电子科技大学

University of Electronic Science and Technology of China

信息論基礎

－无失真信源编码

鄺育軍

通信與信息工程學院

kyj@uestc.edu.cn

13648007026, 61830597

移動互聯實驗室

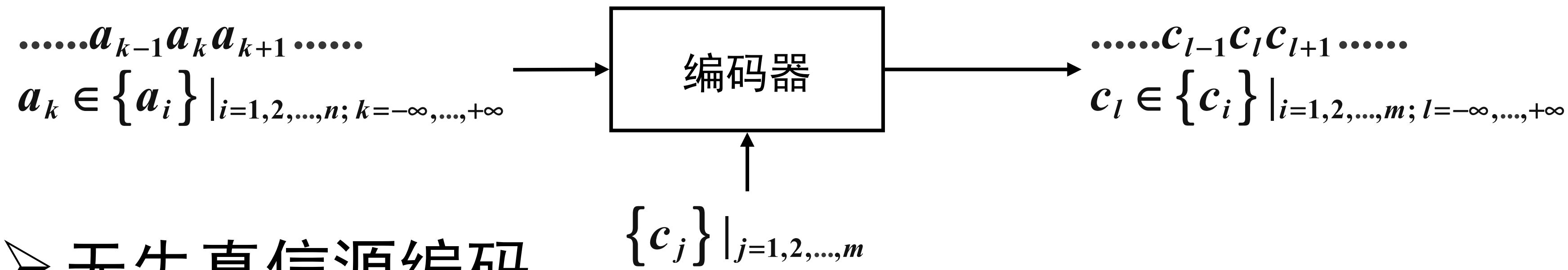
MOBILELINK

内容提要

- 引子
- 即时码与克拉夫特不等式
- 最优码与最优码长的界
- 香农码
- 霍夫曼码
- 费诺码
- 小结

引子

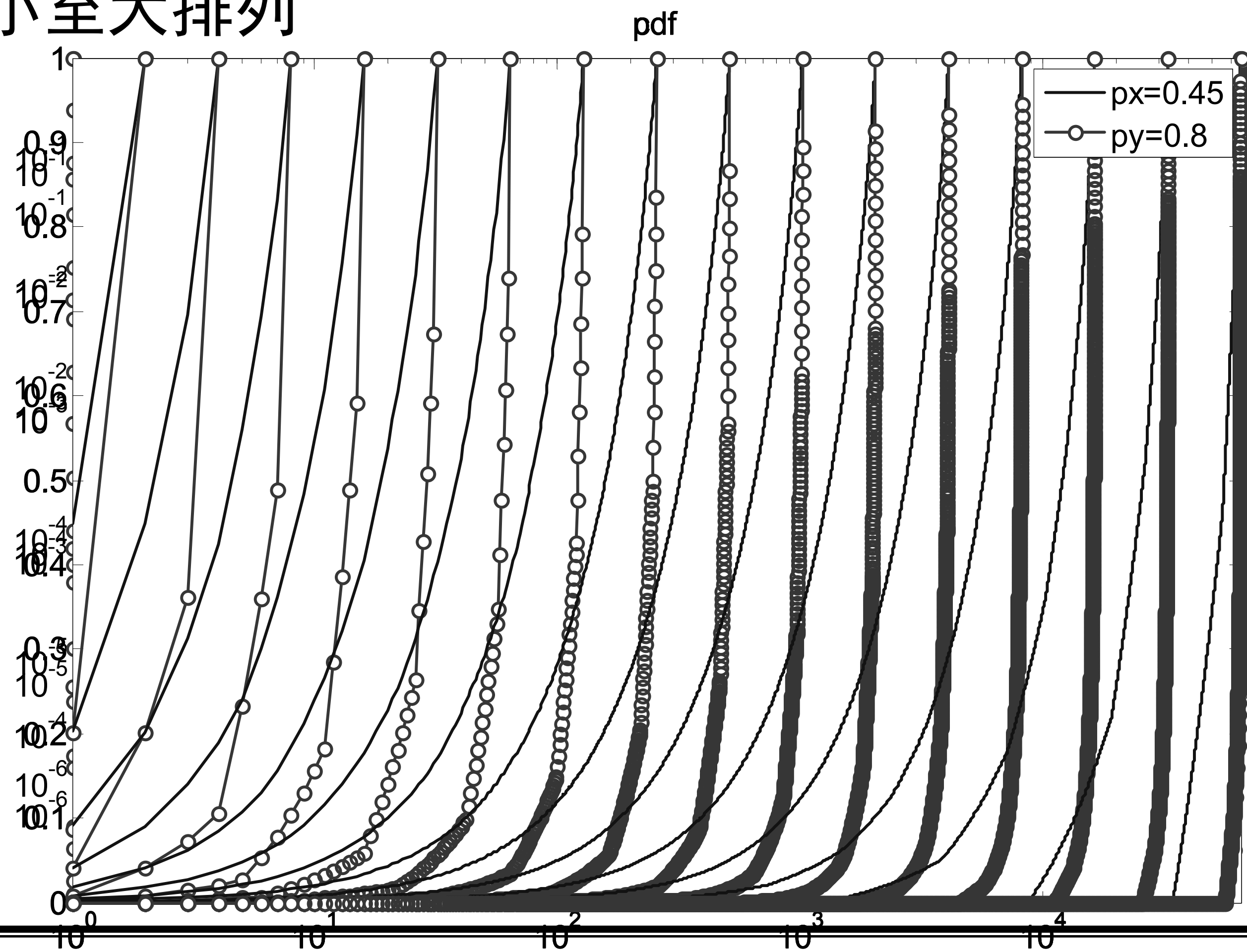
- 数学意义上，信源编码是信源符号序列a_i到码序列(码字)c_i之间的映射。
- 信源编码的模型



- 无失真信源编码
  - 选择适合信道传输的码，一般选二元码
  - 寻求一种将信源符号序列变换为码字的系统方法，这种方法至少要保证典型序列与码字之间的一一对应关系
    - *不造成信息量的损失(正规编码器)*

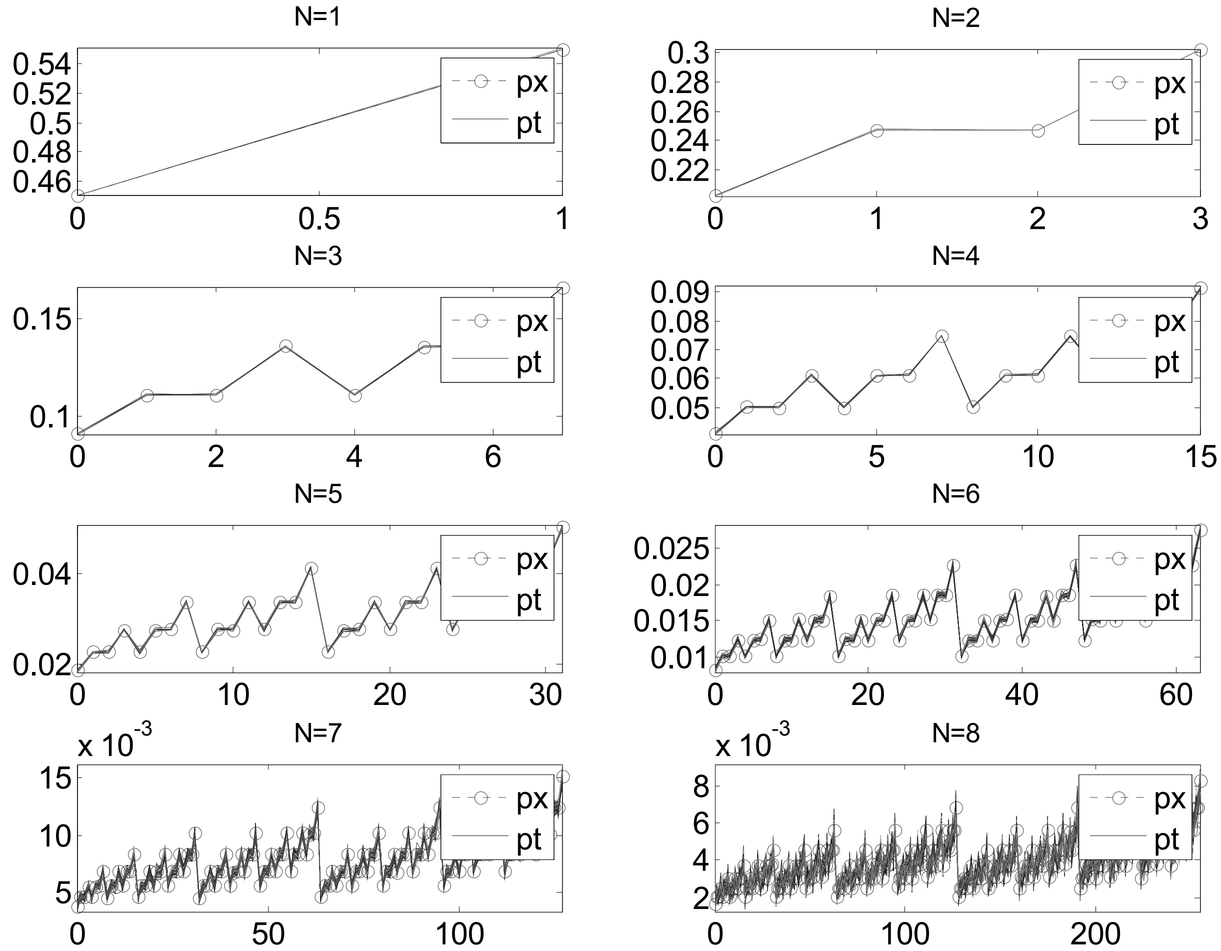
引子-续

- 非等概信源的联合分布概率渐进均分特性
  以下各图pdf均自小至大排列
  - 0.5/0.9
    .9/.1=9
    X的pdf较均匀
  - 0.4/0.6
    二者的分布一致
  - 0.45/0.8
    .55/.45=1.22
    .8/.2=4



引子-续

- 平稳信源的前后转移概率分析
  - 对X任意符号，其下一符号Y的条件概率与Y的概率一致。



例1

- 对信源

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} x_1 & x_2 & x_3 & x_4 \\ 0.5 & 0.3 & 0.15 & 0.05 \end{Bmatrix}$$

进行无失真信源编码

– 与所有符号x_i有一一对应码字c_i的编码是无失真编码

- *一种无失真等长编码是对信源符号进行4-2进制变换：*

$$x_1 \to c_1 = 00, x_2 \to c_2 = 01, x_3 \to c_3 = 10, x_4 \to c_4 = 11$$

- *编码的码长L=2(bit)，熵为*

$$H(X) = -\sum_{i=1}^{4} P(x_i)\mathrm{lb}P(x_i) = -0.5\mathrm{lb}0.5 - 0.3\mathrm{lb}0.3 - 0.15\mathrm{lb}0.15 \\ -0.05\mathrm{lb}0.05 \approx 1.648(\mathrm{bit})$$

- *编码效率*

$$\eta = H(X) / L = 1.648 / 2 = 82.4\%$$

例1-续

– 为提高编码效率，考虑将整数比特的码长压缩为小数比特的平均码长，对信源符号进行不等长编码。

- *一种无失真不等长编码*

$$x_1 \to c_1 = 0, x_2 \to c_2 = 10, x_3 \to c_3 = 110, x_4 \to c_4 = 111$$

- *平均码长*

$$L = \sum_{i=1}^{n} P(x_i) l_i = 0.5\times 1 + 0.3\times 2 + 0.15\times 3 + 0.05\times 3$$
$$= 1.7(bit)$$

- *编码效率*

$$\eta = \frac{H(X)}{L} = \frac{1.648}{1.7} \approx 96.94\%$$

例1-续

- 对信源二次扩展

$$\begin{bmatrix} X^2 \\ P(X^2) \end{bmatrix} = \left\{ \begin{matrix} x_1x_1 & x_1x_2 & x_1x_3 & x_1x_4 & x_2x_1 & x_2x_2 & x_2x_3 & x_2x_4 \\ 0.25 & 0.15 & 0.075 & 0.025 & 0.15 & 0.09 & 0.045 & 0.015 \end{matrix} \right.$$

$$\left. \begin{matrix} x_3x_1 & x_3x_2 & x_3x_3 & x_3x_4 & x_4x_1 & x_4x_2 & x_4x_3 & x_4x_4 \\ 0.075 & 0.045 & 0.0225 & 0.0075 & 0.025 & 0.015 & 0.0075 & 0.0025 \end{matrix} \right\}$$

– 一种无失真不等长编码

$$\begin{aligned}
&a_1 = x_1x_1 \to c_1 = 00, a_2 = x_1x_2 \to c_2 = 100 \\
&a_3 = x_1x_3 \to c_3 = 1100, a_4 = x_1x_4 \to c_4 = 11100 \\
&a_5 = x_2x_1 \to c_5 = 101, a_6 = x_2x_2 \to c_6 = 010 \\
&a_7 = x_2x_3 \to c_7 = 0110, a_8 = x_2x_4 \to c_8 = 111100 \\
&a_9 = x_3x_1 \to c_9 = 1101, a_{10} = x_3x_2 \to c_{10} = 0111 \\
&a_{11} = x_3x_3 \to c_{11} = 111110, a_{12} = x_3x_4 \to c_{12} = 1111110 \\
&a_{13} = x_4x_1 \to c_{13} = 11101, a_{14} = x_4x_2 \to c_{14} = 111101 \\
&a_{15} = x_4x_3 \to c_{15} = 11111110, a_{16} = x_4x_4 \to c_{16} = 11111111
\end{aligned}$$

例1-续

– 编码的平均码长

$$\begin{aligned} L=\sum_{i=1}^{16}P(a_i)l_i &= 0.25\times 2+0.15\times 3+0.075\times 4+0.025\times 5 \\ &+0.15\times 3+0.09\times 3+0.045\times 4+0.015\times 6 \\ &+0.075\times 4+0.045\times 4+0.0225\times 6+0.0075\times 7 \\ &+0.025\times 5+0.015\times 6+0.0075\times 8+0.0025\times 8 \\ &\approx 3.328(bit) \end{aligned}$$

– 编码效率

$$\eta=\frac{2H(X)}{L}=\frac{2\times 1.648}{3.328}=\frac{3.296}{3.328}\approx 99.04\%$$

无失真信源编码—续

- 不等长编码的结论与问题
  - 不等长编码的平均码长可以是小数比特
  - 大概率符号序列编为短码、小概率符号序列编为长码有助于压缩平均码长
  - 信源的N次扩展有助于压缩平均码长，随N的增大，可以预见码率可能渐进达到信源编码的下界
  - 出现无失真译码问题——由于码长不等，接收端如何从码序列串中唯一分割出对应与每一个符号序列的码字

一、即时码与克拉夫特不等式

- 1、即时码与延时码
  - 如果不等长编码使接收端能从码序列串中唯一分割出对应与每一个符号序列的码字，该不等长编码为唯一可译码。
  - 唯一可译码中，能在对应于每一个符号序列的码字结束时立即译出的为即时码，要等到对应与下一个符号序列的码字出现才能译出的为延时码。

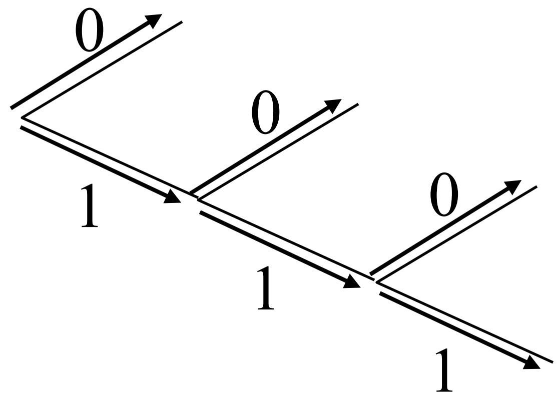
		A	B	C
x_1	0.50	$c_1=0$	$c_1=0$	$c_1=0$
x_2	0.30	$c_2=1$	$c_2=01$	$c_2=10$
x_3	0.15	$c_3=00$	$c_3=011$	$c_3=110$
x_4	0.05	$c_4=01$	$c_4=0111$	$c_4=111$

  - 码A不是唯一可译码，存在译码二义性，码B和码C是唯一可译码；码B是延时码，需等到对应与下一个符号的码字开头0才能确定本码字的结束，存在译码延时；码C是即时码。

一、即时码与克拉夫特不等式

- 码C的特点——任何一个码字都不是其它码字的前缀，是“自我间断”的即时码。
  - 即时码可以用树图来构造。
  - 码C所对应的二元码树图
  - 从树根开始到每一个终节点的联枝代表一个码字：

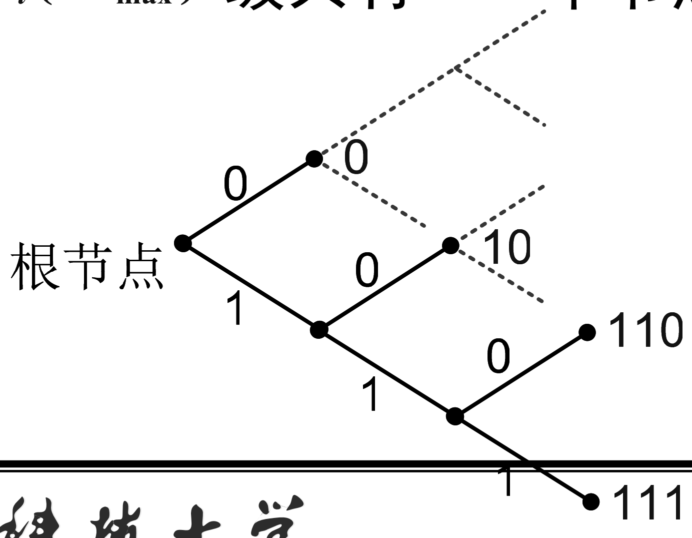
    0,10,110,111

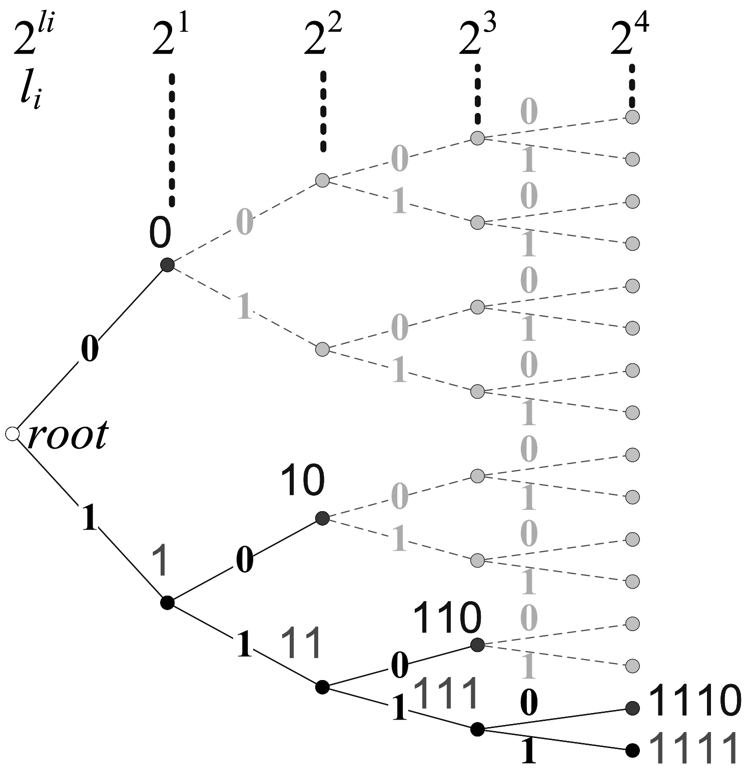


一、即时码与克拉夫特不等式

➢2、克拉夫特不等式

- 二元即时码的码长l_i必定满足克拉夫特不等式：$\sum_{i=1}^{n^N} 2^{-l_i} \le 1$
  l_i i=1,2, …, n^N为二元即时码第i个码字的长度
- 更一般形式：$\sum_{i=1}^{n^N} D^{-l_i} \le 1$
- 考虑一个 $l_{\max}$ 级二元满树，在第 $l_{\max}$ 级共有 $2^{l_{\max}}$ 个节点，在第 $l_i(< l_{\max})$ 级共有 2^{l_i} 个节点





一、即时码与克拉夫特不等式

➢2、克拉夫特不等式-*续*

– 根据即时码的定义，第i个码字后的节点不能再用，故不能用的第 $l_{\max}$ 级节点数为 $2^{l_{\max}-l_i}$。

- *构造即时码的二元码树图上总共不用的第* $l_{\max}$ *级节点总数*

$$\sum_{i=1}^{n^N} 2^{l_{\max}-l_i} \le 2^{l_{\max}} \quad \Rightarrow \quad \sum_{i=1}^{n^N} 2^{-l_i} \le 1$$

证毕。

二、最优码与最优码长的界

➢1、最优码

– L在 $\sum_{i=1}^{n^N} 2^{-l_i} \le 1$ 限制下的条件极值

令

$$\frac{\partial}{\partial l_k}\left\{\sum_{i=1}^{n^N} P(a_i)\, l_i + \lambda\left[\sum_{i=1}^{n^N} 2^{-l_i} - 1\right]\right\} = 0 \qquad k = 1, 2, \cdots, n^N$$

可得

$$\Rightarrow P(a_k) - \lambda 2^{-l_k} \ln 2 = 0$$

$$\Rightarrow 2^{-l_k} = \frac{P(a_k)}{\lambda \ln 2} \qquad k = 1, 2, \cdots, n^N$$

二、最优码与最优码长的界一续

➢1、最优码一续

$$\sum_{k=1}^{n^N} 2^{-l_k} = \sum_{k=1}^{n^N} \frac{P(a_k)}{\lambda \ln 2} = \frac{1}{\lambda \ln 2} \le 1$$

$$\lambda \ge 1/\ln 2$$

$$2^{-l_k} \le P(a_k) \qquad k = 1, 2, \cdots, n^N$$

$$l_k \ge -\mathrm{lb}P(a_k) \qquad k = 1, 2, \cdots, n^N$$

$$L = \sum_{i=1}^{n^N} P(a_k) l_k \ge -\sum_{k=1}^{n^N} P(a_k)\mathrm{lb}P(a_k) = H(X_1 X_2 \cdots X_N)$$

二、最优码与最优码长的界—续

- 2、最优码长的界
  - 选取 $l_i = \lceil -\mathrm{lb}P(a_i) \rceil \quad i = 1,2,\cdots,n^N$
  - 即：

$$-\mathrm{lb}P(a_i) \le l_i < -\mathrm{lb}P(a_i)+1 \quad i = 1,2,\cdots,n^N$$

$$-\sum_{i=1}^{n^N} P(a_i)\mathrm{lb}P(a_i) \le \sum_{i=1}^{n^N} P(a_i)l_i < -\sum_{i=1}^{n^N} P(a_i)\mathrm{lb}P(a_i)+1$$

$$\Rightarrow H(X_1X_2\cdots X_N) \le L < H(X_1X_2\cdots X_N)+1$$

$$\Rightarrow H_N(X_1X_2\cdots X_N) \le \frac{L}{N} < H_N(X_1X_2\cdots X_N)+\frac{1}{N}$$

  - 任意给定ε，取N= 1/ε，只要N足够大，有

$$H_N(X_1X_2\cdots X_N) \le L/N < H_N(X_1X_2\cdots X_N)+\varepsilon$$

  - 对于N次扩展信源，有

$$H(X) \le L/N < H(X)+\varepsilon$$

三、香农码

- 香农码直接基于最优码长的界，是一种采用即时码实现的无失真不等长编码。
- 香农码的编码步骤：
  - ①将符号序列按概率降序排列
  - ②确定第i个码字的码长l_i

$$l_i = \lceil -\mathrm{lb} P(a_i) \rceil \qquad i = 1, 2, \cdots, n^N$$

  - ③令$P(a_0)=0$，计算第i-1个码字的累加概率
  - ④将$P_a(a_i)$用二进制表示，取小数点后l_i位作为符号序列的码字

例1

- 编香农码，并计算编码效率

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} x_1 & x_2 & x_3 & x_4 \\ 0.5 & 0.3 & 0.15 & 0.05 \end{Bmatrix}$$

– ①将符号按概率降序排列

x_i	$P(x_i)$	I_i	$P_a(x_i)$	C_i
x_1	0.5			
x_2	0.3			
x_3	0.15			
x_4	0.05			

例1-续

– ②确定第i个码字的码长l_i

$$-\mathrm{lb}P(x_1)=\mathrm{lb}2=1 \quad l_1=1$$
$$-\mathrm{lb}P(x_2)=-\mathrm{lb}0.3\approx 1.74,\text{取}l_2=2$$
$$-\mathrm{lb}P(x_3)=-\mathrm{lb}0.15\approx 2.74,\text{取}l_3=3$$
$$-\mathrm{lb}P(x_4)=-\mathrm{lb}0.05\approx 4.32,\text{取}l_4=5$$

x_i	$P(x_i)$	l_i	$P_a(x_i)$	C_i
x_1	0.5	1		
x_2	0.3	2		
x_3	0.15	3		
x_4	0.05	5		

例1-续

– ③令$P(x_0)=0$，计算第j-1个码字的累加概率

$$P_a(x_i)=\sum_{j=0}^{i-1}P(x_j)=P_a(x_{i-1})+P(x_{i-1})\quad i=1,2,3,4$$

$$\begin{aligned}P_a(x_1)&=0\\P_a(x_2)&=0+0.5=0.5\\P_a(x_3)&=0.5+0.3=0.8\\P_a(x_4)&=0.8+0.15=0.95\end{aligned}$$

x_i	$P(x_i)$	l_i	$P_a(x_i)$	C_i
x_1	0.5	1	0	
x_2	0.3	2	0.5	
x_3	0.15	3	0.8	
x_4	0.05	5	0.95	

例1-续

– ④将$P_a(x_i)$用二进制表示，取小数点后l_i位作为符号的码字

$$P_a(x_1)=0.0=(0.0)_2 \rightarrow c_1=0$$
$$P_a(x_2)=0.5=(0.10)_2 \rightarrow c_2=10$$
$$P_a(x_3)=0.8=(0.110\ldots)_2 \rightarrow c_3=110$$
$$P_a(x_4)=0.95=(0.11110\ldots)_2 \rightarrow c_4=11110$$

– 编码结果

x_i	$P(x_i)$	l_i	$P_a(x_i)$	C_i
x_1	0.5	1	0	0
x_2	0.3	2	0.5	10
x_3	0.15	3	0.8	110
x_4	0.05	5	0.95	11110

例1-续

- 熵为：

$$
\begin{aligned}
H(X) &= -\sum_{i=1}^{4} P(x_i)\,\mathrm{lb}P(x_i) \\
&= -0.5\mathrm{lb}0.5 - 0.3\mathrm{lb}0.3 - 0.15\mathrm{lb}0.15 - 0.05\mathrm{lb}0.05 \\
&\approx 1.648(\mathrm{bit})
\end{aligned}
$$

- 平均码长：

$$
L = \sum_{i=1}^{4} P(x_i)\, l_i = 0.5\times 1 + 0.3\times 2 + 0.15\times 3 + 0.05\times 5 = 1.8(\mathrm{bit})
$$

- 编码效率：

$$
\eta = \frac{H(X)}{L} = \frac{1.648}{1.8} \approx 91.56\%
$$

例2

- 对如下信源及其二次扩展信源编香农码，并计算编码效率

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} x_1 & x_2 & x_3 \\ 0.5 & 0.3 & 0.2 \end{Bmatrix}$$

– (1)对未扩展信源编码

$$-\mathrm{lb}P(x_1) = -\mathrm{lb}0.5 = 1 \quad l_1 = 1$$
$$-\mathrm{lb}P(x_2) = -\mathrm{lb}0.3 = 1.737, \text{取} l_2 = 2$$
$$-\mathrm{lb}P(x_3) = -\mathrm{lb}0.2 = 2.32, \text{取} l_3 = 3$$

– 编码结果

x_i	$P(x_i)$	l_i	$P_a(x_i)$	C_i
x_1	0.5	1	0	0
x_2	0.3	2	0.5	10
x_3	0.2	3	0.8	110

$$P_a(x_1) = 0$$
$$P_a(x_2) = 0 + 0.5 = 0.5$$
$$P_a(x_3) = 0.5 + 0.3 = 0.8$$

$$\Rightarrow \begin{cases} P_a(x_1) = 0.0 = (0.0)_2 & \rightarrow c_1 = 0 \\ P_a(x_2) = 0.5 = (0.10)_2 & \rightarrow c_2 = 10 \\ P_a(x_3) = 0.8 = (0.1100\ldots)_2 & \rightarrow c_3 = 110 \end{cases}$$

例2-续

- 熵为：

$$H(X) = -\sum_{i=1}^{3} P(x_i)\,\mathrm{lb}P(x_i)$$
$$= -0.5\mathrm{lb}0.5 - 0.3\mathrm{lb}0.3 - 0.2\mathrm{lb}0.2$$
$$\approx 1.4855(\mathrm{bit})$$

- 平均码长：

$$L = \sum_{i=1}^{3} P(x_i)\, l_i = 0.5\times1 + 0.3\times2 + 0.2\times3 = 1.7(\mathrm{bit})$$

- 编码效率

$$\eta = \frac{H(X)}{L} = \frac{1.4855}{1.7} \approx 87.38\%$$

例2-续

– (2)对二次扩展信源编码

$$\begin{bmatrix} X^2 \\ P(X^2) \end{bmatrix} = \begin{Bmatrix} x_1x_1 & x_1x_2 & x_1x_3 & x_2x_1 & x_2x_2 & x_2x_3 & x_3x_1 & x_3x_2 & x_3x_3 \\ 0.25 & 0.15 & 0.1 & 0.15 & 0.09 & 0.06 & 0.1 & 0.06 & 0.04 \end{Bmatrix}$$

$$-\mathrm{lb}P(x_1x_1) = -\mathrm{lb}0.25 = 2 \quad l_1 = 2$$
$$-\mathrm{lb}P(x_1x_2) = -\mathrm{lb}0.15 \approx 2.74 \quad 取 l_2 = 3$$
$$-\mathrm{lb}P(x_2x_1) = -\mathrm{lb}0.15 \approx 2.74 \quad 取 l_3 = 3$$
$$-\mathrm{lb}P(x_1x_3) = -\mathrm{lb}0.1 \approx 3.32 \quad 取 l_4 = 4$$
$$-\mathrm{lb}P(x_3x_1) = -\mathrm{lb}0.1 \approx 3.32 \quad 取 l_5 = 4$$
$$-\mathrm{lb}P(x_2x_2) = -\mathrm{lb}0.09 \approx 3.47 \quad 取 l_6 = 4$$
$$-\mathrm{lb}P(x_2x_3) = -\mathrm{lb}0.06 \approx 4.06 \quad 取 l_7 = 5$$
$$-\mathrm{lb}P(x_3x_2) = -\mathrm{lb}0.06 \approx 4.06 \quad 取 l_8 = 5$$
$$-\mathrm{lb}P(x_3x_3) = -\mathrm{lb}0.04 \approx 4.64 \quad 取 l_9 = 5$$

$$P_a(x_1x_1) = 0$$
$$P_a(x_1x_2) = 0+0.25=0.25$$
$$P_a(x_2x_1) = 0.25+0.15=0.4$$
$$P_a(x_1x_3) = 0.4+0.15=0.55$$
$$P_a(x_3x_1) = 0.55+0.1=0.65$$
$$P_a(x_2x_2) = 0.65+0.1=0.75$$
$$P_a(x_2x_3) = 0.75+0.09=0.84$$
$$P_a(x_3x_2) = 0.84+0.06=0.9$$
$$P_a(x_3x_3) = 0.9+0.06=0.96$$

例2-续

$$
\begin{aligned}
&P_a(x_1x_1) = 0\\
&P_a(x_1x_2) = 0+0.25=0.25\\
&P_a(x_2x_1) = 0.25+0.15=0.4\\
&P_a(x_1x_3) = 0.4+0.15=0.55\\
&P_a(x_3x_1) = 0.55+0.1=0.65\\
&P_a(x_2x_2) = 0.65+0.1=0.75\\
&P_a(x_2x_3) = 0.75+0.09=0.84\\
&P_a(x_3x_2) = 0.84+0.06=0.9\\
&P_a(x_3x_3) = 0.9+0.06=0.96
\end{aligned}
$$

$$
\Rightarrow \begin{cases}
P_a(x_1x_1) = 0.0 = (0.00)_2 & \to c_1 = 00\\
P_a(x_1x_2) = 0.25 = (0.010)_2 & \to c_2 = 010\\
P_a(x_2x_1) = 0.4 = (0.011\ldots)_2 & \to c_3 = 011\\
P_a(x_1x_3) = 0.55 = (0.1000\ldots)_2 & \to c_4 = 1000\\
P_a(x_3x_1) = 0.65 = (0.1010\ldots)_2 & \to c_5 = 1010\\
P_a(x_2x_2) = 0.75 = (0.1100\ldots)_2 & \to c_6 = 1100\\
P_a(x_2x_3) = 0.84 = (0.11010\ldots)_2 & \to c_7 = 11010\\
P_a(x_3x_2) = 0.9 = (0.11100\ldots)_2 & \to c_8 = 11100\\
P_a(x_3x_3) = 0.96 = (0.11110\ldots)_2 & \to c_9 = 11110
\end{cases}
$$

例2-续

a_i	$P(a_i)$	l_i	$P_a(a_i)$	C_i
x_1x_1	0.25	2	0	00
x_1x_2	0.15	3	0.25	010
x_2x_1	0.15	3	0.4	011
x_1x_3	0.1	4	0.55	1000
x_3x_1	0.1	4	0.65	1010
x_2x_2	0.09	4	0.75	1100
x_2x_3	0.06	5	0.84	11010
x_3x_2	0.06	5	0.9	11100
x_3x_3	0.04	5	0.96	11110

- 平均码长：

$$
\begin{aligned}
L &= \sum_{i=1}^{9} P(a_i) l_i \\
&= 0.25\times 2 + 2\times 0.15\times 3 + 2\times 0.1\times 4 \\
&\quad + 0.09\times 4 + (2\times 0.06 + 0.04)\times 5 \\
&= 3.36(\text{bit})
\end{aligned}
$$

- 编码效率

$$
\eta = \frac{2H(X)}{L} = \frac{2\times 1.4855}{3.36} = \frac{2.971}{3.36} \approx 88.42\%
$$

例3

- 对如下信源及其二次扩展信源编香农码，并计算编码效率

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} x_1 & x_2 & x_3 \\ 0.5 & 0.4 & 0.1 \end{Bmatrix}$$

– (1)对未扩展信源编码

$$-\mathrm{lb}P(x_1) = -\mathrm{lb}0.5 = 1 \quad l_1 = 1$$
$$-\mathrm{lb}P(x_2) = -\mathrm{lb}0.4 = 2,\text{取}l_2 = 2$$
$$-\mathrm{lb}P(x_3) = -\mathrm{lb}0.1 = 3.32,\text{取}l_3 = 4$$

$$P_a(x_1) = 0$$
$$P_a(x_2) = 0 + 0.5 = 0.5$$
$$P_a(x_3) = 0.5 + 0.4 = 0.9$$

$$\Rightarrow \begin{cases} P_a(x_1) = 0.0 = (0.0)_2 & \to c_1 = 0 \\ P_a(x_2) = 0.5 = (0.10)_2 & \to c_2 = 10 \\ P_a(x_3) = 0.9 = (0.1110\ldots)_2 & \to c_3 = 1110 \end{cases}$$

– 编码结果

x_i	$P(x_i)$	l_i	$P_a(x_i)$	C_i
x_1	0.5	1	0	0
x_2	0.4	2	0.5	10
x_3	0.1	4	0.9	1110

例3-续

- 熵为：

$$
\begin{aligned}
H(X) &= -\sum_{i=1}^{3} P(x_i)\,\mathrm{lb}P(x_i) \\
&= -0.5\mathrm{lb}0.5 - 0.4\mathrm{lb}0.4 - 0.1\mathrm{lb}0.1 \\
&\approx 1.3610(\mathrm{bit})
\end{aligned}
$$

- 平均码长：

$$L = \sum_{i=1}^{3} P(x_i)\,l_i = 0.5\times1 + 0.4\times2 + 0.1\times4 = 1.7(\mathrm{bit})$$

- 编码效率

$$\eta = \frac{H(X)}{L} = \frac{1.3610}{1.7} \approx 80.06\%$$

例3-续

– (2)对二次扩展信源编码

$$\begin{bmatrix} X^2 \\ P(X^2) \end{bmatrix} = \begin{Bmatrix} x_1x_1 & x_1x_2 & x_1x_3 & x_2x_1 & x_2x_2 & x_2x_3 & x_3x_1 & x_3x_2 & x_3x_3 \\ 0.25 & 0.2 & 0.05 & 0.2 & 0.16 & 0.04 & 0.05 & 0.04 & 0.01 \end{Bmatrix}$$

$$-\mathrm{lb}P(x_1x_1) = -\mathrm{lb}0.25 = 2 \quad l_1 = 2$$
$$-\mathrm{lb}P(x_1x_2) = -\mathrm{lb}0.20 \approx 2.32 \quad 取 l_2 = 3$$
$$-\mathrm{lb}P(x_2x_1) = -\mathrm{lb}0.20 \approx 2.32 \quad 取 l_3 = 3$$
$$-\mathrm{lb}P(x_2x_2) = -\mathrm{lb}0.16 \approx 2.64 \quad 取 l_4 = 3$$
$$-\mathrm{lb}P(x_1x_3) = -\mathrm{lb}0.05 \approx 4.32 \quad 取 l_5 = 5$$
$$-\mathrm{lb}P(x_3x_1) = -\mathrm{lb}0.05 \approx 4.32 \quad 取 l_6 = 5$$
$$-\mathrm{lb}P(x_2x_3) = -\mathrm{lb}0.04 \approx 4.64 \quad 取 l_7 = 5$$
$$-\mathrm{lb}P(x_3x_2) = -\mathrm{lb}0.04 \approx 4.64 \quad 取 l_8 = 5$$
$$-\mathrm{lb}P(x_3x_3) = -\mathrm{lb}0.01 \approx 6.64 \quad 取 l_9 = 7$$

$$P_a(x_1x_1) = 0$$
$$P_a(x_1x_2) = 0+0.25=0.25$$
$$P_a(x_2x_1) = 0.25+0.2=0.45$$
$$P_a(x_2x_2) = 0.45+0.2=0.65$$
$$P_a(x_1x_3) = 0.65+0.16=0.81$$
$$P_a(x_3x_1) = 0.81+0.05=0.86$$
$$P_a(x_2x_3) = 0.86+0.05=0.91$$
$$P_a(x_3x_2) = 0.91+0.04=0.95$$
$$P_a(x_3x_3) = 0.95+0.04=0.99$$

例3-续

$$
\begin{aligned}
&P_a(x_1x_1)=0\\
&P_a(x_1x_2)=0+0.25=0.25\\
&P_a(x_2x_1)=0.25+0.2=0.45\\
&P_a(x_2x_2)=0.45+0.2=0.65\\
&P_a(x_1x_3)=0.65+0.16=0.81\\
&P_a(x_3x_1)=0.81+0.05=0.86\\
&P_a(x_2x_3)=0.86+0.05=0.91\\
&P_a(x_3x_2)=0.91+0.04=0.95\\
&P_a(x_3x_3)=0.95+0.04=0.99
\end{aligned}
$$

$$
\Rightarrow \begin{cases}
P_a(x_1x_1)=0.0=(0.00)_2 & \to c_1=00\\
P_a(x_1x_2)=0.25=(0.010)_2 & \to c_2=010\\
P_a(x_2x_1)=0.45=(0.011\ldots)_2 & \to c_3=011\\
P_a(x_2x_2)=0.65=(0.101\ldots)_2 & \to c_4=101\\
P_a(x_1x_3)=0.81=(0.11001\ldots)_2 & \to c_5=11001\\
P_a(x_3x_1)=0.86=(0.11011\ldots)_2 & \to c_6=11011\\
P_a(x_2x_3)=0.91=(0.11101\ldots)_2 & \to c_7=11101\\
P_a(x_3x_2)=0.95=(0.11110\ldots)_2 & \to c_8=11110\\
P_a(x_3x_3)=0.99=(0.1111110\ldots)_2 & \to c_9=1111110
\end{cases}
$$

例3-续

a_i	$P(a_i)$	l_i	$P_a(a_i)$	C_i
x_1x_1	0.25	2	0	00
x_1x_2	0.20	3	0.25	010
x_2x_1	0.20	3	0.45	011
x_2x_2	0.16	3	0.65	101
x_1x_3	0.05	5	0.81	11001
x_3x_1	0.05	5	0.86	11011
x_2x_3	0.04	5	0.91	11101
x_3x_2	0.04	5	0.95	11110
x_3x_3	0.01	7	0.99	1111110

- 平均码长：

$$
\begin{aligned}
L &= \sum_{i=1}^{9} P(a_i) l_i \\
&= 0.25\times 2+(2\times 0.20+0.16)\times 3 \\
&\quad +2\times(0.05+0.04)\times 5+0.01\times 7 \\
&= 3.15(\text{bit})
\end{aligned}
$$

- 编码效率

$$
\eta=\frac{2H(X)}{L}=\frac{2\times 1.36096}{3.15}=\frac{2.7219}{3.15} \approx 86.41\%
$$

例4-香农编码的缺陷分析

x_i	$P(x_i)$	l_i	$P_a(x_i)$	C_i
x_1	0.5	1	0	0
x_2	0.3	2	0.5	10
x_3	0.2	3/2	0.8	110/11
x_1	0.5	1	0	0
x_2	0.4	2	0.5	10
x_3	0.1	4/2	0.8	1110/11
x_1	0.5	1	0	0
x_2	0.3	2	0.5	10
x_3	0.2	5/2	0.8	11110/11

- 平均码长及编码效率
  - 概率分布1

$$H(X)=1.4855, L=1.7(\text{bit}), \eta \approx 0.8738\%$$

$$H(X)=1.4855, L=1.5(\text{bit}), \eta \approx 99.03\%$$

  - 概率分布2

$$H(X)=1.3610, L=1.7(\text{bit}), \eta \approx 80.06\%$$

$$H(X)=1.3610, L=1.5(\text{bit}), \eta \approx 90.73\%$$

  - 概率分布3

$$H(X)=1.2345, L=1.65(\text{bit}), \eta \approx 74.82\%$$

$$H(X)=1.3610, L=1.5(\text{bit}), \eta \approx 82.30\%$$

例4-二次扩展

a_i	$P(a_i)$	l_i	$P_a(a_i)$	C_i
x_1x_1	0.25	2	0	00
x_1x_2	0.20	3	0.25	010
x_2x_1	0.20	3	0.45	011
x_2x_2	0.16	3	0.65	101
x_1x_3	0.05	5/4	0.81	11001/1100
x_3x_1	0.05	5/4	0.86	11011/1101
x_2x_3	0.04	5/4	0.91	11101/1110
x_3x_2	0.04	5	0.95	11110/11110
x_3x_3	0.01	7/5	0.99	1111110/11111

$$H(X) = 2.722(\text{bit})$$
$$L = 3.13(\text{bit}), \eta \approx 86.96\%,$$
$$L = 2.99(\text{bit}), \eta \approx 91.03\%$$

a_i	$P(a_i)$	L_i	$P_a(a_i)$	C_i
x_1x_1	0.25	2	0	00
x_1x_2	0.225	3	0.25	010
x_2x_1	0.225	3	0.475	011
x_2x_2	0.2025	3	0.7	101
x_1x_3	0.025	6/4	0.9025	111001
x_3x_1	0.025	6/4	0.9275	111011
x_2x_3	0.0225	6/4	0.9525	111100
x_3x_2	0.0225	6/5	0.975	111110
x_3x_3	0.0025	9/5	0.9975	111111110

$$H(X) = 2.469(\text{bit})$$
$$L = 3.05(\text{bit}), \eta \approx 80.95\%,$$
$$L = 2.8725(\text{bit}), \eta \approx 85.95\%$$

三、香农码一续

➢ 香农码的特点

- 编码过程中先确定码长，后确定码字，保证大概率符号序列编为短码，小概率符号序列编为长码
- 第一个符号序列的累加概率为0，其对应的码字总是0、00、…形式
- 具有唯一性
- 平均码长不超过联合熵一个比特，N越大平均码长越接近联合熵

四、霍夫曼码

- 霍夫曼码也是采用即时码实现的无失真不等长编码。
  - 通常霍夫曼码的编码效率高于香农码。
  - 霍夫曼码的编码步骤
    - *①将符号序列按概率降序排列*
    - *②为概率最小的符号序列分配一个码元1，概率次小的符号序列分配一个码元0*
    - *③将概率最小的两个符号序列合并成一个新的符号序列，用两者概率之和作为新符号序列的概率*
    - *重复① ② ③步骤，直到合并出一个以1为概率的新符号序列，结束编码*

例1

- 对如下信源进行编霍夫曼码，并计算编码效率。

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} x_1 & x_2 & x_3 & x_4 \\ 0.5 & 0.3 & 0.15 & 0.05 \end{Bmatrix}$$

– ①将符号按概率降序排列

– ②为概率最小的符号分配一个二进制码元1，概率次小的符号分配一个二进制码元0

– ③将概率最小的两个符号合并成一个新的符号，用两者概率之和作为新符号的概率

– 重复① ② ③步骤，直到合并出一个以1为概率的新符号，结束编码

x_i	$P(x_i)$
x_1	0.5
x_2	0.3
x_3	0.15
x_4	0.05

x_3: 0, x_4: 1 → x_3' 0.2

x_i	$P(x_i)$
x_1	0.5
x_2	0.3
x_3'	0.2

x_2: 0, x_3': 1 → x_2' 0.5

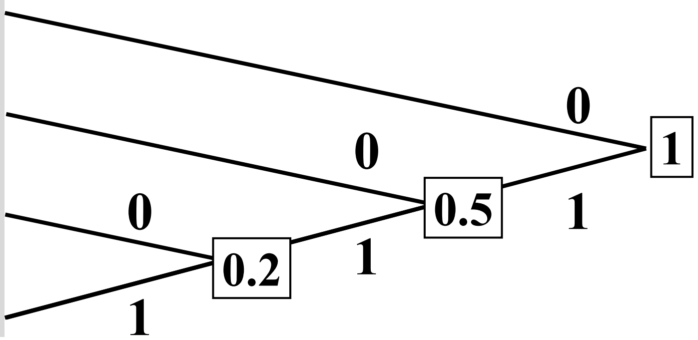
x_i	$P(x_i)$
x_1	0.5
x_2'	0.5

x_1: 0, x_2': 1 → x_1' 1

例1-续

– 整个编码过程汇总：

x_i	$P(x_i)$
x_1	0.5
x_2	0.3
x_3	0.15
x_4	0.05



c_i	l_i
0	1
10	2
110	3
111	3

– 平均码长：

$$L=\sum_{i=1}^{4}P(x_i)l_i=0.5\times1+0.3\times2+0.15\times3+0.05\times3=1.7(\text{bit})$$

– 编码效率：

$$\eta=H(X)/L=1.648/1.7\approx96.94\%$$

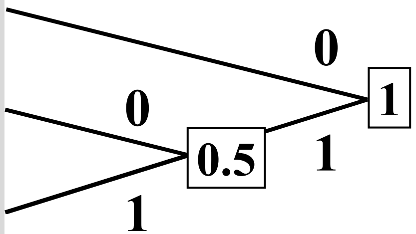
例2

- 对如下信源和其二次扩展信源分别编霍夫曼码，并计算编码效率。

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} x_1 & x_2 & x_3 \\ 0.5 & 0.4 & 0.1 \end{Bmatrix}$$

– (1)对信源编码

x_i	$P(x_i)$
x_1	0.5
x_2	0.4
x_3	0.1



c_i	l_i
0	1
10	2
11	2

– 信源熵

$$H(X) = -\sum_{i=1}^{3} P(x_i)\mathrm{lb}P(x_i)$$
$$= -0.5\mathrm{lb}0.5 - 0.4\mathrm{lb}0.4 - 0.1\mathrm{lb}0.1$$
$$\approx 1.361(\mathrm{bit})$$

– 平均码长

$$L = \sum_{i=1}^{3} P(x_i) l_i$$
$$= 0.5\times1 + 0.4\times2 + 0.1\times2$$
$$= 1.5(\mathrm{bit})$$

– 编码效率

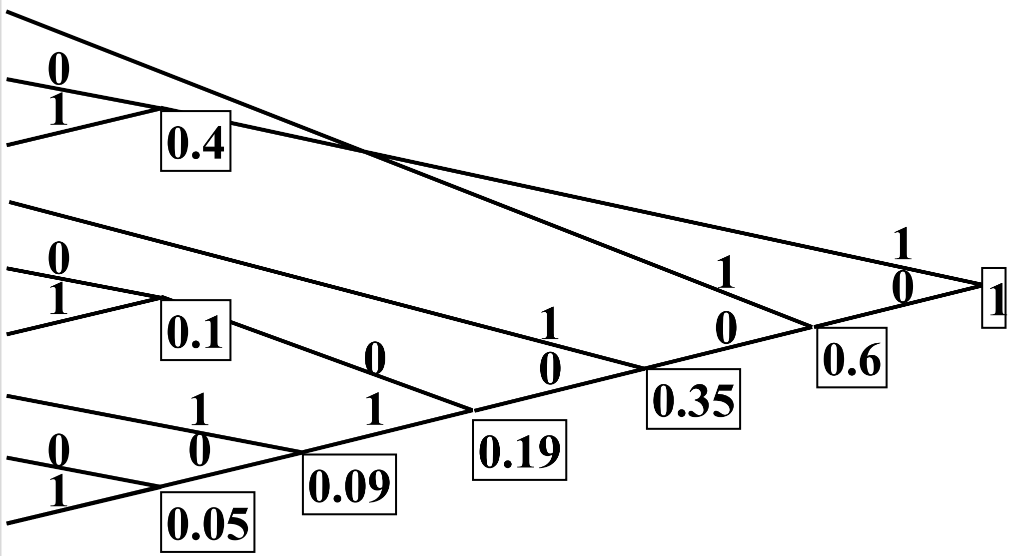
$$\eta = \frac{H(X)}{L} = \frac{1.361}{1.5} \approx 90.73\%$$

例2-续

– (2)对二次扩展信源编码

$$\begin{bmatrix} X^2 \\ P(X^2) \end{bmatrix} = \begin{Bmatrix} x_1x_1 & x_1x_2 & x_1x_3 & x_2x_1 & x_2x_2 & x_2x_3 & x_3x_1 & x_3x_2 & x_3x_3 \\ 0.25 & 0.2 & 0.05 & 0.2 & 0.16 & 0.04 & 0.05 & 0.04 & 0.01 \end{Bmatrix}$$

a_i	$P(a_i)$
x_1x_1	0.25
x_1x_2	0.2
x_2x_1	0.2
x_2x_2	0.16
x_1x_3	0.05
x_3x_1	0.05
x_2x_3	0.04
x_3x_2	0.04
x_3x_3	0.01



c_i	l_i
01	2
10	2
11	2
001	3
00000	5
00001	5
00011	5
000100	6
000101	6

例2-续

- 平均码长

$$L=\sum_{i=1}^{9}P(a_i)l_i$$
$$=(0.25+2\times0.2)\times2+0.16\times3+(2\times0.05+0.04)\times5+(0.04+0.01)\times6$$
$$=2.78(\text{bit})$$

- 编码效率

$$\eta=\frac{2H(X)}{L}$$
$$=\frac{2\times1.361}{2.78}$$
$$\approx97.91\%$$

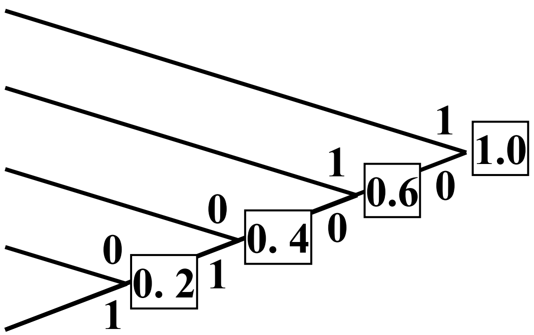
例3

- 用两种方法进行霍夫曼编码，并计算平均码长。

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \end{Bmatrix}$$

– (1)将合并后的新符号排在其它相同概率符号之后

x_i	$P(x_i)$	c_i	l_i
x_1	0.4	1	1
x_2	0.2	01	2
x_3	0.2	000	3
x_4	0.1	0010	4
x_5	0.1	0011	4



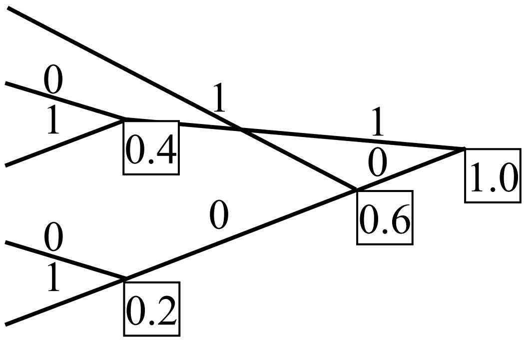
- *平均码长*

$$L = \sum_{i=1}^{5} P(x_i) l_i = 0.4\times1 + 0.2\times2 + 0.2\times3 + 2\times0.1\times4 = 2.2(\text{bit})$$

例3-*续*

– (2)将合并后的新符号排在其它相同概率符号之前

x_i	$P(x_i)$
x_1	**0.4**
x_2	**0.2**
x_3	**0.2**
x_4	**0.1**
x_5	**0.1**



c_i	l_i
00	**2**
10	**2**
11	**2**
010	**3**
011	**3**

- *平均码长*

$$L=\sum_{i=1}^{5}P(x_i)l_i=0.4\times2+0.2\times2+0.2\times2+2\times0.1\times3=2.2(\text{bit})$$

- *采用不同排列方法编出的霍夫曼码，其码字和码长可能完全不相同，但平均码长相等，编码效率不会因排列方法而改变。*

四、霍夫曼码一续

- 霍夫曼码的特点
  - 编码过程中先确定码字，后确定码长
  - 用局部累加概率代替累加概率，多次重新排列合并累加的过程是优化过程
  - 每次合并伴之分配码元保证大概率符号序列编为短码，小概率符号序列编为长码
  - 不具有唯一性，但不同霍夫曼码的编码效率相同
  - 平均码长不超过联合熵一个比特，N越大平均码长越接近联合熵

五、费诺码

- 费诺码基于概率匹配的思想，也是采用即时码实现的无失真不等长编码。
  - 费诺码的编码步骤
    - *①将符号序列按概率降序排列*
    - *②按二进制进行分组，使每组概率尽可能相等*
    - *③给每个分组分配一个二进制码元*
    - *对每个分组重复②③步骤，直到不可分为止*

例1

➢ 编费诺码，并计算编码效率。

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} x_1 & x_2 & x_3 & x_4 \\ 0.5 & 0.3 & 0.15 & 0.05 \end{Bmatrix}$$

– 建立分组表

x_i	$P(x_i)$	分组1	分组2	分组3	C_i
x_1	0.5				
x_2	0.3				
x_3	0.15				
x_4	0.05				

例1-续

– 第一次分组：

$$P(x_1) = 0.5$$
$$P(x_2) + P(x_3) + P(x_4) = 0.3 + 0.15 + 0.05 = 0.5$$

x_i	$P(x_i)$	分组1	分组2	分组3	C_i
x_1	0.5	0			
x_2	0.3	1			
x_3	0.15				
x_4	0.05				

例1-续

– 第二次分组：

$$P(x_2) = 0.3$$
$$P(x_3) + P(x_4) = 0.15 + 0.05 = 0.2$$

x_i	$P(x_i)$	分组1	分组2	分组3	C_i
x_1	0.5	0			
x_2	0.3	1	0		
x_3	0.15		1		
x_4	0.05				

例1-续

– 第三次分组：

$$P(x_3)=0.15$$
$$P(x_4)=0.05$$

– 平均码长

$$L=\sum_{i=1}^{4}P(x_i)l_i = 0.5\times1+0.3\times2 +0.15\times3+0.05\times3 =1.7(\text{bit})$$

– 编码效率

$$\eta=\frac{H(X)}{L}=\frac{1.648}{1.7}\approx 96.94\%$$

– 编码结果

x_i	$P(x_i)$	分组1	分组2	分组3	C_i
x_1	0.5	0			0
x_2	0.3	1	0		10
x_3	0.15		1	0	110
x_4	0.05			1	111

例2

- 分别对信源和其二次扩展信源编费诺码，并计算编码效率。

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} x_1 & x_2 & x_3 \\ 0.5 & 0.4 & 0.1 \end{Bmatrix}$$

– (1)对单符号离散信源编码

- *第一次分组：*

$$P(x_1)=0.5$$
$$P(x_2)+P(x_3)=0.4+0.1=0.5$$

- *第二次分组：*

$$P(x_2)=0.4$$
$$P(x_3)=0.1$$

– 编码结果

x_i	$P(x_i)$	分组1	分组2	C_i
x_1	0.5	0		0
x_2	0.4	1	0	10
x_3	0.1		1	11

– 平均码长

$$L=\sum_{i=1}^{3} P(x_i)l_i=0.5\times1+0.4\times2+0.1\times2$$
$$=1.5(\text{bit})$$

– 编码效率

$$\eta=H(X)/L=1.361/1.5\approx 90.73\%$$

例2-续

– (2)对二次扩展信源编码

$$\begin{bmatrix} X^2 \\ P(X^2) \end{bmatrix} = \begin{Bmatrix} x_1x_1 & x_1x_2 & x_1x_3 & x_2x_1 & x_2x_2 & x_2x_3 & x_3x_1 & x_3x_2 & x_3x_3 \\ 0.25 & 0.2 & 0.05 & 0.2 & 0.16 & 0.04 & 0.05 & 0.04 & 0.01 \end{Bmatrix}$$

- *第一次分组：*

$$P(x_1x_1)+P(x_1x_2)=0.25+0.2$$
$$=0.45$$
$$P(x_2x_1)+P(x_2x_2)+\cdots$$
$$+P(x_3x_3)$$
$$=0.2+0.16+0.05+$$
$$0.05+0.04+$$
$$0.04+0.01$$
$$=0.55$$

- *第二次分组：*

$$P(x_1x_1)=0.25$$
$$P(x_1x_2)=0.2$$
$$P(x_2x_1)=0.2$$
$$P(x_2x_2)+P(x_1x_3)+P(x_3x_1)$$
$$+P(x_2x_3)+P(x_3x_2)+P(x_3x_3)$$
$$=0.16+0.05+0.05+$$
$$0.04+0.04+0.01$$
$$=0.35$$

例2-续

- 第三次分组：

$$P(x_2x_2)=0.16$$
$$P(x_1x_3)+P(x_3x_1)+P(x_2x_3)$$
$$+P(x_3x_2)+P(x_3x_3)$$
$$=0.05+0.05+0.04+0.04+0.01$$
$$=0.19$$

- 第四次分组：

$$P(x_1x_3)+P(x_3x_1)$$
$$=0.05+0.05$$
$$=0.1$$
$$P(x_2x_3)+P(x_3x_2)+P(x_3x_3)$$
$$=0.04+0.04+0.01$$
$$=0.09$$

- 第五次分组：

$$P(x_1x_3)=0.5$$
$$P(x_3x_1)=0.05$$
$$P(x_2x_3)=0.04$$
$$P(x_3x_2)+P(x_3x_3)$$
$$=0.04+0.01$$
$$=0.05$$

- 第六次分组：

$$P(x_3x_2)=0.04$$
$$P(x_3x_3)=0.01$$

例2-续

<table>
<tr><th>a_i</th><th>$P(a_i)$</th><th>分组1</th><th>分组2</th><th>分组3</th><th>分组4</th><th>分组5</th><th>分组6</th><th>C_i</th></tr>
<tr><td>x_1x_1</td><td>0.25</td><td rowspan="2">0</td><td>0</td><td></td><td></td><td></td><td></td><td>00</td></tr>
<tr><td>x_1x_2</td><td>0.20</td><td>1</td><td></td><td></td><td></td><td></td><td>01</td></tr>
<tr><td>x_2x_1</td><td>0.20</td><td rowspan="7">1</td><td>0</td><td></td><td></td><td></td><td></td><td>10</td></tr>
<tr><td>x_2x_2</td><td>0.16</td><td rowspan="6">1</td><td>0</td><td></td><td></td><td></td><td>110</td></tr>
<tr><td>x_1x_3</td><td>0.05</td><td rowspan="5">1</td><td rowspan="2">0</td><td>0</td><td></td><td>11100</td></tr>
<tr><td>x_3x_1</td><td>0.05</td><td>1</td><td></td><td>11101</td></tr>
<tr><td>x_2x_3</td><td>0.04</td><td rowspan="3">1</td><td>0</td><td></td><td>11110</td></tr>
<tr><td>x_3x_2</td><td>0.04</td><td rowspan="2">1</td><td>0</td><td>111110</td></tr>
<tr><td>x_3x_3</td><td>0.01</td><td>1</td><td>111111</td></tr>
</table>

$$L = \sum_{i=1}^{9} P(a_i) l_i = (0.25 + 2\times 0.2)\times 2 + 0.16\times 3 + (2\times 0.05 + 0.04)\times 5$$
$$+ (0.04 + 0.01)\times 6 = 2.78(\text{bit})$$

$$\eta = 2H(X) / L = 2\times 1.361 / 2.78 \approx 97.91\%$$

- *二次扩展后的编码效率提高*

例3

➢ 用两种方法进行费诺编码，并计算平均码长。

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \end{Bmatrix}$$

– (1)将大概率分组排在前面

- *第一次分组：*

$$P(x_1)+P(x_2)$$
$$=0.4+0.2$$
$$=0.6$$
$$P(x_3)+P(x_4)+P(x_5)$$
$$=0.2+0.1+0.1$$
$$=0.4$$

- *第二次分组：*

$$P(x_1)=0.4$$
$$P(x_2)=0.2$$
$$P(x_3)=0.2$$
$$P(x_4)+P(x_5)$$
$$=0.1+0.1$$
$$=0.2$$

- *第三次分组：*

$$P(x_4)=0.1$$
$$P(x_5)=0.1$$

例3-续

- 编码结果

x_i	$P(x_i)$	分组1	分组2	分组3	C_i
x_1	0.4	0	0		00
x_2	0.2		1		01
x_3	0.2	1	0		10
x_4	0.1		1	0	110
x_5	0.1			1	111

- 平均码长

$$L=\sum_{i=1}^{5}P(x_i)l_i=(0.4+2\times0.2)\times2+2\times0.1\times3=2.2(\text{bit})$$

例3-续

– (2)将小概率分组排在前面

- *第一次分组：*

$$P(x_1)=0.4$$
$$P(x_2)+P(x_3)+P(x_4)+P(x_5)$$
$$=0.2+0.2+0.1+0.1$$
$$=0.6$$

- *第二次分组：*

$$P(x_2)=0.2$$
$$P(x_3)+P(x_4)+P(x_5)$$
$$=0.2+0.1+0.1$$
$$=0.4$$

- *第三次分组：*

$$P(x_3)=0.2$$
$$P(x_4)+P(x_5)$$
$$=0.1+0.1$$
$$=0.2$$

- *第四次分组：*

$$P(x_4)=0.1$$
$$P(x_5)=0.1$$

例3-续

- 编码结果

x_i	$P(x_i)$	分组1	分组2	分组3	分组4	C_i
x_1	0.4	0				0
x_2	0.2	1	0			10
x_3	0.2		1	0		110
x_4	0.1			1	0	1110
x_5	0.1				1	1111

$$L=\sum_{i=1}^{5}P(x_i)l_i=0.4\times1+0.2\times2+0.2\times3+2\times0.1\times4=2.2(\text{bit})$$

– 采用不同排列方法编出的费诺码，其码字和码长可能完全不相同，但平均码长相等，编码效率不会因排列方法而改变。

五、费诺码一续

- 费诺码的特点
  - 编码过程中先确定码字，后确定码长
  - 大概率符号序列分解次数少，编为短码，小概率符号序列分解次数多，编为长码
  - 不具有唯一性，但不同费诺码的编码效率相同
  - 平均码长不超过联合熵一个比特，N越大平均码长越接近联合熵

第四章—小结

- 信源编码是信源符号序列到码序列(码字)之间的映射
- 无失真信源编码原则
  - 选择适合信道传输的码
  - 至少要保证典型序列与码字间一一对应
  - 与所有符号有一一对应码字的编码是无失真编码
- 等长编码的平均码长是整数
- 不等长编码的平均码长可以是小数比特
  - 大概率符号序列用短码、小概率符号序列用长码
- 对信源的N次扩展进行编码有助压缩平均码长，且随N增大码率可逼近信源编码下界
- 唯一可译码无码间二义性；即时码无前后码二义性；延迟码有前后码二义性
- 二元即时码的码长满足克拉夫特不等式
  - 最佳编码的平均码长不超过联合熵1比特，且随扩展阶数N增大逐步逼进联合熵

本章小结

- 香农码、霍夫曼码、费诺码均为一种无失真不等长最佳即时码
  - 香农码
    - *依概率降序排列-->从最大概率开始计算累加概率-->以累加概率的二进制小数的小数部分序列作为码字，序列长度为对应符号序列自信息量的上取整*
  - 霍夫曼码效率略高于香农码
    - *依概率降序排列-->对最小概率和次最小概率符号分别分配码元1和0-->将它们合并为新符号序列，概率为原序列概率之和-->重复上述步骤，直至出现概率为1的序列*
    - *新序列与相同概率的原序列的排列顺序不同可能导致不同的编码，但平均码长和编码效率相同*
  - 费诺码基于概率匹配的思想
    - *依概率降序排列-->按二进制进行分组，使每组概率尽可能相等-->给每个分组分配一个二进制码元-->对每个分组重复②③步骤，直到不可分为止*
    - *费诺码与霍夫曼码类似，均为先确定码字后确定码长，不具唯一性，但平均码长和编码效率相同*

本章作业

- (1)在MATLAB下实现香农编码方法，要求
  - 函数形式
    - *[codes, entropy, ratio]=shann_code(P)*
      - 输入参数为概率向量或矩阵;当P为矩阵时，每一行为概率向量(一种信源)；
      - 输出函数数量可变：

        当无参数时，通过fprintf输出编码，各信源的熵，以及编码效率

        当1个参数时，等效为[codes]=shann_code(P)，fprintf各信源的熵及编码效率

        当2个参数时，等效为[codes, entropy]=shann_code(P)，fprintf各信源的编码效率
  - 功能描述
    - *函数自动对P进行合法性检查，如果不满足概率向量条件，给出提示，并终止运行。*

本章作业

- (2)在MATLAB下实现霍夫曼编码方法，要求
  - 函数形式
    - *[codes, entropy, ratio]=Huffman_code(P)*
      - 输入参数为概率向量或矩阵;当P为矩阵时，每一行为概率向量(一种信源)；
      - 输出函数数量可变：

        当无参数时，通过fprintf输出编码，各信源的熵，以及编码效率

        当1个参数时，等效为[codes]= Huffman_code(P)，fprintf各信源的熵及编码效率

        当2个参数时，等效为[codes, entropy]= Huffman_code(P)，fprintf各信源的编码效率
  - 功能描述
    - *函数自动对P进行合法性检查，如果不满足概率向量条件，给出提示，并终止运行。*

本章作业

- (3)在MATLAB下实现费诺编码方法，要求
  - 函数形式
    - *[codes, entropy, ratio]=Fano_code(P)*
      - 输入参数为概率向量或矩阵;当P为矩阵时，每一行为概率向量(一种信源)；
      - 输出函数数量可变：
        - » 当无参数时，通过fprintf输出编码，各信源的熵，以及编码效率
        - » 当1个参数时，等效为[codes]= Fano_code(P)，fprintf各信源的熵及编码效率
        - » 当2个参数时，等效为[codes, entropy]= Fano_code(P)，fprintf各信源的编码效率
  - 功能描述
    - *函数自动对P进行合法性检查，如果不满足概率向量条件，给出提示，并终止运行。*