信息论基础

李　莹
liying2009@ecust.edu.cn

第三章：信源及信源熵

一：信源的分类及其数学模型

二：离散单符号信源

三：离散多符号信源

1. 预备知识
2. 离散平稳无记忆信源
3. 离散平稳有记忆信源
4. 马尔可夫信源
5. 信源的相关性和剩余度

4. 马尔可夫信源

(1) 定义

(2) 熵率

(3) 马尔可夫信源 ⟶ 马尔可夫链

(4) 马尔可夫链

4. 马尔可夫信源

(1) 定义

● 实际的有记忆信源，符号间的相关性可以追溯到很远，使得熵率的计算比较复杂。

●马尔可夫信源是一类相对简单的有记忆信源。信源在某一时刻发出某一符号的概率，除与该符号有关外，只与此前发出的有限个符号有关。

(2) 熵率

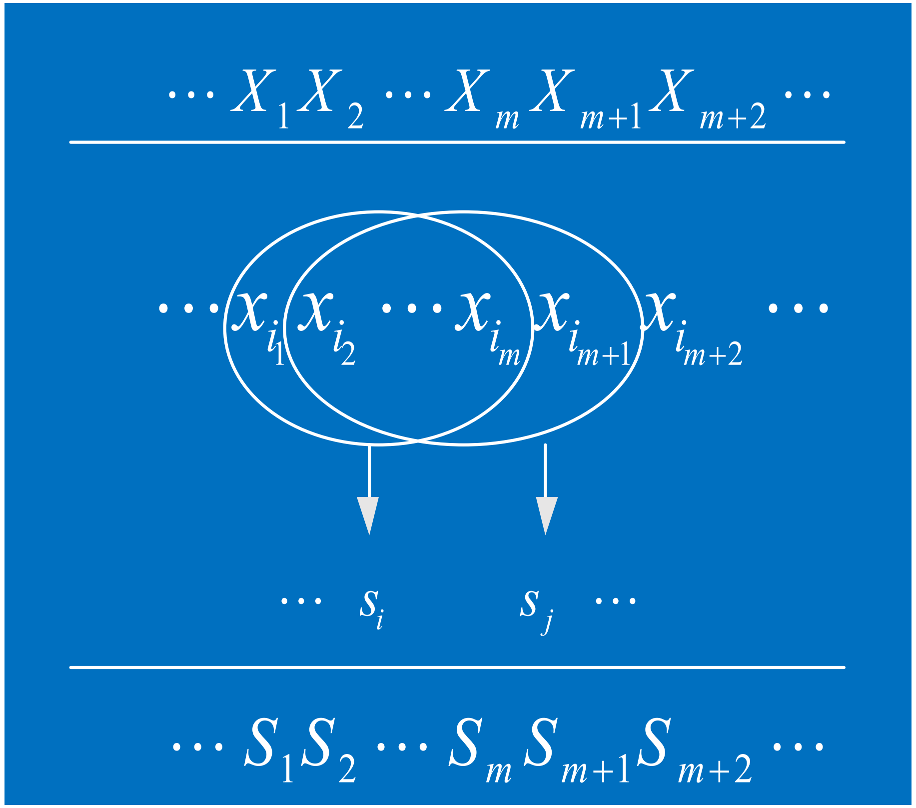
●**对于m阶马尔可夫信源，**

$$
\begin{aligned}
H_\infty &= \lim_{N\to\infty} H(X_N \mid X_1X_2\cdots X_{N-1}) \\
&= \lim_{N\to\infty} H(X_N \mid X_{N-m}X_{N-m+1}\cdots X_{N-1}) \\
&= H(X_{m+1} \mid X_1X_2\cdots X_m) \\
&= H_{m+1}
\end{aligned}
$$

●如何计算条件熵？

条件概率 $P(X_{m+1} \mid X_1X_2\cdots X_m)$ 通常是已知的，我们需要求解的是联合概率$P(X_1X_2\cdots X_m)$。

(3) 马尔可夫信源 ⟶ 马尔可夫链



$$
\begin{aligned}
& p(x_{i_{m+1}} \mid x_{i_1} x_{i_2} \cdots x_{i_m}) \\
&= p(x_{i_{m+1}} \mid s_i) \\
&= p(s_j \mid s_i)
\end{aligned}
$$

例3：

设一个二元一阶马尔可夫信源，信源符号集为 $X=\{0, 1\}$

输出符号的条件概率为

$$p(0|0)=0.25, p(1|0)=0.75, p(0|1)=0.5, p(1|1)=0.5,$$

用状态转移图来描述该信源。

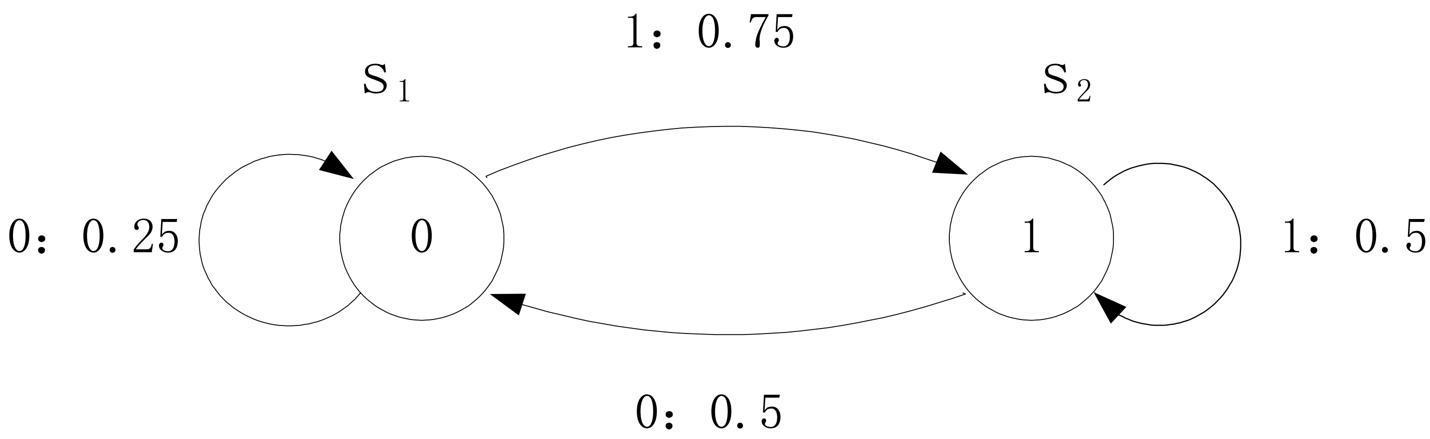


图3.2 一阶马尔可夫信源状态转移图

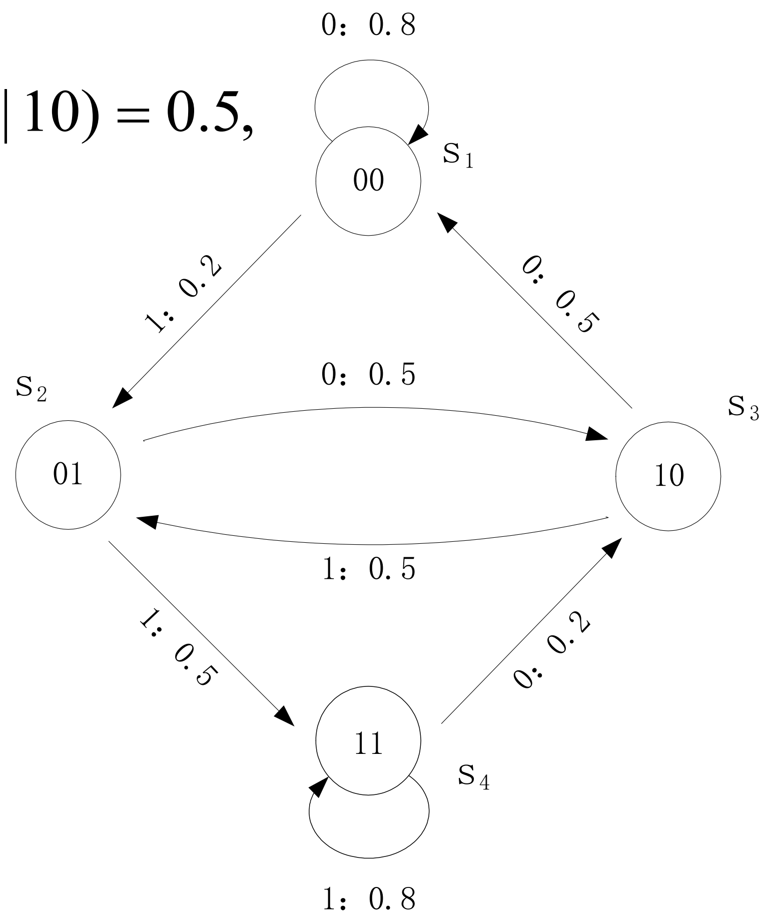
例4：

设一个二元二阶马尔可夫信源，信源符号集为$X=\{0,$ 输出符号的条件概率为

$$p(0\,|\,00)=p(1\,|\,11)=0.8,$$
$$p(0\,|\,01)=p(0\,|\,10)=p(1\,|\,01)=p(1\,|\,10)=0.5,$$
$$p(1\,|\,00)=p(0\,|\,11)=0.2$$

求该信源的状态转移图。

$$\mathbf{P}=\begin{bmatrix}0.8 & 0.2 & 0 & 0\\ 0 & 0 & 0.5 & 0.5\\ 0.5 & 0.5 & 0 & 0\\ 0 & 0 & 0.2 & 0.8\end{bmatrix}$$



●对于 m 阶马尔可夫信源，

$$
\begin{aligned}
H_{m+1} &= H(X_{m+1} \mid X_1 X_2 \cdots X_m) \\
&= E\left[-\log p(x_{i_{m+1}} \mid x_{i_1} x_{i_2} \cdots x_{i_m})\right] \\
&= E\left[-\log p(x_{i_{m+1}} \mid s_i)\right] \\
&= -\sum_{i=1}^{q^m} \sum_{i_{m+1}=1}^{q} p(s_i) p(x_{i_{m+1}} \mid s_i) \log p(x_{i_{m+1}} \mid s_i) \\
&\boxed{= \sum_i p(s_i) H(X \mid s_i)} \\
&= -\sum_i \sum_j p(s_i) p(s_j \mid s_i) \log p(s_j \mid s_i)
\end{aligned}
$$

(4) 马尔可夫链

- 有限状态马尔可夫链
- 状态转移概率
- 齐次马尔可夫链
- Chapman-Kolmogorov方程
- 马尔可夫链的平稳分布

- **马尔可夫链：**

设 $\{X_n, n\geq 1\}$ 为一随机序列，如果对所有 $n\geq 1$，有

$$P\left\{X_n = s_{i_n} \mid X_{n-1} = s_{i_{n-1}}, X_{n-2} = s_{i_{n-1}}, \cdots, X_1 = s_{i_1}\right\} = P\left\{X_n = s_{i_n} \mid X_{n-1} = s_{i_{n-1}}\right\}$$

则称 $\{X_n, n\geq 1\}$ 为马尔可夫链。

- 如果马尔可夫链的状态空间 $S=\{s_1, s_2, \cdots, s_J\}$ 有限，则被称为**有限状态马尔可夫链**；如果状态空间 $S=\{s_1, s_2, \cdots\}$ 是无穷集合，则被称为可数无穷状态的马尔可夫链。

- 状态转移概率（描述马氏链最重要的参数）：

$$p_{ij}(m,n) = P\left\{X_n = s_j \mid X_m = s_i\right\} = P\{X_n = j \mid X_m = i\}$$

- 状态转移概率的性质：

$$(i) \qquad 0 \le p_{ij}(m,n) \le 1$$

$$(ii) \qquad \sum_{j \in S} p_{ij}(m,n) = 1$$

- 一步转移概率： $p_{ij}(m,m+1) \Rightarrow p_{ij}(m)$
- k步转移概率： $p_{ij}^{(k)}(m) = P\{X_{m+k} = j \mid X_m = i\}$

● 齐次马尔可夫链：

如果马氏链状态转移概率与起始时刻无关，即对任意m，有

$p_{ij}(m) = P(X_{m+1} = s_j \mid X_m = s_i) = p_{ij}$，则称为时齐马尔可夫链或齐次马尔可夫链，也称为具有平稳转移概率的马尔可夫链。

● 齐次马氏链可以用转移概率矩阵或状态转移图来描述。

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1J} \\ p_{21} & p_{22} & \cdots & p_{2J} \\ \cdots & \cdots & \cdots & \cdots \\ p_{J1} & p_{J2} & \cdots & p_{JJ} \end{bmatrix}$$

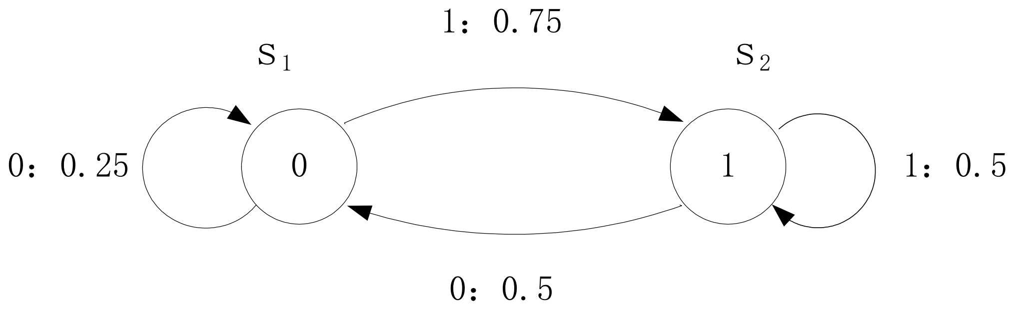


图3.2 一阶马尔可夫信源状态转移图

- **Chapman-Kolmogorov方程：**

$$p_{ij}^{(m+r)} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(r)}$$

或用矩阵表示为

$$\boldsymbol{P}^{(m+r)} = \boldsymbol{P}^{(m)} \boldsymbol{P}^{(r)}$$

$$P(X_k = s_j) = \sum_i P(X_k = s_j, X_0 = s_i) = \sum_i p_{0i} p_{ij}^{(k)}$$

$$\boldsymbol{p_{0i} = P(X_0 = s_i)}$$

$$\lim_{k \to \infty} p_{ij}^{(k)} = W_j$$

$$\lim_{k \to \infty} P(X_k = s_j) = \lim_{k \to \infty} \sum_i p_{0i} p_{ij}^{(k)} = \lim_{k \to \infty} \sum_i p_{0i} W_j = W_j$$

● **遍历性：**

若齐次马尔可夫链，$\forall i, j \in \{1, 2, ..., J\}$ 存在不依赖于 i 的极限

$$\lim_{n\to\infty} p_{ij}^{(n)} = W_j$$

且满足

$$W_j \geq 0, \qquad \sum_{i=1}^{J} W_j = 1$$

则称其具有遍历性（各态历经性）。W_j 为平稳分布。

$$W_j = \sum_{i=1}^{J} W_i p_{ij}$$

- **定理1：**

W_j是满足方程组$\boldsymbol{W}\ \boldsymbol{P}=\boldsymbol{W}$和 $W_j \geq 0,\ \sum_{i=1}^{J} W_j = 1$的唯一解。

- **定理2：**

设 $\boldsymbol{P}$ 为马氏链的状态转移矩阵，则该马氏链平稳分布存在的充要条件是，存在一个正整数 N，使矩阵 $\boldsymbol{P}^N$ 中的所有元素均大于零。

例5：求二阶马尔可夫信源的极限熵。

$$\mathbf{P}=\begin{bmatrix}0.8 & 0.2 & 0 & 0\\ 0 & 0 & 0.5 & 0.5\\ 0.5 & 0.5 & 0 & 0\\ 0 & 0 & 0.2 & 0.8\end{bmatrix}$$

解：

1）首先根据定理2检查该信源是否存在稳态分布：

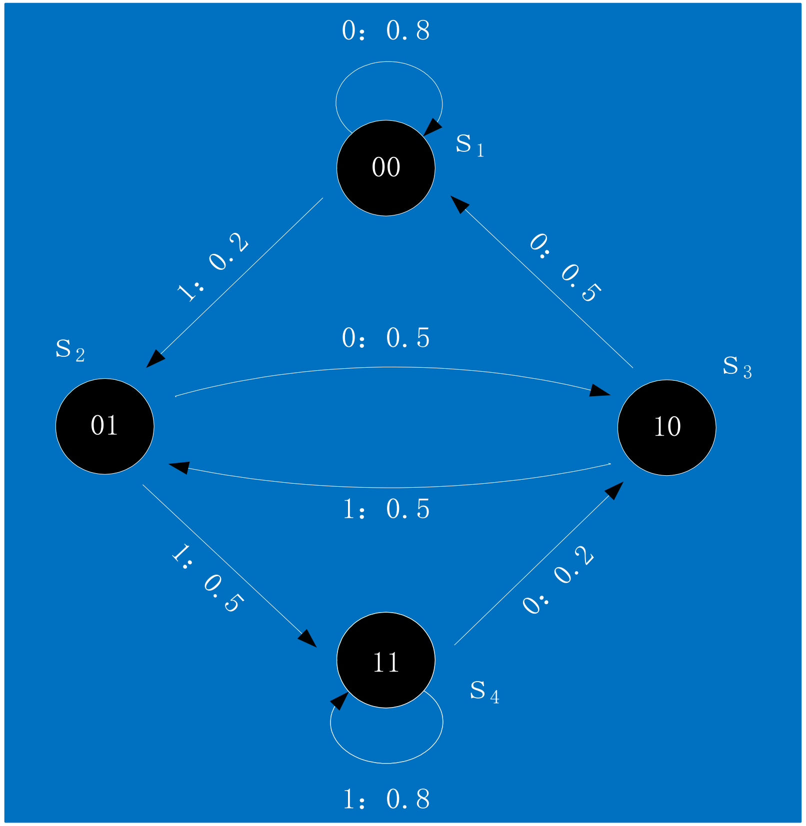
$$\mathbf{P}^2=\begin{bmatrix}* & * & 0 & 0\\ 0 & 0 & * & *\\ * & * & 0 & 0\\ 0 & 0 & * & *\end{bmatrix}\bullet\begin{bmatrix}* & * & 0 & 0\\ 0 & 0 & * & *\\ * & * & 0 & 0\\ 0 & 0 & * & *\end{bmatrix}=\begin{bmatrix}* & * & * & *\\ * & * & * & *\\ * & * & * & *\\ * & * & * & *\end{bmatrix}$$

➡ 所有元素均大于0，稳态分布存在。

2）设状态的平稳分布为 $\mathbf{W}=[W_1 \quad W_2 \quad W_3 \quad W_4]$，根据定理1有

$$\begin{cases}[W_1 \quad W_2 \quad W_3 \quad W_4]\begin{bmatrix}0.8 & 0.2 & 0 & 0\\ 0 & 0 & 0.5 & 0.5\\ 0.5 & 0.5 & 0 & 0\\ 0 & 0 & 0.2 & 0.8\end{bmatrix}=[W_1 \quad W_2 \quad W_3 \quad W_4]\\ W_1+W_2+W_3+W_4=1\end{cases}$$

$$\begin{cases}0.8W_1+0.5W_3=W_1\\ 0.2W_1+0.5W_3=W_2\\ 0.5W_2+0.2W_4=W_3\\ 0.5W_2+0.8W_4=W_4\\ W_1+W_2+W_3+W_4=1\end{cases}$$



$$\Rightarrow \begin{cases} W_1 = p(s_1) = 5/14 \\ W_2 = p(s_2) = 1/7 \\ W_3 = p(s_3) = 1/7 \\ W_4 = p(s_4) = 5/14 \end{cases}$$

3）求熵率：

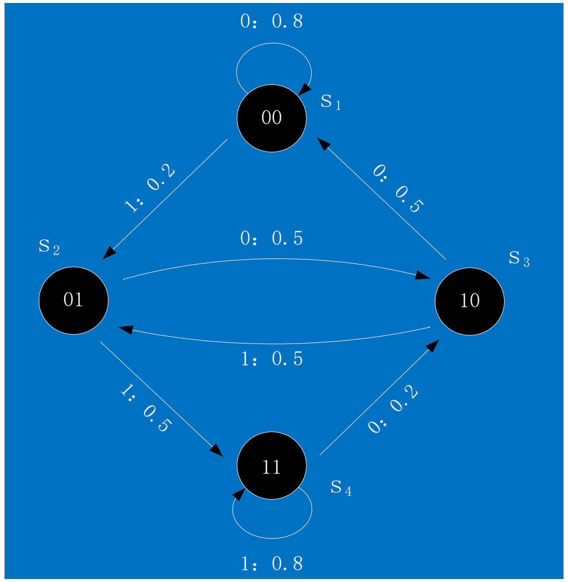
$$H_\infty = H_3 = \sum_i p(s_i) H(X \mid s_i)$$

$$= \frac{5}{14} H(0.8,\ \ 0.2) + \frac{1}{7} H(0.5,\ \ 0.5) + \frac{1}{7} H(0.5,\ \ 0.5) + \frac{5}{14} H(0.8,\ \ 0.2)$$

$$= 0.80 \qquad bit / symbol$$

<补充> 如何求信源发出的符号的极限概率？

$$p(x_i)=\sum_{j=1}^{J}p(s_j)p(x_i\mid s_j)\qquad i=1,\cdots,q$$



符号的平稳概率分布为：

$$p(0)=0.8p(s_1)+0.5p(s_2)+0.5p(s_3)+0.2p(s_4)=0.5$$

$$p(1)=0.2p(s_1)+0.5p(s_2)+0.5p(s_3)+0.8p(s_4)=0.5$$

如果不考虑符号间的相关性，则由符号的平稳概率分布可得信源熵 $H(X)$ = 1比特/符号，而考虑符号间的相关性后，该信源的熵率 0.80 比特/符号

例1：设有一马氏链，其状态转移矩阵为：

$$\mathbf{P}=\begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

问是否存在稳态分布。如果存在，求其极限熵。

$$\mathbf{P}^2 = \begin{bmatrix} 0 & 0 & * \\ * & * & * \\ * & * & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & * \\ * & * & * \\ * & * & 0 \end{bmatrix} = \begin{bmatrix} * & * & 0 \\ * & * & * \\ * & * & * \end{bmatrix} \qquad \mathbf{P}^3 = \begin{bmatrix} * & * & 0 \\ * & * & * \\ * & * & * \end{bmatrix} \begin{bmatrix} 0 & 0 & * \\ * & * & * \\ * & * & 0 \end{bmatrix} = \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}$$

$$\begin{cases} \begin{bmatrix} W_1 & W_2 & W_3 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} W_1 & W_2 & W_3 \end{bmatrix} \\ W_1 + W_2 + W_3 = 1 \end{cases} \Rightarrow \begin{cases} W_1 = 1/3 \\ W_2 = 2/7 \\ W_3 = 8/21 \end{cases}$$

$$H_\infty = \frac{1}{3} H(1) + \frac{2}{7} H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) + \frac{8}{21} H(\frac{1}{2}, \frac{1}{2})$$

5. 信源的相关性和剩余度

- **信源的相关性就是信源符号间的依赖程度。**
- **对于不同平稳信源可以分别计算它的熵（设信源有q个符号）：**

$$H_0 = \log q$$ （独立等概信源）

$$H_1 = H(X_1)$$ （无记忆信源）

$$H_2 = H(X_2 \mid X_1)$$ （一阶马尔可夫信源）

$$H_{m+1} = H(X_{m+1} \mid X_1 X_2 \cdots X_m)$$ （m阶马尔可夫信源）

$$H_\infty = \lim_{N\to\infty} H(X_N \mid X_1 X_2 \cdots X_{N-1})$$ （记忆长度无限的信源）

- 对同一信源，采用不同的模型（假定相关程度不同），计算得到的熵的关系为

$$H_0 \geq H_1 \geq H_2 \geq \cdots \geq H_{m+1} \geq \cdots \geq H_\infty$$

- 结论：

符号间相关性越大，熵越小。

- **定义1：熵的相对率**

$$\eta = \frac{H_\infty}{H_0}$$

- **定义2：信源的剩余度**

$$\gamma = 1 - \eta = 1 - \frac{H_\infty}{H_0} = 1 - \frac{H_\infty}{\log q}$$

英文信源：

H_0=4.76

H_1=4.03

H_2=3.32

H_3=3.1

H_5=1.65

H_∞=1.4

$$\eta = \frac{H_\infty}{H_0} = \frac{1.4}{4.76} = 0.29$$

$$\gamma = 1 - \eta = 0.71$$

5. 信源的相关性和剩余度

	H_0	H_1	H_2	H_3	…	H_∞	η	γ
英文	4.7	4.03	3.32	3.1		1.4	0.29	0.71
法文	4.7					3	0.63	0.37
德文	4.7					1.08	0.23	0.77
西班牙文	4.7					1.97	0.42	0.58
中文（按8千汉字计算）	≈ 13	9.41	8.1	7.7		4.1	0.315	0.685

5. 信源的相关性和剩余度

例 3.7： 计算汉字的剩余度。假设常用汉字约为 10000个，其中 140 个汉字出现的概率占 50%， 625个汉字（含140个）出现的概率占 85%，2400个汉字（含625个）出现的概率占 99.7%，其余 7600个汉字出现的概率占 0.3%，不考虑符号间的相关性，只考虑它的概率分布，在这一级近似下计算汉字的剩余度。

解：为了计算方便，假设每类中汉字出现是等概的，得表

类别	汉字个数	所占概率	每个汉字的概率
1	140	0.5	0.5/140
2	625-140=485	0.85-0.5=0.35	0.35/485
3	2400-625=1775	0.997-0.85=0.147	0.147/1775
4	7600	0.003	0.003/7600

$H_1=H(X)=9.773$ bit/汉字

$H_0=13.288$ bit/汉字

$$\gamma = 1 - \frac{H_1}{H_0} = 0.264$$