INFORMATION THEORY & CODING

Week 6 : Source Coding 2

Dr. Rui Wang

Department of Electrical and Electronic Engineering
Southern Univ. of Science and Technology (SUSTech)

Email: wang.r@sustech.edu.cn

October 25, 2022

Review Summary

- **Classes of codes**

  Prefix codes $\Rightarrow$ Uniquely decodable codes $\Rightarrow$ Nonsingular codes

- **Kraft inequality**

  Prefix codes $\Leftrightarrow \sum D^{-\ell_i} \leq 1$.

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Outline

- **Extended Kraft inequality for prefix code**
- **Kraft inequality for uniquely decodable code**

  Uniquely decodable code does NOT provide more choices than prefix code

- **Bounds on optimal expected length**

  Entropy length is achievable when jointly encoding a random sequence.

Extended Kraft Inequality

Theorem 5.5.1 (Extended Kraft Inequality)

Kraft inequality holds also for all countably infinite set of codewords, i.e., the codeword lengths satisfy the extended Kraft inequality,

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

Conversely, given any $\ell_1, \ell_2, \ldots$ satisfying the extended Kraft inequality, we can construct a prefix code with these codeword lengths.

$1101 \Rightarrow 0.1101 = 2^{-1}+2^{-2}+2^{-4} = \frac{13}{16}$

$\Downarrow$

$[0.1101, 0.1101+0.0001)$.

Consider a D-ary code $y_1 y_2 \cdots y_L$

$\rightarrow 0.y_1 y_2 \cdots y_l = y_1 \cdot D^{-1} + y_2 \cdot D^{-2} + \cdots + y_l D^{-l}$

$= \sum_{i=1}^{l} y_i D^{-i}$

$\rightarrow [0.y_1 y_2 \cdots y_l,\ 0.y_1 y_2 \cdots y_l + D^{-l})$

proof:

suppose the i-th codeword has a length l_i,

it can be represented by $y_1 y_2 \cdots y_{l_i} \xrightarrow{\text{映射}}$

$0.y_1 y_2 \cdots y_{l_i} = \sum_{j=1}^{l_i} y_j D^{-j}$

$\Downarrow$

length D^{-l_i}: $[0.y_1 y_2 \cdots y_{l_i},\ 0.y_1 y_2 \cdots y_{l_i} + D^{-l_i}) \rightarrow I_i$

j-th codeword: l_j, $y_1' y_2' \cdots y_j' \Rightarrow [0.y_1' \cdots y_{l_j}',\ 0.y_1' \cdots y_{l_j}' + D^{-l_j}) \rightarrow I_j$

If $0.y_1 y_2 \cdots y_{l_i} \in [0.y_1' \cdots y_{l_j}',\ 0.y_1' \cdots y_{l_j}' + D^{-l_j})$

Because of prefix code, $I_i \cap I_j = \phi, \forall i \neq j$

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

泊松分布: $P_r[X=k] = \frac{\lambda^k}{k!} e^{-\lambda}, k=0,1,2,\cdots$

构造: $l_1, l_2, \cdots l_\infty$, $\sum_{i=1}^{\infty} D^{-l_i} \leq 1$

$l_1 \leq l_2 \leq l_3 \cdots \leq \cdots$

$[0, D^{-l_1}) \Rightarrow 0.000\cdots \Rightarrow \underbrace{00\cdots 0}_{l_1}$

$[D^{-l_1}, D^{-l_1}+D^{-l_2}) \Rightarrow 0.\underbrace{00\cdots 01}_{l_1} \Rightarrow \underbrace{00\cdots 1}_{l_1}\underbrace{0\cdots 0}_{l_2-l_1}$

$[D^{-l_1}+D^{-l_2}, D^{-l_1}+D^{-l_2}+D^{-l_3}) \Rightarrow 0.0\cdots \underset{l_1}{1} 0\cdots \underset{l_2}{1} \Rightarrow \underbrace{0\cdots 1}_{l_1}\underbrace{0\cdots 1}_{l_2-l_1}$

Extended Kraft Inequality

Theorem 5.2.2 (Extended Kraft Inequality)

Kraft inequality holds also for all countably infinite set of codewords.

Proof.

Consider the ith codeword $y_1 y_2 \cdots y_{\ell_i}$. Let $0.y_1 y_2 \cdots y_{\ell_i}$ be the real number given by the D-ary expansion

$$0.y_1 y_2 \cdots y_{\ell_i} = \sum_{j=1}^{\ell_i} y_j D^{-j},$$

which corresponds to the interval

prefix code

$$[0.y_1 y_2 \cdots y_{\ell_i}, 0.y_1 y_2 \cdots y_{\ell_i} + \frac{1}{D^{\ell_i}}).$$

□

Extended Kraft Inequality

Theorem 5.2.2 (Extended Kraft Inequality)

Kraft inequality holds also for all countably infinite set of codewords.

Proof. (cont.)

By the prefix condition, these intervals are disjoint in the unit interval $[0, 1]$. Thus, the sum of their lengths is ≤ 1. This proves that

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

For converse, reorder indices in increasing order and assign intervals as we walk along the unit interval. $\square$

Kraft Inequaltiy for Uniquely Decodable Codes

由之前可知 prefix code is a subset of uniquely decodable code.

Theorem 5.2.3 (McMillan)

说明 uniquely decodable code 在 code length 方面, prefix code 都能满足.

The codeword lengths of any uniquely decodable D-ary code must satisfy the Kraft inequality

$$\sum D^{-\ell_i} \leq 1.$$

Conversely, given a set of codeword lengths that satisfy this inequality, it is possible to construct a uniquely decodable code with these codeword lengths.

Proof.

Consider C^k, the k-th extension of the code by k repetitions. Let the codeword lengths of the symbols $x \in \mathcal{X}$ be $\ell(x)$. For the k-th extension code, we have

$$\ell(x_1, x_2, \ldots, x_k) = \sum_i^k \ell(x_i).$$

很巧妙的思想

proof: uniquely decodable $\Rightarrow \sum D^{-l_i} \leq 1$

① Consider arbitrary source sequence $x^k = x_1 x_2 \cdots x_k$ with UD.

$l(x^k) = l(x_1) + l(x_2) + \cdots + l(x_k)$

$\rightarrow \left[\sum_{x \in \mathcal{X}} D^{-l(x)}\right]^k = \sum_{x_1 \in \mathcal{X}} D^{-l(x_1)} \sum_{x_2 \in \mathcal{X}} D^{-l(x_2)} \cdots \sum_{x_k \in \mathcal{X}} D^{-l(x_k)}$

$= \sum_{x_1} \sum_{x_2} \cdots \sum_{x_k} D^{-[l(x_1) + l(x_2) + \cdots + l(x_k)]}$

Suppose the maximum codeword length is l_{max}

$l(x_1) + l(x_2) + \cdots + l(x_k) = \sum_{i=1}^{k} l(x_i) \in [1, k l_{max}]$

let $a(m)$ be the number of $(x_1, x_2 \cdots x_k)$ with $\sum_{i=1}^{k} l(x_i) = m$

$\rightarrow = a(1) D^{-1} + a(2) D^{-2} + \cdots + a(m) D^{-m} + \cdots + a(k l_{max}) D^{-k l_{max}}$

$= \sum_{m=1}^{k l_{max}} a(m) D^{-m}$

$a(m) \leq D^m$: 长度为m的D进制编码 (仔细理解!)

$\rightarrow \leq \sum_{m=1}^{k l_{max}} D^m D^{-m} = k l_{max}$

$\rightarrow \left[\sum_{x \in \mathcal{X}} D^{-l(x)}\right]^k \leq k l_{max} \rightarrow \sum_{x \in \mathcal{X}} D^{-l(x)} \leq (k l_{max})^{\frac{1}{k}}$

when $k \to \infty$, $\sum_{x \in \mathcal{X}} D^{-l(x)} \leq 1$.

Kraft Inequaltiy for Uniquely Decodable Codes

Theorem 5.5.1 (McMillan)

The codeword lengths of any uniquely decodable D-ary code must satisfy the Kraft inequality

$$\sum D^{-\ell_i} \leq 1.$$

Proof. (cont.)

Consider

$$\begin{aligned}
\left(\sum_{x\in\mathcal{X}} D^{-\ell(x)}\right)^k &= \sum_{x_1\in\mathcal{X}}\sum_{x_2\in\mathcal{X}}\cdots\sum_{x_k\in\mathcal{X}} D^{-\ell(x_1)}D^{-\ell(x_2)}\ldots D^{-\ell(x_k)} \\
&= \sum_{x_1,x_2,\cdots x_k\in\mathcal{X}^k} D^{-\ell(x_1)}D^{-\ell(x_2)}\ldots D^{-\ell(x_k)} \\
&= \sum_{x^k\in\mathcal{X}^k} D^{-\ell(x^k)}
\end{aligned}$$

Kraft Inequaltiy for Uniquely Decodable Codes

Theorem 5.5.1 (McMillan)

The codeword lengths of any uniquely decodable D-ary code must satisfy the Kraft inequality

$$\sum D^{-\ell_i} \leq 1.$$

Proof. (cont.)

Let $\ell_{\max}$ be the maximum codeword length and $a(m)$ is the number of source sequences x^k mapping into codewords of length m. Unique decodability implies that $a(m) \leq D^m$. We have

$$\Big(\sum_{x\in\mathcal{X}} D^{-\ell(x)}\Big)^k = \sum_{x^k\in\mathcal{X}^k} D^{-\ell(x^k)} = \sum_{m=1}^{k\ell_{\max}} a(m)D^{-m}$$

$$\leq \sum_{m=1}^{k\ell_{\max}} D^m D^{-m}$$

$$= k\ell_{\max}$$

Kraft Inequaltiy for Uniquely Decodable Codes

Theorem 5.5.1 (McMillan)

The codeword lengths of any uniquely decodable D-ary code must satisfy the Kraft inequality

$$\sum D^{-\ell_i} \leq 1.$$

Proof. (cont.)

$$\left(\sum_{x\in\mathcal{X}} D^{-\ell(x)}\right)^k \leq k\ell_{\text{max}}.$$

Hence,

$$\sum_j D^{-\ell_j} \leq (k\ell_{\text{max}})^{1/k}$$

holds for all k. Since the RHS$\rightarrow 1$ as $k \rightarrow \infty$, we prove the Kraft inequality. For the converse part, we can construct a prefix code as in **Theorem 5.2.1**, which is also uniquely decodable. □

Optimal Codes

Problem To find the set of lengths $\ell_1, \ell_2, \ldots, \ell_m$ satisfying the Kraft inequality and whose expected length $L = \sum p_i \ell_i$ is minimized.

Optimization:

minimize $L = \sum p_i \ell_i$

subject to $\sum D^{-\ell_i} \leq 1$ and ℓ_i's are integers.

$X, (P_1\ P_2 \cdots P_m)$

$l_1\ l_2 \cdots l_m$

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

Optimal Code

$\min_{(l_1, l_2 \cdots l_m)} \sum_{i=1}^{m} P_i l_i$ (平均编码长度)

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

哈夫曼编码.

integer Programming

(整数编码).

Optimal Codes

Theorem 5.3.1

The expected length L of any prefix D-ary code for a random variable X is no less than $H_D(X)$, i.e.,

$$L \geq H_D(X),$$

with equality iff $D^{-\ell_i} = p_i$.

Proof.

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

"=" holds if $c = 1$ and $r_i = p_i$.

where $r_i = D^{-\ell_i} / \sum_j D^{\ell_j}$ and $c = \sum D^{-\ell_i} \leq 1$. □

$1、 L = \sum_i p_i l_i \geq H_D(X)$

proof: $L - H_D(X) = \sum_i p_i l_i - \sum_i p_i \log_D \frac{1}{p_i}$

$= -\sum_i p_i \cdot \log_D D^{-l_i} - \sum_i p_i \log_D \frac{1}{p_i}$

$= \sum_i p_i \log_D \frac{p_i}{D^{-l_i}}$

Let $r_i = \frac{D^{-l_i}}{\sum_j D^{-l_j}}$, $\sum r_i = 1$

$\to = \sum_i p_i \log_D \left(\frac{p_i}{D^{-l_i}} \frac{r_i}{r_i}\right)$

$= \sum_i p_i \log_D \frac{p_i}{r_i} + \sum_i p_i \log_D \frac{r_i}{D^{-l_i}}$

$= \sum_i p_i \log_D \frac{p_i}{r_i} - \underbrace{\sum_i p_i \log_D \underbrace{\sum_j D^{-l_j}}_{\leq 1}}_{\leq 0}$ ... $\geq 0.$

$= D(p\|r) + \geq 0 \Rightarrow \geq 0$

$\to L \geq H_D(X)$

$\text{"="} \Leftrightarrow \left\{ \begin{array}{l} D(p\|r) = 0 \Rightarrow p_i = r_i = \frac{D^{-l_i}}{\sum_j D^{-l_j}} \\ \sum_j D^{-l_j} = 1 \end{array} \right\} \Rightarrow p_i = D^{-l_i}$

Optimal Codes

Theorem 5.3.1

The expected length L of any prefix D-ary code for a random variable X is no less than $H_D(X)$, i.e.,

$$L \geq H_D(X),$$

with equality iff $D^{-\ell_i} = p_i$.

Definition

A probability distribution is called D-adic if each of the probabilities is equal to D^{-n} for some n. Thus, we have equality in the theorem iff the distribution of X is D-adic.

Remark

$H_D(X)$ is a lower bound on the optimal code length. The equality holds iff p is D-adic.

从这开始讲.

Bound on the Optimal Code Length

Theorem 5.4.1 (Shannon Codes)

Let $\ell_1^*, \ell_2^*, \ldots, \ell_m^*$ be optimal codeword lengths for a source distribution $\mathbf{p}$ and a D-ary alphabet, and let L^* be the associated expected length of an optimal code ($L^* = \sum p_i \ell_i^*$). Then

$$H_D(X) \le L^* < H_D(X) + 1.$$

Proof.

Take $\ell_i = \lceil -\log_D p_i \rceil$. Since

$$\sum_{i \in \mathcal{X}} D^{-\ell_i} \le \sum p_i = 1,$$

these lengths satisfy Kraft inequality and we can create a prefix code. Thus,

$$\begin{aligned} L^* &\le \sum p_i \lceil -\log_D p_i \rceil \\ &< \sum p_i(-\log_D p_i + 1) \\ &= H_D(X) + 1. \end{aligned}$$

□

Suppose the R.V. PMF is $(p_1, p_2, \cdots p_m)$

Let $l_i = \lceil -\log_D p_i \rceil$ (向上取整)

$$\sum_{i=1}^m D^{-l_i} \le \sum_{i=1}^m D^{-(-\log_D p_i)} = \sum_{i=1}^m p_i = 1$$

$$L^* \le L = \sum_{i=1}^m p_i l_i = \sum_{i=1}^m p_i \lceil -\log_D p_i \rceil$$

$$< \sum_{i=1}^m p_i(-\log_D p_i + 1)$$

$$H_D(X) + 1 = -\sum_i p_i \log_D p_i + \sum_i p_i$$

Consider a sequence of i.i.d. R.V.s

$$X^k = x_1 x_2 \cdots x_k.$$

Scheme1 / code extension.

x	codeword
a_1	$C(a_1)$
a_2	$C(a_2)$
$\vdots$	$\vdots$
a_m	$C(a_m)$

optimal prefix code C_1

$$H_D(X) \le L^* \le H_D(X) + 1$$

$$\rightarrow \underbrace{kH_D(X) \le kL^* \le kH_D(X) + k}_{\text{gap: } k \text{ bits.}}$$

scheme 2

x^k	codeword
$b_1^1 b_2^1 \cdots b_k^1$	$C(b_1^1 b_2^1 \cdots b_k^1)$
$\vdots$	$\vdots$
$b_1^{m^k} b_2^{m^k} \cdots b_k^{m^k}$	$C(b_1^{m^k} b_2^{m^k} \cdots b_k^{m^k})$

m^k

$$H_D(X^k) \le \tilde{L}^* < H_D(X^k) + 1$$

$$\rightarrow k \cdot H_D(X) \le \tilde{L}^* < k H_D(X) + 1$$

$$H_D(X) \le \frac{\tilde{L}^*}{k} < H_D(X) + \frac{1}{k}$$

$$k \to \infty \Rightarrow \frac{\tilde{L}^*}{k} = H_D(X)$$

$\tilde{L}^*$: 表达k个随机变量需要的比特数.

Bound on the Optimal Code Length

Theorem 5.4.2

Consider a system in which we send a sequence of n symbols from $\mathcal{X}$. The symbols are assumed to be i.i.d. according to $p(x)$. The minimum expected codeword length per symbol satisfies

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq L_n^* < \frac{H(X_1, X_2, \ldots, X_n)}{n} + \frac{1}{n}.$$

Proof.

First,

$$L_n = \frac{1}{n} \sum p(x_1, x_2, \ldots, x_n) \ell(x_1, x_2, \ldots, x_n) = \frac{1}{n} E[\ell(X_1, X_2, \ldots, X_n)]$$

We also have

$$H(X_1, X_2, \ldots, X_n) \leq E[\ell(X_1, X_2, \ldots, X_n)] < H(X_1, X_2, \ldots, X_n) + 1.$$

Since $X_1, X_2, \ldots, X_n$ are i.i.d., $H(X_1, X_2, \ldots, X_n) = nH(X)$. $\square$

Textbook

Related Sections : 5.3 - 5.5