第3章 信源压缩编码理论（第二部分）

樊平毅 教授

清华大学电子工程系 WIST LAB.

Email: fpy@tsinghua.edu.cn

2022年9月14日

信源压缩的核心问题

- 对于一个随机信源，我们知道了其熵的计算方法，也了解了随机过程的熵率计算方法

- **信源压缩需要回答几个核心问题：**

（与常见的表示相比，用较少的信息比特可以表示出来）

- 信源完备表示需要的最小信息速率是什么？ 信息熵或熵率
- 信源表示，如果是非完备的表示，需要有多少损失？如何刻画其损失量？
- 信源压缩表示的具体编码方法是什么？ 与信息熵或熵率的差值在什么范围

定义 关于随机变量 X 的信源编码 C 是从 X 的取值空间 $\mathcal{X}$ 到 D^* 的一个映射，其中 D^* 表示 D 元字母表 D 上有限长度的字符串所构成的集合。用 $C(x)$ 表示 x 的码字并用 $l(x)$ 表示 $C(x)$ 的长度。

概念解释

例如，C(红)$=00$，C(蓝)$=11$ 是 $\mathcal{X}=\{$红，蓝$\}$ 关于字母表 $D=\{0,1\}$ 的一个信源编码

定义 设随机变量 X 的概率密度函数为 $p(x)$，定义信源编码 $C(x)$ 的期望长度 (expected length) 为

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

其中 $l(x)$ 表示对应于 x 的码字长度。

统计意义上的定义

例题

设随机变量 X 的分布及其码字分配如下：

$$\Pr(X=1)=\frac{1}{2},\ 码字\ C(1)=0$$

$$\Pr(X=2)=\frac{1}{4},\ 码字\ C(2)=10$$

$$\Pr(X=3)=\frac{1}{8},\ 码字\ C(3)=110$$

$$\Pr(X=4)=\frac{1}{8},\ 码字\ C(4)=111$$

易知 X 的熵 $H(X)$ 为 1.75 比特

期望长度 $L(C)=El(X)$ 亦是 1.75 比特

期望长度正好等于其熵值的编码

任何一个比特序列都可以惟一地解码成为关于 X 中的字符序列

例如，比特串 0110111100110 解码后为 134213。

0，	1
110，	3
111，	4
10，	2
0，	1
110	3

唯一可译码的要求

例题

考虑关于随机变量编码的另一简单例子：

$$\Pr(X=1)=\frac{1}{3},\ \text{码字}\ C(1)=0$$

$$\Pr(X=2)=\frac{1}{3},\ \text{码字}\ C(2)=10$$

$$\Pr(X=3)=\frac{1}{3},\ \text{码字}\ C(3)=11$$

该编码也是惟一可译的

熵为 $\log 3=1.58$ 比特

编码的期望长度 为 1.66 比特

此时 $El(X)>H(X)$

（莫尔斯码） 莫尔斯码是关于英文字母表的一个相当有效的编码方案

定义 如果编码将 X 的取值空间中的每个元素映射成 D^* 中不同的字符串

$$x\neq x'\Rightarrow C(x)\neq C(x')$$

则称这个编码是非奇异的(nonsigular)

唯一可译性成为信源编码的一个基本要求

问题： 如何将一个编码模式的性能讨论纳入概率统计学的框架？

答案：需要引入随机序列的概念，

◆一种是独立同分布的随机序列， （分组码）

◆一种是具有一定相关性的随机序列，如引入Markov 结构。 （卷积码）

定义 编码 C 的扩展(extension) C^* 是从 $\mathcal{X}$ 上的有限长字符串到 $\mathcal{D}$ 上的有限长字符串的映射

$$C(x_1x_2\cdots x_n)=C(x_1)C(x_2)\cdots C(x_n)$$

其中 $C(x_1)C(x_2)\cdots C(x_n)$ 表示相应码字的串联

例题： 0,1 重复编码 若 $C(x_1)=00$，$C(x_2)=11$，则 $C(x_1x_2)=0011$

定义 如果一个编码的扩展编码是非奇异的，则称该编码是唯一可译的(uniquely decodable)。

表明：所有有限长的编码串都是唯一可译的，不会出现歧义的情况

定义 若码中无任何码字是其他码字的前缀，则称该编码为前缀码（prefix code)或即时码(instantaneous code)。

一个码字是其他某个码字头部的一部分

（因为采用的是顺序译码）

简单的处理方式：任何编码码字不包含一个子串是其他码字。这样，在译码过程中，只需比对字符串与码字的编码字母就可，不会发生歧义。

优点：由于何时结束码字都可以瞬时辨认出来，因而无需参考后面的码字就可译出即时码。

例题

即时码是一个自我间断码

设随机变量 X 的分布及其码字分配如下：

$$\Pr(X=1)=\frac{1}{2}，码字\ C(1)=0$$

$$\Pr(X=2)=\frac{1}{4}，码字\ C(2)=10$$

$$\Pr(X=3)=\frac{1}{8}，码字\ C(3)=110$$

$$\Pr(X=4)=\frac{1}{8}，码字\ C(4)=111$$

编码方案所产生的二元串 01011111010

将它分解成 0，10，111，110，10

验证一下是否即时码

观察一下序列分割译码结果

压缩编码需要考虑两个基本需求：

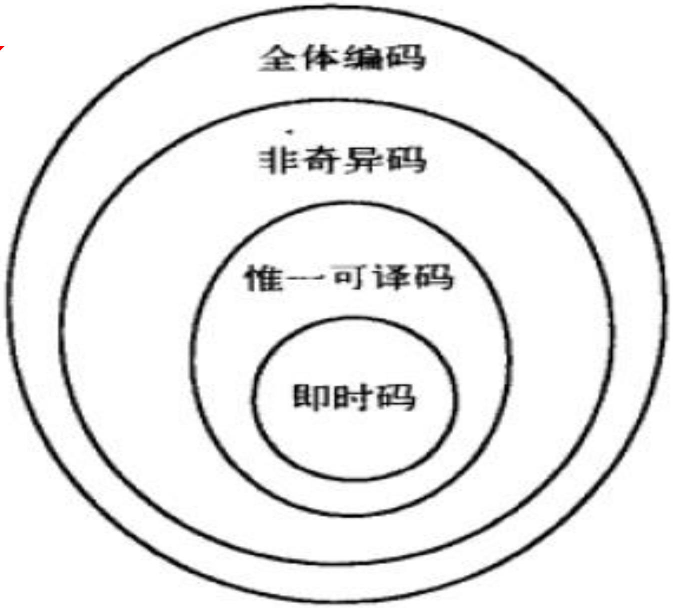
- 码字的唯一可译性
- 平均编码长度尽可能接近信息熵

例题：编码的种类

表 5-1　码的几种类型

X	奇异的	非奇异，但不是惟一可译的	惟一可译，但不是即时的	即时的
1	0	0	10	0
2	0	010	00	10
3	0	01	11	110
4	0	10	110	111

即时特征被破坏



即时码的优势就是译码速度快，
扫描模式即可迅速实现

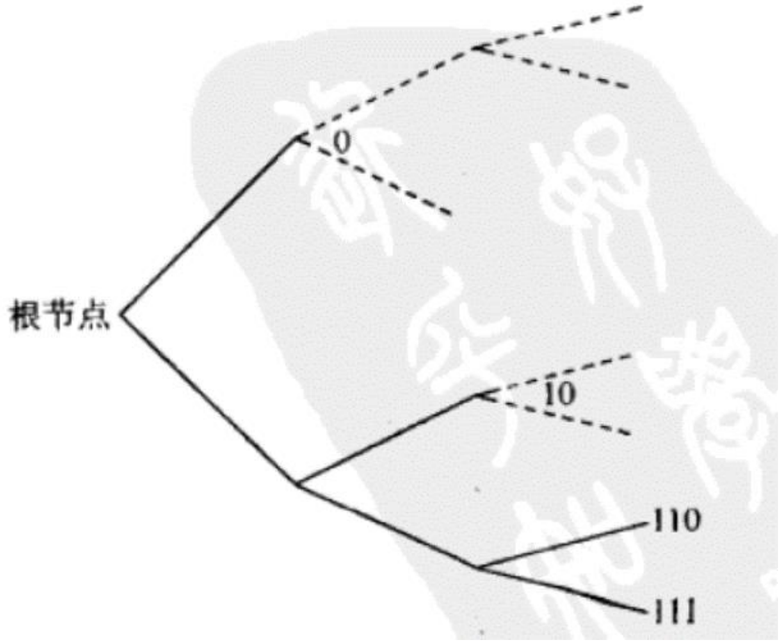
可采用并联译码策略，
分段处理，对多段进行同时译码

思考

挑战：为了保证码字的唯一可译性，理论上的判定准则是什么？

源于观察树的结构，每片叶子都有其独特性，
有自己独特的位置，没有两片叶子是完全相同的

构造一个2叉树的扩张模型



唯一可译性限制条件

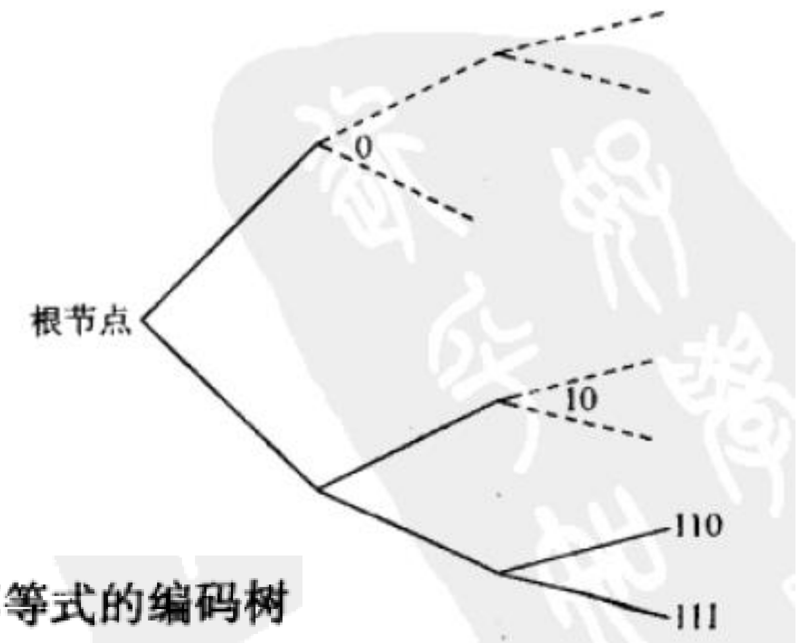
挑战： 为了保证码字的唯一可译性，理论上的判定准则是什么？

定理：(Kraft 不等式) 对于 D 元字母表上的即时码(前缀码)，码字长度 $l_1, l_2, \cdots, l_m$

必定满足不等式

$$\sum_i D^{-l_i} \leqslant 1$$

反之，若给定满足以上不等式的一组码字长度，则存在一个相应的即时码，其码字长度就是给定的长度。

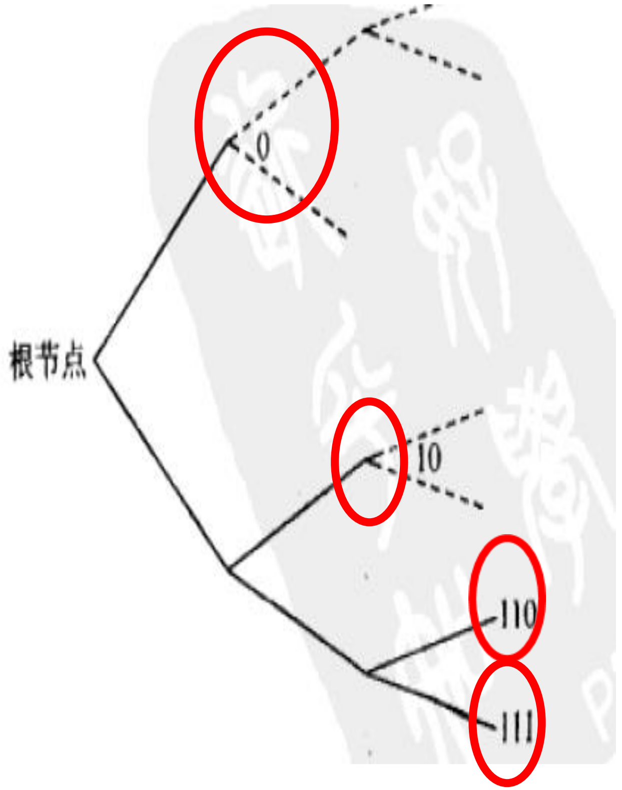


关于 Kraft 不等式的编码树

源于观察树的结构，每片叶子都有其独特性，
有自己独特的位置，没有两片叶子是完全相同的

构造一个2叉树的扩张模型

证明：考虑每一节点均含 D 个子节点的 D 叉树。假定树枝代表码字的字符。例如，源于根节点的 D 条树枝代表着码字第一个字符的 D 个可能值。另外，每个码字均由树的一片叶子表示。因此，始于根节点的路径可描绘出码字中的所有字符。作为例子，对于二叉树情形如图 5.2 所示。码字的前缀条件表明树中无一码字是其他任一码字的祖先。因而，在这样的编码树中，每一码字都去除了它的可能成为码字的所有后代。



0， 1， 2， 3 的即时码编码模式

小技巧

令 l_{max} 为码字集中最长码字长度。考虑在树中 l_{max} 层的所有节点，可知其中有些是码字，有些是码字的后代，而另外的节点既不是码字，也不是码字的后代。在树中 l_i 层的码字拥有 l_{max} 层中的 $D^{l_{max}-l_i}$ 个后代。所有这样的后代集不相交。而且，这些集合中的总节点数必定小于或等于 $D^{l_{max}}$。因此，对所有码字求和，则可得

所有码字后代的总和不超过总约束

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

$$\Downarrow$$

$$\sum D^{-l_i} \leqslant 1$$

这就是 Kraft 不等式

逆命题证明

（1）

反之，若给定任意一组满足 Kraft 不等式的码字长度 $l_1, l_2, \cdots, l_m$，总可以构造出如图~~5-~~2所示的编码树。将第一个深度为 l_1 的节点（依字典序）标为码字 1，同时除去树中属于它的所有后代。然后在剩余的节点中找出第一个深度为 l_2 的节点，将其标为码字 2，同时除去树中所有属于它的所有后代，等等。按此方法继续下去，即可构造出一个码字长度为 $l_1, l_2, \cdots, l_m$ 的前缀码。

（2）

（3）

Kraft 不等式是唯一可译码的充分必要条件，
也是编码规则中最重要的约束

一般性证明

定理 5.2.2(推广的 Kraft 不等式) 对任意构成前缀码的可数无限码字集，码字长度也满足推广的 Kraft 不等式。

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

反之，若给定任意满足推广的 Kraft 不等式的 $l_1, l_2, \cdots$，则可构造出具有相应码长的前缀码

证明： 不妨设 D 元字母表为 $\{0,1,\cdots,D-1\}$，第 i 个码字是 $y_1 y_2 \cdots y_{l_i}$。记 $0.y_1 y_2 \cdots y_{l_i}$ 是以 D 进制表示的实值小数，即

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

由此，这个码字对应于一个区间

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

一般性证明

这是一个实数集合，集合中所有实数的 D 进制表示都以 $0.y_1y_2\cdots y_{l_i}$ 开始。这个集合是单位区间 $[0, 1]$ 的子区间。同时由前缀条件可知，所有这些区间均不相交。因而，它们的区间长度总和小于或等于 1。至此证明了

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

按照长度排序

正如有限情形，只需沿着上述证明的相反思路进行，即可构造出码长为 $l_1, l_2, \cdots$ 且满足 Kraft 不等式的编码。首先将长度下标重新排列，使得 $l_1 \leqslant l_2 \leqslant \cdots$。然后从单位区间的低端开始，依次将单位区间进行分配，即可获得满足条件的码字集。例如，如果想构造一个二元编码使其具有 $l_1=1, l_2=2, \cdots$，那么，将区间 $\left[0, \frac{1}{2}\right)$，$\left[\frac{1}{2}, \frac{1}{4}\right)$，$\cdots$ 分配给字符，使其对应码字 0，10，$\cdots$ □

举例

最有编码的发现规则

- 挑战：

➢ 在唯一可译性基础上，如何找到平均长度最短的编码方式？

➢ 具体的编码规则是什么？

构造一个标准的最优化问题：在所有整数 $l_1, l_2, \cdots, l_m$ 上，最小化

$$L = \sum p_i l_i$$

其约束条件为

$$\sum D^{-l_i} \leqslant 1$$

利用拉格朗日(Lagrange)乘子法

$$J = \sum p_i l_i + \lambda(\sum D^{-l_i})$$

关于 l_i 求微分，可得

$$\frac{\partial J}{\partial l_i} = p_i - \lambda D^{-l_i}\log_e D$$

令偏导数为 0，

$$D^{-l_i} = \frac{p_i}{\lambda \log_e D}$$

将此代入约束条件中以求得合适的 λ，可得 $\lambda = 1/\log_e D$，

$$p_i = D^{-l_i}$$

最优码长为

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

期望码字长度为

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

这就是信息熵的最优性表示

最短唯一可译码的长度下界就是信息熵

定理： 随机变量 X 的任一 D 元即时码的期望长度必定大于或等于熵 $H_D(X)$，

$$L \geqslant H_D(X)$$

当且仅当 $D^{-l_i}=p_i$，等号成立

另一种数学证明方法：

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

$$=-\sum p_i \log_D D^{-l_i}+\sum p_i \log_D p_i$$

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

$$L-H=\sum p_i \log_D \frac{p_i}{r_i}-\log_D c$$

$$=D(\mathbf{p}\,\|\,\mathbf{r})+\log_D \frac{1}{c}$$

$$\geqslant 0$$

由相对熵的非负性以及 $c\leqslant 1$(利用 Kraft 不等式)

因此，$L\geqslant H$，当且仅当 $p_i=D^{-l_i}$

（即对所有的 i，$-\log_D p_i$ 为整数）

等号成立

定义 对于某个 n，如果概率分布的每一个概率值均等于 D^{-n}，则称这个概率分布是 D 进制的(D-adic)

因为 $-\log_D p_i$ 可能不是整数，因此在实际编码过程中，需要取整，从而导致出现冗余部分，对冗余部分进行估计

最优码长的界

现在证明期望描述长度 L 的取值范围在其下界与下界加 1 比特之间，即

$$H(X) \leqslant L < H(X) + 1$$

理论分析

最小化 $L=\sum p_i l_i$，其约束条件为 $l_1, l_2, \cdots, l_m$ 为整数且 $\sum D^{-l_i} \leqslant 1$

$$L - H_D = D(\mathbf{p} \| \mathbf{r}) - \log(\sum D^{-l_i}) \geqslant 0$$

其中 $\mathbf{r}(r_i = D^{-l_i}/\sum D^{-l_i})$

$\boxed{\text{若码长选取 } l_i = \log_D \frac{1}{p_i}\text{，有 } L = H}$

由于 $\log_D \frac{1}{p_i}$ 未必为整数，则通过取整运算

$$l_i = \left\lceil \log_D \frac{1}{p_i} \right\rceil$$

其中 $\lceil x \rceil$ 表示 $\geqslant x$ 的最小整数

这组整数满足 Kraft 不等式

$$\sum D^{-\lceil \log \frac{1}{p_i} \rceil} \leqslant \sum D^{-\log \frac{1}{p_i}} = \sum p_i = 1$$

$$\log_D \frac{1}{p_i} \leqslant l_i < \log_D \frac{1}{p_i} + 1$$

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

定理 设 $l_1^*, l_2^*, \cdots, l_m^*$ 是关于信源分布 p 和一个 D 元字母表的一组最优码长

最优码的相应期望长度($L^* = \sum p_i l_i^*$) 则

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

说明，实际最优码的期望长度比熵大，但不会超出 1 比特的附加位

原因 $\boxed{\log_D \frac{1}{p_i} \text{并非总是整数造成的}}$

对于扩展码， 系统的结果可以改进

定义 L_n 为每输入字符期望码字长度

$$L_n = \frac{1}{n}\sum p(x_1,x_2,\cdots,x_n)l(x_1,x_2,\cdots,x_n) = \frac{1}{n}El(X_1,X_2,\cdots,X_n)$$

将上面推导的界应用于此时的编码

$$H(X_1,X_2,\cdots,X_n)\leqslant El(X_1,X_2,\cdots,X_n)<H(X_1,X_2,\cdots,X_n)+1$$

由于 $X_1,X_2,\cdots,X_n$ 是 i.i.d.

因此 $H(X_1,X_2,\cdots,X_n)=\sum H(X_i)=nH(X)$

$$H(X)\leqslant L_n<H(X)+\frac{1}{n}$$

表明：只要信息编码长度大任何编码模式都是近似最优的

通过使用足够大的分组长度　可以使其每字符期望码长任意地接近 $H(X)$

定理： 每字符最小期望码字长满足

$$\frac{H(X_1,X_2,\cdots,X_n)}{n}\leqslant L_n^*<\frac{H(X_1,X_2,\cdots,X_n)}{n}+\frac{1}{n}$$

进一步，若 $X_1,X_2,\cdots,X_n$ 是平稳随机过程，则

$$L_n^*\rightarrow H(\mathcal{X})$$

其中 $H(\mathcal{X})$ 为随机过程的熵率。

编码方法---

对于一个实际的数据集，人们只能通过经验分布估计真实的分布，此时采用的编码如下：

假定真实分布的概率密度函数是 $p(x)$

估计的概率密度函数 $q(x)$

码长为 $l(x)=\left\lceil \log\frac{1}{q(x)} \right\rceil$

则系统平均编码长度满足

$$H(p)+D(p\|q)\leqslant E_p l(X)<H(p)+D(p\|q)+1$$

对应的偏差量为 相对熵 $D(p\|q)$

证明：

$$\begin{aligned}
El(X) &= \sum_x p(x)\left\lceil \log\frac{1}{q(x)} \right\rceil \\
&< \sum_x p(x)\left(\log\frac{1}{q(x)}+1\right) \\
&= \sum_x p(x)\log\frac{p(x)}{q(x)}\frac{1}{p(x)}+1 \\
&= \sum_x p(x)\log\frac{p(x)}{q(x)}+\sum_x p(x)\log\frac{1}{p(x)}+1 \\
&= D(p\|q)+H(p)+1
\end{aligned}$$

唯一可译码的约束条件-- Kraft不等式

问题： 前面讨论的都是即时码的处理策略，对于更一般的唯一可译码，是否可以改进Kraft 不等式？

（McMillan） 任意唯一可译的 D 元码的码字长度必然满足 Kraft 不等式

$$\sum D^{-l_i} \leqslant 1$$

定理： 反之，若给定满足上述不等式的一组码字长度，则可以构造出具有同样码字长度的唯一可译码

注意：结论没有变化，唯一变化的就是数学证明上的处理技巧

证明：考虑编码 C 的 k 次扩展 C^k（即原先惟一可译码 C 的 k 次串联所形成的码）。由惟一可译性的定义，该码的 k 次扩展是非奇异的。由于所有长度为 n 的不同 D 元串的数目仅为 D^n，故由惟一可译性可知，在码的 k 次扩展中，长度为 n 的码序列数目必定不超过 D^n。由此讨论来证明 Kraft 不等式。

设字符 $x\in\mathcal{X}$ 所对应的码字长度记为 $l(x)$。对于扩展码，码序列的长度为

$$l(x_1,x_2,\cdots,x_k)=\sum_{i=1}^{k}l(x_i)$$

要证明的不等式为

$$\sum_{x\in\mathcal{X}}D^{-l(x)}\leqslant 1$$

证明的技巧就是考虑上式左边量的 k 次幂

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

$$= \sum_{x_1,x_2,\cdots,x_k\in\mathcal{X}^k} D^{-l(x_1)}D^{-l(x_2)}\cdots D^{-l(x_k)}$$

$$= \sum_{x^k\in\mathcal{X}^k} D^{-l(x^k)}$$

现将上式中的各项按码字长度合并同类项，可得

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

其中 $l_{\max}$表示码字长度的最大值

$a(m)$表示所有 m 长码字对应的信源序列x^k 的数目

原编码是惟一可译的

从而对于每个 m 长码字序列

至多存在一个信源序列与其对应

$\boxed{\text{至多存在 } D^m \text{ 个 } m \text{ 长的序列}}$

因此 $a(m)\leqslant D^m$

$$\left(\sum_{x\in\mathcal{X}} D^{-l(x)}\right)^k = \sum_{m=1}^{kl_{\max}} a(m)D^{-m}$$

$$\leqslant \sum_{m=1}^{kl_{\max}} D^m D^{-m}$$

$$= kl_{\max}$$

$$\sum_j D^{-l_j} \leqslant (kl_{\max})^{1/k}$$

唯一可译码不能改进Kraft不等式

$$\sum_j D^{-l_j} \leqslant (kl_{max})^{1/k}$$

由于上述不等式对任意的 k 均成立

当 $k\to\infty$时，不等式仍然成立

$$\sum_j D^{-l_j} \leqslant 1$$

用到 $\frac{\log k}{k}$ ---> 0　$(kl_{max})^{1/k}\to 1$

反之，若给定满足 Kraft 不等式的一组 $l_1, l_2, \cdots, l_m$，利用好即时码的构造树 可以构造出相应的即时码。由于任何即时码都是惟一可译的因而也构造出了惟一可译码。

推论　无限信源字母表$\mathcal{X}$的惟一可译码亦满足 Kraft 不等式。

直接利用

$$\sum_{i=1}^{\infty} D^{-l_i} = \lim_{N\to\infty}\sum_{i=1}^{N} D^{-l_i} \leqslant 1$$

小结

给出了压缩编码的两个基本需求：唯一可译性原则---Kraft 不等式

平均编码长度的下界 H（X）；

- 说明了最优编码的偏差量最多为1比特；
- 随着码字的扩张，偏差量可以近似为0，但译码复杂度增加；

- 系统采用直方图得到的密度函数，进行编码时，偏差量等于相对熵；

- 从即时码退化到唯一可译码，但Kraft不等式约束不能放松；
- **Kraft不等式成为验证唯一可译性的标准规范**