinct.
  - Under this assumption, the definition of unique decodability of $C$ is that whenever

$$u_1 \ldots u_m = v_1 \ldots v_n$$

with $u_1, \ldots, u_m, v_1, \ldots, v_n \in \mathcal{C}$, we have $m = n$ and $u_i = v_i$ for each $i$.

# Uniquely Decodable Codes (Cont.)

- Example 1.6
  - In Example 1.4, the binary coding of a die is not uniquely decodable.
  - Give an example.
  - Can you fix it?
- Theorem 1.7
  - If the code-words $w_i$ in $C$ all have the same length, then $C$ is uniquely decodable.
- If all the code-words in $C$ have the same length $l$, we call $C$ **a block code of length $\boldsymbol{l}$**.

# Uniquely Decodable Codes (Cont.)

- Example 1.8
  - The binary code $C$ given by

    $$s_1 \mapsto w_1 = 0,\ s_2 \mapsto w_2 = 01,\ s_3 \mapsto w_3 = 011$$

  - has variable lengths, but is still uniquely decodable.
  - for example, $\mathbf{t} = 001011010011 = 0.01.011.01.0.011$
  - $\Rightarrow \quad \mathbf{s} = s_1 s_2 s_3 s_2 s_1 s_3.$
- We define
  - $\mathcal{C}_0 = \mathcal{C}$, and
  - $\mathcal{C}_n = \{\, w \in T^+ \mid uw = v \text{ where } u \in \mathcal{C},\ v \in \mathcal{C}_{n-1} \text{ or } u \in \mathcal{C}_{n-1},\ v \in \mathcal{C} \,\} \quad (1.3)$
  - Note: $\mathcal{C}_1 = \{\, w \in T^+ \mid uw = v \text{ where } u, v \in \mathcal{C} \,\}.$

# Uniquely Decodable Codes (Cont.)

- For each $n \geq 1$; we then define $$C_\infty = \bigcup_{n=1}^{\infty} C_n. \quad (1.4)$$
  - Note: if $C_{n-1} = \emptyset$ then $C_n = \emptyset$,
- Example 1.9
  - Let $C$={0, 01, 011} as in Example 1.8. Then
  - $C_1 =$ ? $\quad C_2 =$ ? $\quad C_n =$ ? for all $n \geq 2$ $\quad C_\infty =$ ?
- Theorem 1.10 (The Sardinas-Patterson Theorem)
  - A code $C$ (finite) is uniquely decodable if and only if the sets $C$ and $C_\infty$ are disjoint.
  - A code $C$ (finite or infinite) is uniquely decodable if and only if $C_n \cap C_\infty = \emptyset$ and $C_n = \emptyset$ for some $n \geq 1$.

# Uniquely Decodable Codes (Cont.)

- Example 1.11
  - If $C$ = {0,01,011} as in Examples 1.8 and 1.9, then $C_\infty = \{1, 11\}$ which is disjoint from C.
- Example 1.12
  - Let $C$ be the ternary code {01, 1, 2, 210}. Then $C_1$ = {10},
  - $C_2$ = {0} and $C_3$ = {1}, so $1 \in C \cap C_\infty$ and thus $C$ is not uniquely decodable.
  - Can you find an example of non-unique decodability?
- Example 1.13
  - Find an example where all finite code-sequences are decoded uniquely, but some infinite ones are not.

# 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

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

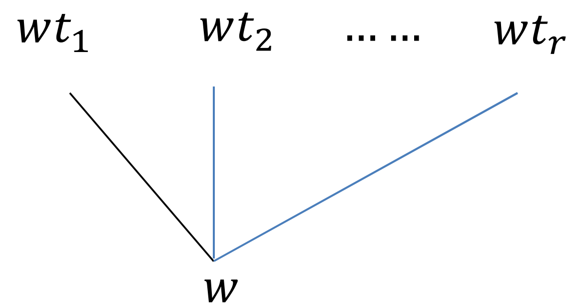
  - the reverse of the code $C$ in Example 1.14
  - this is uniquely decodable
  - It is also instantaneous
- Formal definition
  - A code $C$ is instantaneous if, for each sequence of code-words $w_{i_1} w_{i_2}, \ldots w_{i_n}$, every code-sequence beginning $t = w_{i_1} w_{i_2}, \ldots w_{i_n} \ldots$ 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.

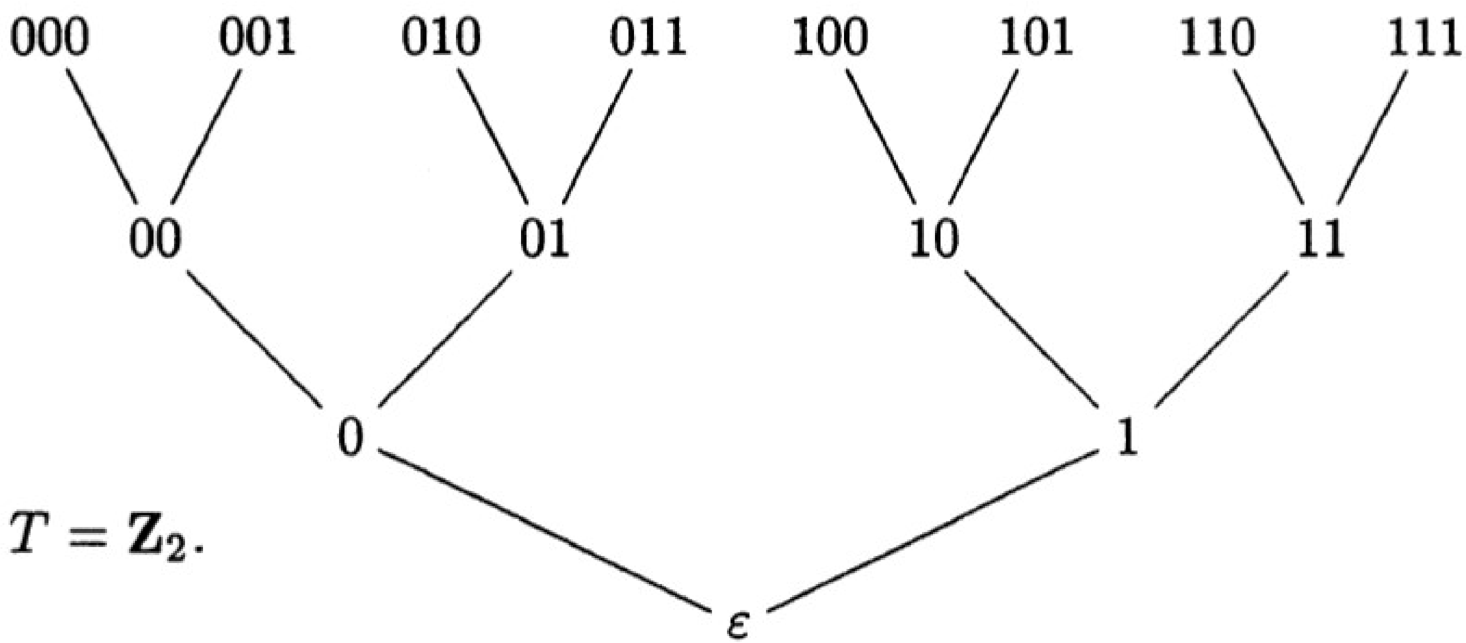
# Prefix Code

- 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, $c_1 = \emptyset$ in the notation
- Theorem 1.17
  - A code $C$ is instantaneous if and only if it is a prefix code.

# 1.4 Constructing Instantaneous Codes

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



# 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 $w_j \in C$ $(i \neq j)$.

# Examples

- 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?

# 1.5 Kraft's Inequality

- Theorem 1.20
  - There is an instantaneous $r$-ary code $C$ with word-lengths $l_1, \dots, l_q$, if and only if

$$\sum_{i=1}^{q} \frac{1}{r^{l_i}} \le 1 . \qquad (1.5)$$

  - Proof
    - $\Leftarrow$

$$r^{l-l_1} < r^l \sum_{i=1}^{q} \frac{1}{r^{l_i}} \le r^l ,$$

$$\sum_{i=1}^{k} r^{l-l_i} < r^l \sum_{i=1}^{q} \frac{1}{r^{l_i}} \le r^l ,$$

    - $\Rightarrow$

$$\sum_{i=1}^{q} r^{l-l_i} \le r^l ,$$

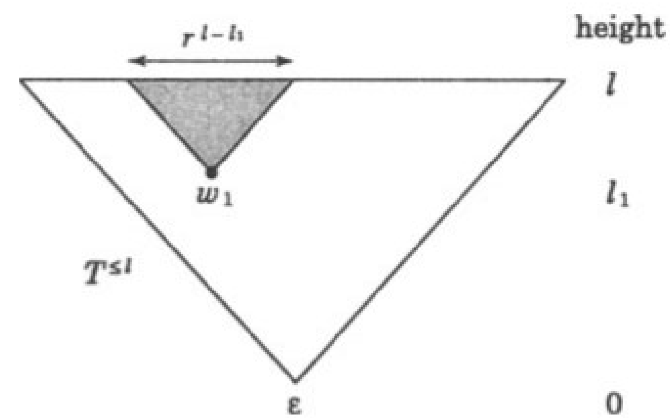



Figure 1.3

# 1.6 McMillan's Inequality

- Theorem 1.21
  - There is a uniquely decodable $r$-ary code $C$ with word-lengths $l_1, \dots, l_q$, if and only if

$$\sum_{i=1}^{q} \frac{1}{r^{l_i}} \leq 1 . \qquad (1.6)$$

- Corollary 1.22
  - There is an instantaneous $r$-ary code with word-lengths $l_1, \dots, l_q$, if and only if there is a uniquely decodable $r$-ary code with these word-lengths .

# 1.7 Comments on Kraft's and McMillan's Inequalities

- Comment 1.23
  - Theorems 1.20 and 1.21 do not say that an $r$-ary code with word-lengths $l_1, \dots, l_q$ is instantaneous or uniquely decodable if and only if $\sum r^{-l_i} \leq 1$
  - Examples: $C = \{0, 01{,}011\}$ and $C = \{0{,}01{,}001\}$
- Comment 1.24
  - Theorems 1.20 and 1.21 assert that if $\sum r^{-l_i} \leq 1$ then there exist codes with these parameters which are instantaneous and uniquely decodable.
  - Example: $C = \{0{,}10{,}110\}$

# Comments (Cont.)

- Comment 1.25
  - If an $r$-ary code $C$ is uniquely decodable, then it need not be instantaneous, but by Corollary 1.22 there must be an instantaneous $r$-ary code with the same word-lengths.
  - Examples: $C = \{0, 01, 11\}$ and $D = \{0, 10, 11\}$
- Comment 1.26
  - The summand $r^{-l_i}$ in $\mathrm{K} = \sum r^{-l_i}$ corresponds to the rather imprecise notion of the "proportion" of the tree $T^*$ above a vertex $w_i$ of height $l_i$.