2022-2023学年秋季学期

信息论与编码
第六讲

复习

$$D \to P_D \to R(D) = \min_{p(y|x)\in P_D} I(X;Y)$$

定义：在许可试验信道集合中，总可以找到某一试验信道P(Y|X)，使信道信息传输率 I(X;Y)达到最小值，记作 R(D)，即，

$$R(D) = \min_{p(y|x)\in P_D} I(X;Y)。$$

- D 失真许可的试验信道：满足保真度准则的信道。

令 $\mathcal{P}_D$ 表示所有 D 失真许可的试验信道的集合 $\mathcal{P}_D = \{p(y|x) : \bar{D} \le D\}$。

称 R(D)为信息率失真函数，简称为率失真函数。

单位：比特/信源符号；D 为允许平均失真度。

信道容量与信息率失真函数的对偶性：

二者都是求平均互信息极值问题：

$$C = \max_{p(x_i)} I(X;Y),\ R(D) = \min_{p(y_j|x_i)\in P_D} I(X;Y)。$$

1) 二者都是求平均互信息量，单位：比特/符号。

- 信道容量：信道固定，计算 I(X;Y)的极大值；其理论依据为 I(X;Y)是输入信源概率分布$\{p(x_i)|i\}$的上凸函数。
- 信息率失真函数：在试验信道（满足保真准则的信道）中求 I(X;Y)的极小值。其理论依据：I(X;Y)为信道转移概率分布$p(y|x)$的下凸函数。

2) 信道容量一旦求出后，就仅与 p(y|x)有关了，与信源无关，反映信道特性。信息率失真函数一旦求出后，只与输入信源的概率分布$\{p(x_i)|i\}$有关了，与信道无关，反映信源特性。

3) 信道容量，解决通信的可靠性问题，是信息传输的理路基础，通过信道编码增加信息冗余度来实现。信息率失真函数是为解决通信的有效性问题，是压缩理论的基础，通过信源编码，减少信息的冗余度来实现。

复习

信源给定后，$p(x_i)$是固定的，故

$$D_{min} = \sum_{i=1}^{n} p(x_i) min\{\sum_{j=1}^{m} p(y_j|x_i)d(x_i,y_j)\}。$$

由于$d(x_i,y_j)$是已知的，于是只需选择试验信道的转移概率$p(y_j|x_i)$，

对于每个x_i，使得$\sum_{j=1}^{m} p(y_j|x_i)d(x_i,y_j)$最小。

由$0 \le p(y_j|x_i) \le 1$，且$\sum_{j=1}^{m} p(y_j|x_i) = 1$，知，

适当选择每个x_i使得

$$\begin{cases} \sum_j p(y_j|x_i) = 1, & \text{对所有}d(x_i,y_j)\text{为最小值的}y_j \\ \sum_j p(y_j|x_i) = 0, & \text{对所有}d(x_i,y_j)\text{不为最小值}y_j \end{cases}$$

$$D_{min} = \sum_{i=1}^{n} p(x_i) \min_j d(x_i,y_j).$$

复习

当 I(X;Y)= 0 时，X 与 Y 互相独立；这等价于通信中断，即，p(y|x)=p(y) 。这样，不同的$p(y_j)$均可使 R(D)=0，但是所造成的$\bar{D}$不同。

选取其中最小者定义为D_{max}，即，

$$D_{max}=\min_{p(y_j)}\bar{D}=\min_{p(y)}\sum_{y}p(y)\left(\sum_{x}p(x)d(x,y)\right)。$$

$$\begin{cases}\text{选取使}\sum_{i}p(x_i)d(x_i,y_j)\text{最小的}p(y)=1.\\ \text{令其它的}\sum_{i}p(x_i)d(x_i,y_j)\text{对应的}p(y)=0.\end{cases}$$

于是有

$$D_{max}=\min_{p(y)}\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)p(y_j)d(x_i,y_j)=\min_{p(y)}\sum_{i=1}^{n}p(x_i)d(x_i,y_j)$$

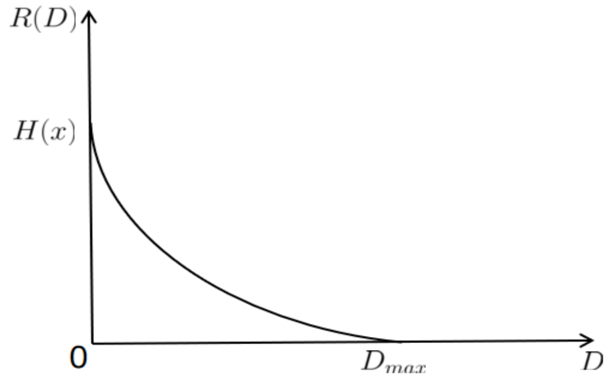
3) R(D)的单调性和连续性

- 连续性：由凸性可得。
- 单调性：严格递减，即，在$D_{min} \leq D \leq D_{max}$范围内，R(D)不可能为常数。

说明：直观地，R(D)的非递增性，表明：允许失真越大，所要求的信息率越小。

$$\begin{cases} R(0) = H(x)\text{：无失真} \\ R(D_{max}) = 0\text{：（完全失真）}X\text{与}Y\text{无关} \end{cases}$$

于是，R(D)的函数图像如图：



4.3 离散信源的信息率失真函数

对于离散信源，计算

$$\left\{\begin{matrix}\text{信息率失真函数}\\ \text{信道容量}\end{matrix}\right\}\text{在约束条件下求平均互信息的权值问题。}$$

取极大值后仅与信道有关。

差别： 约束条件不同

1) 信道容量 C：信源（信道固定），求平均互信息的条件极大值。

2) 信息率失真函数 R(D)：信道（信源固定），已知信源的概率分布和失真度函数条件下，求平均互信息的条件极小值。

取极小值后仅与信源有关。

用 Langrange 乘子法：

- ➢ 原则上可求。
- ➢ 求显式表达式困难。
- ➢ 仅能求出信息率失真函数的参量表达式。

4.3.1信息率失真函数的计算——拉格朗日乘子法

引入变量 S 作为参数，信息率失真函数为 R(S)，失真函数为 D(S)。

设给定信源 X，$P(X)=\{p(x_i):i=1,2,\cdots,n\}$；规定失真函数为$d(x_i,y_j)$,$i=1,2,\cdots,n;j=1,2,\cdots,m$。选定允许平均失真度为$\bar{D}$，则信源 X 的信息率失真函数 R(D)的约束条件：

$$\begin{cases} p(y_j|x_i)\geq 0, \quad i=1,2,\cdots,n;j=1,2,\cdots,m. \\ \sum_{j=1}^{m} p(y_j|x_i)=1, \quad i=1,2,\cdots,n. \\ \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i)p(y_j|x_i)d(x_i,y_j)=\bar{D} \\ p(x)\text{固定} \end{cases}$$

$R(D)=min\,I(X;Y)$，其中

$I(X;Y)=\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i)p(y_j|x_i)\log\frac{p(y_j|x_i)}{\sum_{k=1}^{n} p(x_k)p(y_j|x_k)}$为 p(y|x)的下凸函数。

取拉格朗日常数$\mu_i, i=1,2,\cdots,n$，对应于

$\sum_{j=1}^{m} p(y_j|x_i)=1$，$i=1,2,\cdots,n$。

构成辅助函数

$$F=I(X;Y)-\sum_i \mu_i\left(\sum_{j=1}^{m} p(y_j|x_i)-1\right)-S\left(\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i)p(y_j|x_i)d(x_i,y_j)-\bar{D}\right)$$

计算$\frac{\partial F}{\partial p(y|x)}=0$

{eq1} $$p(x_i)\ln\frac{p(y_j|x_i)}{p(y_j)}-\mu_i-S\cdot p(x_i)d(x_i,y_j)=0$$

{eq*} $$p(y_j|x_i)=p(y_j)\lambda_i\cdot e^{S\cdot d(x_i,y_j)}$$

{eq2} $$\lambda_i=\frac{1}{\sum_{j=1}^{m} p(y_j)e^{S\cdot d(x_i,y_j)}}$$

{eq3} $$\sum_{i=1}^{n}\lambda_i p(x_i)e^{S\cdot d(x_i,y_j)}=1$$

{eq4} $$p(y_j|x_i)=p(y_j)\lambda_i e^{S\cdot d(x_i,y_j)}=\frac{p(y_j)\cdot e^{S\cdot d(x_i,y_j)}}{\sum_{j=1}^{m}p(y_j)e^{S\cdot d(x_i,y_j)}}$$

{eq5} $$D(S)=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)p(y_j)d(x_i,y_j)\cdot\lambda_i\cdot e^{S\cdot d(x_i,y_j)}$$

{eq6} $$R(D)=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)p(y_j)d(x_i,y_j)\cdot\lambda_i\cdot e^{S\cdot d(x_i,y_j)}\ln(\lambda_i\cdot e^{S\cdot d(x_i,y_j)})$$
$$=S\cdot D(S)+\sum_{i=1}^{n}p(x_i)\ln\lambda_i.$$

一般情况下，S无法消失，因此得不到R(D)的显式解。但对某些特定的简单问题，可以消去S。

利用{eq2}-{eq6})式求R(D)。

由{eq3}，解得λ_i;

由{eq2}，解得$p(y_i)$;

再由{eq4},解得$p(y_j|x_i)$;

将上述代入{eq5}和{eq6}，得$D(S)$，由此知，S，$R(D)$;

将 S 代入各式得$\lambda_i, p(y_i), p(y_j|x_i)$。

S 性质：

- 将 S 视为 D 的函数，对 R(D)求导$\frac{dR(D)}{dD}$，整理可得，$\frac{dR(D)}{dD}=S$。由此可知，S 为 R(D)的斜率。

$$D\in(0,D_{max}),\begin{cases}D\to 0\text{时},S\to-\infty\\ D>D_{max},R(D)=0,S=0.\end{cases}\text{，即，}S\in(-\infty,0).\text{。}$$

4.3.2二元离散信源信息率失真函数的计算

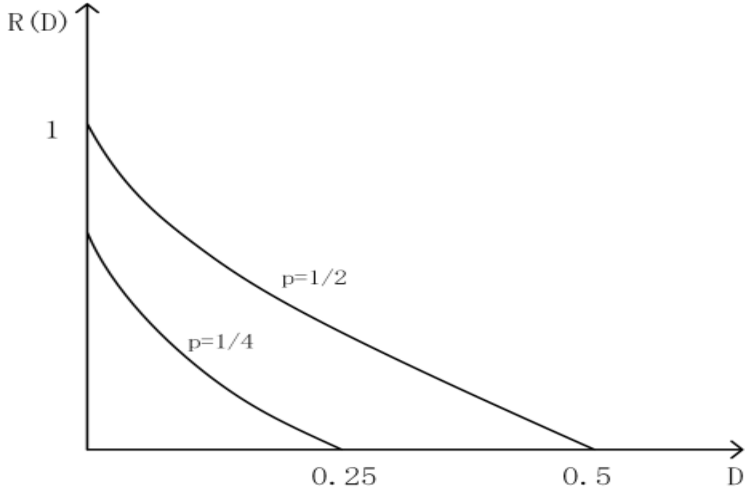
考虑二元离散无记忆信源

$$\begin{pmatrix}X\\P(X)\end{pmatrix}=\begin{pmatrix}0 & 1\\ p & 1-p\end{pmatrix},0\le p\le\tfrac{1}{2}\text{，}\ Y=\{0,1\}\text{。}$$

失真矩阵$D=\begin{pmatrix}0&\alpha\\ \alpha&0\end{pmatrix}$。求 R(D)。

$$R(D)=\begin{cases}H(p)-H(D/\alpha), & 0\le D\le D_{max}=\alpha p\\ 0, & D>\alpha p\end{cases}$$

如图，当$p=\frac{1}{2}$和$p=\frac{1}{4}$时 R(D)的曲线。



- 对于同一平均失真度 D,信源分布越均匀，R(D)越大，信源压缩的可能性越小。
- 若信源分布越不均匀，即，信源的冗余度越大，R(D)越小，信源压缩的可能性越大。

说明：

1) 信源的 R(D)计算困难。
2) 信源的概率空间仅是一种模型，很难准确计算概率分布。
3) 精确计算 R(D)函数是没有必要的。一般，从失真要求出发，估算几个特殊点，如$D=0, D_{max}$以及几个 D 值对应的 R(D)，可以画出 R(D)函数曲线。
4) 信息率失真理论给出，给定失真度 D 条件下，信源输出的信息率所能压缩的极限 R(D)，但没有给出压缩方法，但可作为一尺度，衡量压缩编码方法的压缩效果。

4.4 保真度准则下的信源编码定理

Shannon 第三定理，即为保真准则下的信源偏码定理：

设一离散平稳无记忆信源，输出随机变量序列为$\vec{X}=(X_1X_2\cdots X_L)$。若

该信源的信息率失真函数是 R(D)，并选定有限的失真函数，对于任意允许平均失真度$D\geq 0$和$\forall\epsilon>0$，当信息率 R>R(D)时，只要信源序列长度 L 足够大，一定存在一种编码方式 C,使得译码后的平均失真度$\bar{D}(C)\leq D+\epsilon$。反之，若 R<R(D),则对$\forall C$有$\bar{D}(C)>D$,即，译码平均失真度必大于允许平均失真度。

说明：

1) 这是有失真信源压缩的理论基础：允许失真度 D 确定后，总存在一种编码方法，使编码的信息率 R 可任意接近于 R(D)函数，而且平均失真度$\bar{D}(C) \le D+\epsilon$。反之，若 R<R(D),则编码的平均失真度将大于 D。总之，信息率失真函数是一个界限。

$$\begin{cases} R(D) \le R, & \text{则}\bar{D}(C) \le D+\epsilon\text{，失真可控，即，即使有失真，仍满足要求。} \\ R(D) > R, & \text{不可控，不能满足要求} \end{cases}$$

信源编码

编码就是将信源要发送的消息转变为信道传输的符号的变换，即，

变换 $f: X \times C \longrightarrow W$，其中

- $X = \{x_1, x_2, \ldots, x_n\}$为要发送的消息符号；
- $W = \{w_1, w_2, \ldots, w_m\}$为码字集合；
- $C = \{c_1, c_2, \ldots, c_r\}$为码符集合，即，C 为适合信道传输的符号。

编码分两部分：

1)信源编码：用尽量少的码符表示消息,保证有效性。

2)信道编码：对信源的编码再进行一次编码，提高抗干扰能力。

注：存在一种联合信源信道编码，即，将上述两部分一次完成。

$$\text{信源编码}\begin{cases} \text{离散信源编码} \Rightarrow \text{可做到无失真编码} \Leftarrow \text{(理论基础:)无失真信源编码定理} \\ \text{连续信源编码} \Rightarrow \text{只能做到限失真编码} \Leftarrow \text{(理论基础:)限失真信源编码定理} \\ \text{相关信源编码} \end{cases}$$

- 离散信源可以做到无失真编码。
- 本章主要研究无失真信源编码技术与方法。

- **Shannon第一定理**：信源的信息熵是信源进行无失真编码的理论极限值。即，总能找到一种合适的编码方法使编码后信源的信息传输率R任意地逼近信源的信息熵而不存在任何失真。

因此，在数据压缩技术中无失真的信源编码又称为熵编码。

几种码的定义：

1).二元码：码符集$C=\{0,1\}$，所得码字一定是一些二元序列

2).等长码：所有码字的长度相等(亦称为定长码)

3).变长码：码字长度不同

4).非奇异码：一组码字中所有码字都不相同，即所有信源符号映射到不同的码符号序列

5).奇异码：一组码中有相同的码字

6).同价码：每个码符号所占用的传输时间都相同，如：二元码，等长的同价码的每个码字所占用的传输时间相同

7).唯一可译码: 码的任意一串有限长的码符号序列只能被唯一的译成所对应的信源符号.分为即时码、非即时码

(1)信息率(编码速率)

$$R=(K/N)\log_2 r \quad \text{bit/信源符号}$$

$\log_2 r$ —— 每个码符号的最大熵(bit/码符号)

$K\log_2 r$ —— 每个码符号序列最大熵(bit/码序列)

$(K/N)\log_2 r$— 编码后，平均每个信源符号所能载荷的最大信息量

(2)编码效率

$$\eta=\frac{H(X)}{R}$$

$H(X)$—编码前，平均每个信源符号包含的信息量

R——编码后，平均每个信源符号所能传送的最大信息量

(3) 平均码长为 $\overline{L}_N=\sum_{i=1}^{n^N} p(\vec{\alpha}_i)l_i$

信源编码的要求：

1).适合信道中传输，即用信道传输的码符号编码。

2).码字必须与信源消息符号一一对应。

3).由码字可及时译出相应的消息。

4).高效性。

问题：

1) 实际中，往往连续发送许多消息符号连成一串，对应码字也必然连成一串。如何保证消息串集合与码字串集合一一对应？

2) 即使单个消息与码字一一对应，N 次扩展后是否还能一一对应？

例：

信源编码	P(x)	码1	码2	码3
x_1	$p(x_1)$	00	0	0
x_2	$p(x_2)$	01	01	10
x_3	$p(x_3)$	10	001	00
x_4	$p(x_4)$	11	111	01

码2的二次扩展？

$\alpha_1 = x_1x_1 \quad \alpha_2 = x_1x_2 \quad \alpha_3 = x_1x_3 \quad \alpha_4 = x_1x_4 \quad \alpha_5 = x_2x_1 \quad \cdots \quad \alpha_{16} = x_4x_4$

$w_1w_1 = 00 \quad w_1w_2 = 001 \quad w_1w_3 = 0001 \quad w_1w_4 = 0111 \quad w_2w_1 = 010 \quad \cdots \quad w_4w_4 = 111111$

但实际表明，即使单个消息与码字一一对应，但是 N 次扩展后不一定。如，信源编码 3，$\{x_1x_3, x_3x_1\}$都可译为0000, $\{x_1x_2, x_4x_1\}$都可译为010，导致无法正确译码。

要求：

1）不同的单个消息符号变成不同的单个码字。

2）任意有限长德消息序列所对应的码符序列也各不相同，即，N 次扩展后仍是非奇异码。

给出一组码，如何判断它是不是唯一可译码？给出一组消息符号和码符号集，在何条件下才能编出唯一可译码？

唯一可译码定理

设原始信源符号集合为 $X=\{x_1,x_2,\ldots,x_n\}$，码符号集合 $C=\{c_1,c_2,\ldots,c_r\}$，用 C 对其进行编码，码字集合为 $W=\{w_1,w_2,\ldots,w_m\}$，$|w_i|=n_i, 1\le i\le m$，则存在唯一可译码的充要条件为码长的组合满足 $\sum_{i=1}^{m} r^{-n_i}\le 1$。

(称此不等式为 Kraft 不等式)

注：Kraft 不等式涉及消息符号个数，码符号个数以及码长等三种参数，证明略.

下面分别讨论几类常用唯一可译码

即时码：

一个码中任意码字都不是另一个码字的前缀，或者，任意码字都不是另一个码字的延长，当收到一个码字后可以立即译出其对应的消息符号。

对于唯一可译码，能否即时地译出其对应的消息符号，这对其应用很有影响。

例： 信源$(X,P(x)), X=\{x_1,x_2,x_3,x_4\}, C=\{0,1\}$有如下编码：

码 W1：

$$x_1 \to w_1 = 1,\ x_2 \to w_2 = 10,$$

$$x_3 \to w_3 = 100,\ x_4 \to w_4 = 1000$$

码 W2：

$$x_1 \to w_1 = 1,\ x_2 \to w_2 = 01,$$

$$x_3 \to w_3 = 001,\ x_4 \to w_4 = 0001$$

对于码 W1，收到 100 后仍然不能马上译为x_3,因为有可能是x_4，因此需要收到下一个符号后才能断定；对于码 W2，每收到一个 1，即可马上译码.

如何构造即时码

根据即时码的定义，构造即时码时只要在所有可能的码字符号列中挑选出一些相互不构成前缀或后延关系的码字即可。故要列出所有可能的码字序列，这通过树图来实现。

❍树图构造：

原则：任意码字不是另一个码字的前缀
在所有可能的码符号序列中，挑选出一些相互不构成前缀或后延关系的字作为码字。利用树图，列出所有可能的码符号序列，挑选码字。

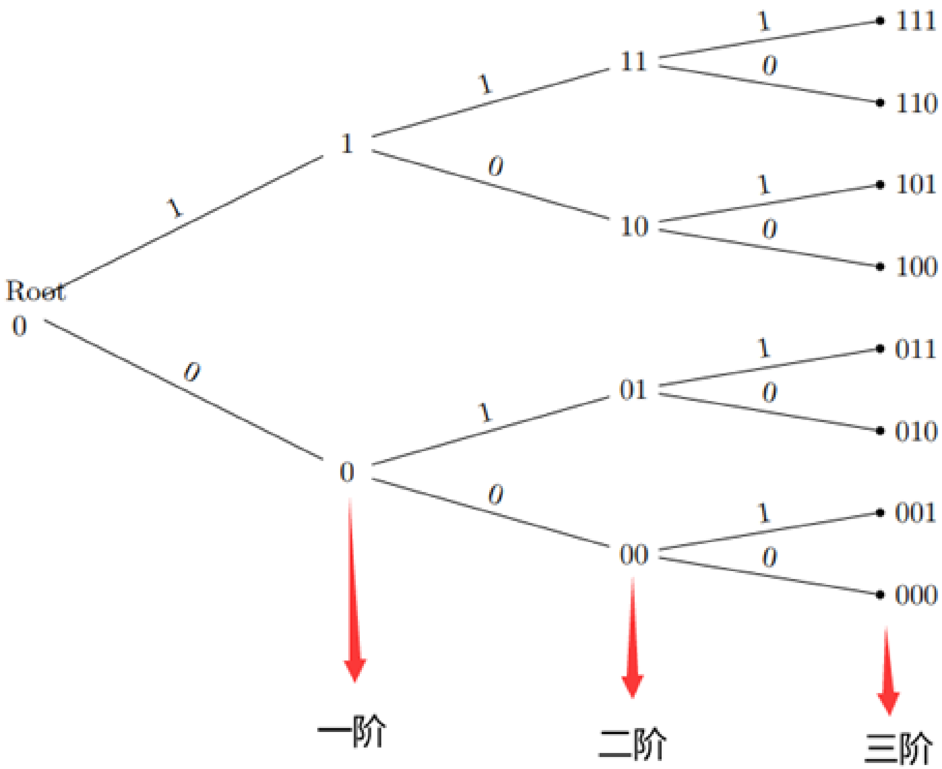
设信源$|X|=n$，码符集为$C=\{c_1,c_2,\cdots,c_r\}$，$|C|=r$。根为起点(未编码)。

step1:由根向下分出树枝：码符集合 C 里有几个码符号就分出几条树枝(r 条)。

step2：枝旁的 0,1 表示对码第一位所赋予的码符号，枝头下的圆圈是一阶节点，其旁边标上只有一位的码符号序列，这时从 Root 到各一阶节点连线旁边的码符组成，一阶节点有$|C|=r$个。

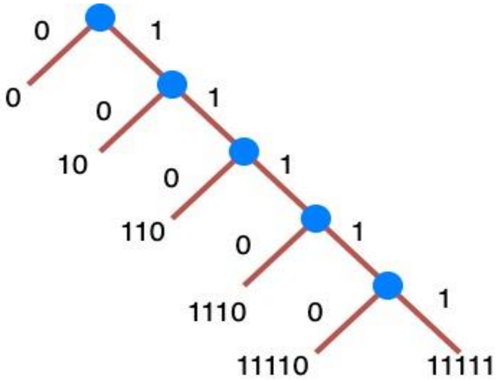
step3：从一阶节点开始的各节点同样分出 r 个树枝(如根一样)，树枝表示码符号序列各步演变的路径。树枝旁的数码表示在上一阶节点上再添加的一个码符号。枝头下的圆圈表示下一阶节点，旁边也标有其所代表的码符号序列，这个序列是由根到本节点沿途枝上的码符号排列而成。

step4：将与被选用的码字对应的节点和路径保留下来（非前缀关系的点），其余节点去掉，得到码树.

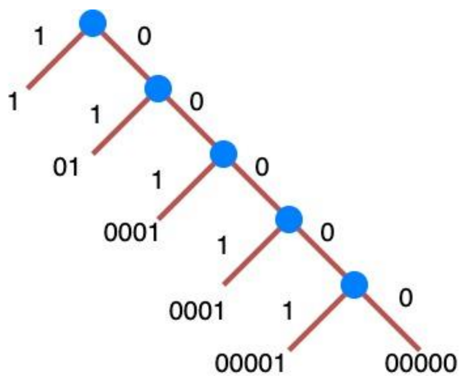


注：即时码不唯一。

例：$X = \{x_1, x_2, x_3, x_4, x_5\}$，$C = \{0, 1\}$。



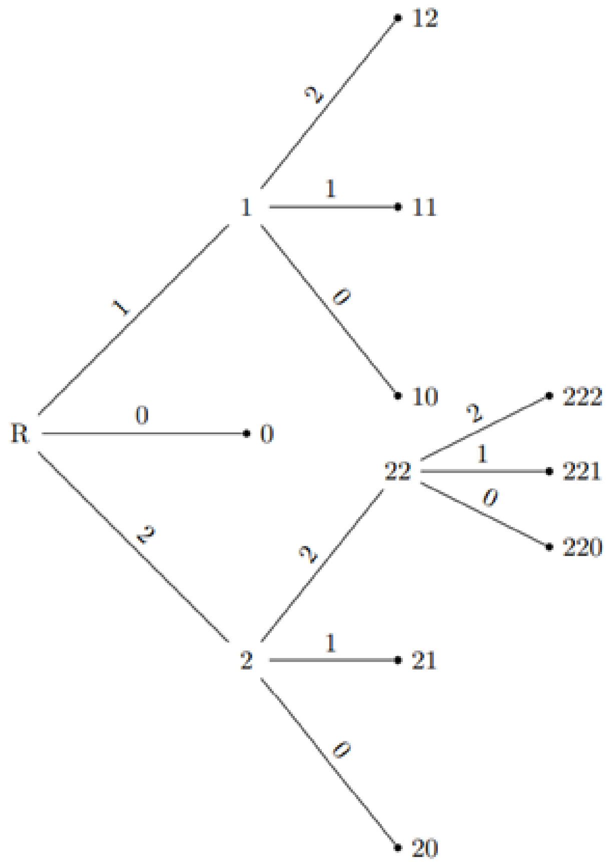
$$W_1 = \{0, 10, 110, 1110, 11110/11111\}$$



$$W_2 = \{1, 01, 001, 0001, 00001/00000\}$$

这说明即时码不唯一。

例：$X=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\}$，$C=\{0,1,2\}$



w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8
0	10	11	12	20	21	220	221

Shannon编码 对于离散无记忆信源

$$\begin{pmatrix} X \\ P(x) \end{pmatrix} = \begin{pmatrix} x_1 & \dots & x_i & \dots & x_n \\ p(x_1) & \dots & p(x_i) & \dots & p(x_n) \end{pmatrix} \quad \sum_{i=1}^{n} p(x_i) = 1$$

二进制 Shannon 码编码步骤如下：

1).将信源符号按概率从大到小排列，不放设，

$$p(x_1) \geq p(x_2) \geq \dots \geq p(x_n)。$$

2).令$p(x_0) = 0$，$p_\alpha(x_j) = \sum_{i=0}^{j-1} p(x_i)$，$j = 1, 2, \cdots, n$，称$p_\alpha(x_j)$为累加概率；

3).确定满足下列不等式的整数l_i,并令l_i为第 i 个码字的长度：

$$\log \frac{1}{p(x_i)} \leq l_i < \log \frac{1}{p(x_i)} + 1，\text{或者} l_i = \lceil \log \frac{1}{p(x_i)} \rceil,$$

其中$\lceil u \rceil$是大于等于 u 的最小整数，即$u \leq \lceil u \rceil < u + 1$。

4).将$p_a(x_j)$用二进制表示，并取小数点后l_i位作为符号x_i的编码。

例　对单符号信源

$$\begin{pmatrix} X \\ p(X) \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & \\ 0.25 & 0.25 & 0.2 & 0.15 & 0.1 & 0.05 \end{pmatrix},$$

二元编码符集$C = \{0, 1\}$。

求：其 Shannon 编码、平均码长、编码信息率和编码效率。

费诺编码：

这是一种基于统计匹配的编码。步骤如下，

1) 将信源消息(符号)按其出现的概率由大到小依次排列；
2) 将依次排列的信源符号按概率值分成两大组，使得两组的概率之和接近相同，并对各组分别赋予一个二进制码元 0 和 1；
3) 将每个大组的信源符号进一步分成两组，使划分后的两个组的概率之和接近相同，并又分别赋予一个二进制符号 0 和 1；
4) 如上重复，直到每个组只剩下一个信源符号为止；
5) 信源符号所对应的码字即为费诺码。

例. 对单符号信源

$$\begin{pmatrix} X \\ p(X) \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ 0.32 & 0.22 & 0.18 & 0.16 & 0.08 & 0.04 \end{pmatrix}$$

进行费诺编码。

例. 对单符号信源 $\begin{pmatrix} X \\ p(X) \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}$

进行费诺编码。

❍霍夫曼编码

这是一种效率比较高的变长无失真信源编码方法。它按照概率进行编码是图像压缩标准JEPG、MPEG中典型算法。

单符号离散无记忆信源空间

$$\begin{pmatrix} X \\ P(x) \end{pmatrix} = \begin{pmatrix} x_1 & \dots & x_i & \dots & x_n \\ p(x_1) & \dots & p(x_i) & \dots & p(x_n) \end{pmatrix}, \quad \sum_{i=1}^{n} p(x_i) = 1,$$

码符集$C = \{0,1\}$，$|C| = r = 2$。

二元**Huffman编码**主要步骤如下：

1) 检查$X = \{x_1, \cdots, x_n\}$中符号的概率之和是否为 1；
2) 将符号按其概率从大到小顺序排列，并从上到下填入表中，概率相同者上下位置可以随意排列。符号之间（行之间）要留些空白，以便于编码过程中填入数字。

3) i) 从下到上取 r 个概率$p_{t_1},p_{t_2},\cdots,p_{t_r}$相加，得$p^{(1)}=\sum_{j=1}^{r}p_{t_j}$;

ii）将$p^{(1)}$与上边没有用的概率(按照第 2 步的方式)再次按由大到小排列。若$p^{(1)}$与上边未用的某概率相同，则与相同概率随意排列（如，放到最后），但接下来再出现概率相同的情况，亦按此排列顺序，避免位置变化。这样得到缩减概率排序 1。

iii）连线$p_{t_1},p_{t_2},\cdots,p_{t_r}$与$p^{(1)}$，并在$p_{t_1},p_{t_2},\cdots,p_{t_r}$旁边的路径横线上标注相应的码符号，如，$c_1$在上线，$c_2$在下线，等。（反序亦可，但一旦选定，后面均要一致按此序。）

4) 在缩减序列 1 中从底向上再截取 r 个概率相加，得$p^{(2)}$。将$p^{(2)}$与上面的概率排序，得到概率缩减序列 2，并同上连线，在右边横线上标出码符c_0、c_1、c_2、…。

5) 重复上述步骤 3)，处理所见序列 2。依次类推，直到最后剩下两个概率，其和为 1。

6) 编码表中是一棵横倒的码树，根 R 在最右边，最后确定信源符号的码字：

如前即时码一样，从树根开始，从右向左沿路径一直找到各符号的概率，把沿途在路径线上碰到的码符号按先后顺序从左向右填入表中，即为该符号的码字。

最后得到的各码字及码长都在编码表中给出。

例 对信源$\begin{pmatrix} X \\ p(X) \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0.4 & 0.1 & 0.2 & 0.1 & 0.2 \end{pmatrix}$,

$C = \{0, 1\}$,

求其二元霍夫曼编码.

说明1、当使用编码表编码，在概率旁边的线上标注码符号时，若与上述步骤中相反，即，上线标1，下线标0，这时所得码字结构与“上线标0，下线标1”完全相同，只是0、1对调。

说明2、若在编码中概率之和与缩减概率序列中其它概率相同的时候，概率之和一律写在上面与写在下面的编码差别不大，但码长更集中在平均码长附近。

r元Huffman编码

$r \neq 2$。在概率求和时，从下向上每次取 r 个概率相加，其它操作与 r=2 时相同。但有时在 Root 前的最后一次概率相加时会发现不够 r 个时，可如下处理：

若最后一次缩减概率序列剩下的概率个数 t<r，可在原始信源中加上 m=r-t 个概率 0 的假信源符号，然后进行编码。这里添加 m 的大小要保证 Root 前恰有 r 个概率。

即，扩充 X，在最初的时候增加合适数量$x_{n+1},\cdots,x_{n+m}$，且对应的概率为$p(x_{n+1})=p(x_{n+2})=\cdots=p(x_{n+m})=0$，再重新编码。

按照线(路径)上标注的码符读取对应的编码.

例 对信源$\begin{pmatrix} X \\ p(X) \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ 0.24 & 0.20 & 0.18 & 0.16 & 0.14 & 0.08 \end{pmatrix}$,

$C = \{0, 1, 2\}$，求其三元霍夫曼编码.

例 对信源$\begin{pmatrix} X \\ p(X) \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0.35 & 0.15 & 0.25 & 0.05 & 0.2 \end{pmatrix}$,

$C = \{0, 1\}$,

求其霍夫曼编码.

例 对信源$\begin{pmatrix} X \\ p(X) \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \end{pmatrix}$,

$C = \{0, 1\}$,

求其霍夫曼编码.