信息论与编码

Information Theory and Coding

西南交通大学

电子信息工程专业

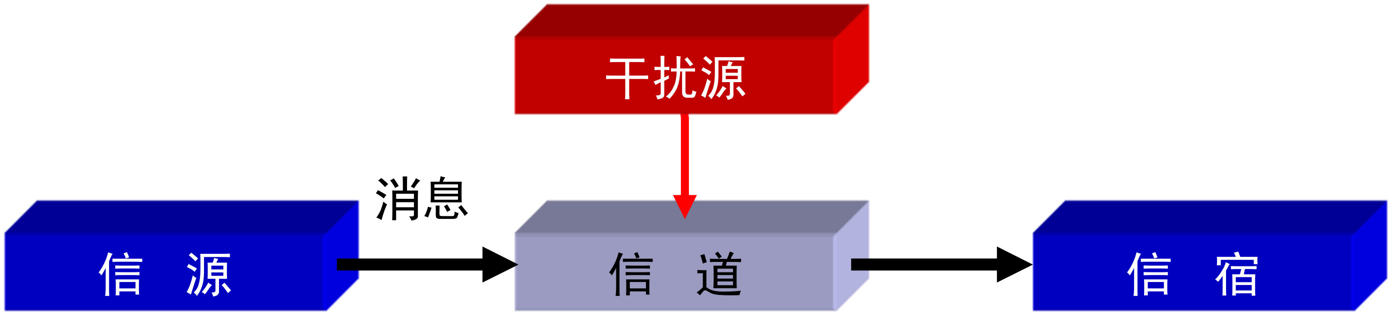
2020

第2章 信源熵

- 主要内容
  - 信源的描述：离散和连续信源
  - 单符号离散信源的信息量、信息熵
  - 离散序列信源的信息熵
  - 互信息量
  - 冗余度

第2章 信源熵

2-1 信源的描述（数学模型）和分类



消息传递的特点：**不确定性**

1）收信人无法判断发送人将要描述何种事物的运动状态的消息；

2）收信人无法判断所描述的事物处于何种状态；

3）由于干扰的存在，收信人无法判断收到的消息的正确可靠程度；

2-1 信源的描述（数学模型）和分类

- 根据信源发出的消息在时间上和幅度上的分布情况，信源可分为两类：
  - **离散信源**：时间**和**幅度都**离散**，如数字、文字、数字信号、数据等
  - **连续信源**：时间**或**幅度**连续分布**，如语音、图像、图形、模拟信号等

2-1 信源的描述（数学模型）和分类

离散信源：信源可能输出的消息数是有限的和可数的，而且每次只输出其中的一个消息。

例题2.1：掷骰子试验，掷一颗质地均匀的骰子，研究其朝上面的点数。

样本空间

离散信源数学模型(信源空间)

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}$$

概率分布

$$\sum_{i=1}^{6} P(X = a_i) = 1, \quad P(X = a_i) \geq 0$$

2-1 信源的描述（数学模型）和分类

连续信源：发出在时间或幅值上是连续分布的连续消息的信源。

例题2.2：语音信号、热噪声信号等某时间的连续取值数据。控制系统中有关电压、温度、压力等测得的连续数据。

连续信源数学模型（信源空间）

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} (a,b) \\ p(x) \end{bmatrix} \text{或} \begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} (-\infty,\infty) \\ p(x) \end{bmatrix}$$

$$\int_a^b p(x)dx = 1 \text{或} \int_{-\infty}^{\infty} p(x)dx = 1$$

X的取值范围

概率密度函数

2-1 信源的描述（数学模型）和分类

- 离散信源
  - 离散无记忆信源
    - 发出单个符号的离散无记忆信源
    - 发出符号序列的离散无记忆信源
  - 离散有记忆信源
    - 发出符号序列的离散有记忆信源
    - 发出符号序列的马尔科夫信源

离散无记忆信源：所发出的各个符号是相互独立的，发出符号序列中的各个符号之间没有统计关联性，各个符号的出现概率是其自身的先验概率。

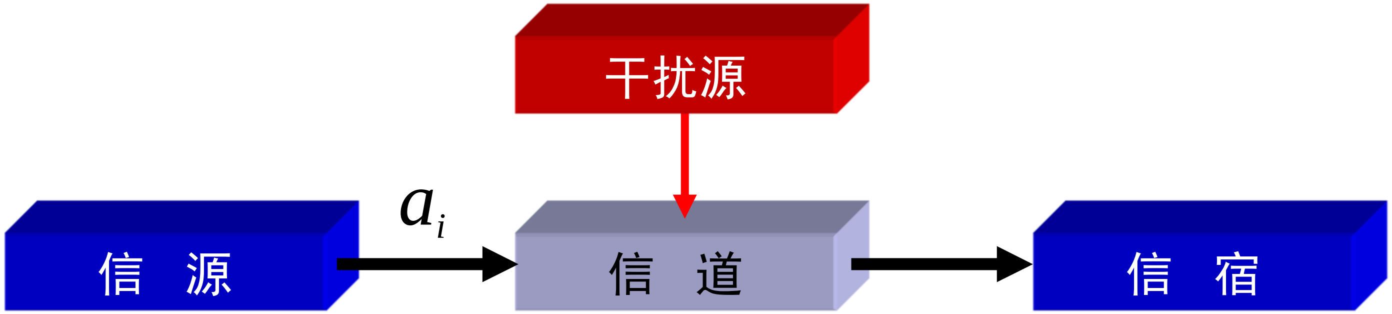
离散有记忆信源：所发出的各个符号的概率是有关联的。

2-1 信源的描述（数学模型）和分类

- 离散信源
  - 离散无记忆信源
    - 发出单个符号的离散无记忆信源
    - 发出符号序列的离散无记忆信源
  - 离散有记忆信源
    - 发出符号序列的离散有记忆信源
    - 发出符号序列的马尔科夫信源

- **发出单个符号的信源**：信源每次只发出一个符号代表一个消息；
- **发出符号序列的信源**：信源每次发出一组含两个以上符号的符号序列代表一个消息；

2-2 单符号离散信源



若某信源可以发出r种不同的符号，$a_1, a_2, \cdots, a_r$，相应的先验概率分别为$p(a_1), p(a_2), \cdots, p(a_r)$.可以用随机变量$X$表示这个信源，其信源空间表示为：

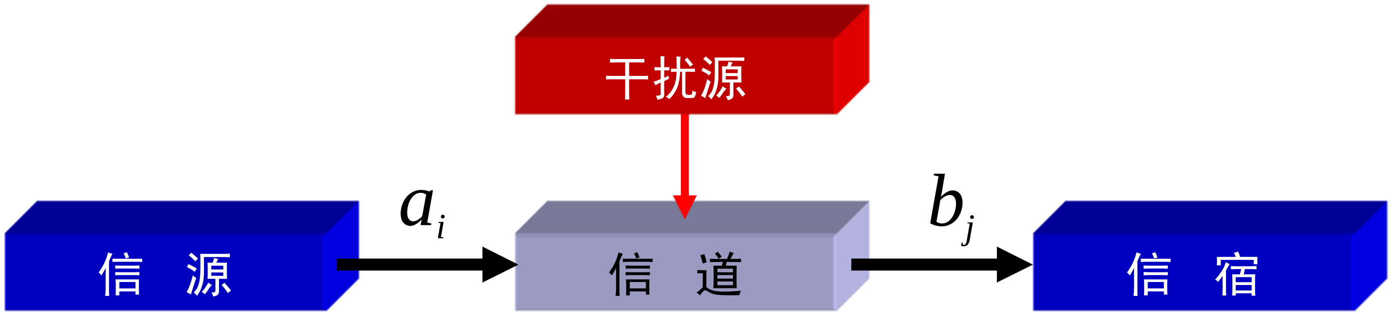
$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_r \\ p(a_1) & p(a_2) & \cdots & p(a_r) \end{bmatrix}$$

其中$0 \le p(a_i) \le 1, \quad (i = 1, 2, \cdots, r)$

$$\sum_{i=1}^{r} p(a_i) = 1$$

2-2 单符号离散信源

2.2.1 自信息量



$I(a_i;b_j)=$[收到b_j前，收信者对信源发出 a_i的不确定性]$-$
[收到b_j后，收信者对信源发出 a_i仍然存在的不确定性]
$=$收信者收到 b_j前后，对信源发出 a_i存在的不确定性的消除

2-2 单符号离散信源

2.2.1 自信息量

[收到b_j后，收信者对信源发出a_i仍然存在的不确定性]$=0$

$I(a_i;a_i)=$[收到a_i前，收信者对信源发出a_i的不确定性]

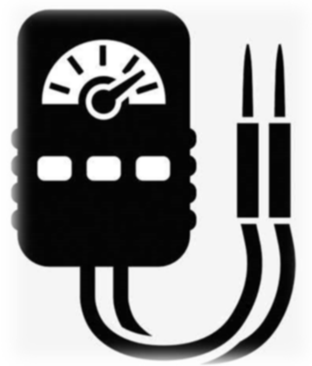
用$I(x_i)$表示信源符号的自信息量：

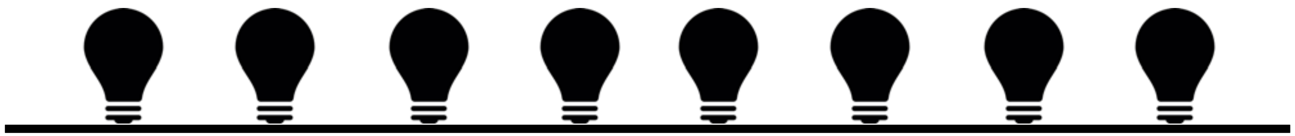
$I(x_i)=$[收到a_i前，收信者对信源发出a_i的不确定性]

2-2 单符号离散信源

思考：

在一条电线上串联了8个小灯泡，每个灯泡坏掉的概率是等概的。假设这条电路中有一个灯泡是坏的，导致电路不能工作。现在利用一台万用表去查找这个坏掉的灯泡，当找到这个坏掉的灯泡时可以获得多少信息量？



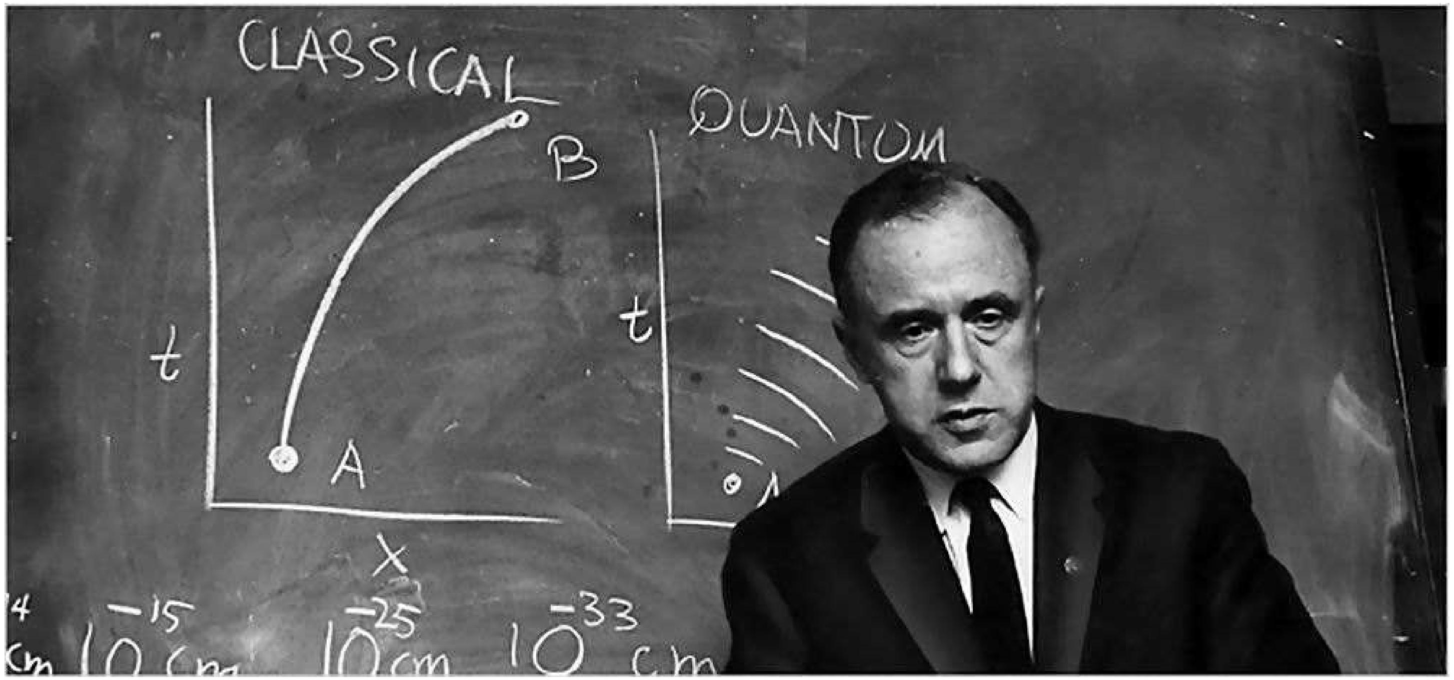


回忆第一章内容：

- ◆第一个重要概念：信息传递的通道中传送的是随机变量的值
- ◆ 第二个重要概念：事件发生的概率越小，此事件含有的信息量就越大。
- ◆第三个重要概念：消息随机变量的随机性越大，此消息随机变量含有的信息量就越大。

2-2 单符号离散信源

2.2.1 自信息量



It from Bit. ——John Archibald Wheeler

万物源自比特 —— 约翰 阿奇博尔德 惠勒

任何事物——任何粒子、任何力场，甚至时空连续本身都源自于信息。

2-2 单符号离散信源

2.2.1 自信息量

自信息量一定是信源发出符号的先验概率的函数：

并且满足：$I(x_i)=f[p(x_i)],(i=1,2,\cdots,r)$

[1] 函数$I(x_i)=f[p(x_i)](i=1,2,\cdots,r)$，是先验概率$p(x_i)$的单调减函数

[2] 信源符号x_i的先验概率$p(x_i)=0$,那么$I(x_i)=f[p(x_i)]=\infty$

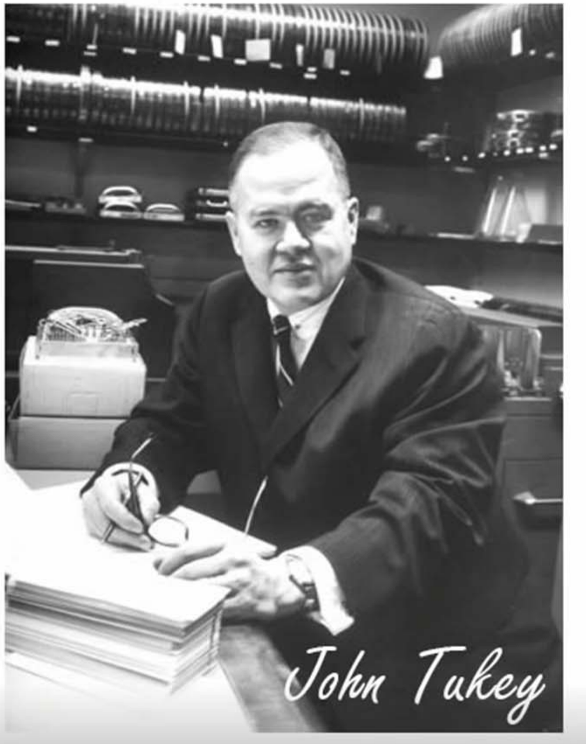
[3] 信源符号x_i的先验概率$p(x_i)=1$,那么$I(x_i)=f[p(x_i)]=0$

[4] 两个统计独立的信源X,Y，$I(x_iy_j)=f[p(x_iy_j)]=I(x_i)+I(y_j)$

2-2 单符号离散信源

Extract from Shannon's paper:

"The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly 'bits', a word suggested by. J. W. Tukey. A device with two stable positions, such as a relay or a flip-flop circuit, can store one bit of information. N such devices can store N bits, since the total number of possible states is 2N and log(2N) = N"



2-2 单符号离散信源

2.2.1 自信息量

定义：设信源输出r个消息$\{a_1, a_2, \cdots, a_r\}$，每个消息的出现概率分别为$\{p(a_1), p(a_2), \cdots, p(a_r)\}$，每个消息的自信息量为：

$$I(x_i) = \log_2\left(\frac{1}{p(x_i)}\right) = -\log_2\left[p(x_i)\right] \text{，其中} i = 1, 2, \cdots, r$$

对数的底	公式	单位	
2	$I(x_i) = -\log_2\left[p(x_i)\right]$	比特	Bit (Binary Digit)
e	$I(x_i) = -\log_e\left[p(x_i)\right]$	奈特	Nat (Nature Digit)
10	$I(x_i) = -\log_{10}\left[p(x_i)\right]$	笛特（哈特）	Det (Decimal Digit)

信息量单位换算：

$$1\text{bit} = \log_{10} 2 \approx 0.301\text{Det} \qquad 1\text{Det} = \log_2 10 \approx 3.322\text{bit}$$

$$1\text{bit} = \log_e 2 \approx 0.693\text{Nat} \qquad 1\text{Nat} = \log_2 e \approx 1.443\text{bit}$$

2-2 单符号离散信源

Extract from Shannon's paper:

"The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly 'bits', a word suggested by. J. W. Tukey. A device with two stable positions, such as a relay or a flip-flop circuit, can store one bit of information. N such devices can store N bits, since the total number of possible states is 2N and log(2N) = N"

一个具有两个稳态的设备，如继电器或双稳态触发器电路，可以储存1比特信息。N个此类设备就可以储存N比特，因为可能状态的总数为2^N,而$\log_2 2^N=N$

2-2 单符号离散信源

例题2.3： 在一个箱子中，有属性相同的红、黄、蓝三种颜色的彩球，共36个, 其中红球18个，黄球12个，蓝球6个，任取一球将作为实验结果。如果事件A,B,C分别表示摸出的是红球、黄球和蓝球，试计算事件A,B,C发生后所提供的信息量。

2-2 单符号离散信源

解：根据题意，该信源的信源空间为：

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} A & B & C \\ 18/36 & 12/36 & 6/36 \end{bmatrix}$$

事件A发生时获得的信息量为：

$$I(A) = -\log p(A) = -\log(18/36) = 1 \quad (\text{bit})$$

事件B发生时获得的信息量为：

$$I(B) = -\log p(B) = -\log(12/36) = 1.585 \quad (\text{bit})$$

事件C发生时获得的信息量为：

$$I(C) = -\log p(C) = -\log(6/36) = 2.585 \quad (\text{bit})$$

2-2 单符号离散信源

● 联合自信息量

➢ 如果两符号x_i，y_j 同时出现，可用联合概率 $p(x_i y_j)$ 来表示，此时的联合自信息量为：

$$I(x_i y_j) = -\log p(x_i y_j)$$

● 条件自信息量

➢ 如果符号x_i, y_j不相互独立，有一定的联系，则可用条件概率 $p(x_i \mid y_j)$（在符号y_j 出现的条件下，符号x_i发生的条件概率）来表示，这样，条件自信息量可定义为：

$$I(x_i \mid y_j) = -\log p(x_i \mid y_j)$$

2-2 单符号离散信源

例题 2.4:

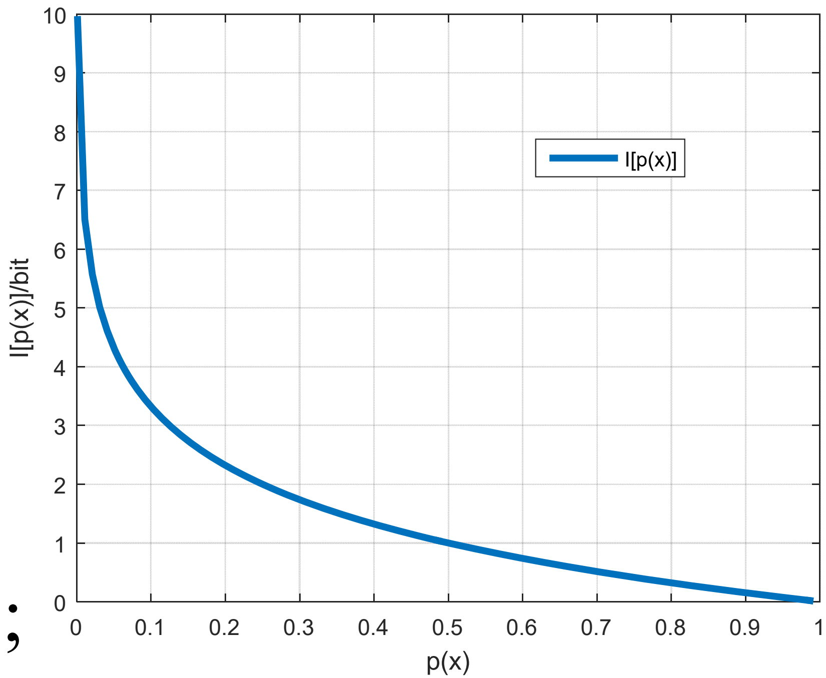
(1) 英文字母中“A”出现概率0.064，“C”出现概率为0.022，计算两个字母出现的自信息量；

(2) 如果字母前后出现相互独立，计算“AC”和“CA”出现的自信息量；

(3) 如果字母前后出现非相互独立，“A”出现后立即出现“C”的概率为0.04，计算“AC”出现的自信息量。

2-2 单符号离散信源

$$I(x)=\log_2\left(\frac{1}{p(x)}\right)$$

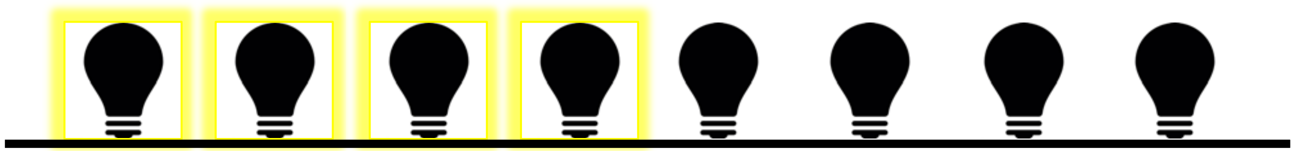


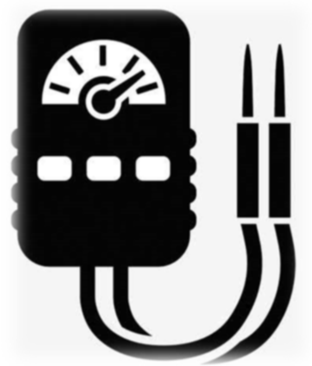
- **自信息量的性质**
  - (1) $p(x_i)=1,\ I(x_i)=0$;
  - (2) $p(x_i)=0,\ I(x_i)=\infty$;
  - (3) 非负性;
  - (4) 单调递减性：若$p(x_1)<p(x_2)$，则 $I(x_1)>I(x_2)$
  - (5) 可加性：若两符号x_i，y_j同时出现，且相互独立时，则有 $I(x_i,y_j)=I(x_i)+I(y_j)$

2-2 单符号离散信源

思考:

在一条电线上串联了8个小灯泡，每个灯泡坏掉的概率是等概的。假设这条电路中有一个灯泡是坏的，导致电路不能工作。现在利用一台万用表去查找这个坏掉的灯泡，当找到这个坏掉的灯泡时可以获得多少信息量?





2-2 单符号离散信源

2.2.2 信源的信息熵

熵（entropy）：指的是体系的混乱的程度，它在控制论、概率论、数论、天体物理、生命科学等领域都有重要应用，在不同的学科中也有引申出的更为具体的定义，是各领域十分重要的参量。熵由鲁道夫·克劳修斯（Rudolf Clausius）提出，并应用在热力学中。

1822年1月2日 - 1888年8月24日），德国物理学家和数学家，热力学的主要奠基人之一

热力学第一定律:

能量守恒定律。 (宇宙中能量守恒)

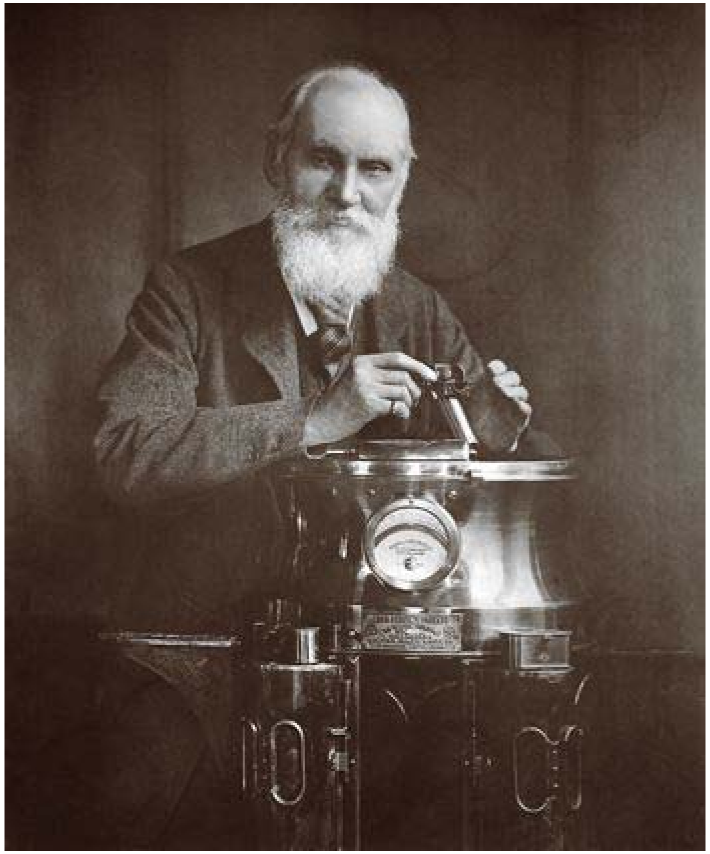
热力学第二定律(3种描述):

- ✓ **克劳修斯表述为热量可以自发地从温度高的物体传递到温度低的物体，但不可能自发地从温度低的物体传递到温度高的物体；**
- ✓ **开尔文-普朗克表述为不可能从单一热源吸取热量，并将这热量完全变为功，而不产生其他影响。**
- ✓ **以及熵增表述：孤立系统的熵永不减小。**

(宇宙中熵增加)

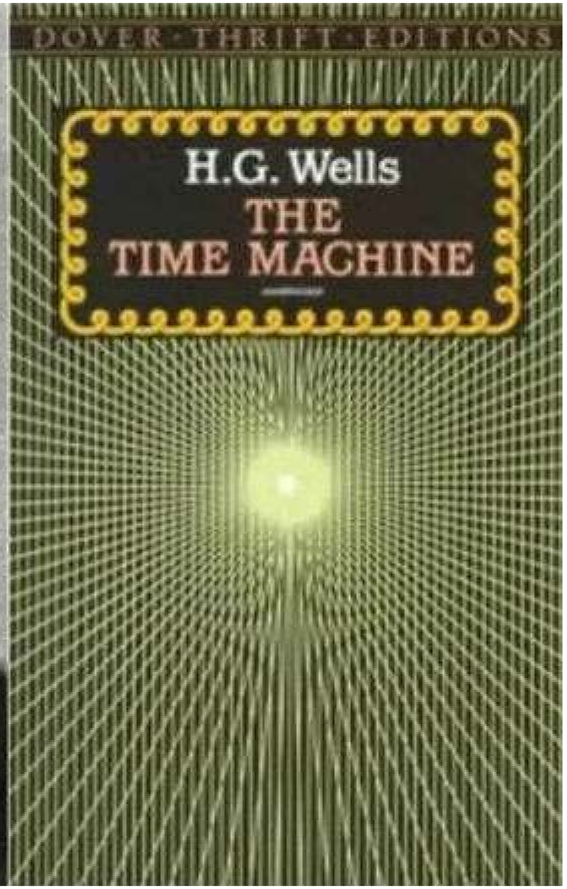
热力学第三定律:

通常表述为绝对零度时，所有纯物质的完美晶体的熵值为零，或者绝对零度(T=0K)不可达到。



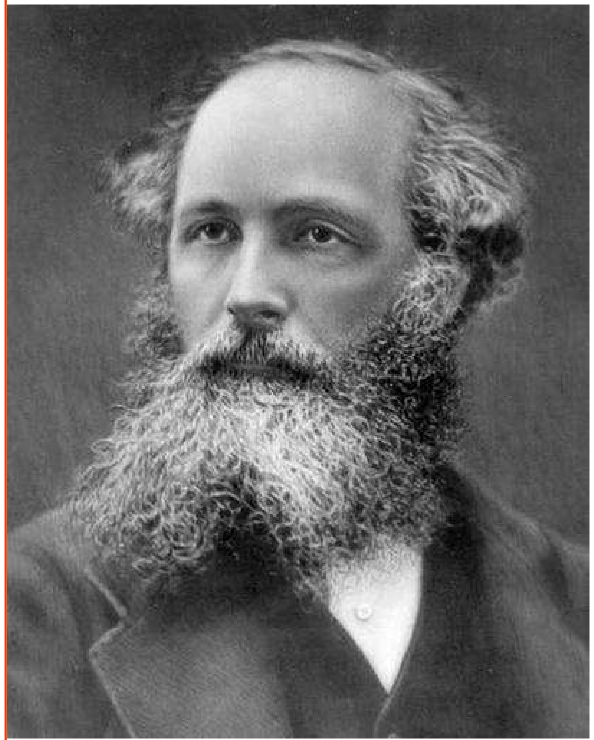
开尔文勋爵（威廉·汤姆森），描述了一番暗淡的前景，使得热力学第二定律吸引了公众的想象力。
他在1862年宣称：“虽然机械能不灭，但一个普遍趋势是机械能会**耗散**，导致在整个物质宇宙内，运动会停止，势能会耗竭，而热能则会逐渐增加和扩散。这样最终整个宇宙会归于一个静止和死寂的状态。”

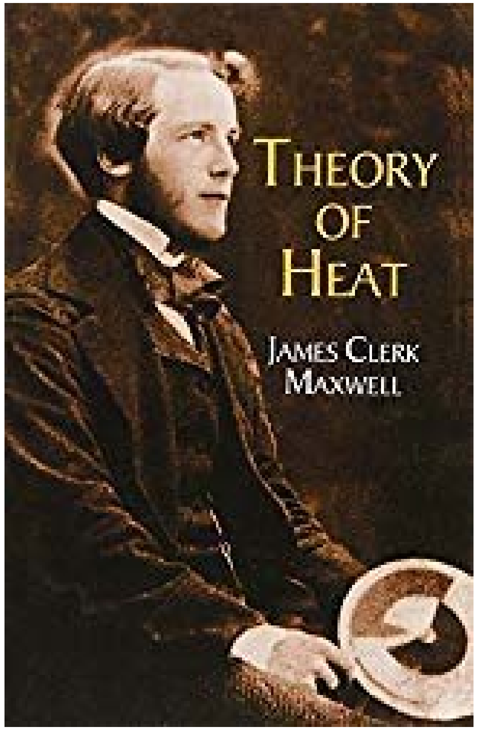
威廉 汤姆逊（William Thomson 1824年—1907年），受勋后的名为开尔文勋爵（Lord Kelvin），英国著名数学物理学家、工程师，在热力学、电磁学等领域有重大贡献，领导建立了世界上第一条大西洋海底电缆。



H.G 威尔斯 的科幻小说《时间机器》里，熵主宰了宇宙的最终命运：生命衰灭，太阳死亡，“令人厌恶的毁灭气息笼罩着整个世界”。

赫伯特·乔治·威尔斯（Herbert George Wells，1866年9月21日—1946年8月13日），又译作赫伯特·乔治·韦尔斯，英国著名小说家，新闻记者、政治家、社会学家和历史学家。

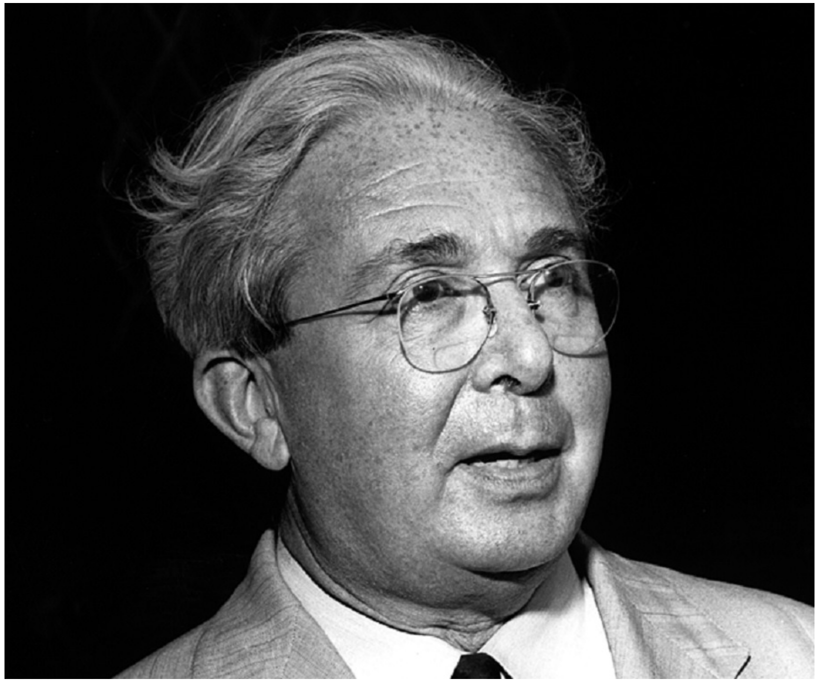




1871年，在《热理论》一书的末章《热力学第二定律的限制》中，麦克斯韦设计了一个假想的存在物，即著名的“麦克斯韦妖”（Maxwell's demon）

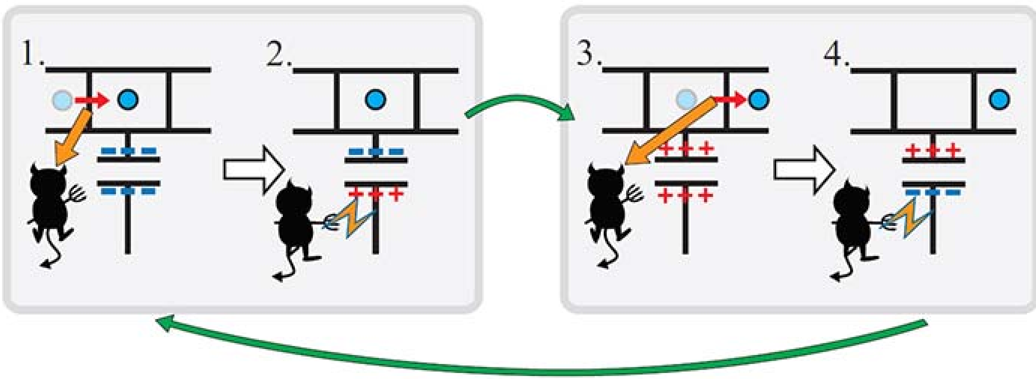
詹姆斯·克拉克·麦克斯韦（**James Clerk Maxwell，1831∽1879**），出生于苏格兰爱丁堡，英国物理学家、数学家。经典电动力学的创始人，统计物理学的奠基人之一。

1847年进入爱丁堡大学学习数学和物理，毕业于剑桥大学。他成年时期的大部分时光是在大学里当教授，最后是在剑桥大学任教。**1873**年出版的《论电和磁》，也被尊为继牛顿《自然哲学的数学原理》之后的一部最重要的物理学经典。麦克斯韦被普遍认为是对物理学最有影响力的物理学家之一。没有电磁学就没有现代电工学，也就不可能有现代文明。



1929年：
可以设想，这个小妖怪每次处理一个粒子，就是做了一次信息与能量的转换。
精确计算每一次度量和记忆，这种转换是可以计算的，每次获取一单位的信息总带来一定的熵增加+klog2单位。在每一个循环结束的时候，那么记忆不得不被清除。

利奥·西拉德 （Leo Szilard，1898年2月11日—1964年5月30日），出生于匈牙利的布达佩斯，美籍匈牙利核物理学家、美国芝加哥大学教授，曾参与美国曼哈顿计划，第二次世界大战后积极倡导核能的和平利用、反对使用核武器。



1952年，兰道尔加入IBM公司工作，1961年，兰道尔在《IBM研究通讯》上发表了一篇令他青史留名的论文，这篇论文的题目是《不可逆性与计算过程中的热量产生问题》。在这篇论文中，兰道尔指出了一件以前从来没人发现的事情：经典计算机要擦除一个经典比特的信息，其所消耗的最小能量是：

$$k_B \ln 2$$

其中k_B 是玻尔兹曼常数$k_B = 1.380649 \times 10^{-23}$ J/K 。

Irreversibility and Heat Generation in the Computing Process

Abstract: It is argued that computing machines inevitably involve devices which perform logical functions that do not have a single-valued inverse. This logical irreversibility is associated with physical irreversibility and requires a minimal heat generation, per machine cycle, typically of the order of *kT* for each irreversible function. This dissipation serves the purpose of standardizing signals and making them independent of their exact logical history. Two simple, but representative, models of bistable devices are subjected to a more detailed analysis of switching kinetics to yield the relationship between speed and energy dissipation, and to estimate the effects of errors induced by thermal fluctuations.

How many Bits in the Universe?

How many bits of information can be contained by the whole universe?

2-2 单符号离散信源

2.2.2 信源的信息熵

自信息函数的不足：

[1] 信源发符号x_i不是确定事件，是概率为$p(x_i)$的随机事件，相应的自信息函数$I(x_i)$也是一个以$p(x_i)$为概率的随机性的量

[2] $I(x_i)$只能表示信源发某一个特定的符号x_i所提供的信息量，不能成为信源总体的信息度量

2-2 单符号离散信源

2.2.2 信源的信息熵

定义：

将自信息量的数学期望称为信源的平均自信息量$H(X)$

$$H(X)=E\left[I(X)\right]=\sum_{i=1}^{r}p(x_i)I(x_i)$$

$$=-\sum_{i=1}^{r}p(x_i)\log p(x_i) \quad 比特/符号$$

将$H(X)$称为信源X的信息熵

2-2 单符号离散信源

2.2.2 信源的信息熵

例题2.5: 箱子内有100个属性相同的球，其中80个红球，20个白球。随意取一个球，试计算所提供的平均信息量。

解：根据题意，该信源的信源空间为：

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ p(x_1) & p(x_2) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ 80/100 & 20/100 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ 4/5 & 1/5 \end{bmatrix}$$

平均信息量为：

$$E(I(X)) = -\sum_{i=1}^{2} p(x_i)\log p(x_i)) = 0.7219 \ \ (\text{bit/symbol})$$

2-2 单符号离散信源

2.2.2 信源的信息熵

例题2.6： 二元信源是离散信源的一个特例，该信源X的输出符号只有两个，设置为0，1，输出符号为0，1的概率分别为p、q，并且p+q=1. 求二元信源的信息熵.

解：根据题意，该信源的信源空间为：

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ p(x_1) & p(x_2) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & q \end{bmatrix}$$

信源的信息熵为：

$$\begin{aligned} H(X) &= -p\log p - q\log q \\ &= -p\log p - (1-p)\log(1-p) = H(p) \end{aligned}$$

2-2 单符号离散信源

分析：信源的信息熵H(X)是概率p的函数，通常用H(p)表示，p取值为[0,1]区间.

思考绘制H(p)函数曲线

提示：

对于任意小的正数ε:

$$\ln\varepsilon = 2\left\{\frac{\varepsilon-1}{\varepsilon+1}+\frac{1}{3}\left(\frac{\varepsilon-1}{\varepsilon+1}\right)^3+\frac{1}{5}\left(\frac{\varepsilon-1}{\varepsilon+1}\right)^5+\dots\right\}$$

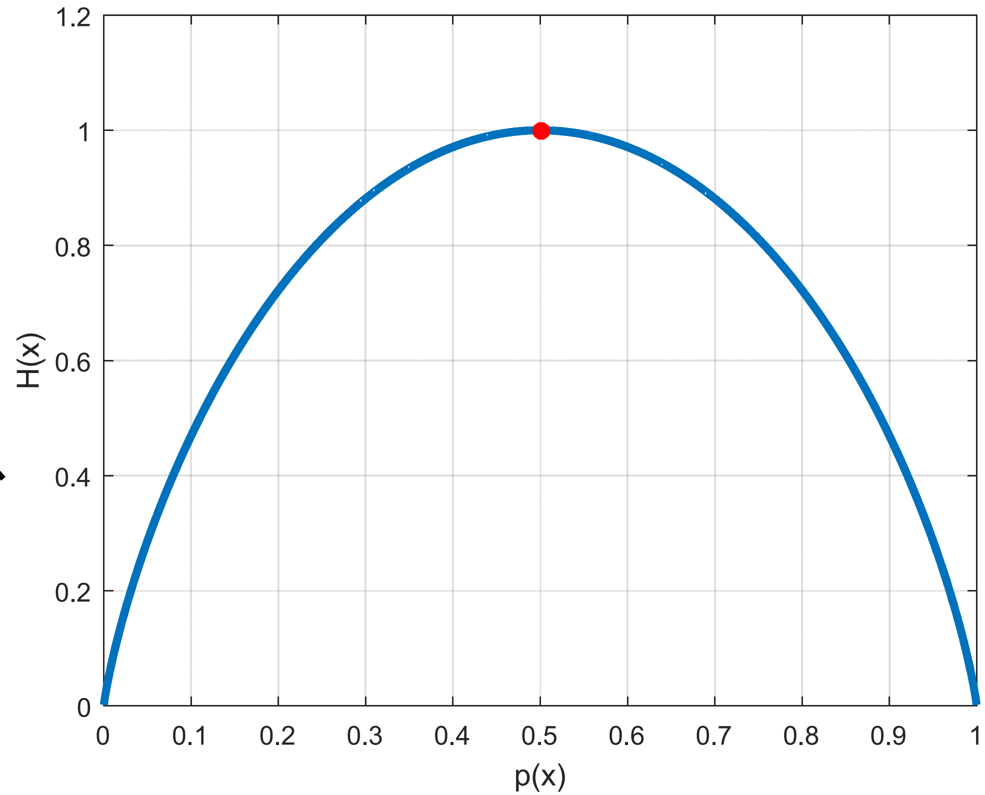
当$\varepsilon\to 0$，则$\lim\limits_{\varepsilon\to 0}\left(\varepsilon\ln\varepsilon\right)=0$

2-2 单符号离散信源

分析： 信源的信息熵H(X)是概率p的函数，通常用H(p)表示，p取值为[0,1]区间.

H(p)函数曲线如右图所示：

- 若二元信源的输出符号是确定的，即p=1或q=0，则该信源不提供任何信息；
- 若二元信源符号0或1以等概率发生时，信源熵达到极大值，等于1 bit信息量。



2-2 单符号离散信源

2.2.2 信源的信息熵

信息熵具有以下三种物理含义：

[1] $H(X)$表示信源输出前，信源的平均不确定性；

[2] $H(X)$表示信源输出后，每个消息所提供的平均信息量；

[3] $H(X)$可以表示随机变量X的随机性；

回顾：

◆第一个重要概念：信息传递的通道中传送的是随机变量的值

◆ 第二个重要概念：事件发生的概率越小，此事件含有的信息量就越大。

◆第三个重要概念：消息随机变量的随机性越大，此消息随机变量含有的信息量就越大。

2-2 单符号离散信源

2.2.3 信源的信息熵的性质

定义：熵函数$H(\mathbf{P})$是概率矢量$\mathbf{P}$的一种特殊的函数，

其函数形式为：$H(\mathbf{P})=H(p_1,p_2,\cdots,p_r)=-\sum_{i=1}^{r}p_i\log p_i$

信源X的信息熵具有以下三种表达方式：

[1] 指明是信源X的信息熵时（即离散信源X）可以用$H(X)$表示

[2] 要表明概率分布是$(p_1,p_2,\cdots,p_r)$的信源的信息熵时，可以采用熵函数的形式$H(p_1,p_2,\cdots,p_r)$

[3] 要表明信源的r个概率分量构成的概率矢量为$\mathbf{P}=(p_1,p_2,\cdots,p_r)$时，信源的信息熵可以采用熵函数的形式$H(\mathbf{P})$

2-2 单符号离散信源

性质1：对称性

$$H(X)=H(p_1,p_2,\cdots,p_r)=H(p_2,p_1,\cdots,p_r)=\cdots$$

性质2：确定性

$$\begin{aligned}H(X)&=H(0,0,\cdots,1,\cdots,0)\\&=[0\log 0+0\log 0+\cdots+1\log 1+0\log 0]\\&=0\end{aligned}$$

性质3：非负性（离散信源）

$$H(X)=H(p_1,p_2,\cdots,p_r)\geq 0$$

性质4：扩展性

$$\lim_{\varepsilon\to 0}H_{q+1}(p_1,p_2,\cdots,p_q-\varepsilon,\varepsilon)=H_q(p_1,p_2,\cdots,p_q)$$

2-2 单符号离散信源

性质5：连续性

$$\lim_{\varepsilon \to 0} H(p_1, p_2, \cdots, p_{q-1} - \varepsilon, p_q + \varepsilon) = H_q(p_1, p_2, \cdots, p_q)$$

性质6：递推性

$$\begin{aligned}&H(p_1, p_2, \cdots, p_{n-1}, q_1, q_2, \cdots, q_m)\\&= H(p_1, p_2, \cdots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \cdots, \frac{q_m}{p_n}\right)\end{aligned}$$

2-2 单符号离散信源

性质7：香农辅助定理

对于任意r维概率矢量$\mathbf{P}=(p_1,p_2,\cdots,p_r)$和

$\mathbf{Q}=(q_1,q_2,\cdots,q_r)$

有下列不等式成立：

$$H(p_1,p_2,\cdots,p_r)=-\sum_{i=1}^{r}p_i\log p_i\le-\sum_{i=1}^{r}p_i\log q_i$$

香农辅助定理说明：符号种数为r的信源X的熵函数，一定不大于符号数同样为r的另一信源中每一个符号的自信息量在信源X的概率空间中的统计平均值

2-2 单符号离散信源

性质8：最大熵定理

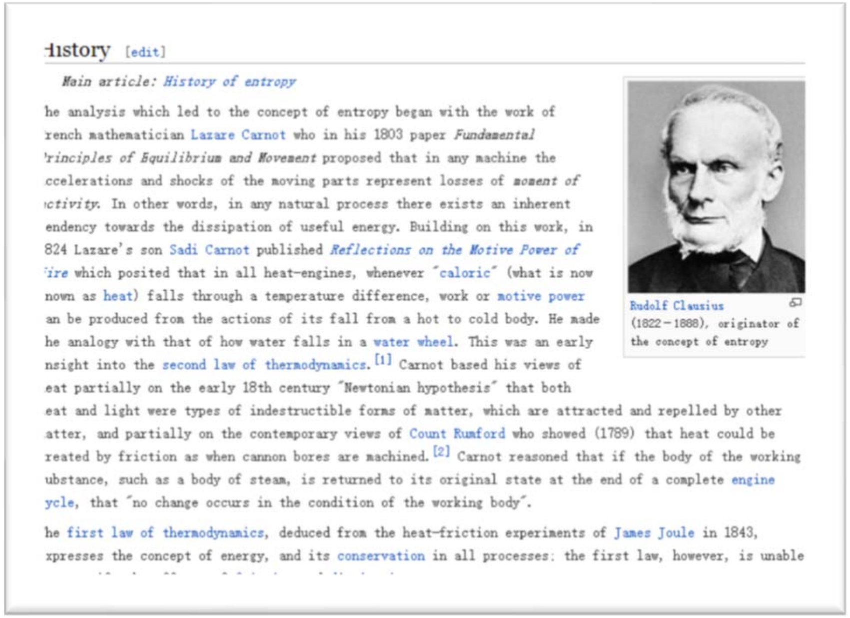
$$H(p_1, p_2, \cdots, p_n) \le H(1/n, 1/n, \cdots, 1/n)$$

$$= \log n = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} \quad （比特/符号）$$

性质9：可加性

若信源X和Y统计独立，则$H(XY)=H(X)+H(Y)$

2-2 单符号离散信源



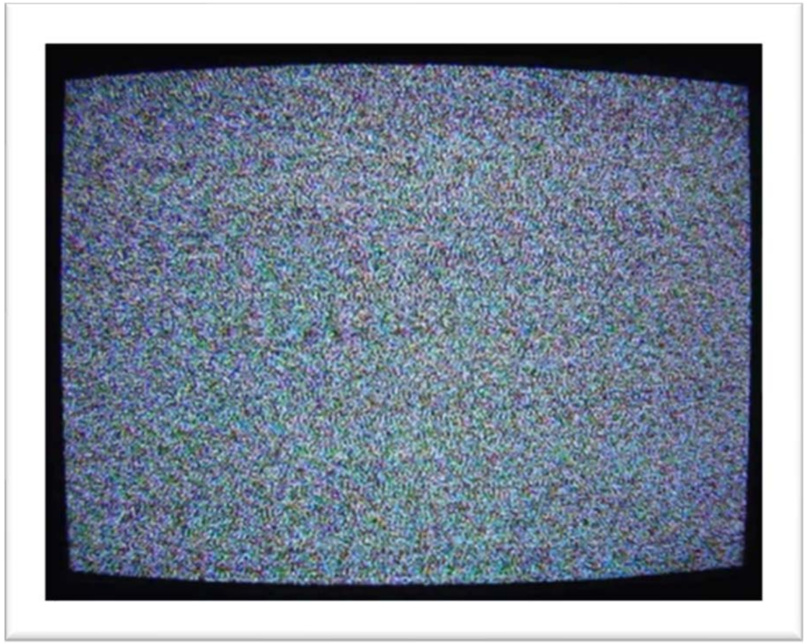
ıistory [edit]

Main article: History of entropy

he analysis which led to the concept of entropy began with the work of rench mathematician Lazare Carnot who in his 1803 paper *Fundamental 'rinciples of Equilibrium and Movement* proposed that in any machine the ccelerations and shocks of the moving parts represent losses of *moment of ıctivity*. In other words, in any natural process there exists an inherent endency towards the dissipation of useful energy. Building on this work, in 824 Lazare's son Sadi Carnot published *Reflections on the Motive Power of 'ire* which posited that in all heat-engines, whenever "caloric" (what is now nown as heat) falls through a temperature difference, work or motive power an be produced from the actions of its fall from a hot to cold body. He made he analogy with that of how water falls in a water wheel. This was an early nsight into the second law of thermodynamics.[1] Carnot based his views of eat partially on the early 18th century "Newtonian hypothesis" that both eat and light were types of indestructible forms of matter, which are attracted and repelled by other atter, and partially on the contemporary views of Count Rumford who showed (1789) that heat could be reated by friction as when cannon bores are machined.[2] Carnot reasoned that if the body of the working ubstance, such as a body of steam, is returned to its original state at the end of a complete engine ycle, that "no change occurs in the condition of the working body".

Rudolf Clausius (1822–1888), originator of the concept of entropy

he first law of thermodynamics, deduced from the heat-friction experiments of James Joule in 1843, xpresses the concept of energy, and its conservation in all processes; the first law, however, is unable



思考题：

设有12个体积、颜色均相同的小球，其中一个与其它球不同（或者轻或者重），现采用一个无砝码的天平来测量，为了在天平上称出哪一个球与其它球不同，且判断其重量是比其它球轻还是重，请问至少必须称多少次?

提示：

1） 当你发现这个小球后（无论轻重）你获得了多少信息量?

2） 考虑扔一个硬币你得到多少信息？一次扔两个硬币？扔一个有4面的骰子?

3） 考虑如果第一次称量时，一下两种情况可分别获得多少信息量：

a）天平每边放6个小球；

b）天平每边放4个小球，剩余4个小球不称量；

2-3 离散序列信源

定义：由多个符号组成的时间（或空间）序列才能代表一个完整的消息的信源称为离散多符号信源

多符号离散信源
- 离散非平稳信源
- 离散平稳信源
  - 离散平稳无记忆信源
  - 离散平稳有记忆信源

离散平稳有记忆信源
- 非马尔可夫信源
- 马尔可夫(Markov)信源

2-3 离散序列信源

2.3.1 离散平稳信源

定义：

设Q,T为任意时刻，若信源X的概率分布与时间无关，即有

$$P(X_Q) = P(X_T)$$

$$P(X_Q X_{Q+1}) = P(X_T X_{T+1})$$

$$\cdots$$

$$P(X_Q X_{Q+1} \cdots X_{Q+N}) = P(X_T X_{T+1} \cdots X_{T+N})$$

将信源X称为N维离散平稳信源

2-3 离散序列信源

2.3.2 离散平稳无记忆信源的序列熵（信息熵）

定义：

如果N维离散平稳信源 $\mathbf{X}=X_1X_2\cdots X_N$中，各时刻随机变量$X_k\,(k=1,2,\cdots N)$之间统计独立，则把信 源$X$称为离散平稳无记忆信源，将 信源$\mathbf{X}=X_1X_2\cdots X_N$称为$N$维离散平稳无记忆信源

2-3 离散序列信源

把N维离散平稳无记忆信源$\mathbf{X} = X_1 X_2 \cdots X_N$称为离散平稳无记忆信源$\mathbf{X}$的$N$次扩展信源，记为

$$\mathbf{X}^N = X_1 X_2 \cdots X_N$$

把信源输出的序列看成是一组一组发出的。

例如：电报系统中，可以认为每两个二进制数字组成一组。这样信源输出的是由2个二进制数字组成的一组符号。这时可以将它们等效看成一个新的信源，它由四个符号00，01，10，11组成，把该信源称为**二进制无记忆信源的二次扩展。**

2-3 离散序列信源

例如： 如果把每三个二进制数字组成一组，这样长度为3的二进制序列就有8种不同的序列，可等效成一个具有8个符号的信源，把它称为**二进制无记忆信源的三次扩展信源。**

二进制无记忆信源的N次扩展：把每N个二进制数字组成一组，则信源等效成一个具有2^N个符号的新信源，把它称为**二进制无记忆信源的N次扩展信源。**

2-3 离散序列信源

■ N次扩展信源的数学模型

由于单符号离散信源的数学模型为：

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_r \\ p(a_1) & p(a_2) & \cdots & p(a_r) \end{bmatrix} \text{其中} \sum_{i=1}^{r} p(a_i) = 1$$

那么信源的N次扩展信源用X^N表示，它是具有r^N个符号的新的离散信源，它的数学模型为：

$$\begin{bmatrix} X^N \\ P(\mathbf{X}) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{r^N} \\ p(\alpha_1) & p(\alpha_2) & \cdots & p(\alpha_{r^N}) \end{bmatrix} \text{其中} \sum_{i=1}^{r^N} p(\alpha_i) = 1$$

$$\text{且} \alpha_i = a_{i1}a_{i2}a_{i3}\cdots a_{iN}, p(\alpha_i) = p(a_{i1}a_{i2}a_{i3}\cdots a_{iN})$$

2-3 离散序列信源

■ N次扩展信源的熵

$$H(\mathbf{X}) = H(X^N) = -\sum_{i=1}^{r^N} p(\alpha_i)\log p(\alpha_i) \;\; bit/Message$$

根据信息熵的定义，可以证明：

$$H(\mathbf{X}) = H(X^N)$$

$$= H(X_1 X_2 \cdots X_N) = H(X_1) + H(X_2) + \cdots + H(X_N)$$

考虑到信源的平稳特性

$$H(\mathbf{X}) = H(X^N) = H(X_1 X_2 \cdots X_N) = NH(X)$$

一般加法公式

设 A 和 B 是两个任意的事件，则

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

例　袋中装有 2 个红球，3 个白球，4 个黑球，从中每次任取一球，并放回，连续取两次，求取得的两球中无黑球或无红球的概率.

乘法公式

定义 对于事件 A 和 B，若 $P(A)>0$，则称

$$P(B|A)=\frac{P(AB)}{P(A)}$$

为在事件 A 发生的条件下事件 B 发生的条件概率

（conditional probability）.

乘法公式

推广　概率的乘法公式可以推广到有限多个事件的情形，即

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2\mid A_1)P(A_3\mid A_1A_2)\cdots P(A_n\mid A_1A_2\cdots A_{n-1})$$

例　设 100 件产品中有 5 件是不合格品，用下列两种方法抽取 2 件，求 2 件都是合格品的概率：

（1）不放回顺序抽取；

（2）放回顺序抽取.

全概率公式

定义　对于集合 S，集合 S 的一列非空子集 $A_1, A_2, \cdots, A_n$ 称为 S 的划分，如果这一列子集满足

（1）$A_i A_j = \phi$，$i \neq j$，$i, j = 1, 2, \cdots, n$；

（2）$A_1 \cup A_2 \cup \cdots \cup A_n = S$.

定理（全概率公式）　设样本空间为 Ω，$A_1, A_2, \cdots, A_n$ 是 Ω 的一个划分，且 $P(A_i) > 0$，$i = 1, 2, \cdots, n$，则事件 B 发生的概率为

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B \mid A_i)$$

列表确定联合事件概率和边际概率

概率	事件		
事件	B1	B2	总计
A1	**P(A1&B1)**	**P(A1&B2)**	**P(A1)**
A2	**P(A2&B1)**	**P(A2&B2)**	**P(A2)**
总计	**P(B1)**	**P(B2)**	1

贝叶斯公式

定理　设样本空间为Ω，$A_1, A_2, \cdots, A_n$是Ω的一个划分，且$P(A_i) > 0$，$i = 1, 2, \cdots, n$，对于事件B，$P(B) > 0$，则事件B发生的前提下，事件A_i发生的概率是

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{\sum_{j=1}^{n} P(A_j)P(B \mid A_j)}, \quad i = 1, 2, \cdots, n$$

称为贝叶斯公式，又称为逆概率公式.

2-3 离散序列信源

例题2.7：如果离散平稳无记忆信源X的信源空间为：

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$$，计算信源X的二次扩展信源$\mathbf{X} = X_1 X_2$的信源熵

2-3 离散序列信源

解：信源 X 的概率分布为：

			x_j	
	$p(x_i x_j)$	a_1	a_2	a_3
	a_1	1/16	1/8	1/16
x_i	a_2	1/8	1/4	1/8
	a_3	1/16	1/8	1/16

二次扩展信源空间为：

$$\begin{bmatrix} X^2 \\ P(X^2) \end{bmatrix} = \begin{bmatrix} a_1a_1 & a_1a_2 & a_1a_3 & a_2a_1 & a_2a_2 & a_2a_3 & a_3a_1 & a_3a_2 & a_3a_3 \\ 1/16 & 1/8 & 1/16 & 1/8 & 1/4 & 1/8 & 1/16 & 1/8 & 1/16 \end{bmatrix}$$

二次扩展信源的信息熵为：

$$H(X^2) = H(X_1X_2) = -\sum_{i=1}^{3}\sum_{j=1}^{3} p(a_ia_j)\log p(a_ia_j) = 3\,(\text{bit/message})$$

2-3 离散序列信源

注意：

N维离散平稳信源 $\mathbf{X}=X_1X_2\cdots X_N$的信息熵的单位应该是信源$\mathbf{X}=X_1X_2\cdots X_N$每发一条消息所提供的 平均信息量，而信源 $\mathbf{X}=X_1X_2\cdots X_N$每发一条消息有 N个符号组成，因此单位应 该：

比特 / N信符，或者是：比特 /消息

2-3 离散序列信源

2.3.3 离散平稳有记忆信源的序列熵（信息熵）

【一】**定义**：

如果N维离散平稳信源$\mathbf{X}=X_1X_2\cdots X_N$中，各时刻随机变量$X_k(k=1,2,\cdots N)$之间并非统计独立，而是统计联系的，则把信源$X$称为离散平稳有记忆信源，将信源$\mathbf{X}=X_1X_2\cdots X_N$称为$N$维离散平稳有记忆信源。

（把N维离散平稳有记忆信源$\mathbf{X}=X_1X_2\cdots X_N$某时刻前已知的随机变量个数，即单位时间数称为离散平稳有记忆信源$\mathbf{X}$的记忆长度）

【二】下面以N=2时为例，分析二维离散平稳有记忆信源的信息熵：

（1） **联合熵**　　（2） **条件熵**

2-3 离散序列信源

（1） 联合熵

已知2维离散平稳有记忆信源 $\mathbf{X}=X_1X_2$

其中离散平稳稳有记忆信 源X的信源空 间

$$\begin{bmatrix} X \\ P \end{bmatrix}=\begin{bmatrix} a_1 & a_2 & \cdots & a_r \\ p(a_1) & p(a_2) & \cdots & p(a_r) \end{bmatrix} \text{且} 0\le p(a_i)\le 1, \sum_{i=1}^{r} p(a_i)=1$$

因此，信源 $\mathbf{X}=X_1X_2$的信源空 间

$$\begin{bmatrix} \mathbf{X} \\ P \end{bmatrix}=\begin{bmatrix} a_1a_1 & a_1a_2 & \cdots & a_ra_r \\ p(a_1a_1) & p(a_1a_2) & \cdots & p(a_ra_r) \end{bmatrix} \text{且} \sum_{i=1}^{r}\sum_{j=1}^{r} p(a_ia_j)=1$$

根据信息熵的定义：

$$H(X_1X_2)=-\sum_{i=1}^{r}\sum_{j=1}^{r} p(a_ia_j)\log p(a_ia_j)$$

将$H(X_1X_2)$称为X_1X_2的联合熵

2-3 离散序列信源

联合熵的计算：

$$H(\mathbf{X}) = H(X_1 X_2)$$

$$= -\sum_{i=1}^{r}\sum_{j=1}^{r} p(a_i a_j) \log p(a_i a_j)$$

$$= -\sum_{i=1}^{r}\sum_{j=1}^{r} p(a_i a_j) \log p(a_i) p(a_j / a_i)$$

$$= -\sum_{i=1}^{r}\sum_{j=1}^{r} p(a_i a_j) \log p(a_i) - \sum_{i=1}^{r}\sum_{j=1}^{r} p(a_i a_j) \log p(a_j / a_i)$$

2-3 离散序列信源

(2) 条件熵

可以先求出已知前面一个符号$X_1 = a_i$时，信源输出下一个符号的平均不确定性：

$$H(X_2 / X_1 = a_i) = -\sum_{j=1}^{r} p(a_j / a_i) \log p(a_j / a_i)$$

再对所有a_i的可能取值进行统计平均，得到：

$$\begin{aligned} H(X_2 / X_1) &= \sum_{i=1}^{r} p(a_i) H(X_2 / X_1 = a_i) \\ &= -\sum_{i=1}^{r} \sum_{j=1}^{r} p(a_i) p(a_j / a_i) \log p(a_j / a_i) \\ &= -\sum_{i=1}^{r} \sum_{j=1}^{r} p(a_i a_j) \log p(a_j / a_i) \end{aligned}$$

将$H(X_2 / X_1)$称为条件熵

2-3 离散序列信源

联合熵的计算：

$$H(\mathbf{X})=H(X_1X_2)$$

$$=-\sum_{i=1}^{r}\sum_{j=1}^{r}p(a_ia_j)\log p(a_ia_j)=-\sum_{i=1}^{r}\sum_{j=1}^{r}p(a_ia_j)\log p(a_i)p(a_j/a_i)$$

$$=-\sum_{i=1}^{r}\sum_{j=1}^{r}p(a_ia_j)\log p(a_i)-\sum_{i=1}^{r}\sum_{j=1}^{r}p(a_ia_j)\log p(a_j/a_i)$$

$$=-\sum_{i=1}^{r}\sum_{j=1}^{r}p(a_i)p(a_j/a_i)\log p(a_i)+H(X_2/X_1)$$

$$=-\sum_{i=1}^{r}p(a_i)\sum_{j=1}^{r}p(a_j/a_i)\log p(a_i)+H(X_2/X_1)$$

$$=-\sum_{i=1}^{r}p(a_i)\log p(a_i)+H(X_2/X_1)$$

$$=H(X_1)+H(X_2/X_1)=H(X_2)+H(X_1/X_2)$$

结论：两个有相互依赖关系的随机变量X_1和X_2所组成的随机矢量$\mathbf{X}=X_1X_2$的联合熵$\boldsymbol{H}(\boldsymbol{X})$,等于第一个随机变量的熵$\boldsymbol{H}(X_1)$与第一个随机变量$X_1$已知的前提下，第二个随机变量$X_2$的条件熵$\boldsymbol{H}(X_2/X_1)$之和。

2-3 离散序列信源

【三】有记忆与无记忆信源熵的差别

比较二维离散平稳有记忆信源$\mathbf{X}=X_1X_2$的信息熵$H(\mathbf{X})$

和二维离散平稳无记忆信源$X^2=X_1X_2$的信息熵$H(X^2)$

$$H(X^2)=H(X_1)+\boxed{H(X_2)}$$

$$H(\mathbf{X})=H(X_1)+\boxed{H(X_2/X_1)}$$

结论：$H(\mathbf{X})=H(X_1)+H(X_2/X_1)$

$$\le H(X_1)+H(X_2)=H(X^2)$$

2-3 离散序列信源

【四】平均符号熵

用于评估离散平稳信源X每发一个符号所提供的平均信息量—**平均符号熵** $H_N(\mathbf{X})=H(X_1X_2\ldots X_N)/N$

$$H_2(\mathbf{X})=H_2(X_1X_2)=H(X_1X_2)/2=\{H(X_1)+H(X_2/X_1)\}/2$$

$$H_2(X^2)=H_2(X_1X_2)=H(X_1X_2)/2=\{H(X_1)+H(X_2)\}/2$$
$$=2H(X)/2=H(X)$$

同时 $$H_2(\mathbf{X})\le H_2(X^2)=H(X)$$

2-3 离散序列信源

例题2.8：

已知离散平稳有记忆信源信源信源空间为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ 11/36 & 4/9 & 1/4 \end{bmatrix}$$

现信源发出二重符号序列(a_i, a_j)，两个符号之间用条件概率表示其关联性，$p(a_j/a_i)$如下表，求信源的序列熵和平均符号熵。

			a_j	
	$p(a_j \mid a_i)$	a_1	a_2	a_3
	a_1	9/11	2/11	0
a_i	a_2	1/8	3/4	1/8
	a_3	0	2/9	7/9

2-3 离散序列信源

解：根据题意，二次扩展信源的概率分布为：

			a_j	
	$p(a_i a_j)$	a_1	a_2	a_3
	a_1	$1/4$	$1/18$	0
a_i	a_2	$1/18$	$1/3$	$1/18$
	a_3	0	$1/18$	$7/36$

$$(p(a_1 a_1) = p(a_1/a_1)p(a_1)$$
$$=1/4)$$

信源条件熵为：

$$H(X_2/X_1) = -\sum_{i=1}^{3}\sum_{j=1}^{3} p(a_i a_j)\log p(a_j/a_i)$$
$$= 0.8717 \quad \text{(bit/symbol)}$$

2-3 离散序列信源

单符号信源熵为：

$$H(X_1) = -\sum_{i=1}^{3} p(a_i)p(a_i) = 1.5426 \ (\text{bit/symbol})$$

发二重符号的序列熵为：

$$H(\mathbf{X}) = H(X_1X_2) = H(X_1) + H(X_2/X_1)$$
$$= 2.4143 \ (\text{bit/sequence})$$

平均符号熵为：

$$H_2(\mathbf{X}) = \frac{1}{2}H(\mathbf{X}) = 1.2071 \ (\text{bit/symbol})$$

本题结论：$H_2(\mathbf{X}) \le H(X)$

2-3 离散序列信源

【五】N维离散平稳有记忆信源的序列熵

信源输出为N长序列时，信源的序列熵为：

$$
\begin{aligned}
H(\mathbf{X}) &= H(X_1 X_2 \cdots X_N) \\
&= H(X_1) \\
&+ H(X_2 / X_1) \\
&+ H(X_3 / X_1 X_2) + \\
&\cdots \\
&+ H(X_N / X_1 X_2 \cdots X_{N-1})
\end{aligned}
$$

2-3 离散序列信源

$$H(X_1X_2\cdots X_N)=-\sum_{i_1}\sum_{i_2}\cdots\sum_{i_N}p(x_{i_i}x_{i_2}\cdots x_{i_N})\log_2 p(x_{i_i}x_{i_2}\cdots x_{i_N})$$

$$=-\sum_{i_1}\sum_{i_2}\cdots\sum_{i_N}p(x_{i_i}x_{i_2}\cdots x_{i_N})\log_2[p(x_{i_i})p(x_{i_2}/x_{i_i})\cdots p(x_{i_N}/x_{i_i}x_{i_2}\cdots x_{i_{N-1}})]$$

$$=-\sum_{i_1}\left[\sum_{i_2}\cdots\sum_{i_N}p(x_{i_i}x_{i_2}\cdots x_{i_N})\right]\log_2 p(x_{i_1})$$

$$-\sum_{i_1}\sum_{i_2}\left[\cdots\sum_{i_N}p(x_{i_i}x_{i_2}\cdots x_{i_N})\right]\log_2 p(x_{i_2}/x_{i_i})$$

$$\cdots-\sum_{i_1}\sum_{i_2}\cdots\sum_{i_N}p(x_{i_i}x_{i_2}\cdots x_{i_N})\log_2 p(x_{i_N}/x_{i_i}x_{i_2}\cdots x_{i_{N-1}})$$

$$=-\sum_{i_1}p(x_{i_1})\log_2 p(x_{i_1})-\sum_{i_1}\sum_{i_2}p(x_{i_1}x_{i_2})p(x_{i_2}/x_{i_i})-$$

$$\cdots-\sum_{i_1}\sum_{i_2}\cdots\sum_{i_N}p(x_{i_i}x_{i_2}\cdots x_{i_N})\log_2 p(x_{i_N}/x_{i_i}x_{i_2}\cdots x_{i_{N-1}})$$

$$=H(X_1)+H(X_2/X_1)+\cdots+H(X_N/X_1X_2\cdots X_{N-1})$$

2-3 离散序列信源

【六】结论

结论1：

$H(X_N / X_1X_2\cdots X_{N-1})$是序列长度$N$的单调非增函数

$$
\begin{aligned}
&H(X_N / X_1X_2\cdots X_{N-1}) \le H(X_N / X_2X_3\cdots X_{N-1}) \\
&= H(X_{N-1} / X_1X_2\cdots X_{N-2}) \le H(X_{N-1} / X_2X_3\cdots X_{N-2}) \\
&\vdots \\
&\le H(X_2 / X_1)
\end{aligned}
$$

结论1说明：

各维条件熵随着前面已知的随机变量的个数，即单位时间数的增长而减小。或在发某个符号之前关于该符号的预备知识越多，发出该符号所提供的平均信息量也就越小。

2-3 离散序列信源

结论2： N给定时，平均符号熵大于条件熵

$$H_N(\mathbf{X}) \geq H(X_N / X_1 X_2 \cdots X_{N-1})$$

结论2表明：

N维离散平稳有记忆信源$\mathbf{X}=X_1X_2\ldots\ldots X_N$的平均符号熵$H_N(\mathbf{X})$，一定不小$(N\text{-}1)$维条件熵。

2-3 离散序列信源

结论3： 平均符号熵随N的增加是递减的

$H_N(\mathbf{X})$是序列长度N的单调非增函数

$$H_N(\mathbf{X}) \le H_{N-1}(\mathbf{X}) \le H_{N-2}(\mathbf{X}) \le \cdots \le H(X)$$

结论3说明：

N 维离散平稳有记忆信源$\mathbf{X}=X_1X_2\ldots\ldots X_N$的平均符号熵$H_N(\mathbf{X})$的最大值是信源在起始时刻的信息熵$H(X)$。

平均符号熵$H_N(\mathbf{X})$是记忆长度N的单调非增函数，随着N 增加而减小。

2-3 离散序列信源

结论4： 如果H(X1)<无穷大，

当$N \to \infty$时，

$$H_\infty(\mathbf{X}) = \lim_{N\to\infty} H_N(\mathbf{X}) = \lim_{N\to\infty} H(X_N / X_1 X_2 \cdots X_{N-1})$$

$H_\infty(\mathbf{X})$称为极限熵，又称为极限信息量

结论4说明：

对于记忆长度N足够大（当$N \to \infty$时），离散平稳有记忆信源$\mathbf{X}$每发一个符号提供的平 均信息量，即极限熵$H_\infty(\mathbf{X})$，等于条件熵$H(X_N / X_1 X_2 \cdots X_{N-1})$在$N \to \infty$时的极限熵。

2-3 离散序列信源

例题2.9：

(1) 为了使电视图像获得良好的清晰度和规定的适当的对比度，需要用5×10^5个象素和8个不同亮度电平，求传递此图像所需的信息率（比特/秒）。并设每秒要传送30帧图像，所有象素是独立变化的，且所有亮度电平等概率出现。

(2) 设某彩色电视系统，除了满足对于黑白电视系统的上述要求外，还必须有30个不同的色彩度，试计算传输该彩色系统的信息率为传输黑白系统的信息率的多少倍?

2-3 离散序列信源

解题（1）

电视机的每个像素亮度作为信源，则信源空间为：

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_8 \\ 1/8 & 1/8 & \cdots & 1/8 \end{bmatrix} \text{其中} \sum_{i=1}^{8} p(a_i) = 1$$

每个像素亮度含有的信息量为

$$H(X) = \log 8 = 3bit/\ pixel$$

每帧图像数就是每一个离散亮度信源的无记忆N次扩展信源

因此每帧图像含有的信息量为

$$H(X^N) = NH(X) = 5\times 10^5 \times \log 8 = 1.5\times 10^6 bit/\ frame$$

电视机每秒30帧图像，因此每秒传递的信息率为

$$R = 30\,frame/s \times 1.5\times 10^6 bit/\ frame = 4.5\times 10^7 bit/s$$

2-3 离散序列信源

2.3.4 马尔可夫信源的极限熵

【一】马尔可夫信源定义

定义：

离散平稳有记忆信源 $X:\{a_1 a_2 \cdots a_r\}$发出时间域延伸到无穷远的随机变量序列$\mathbf{X} = \cdots X_1 X_2 \cdots X_m X_{m+1} \cdots$
如果$\mathbf{X}$中的任一时刻$(m+1)$的随机变量X_{m+1}只依赖于它前面已发的m个随机变量$X_1 X_2 \cdots X_m$，与更前面的随机变量无关，就将这种信源称为m阶马尔可夫$(Markov)$信源.(简称为m阶M信源)

2-3 离散序列信源

（1）信源的符号集和状态

- 有一类信源，输出的符号序列中符号之间的依赖关系是有限的，即任何时刻信源符号发生的概率只与前面已经发出的m个符号有关，而与更前面发出的符号无关。
- 设符号集为X和状态为S。信源输出的信息符号还与信源所处的状态有关。

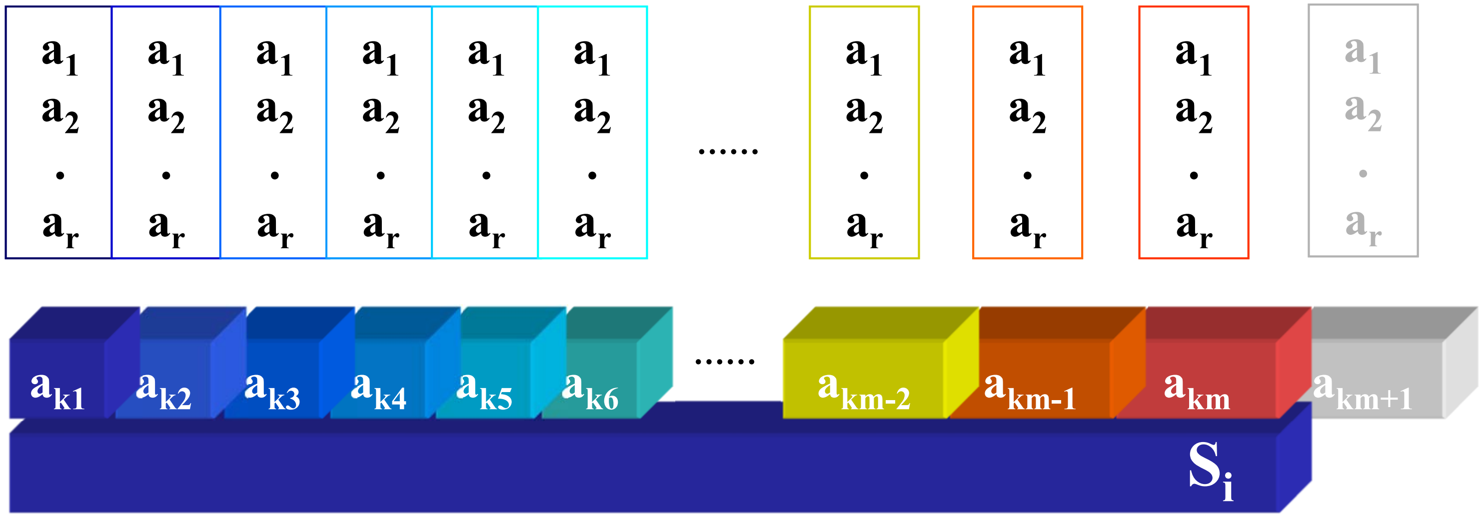
$$S_i=\{a_{k1},a_{k2},\cdots,a_{km}\}\text{可以看成是状态 } S_i,$$

$$\text{其中}a_{k1},a_{k2},\cdots,a_{km}\in X=\{a_1,a_2,\cdots,a_r\}$$

$$k1,k2,\cdots,km=1,2,\cdots,r \qquad i=1,2,\cdots,\quad r^m$$

- 每一时刻信源发出一个符号后，所处的状态将发生转移。

2-3 离散序列信源



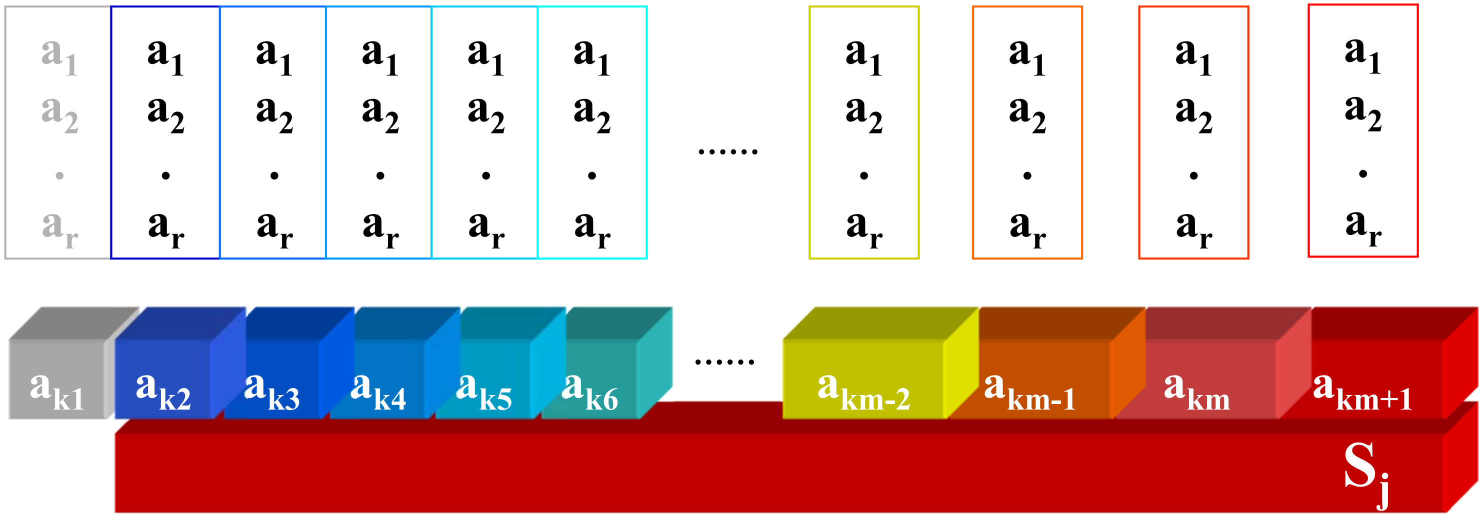
时刻1到时刻m的随机变量序列提供的某一具体消息：

$S_i=\{a_{k1},a_{k2},\cdots,a_{km}\}$可以看成是状态 S_i

其中$a_{k1},a_{k2},\cdots,a_{km}\in\{a_1,a_2,\cdots,a_r\}$

$k1,k2,\cdots,km=1,2,\cdots,r \quad i=1,2,\cdots,\ r^m$

2-3 离散序列信源



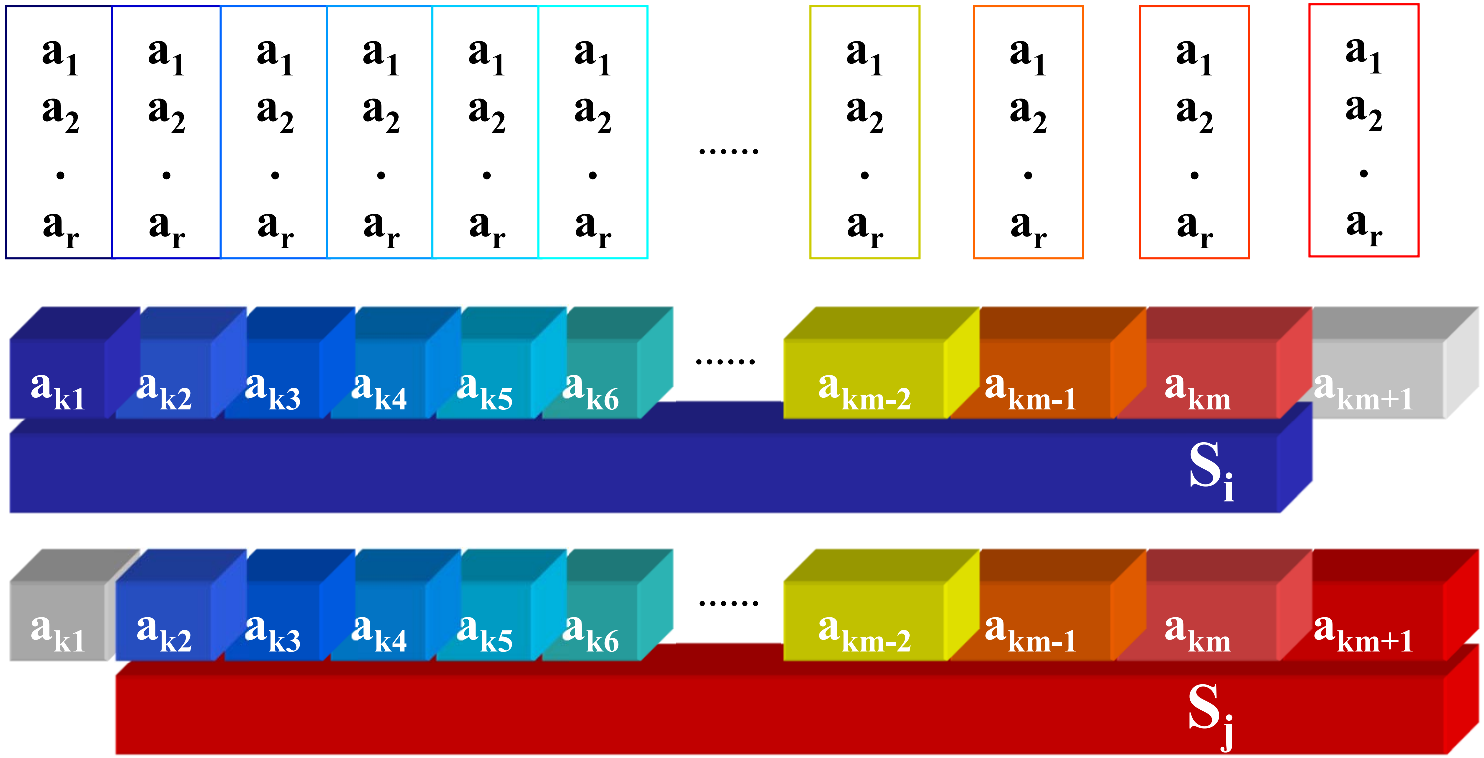
时刻2到时刻$m+1$的随机变量序列提供的某一具体消息：

$S_j=\{a_{k2},a_{k3},\cdots,a_{km+1}\}$可以看成是状态$S_j$

其中$a_{k2},a_{k3},\cdots,a_{km},a_{km+1}\in\{a_1,a_2,\cdots,a_r\}$

$k2,\cdots,km,km+1=1,2,\cdots,r \qquad j\in J\{j_1,j_2,\cdots,j_r\}$

2-3 离散序列信源



■ 若信源处于某一状态Si，当它发出一个符号后a_{km+1}，所处的状态就变为Sj；

■ 任何时刻信源处在什么状态完全由前一时刻的状态和发出的符号决定；

2-3 离散序列信源

例题2.10:

设一个二元一阶马尔科夫信源，信源符号集为X={0,1},信源输出符号的条件概率为:

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

请计算状态转移概率，并列写状态转移概率矩阵。

解:

已知r=2,m=1. 因此r^m=2 有两个状态 分别是 $S_1=0$, $S_2=1$

可由信源输出符号的条件概率得到信源状态转移概率

$$p(S_1|S_1)=0.25,\ p(S_1|S_2)=0.5,\ p(S_2|S_1)=0.75,\ p(S_2|S_2)=0.5,$$

状态转移概率矩阵

$$P=\begin{bmatrix} 0.25 & 0.75 \\ 0.5 & 0.5 \end{bmatrix}$$

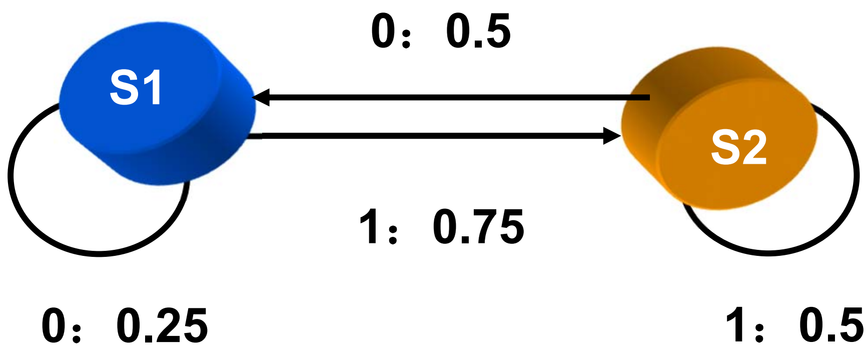
注意: 对于一阶马尔可夫信源，其状态转移概率和信源输出符号的条件概率相同

2-3 离散序列信源

定义：马尔可夫链的状态转移图。每个圆圈代表一种状态，状态之间的有向线代表某一状态向另一状态的转移。有向线一侧的符号和数字分别代表发出的符号和条件概率。

因此可以得到上一例题的马尔可夫状态转移图

$$P=\begin{bmatrix}0.25 & 0.75\\ 0.5 & 0.5\end{bmatrix}$$



2-3 离散序列信源

【二】马尔可夫信源数学描述

信源输出依赖长度为$m+1$的随机序列就转化为对应的状态序列，这种状态序列符合简单的马尔可夫链的性质。

(1)确定离散平稳有记忆信源X 的符号集，再确定**m**阶数(**m>0**)。

根据r, m得到m阶M信源所有 r^m种消息（状态）

即消息： $S_i = \{a_{k1}, a_{k2}, \cdots, a_{km}\}$

其中$a_{k1}, a_{k2}, \cdots, a_{km} \in \{a_1, a_2, \cdots, a_r\}$

$k1, k2, \cdots, km = 1,2,\cdots, r$

$i = 1,2,\cdots, \quad r^m$

消息集合： $S:\{S_1, S_2, \cdots, S_{r^m}\}$

2-3 离散序列信源

- m阶马尔可夫离散信源数学模型：

  m阶马尔可夫离散信源的数学模型可由一组信源符号集和一组条件概率确定：

$$\begin{bmatrix} X \\ P(X_{m+1}) \end{bmatrix} = \begin{bmatrix} a_1, a_2, \cdots, a_r \\ p(a_{k\,m+1} / a_{k1} a_{k2} \cdots a_{k\,m}) \end{bmatrix}$$

满足 $\sum_{km+1=1}^{r} p(\mathrm{a}_{k\,m+1} / a_{k1} a_{k2} \cdots a_{k\,m})=1$

其中 $k1, k2, \cdots, km, km+1 = 1, 2, \cdots, r$

2-3 离散序列信源

(2) 给定 r^{2m} 个由任一消息（状态）s_i 到下一个消息（状态）s_j 的：

一步转移概率：

$$p(a_{km+1}/a_{k1},a_{k2},\cdots,a_{km})=p(a_{km+1}/S_i)=p(S_j/S_i)$$

$$且 0\le p(S_j/S_i)\le 1,\ \sum_{j\in J}p(S_j/S_i)=1$$

其中：$k1,k2,\cdots,km+1=1,2,\cdots,r$

$$i=1,2,\cdots,r^m,\ j\in J\{j_1,j_2,\cdots,j_r\}$$

n步转移概率：

$$p\{S(m+n)=j/S(m)=i\}=p_{ij}^{(n)}(m)$$

其中：$i=1,2,\cdots,r^m,\ j\in J$

2-3 离散序列信源

(3) 状态转移矩阵

由于系统在任一时刻处于状态空间$S:\{S_1,S_2,\cdots,S_{r^m}\}$中任意一个状态，将给定的状态一步转移概率$p(S_j/S_i)$按照对应关系，排成一个状态一步转移矩阵$\mathbf{P}$:

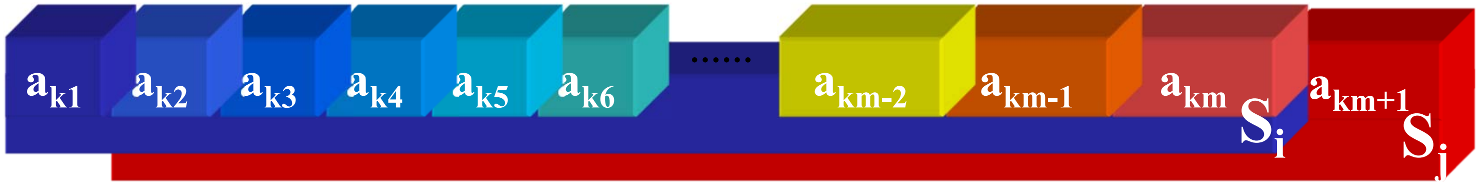
$$\mathbf{P}=\{p_{ij},i,j=1,2,\cdots,r^m\}$$

或$$\mathbf{P}=\begin{bmatrix} p_{11} & \cdots & p_{1r^m} \\ \vdots & \ddots & \vdots \\ p_{r^m1} & \cdots & p_{r^mr^m} \end{bmatrix}$$

$$\mathbf{P}^{(n)}=[\mathbf{P}]^n$$

2-3 离散序列信源

(4) 状态极限概率与m阶M信源极限熵



$$H_\infty = \lim_{N\to\infty} H(X_N / X_1 X_2 \cdots X_{N-1})$$

$$= \lim_{N\to\infty} \left\{ -\sum_{k1=1}^{r} \cdots \sum_{kN=1}^{r} p(a_{k1} a_{k2} \cdots a_{kN}) \log p(a_{kN} / a_{k1} a_{k2} \cdots a_{kN-1}) \right\}$$

$$= -\sum_{k1=1}^{r} \cdots \sum_{km+1=1}^{r} p(a_{k1} a_{k2} \cdots a_{km+1}) \log p(a_{km+1} / a_{k1} a_{k2} \cdots a_{km})$$

$$= -\sum_{k1=1}^{r} \cdots \sum_{km+1=1}^{r} p(a_{k1} a_{k2} \cdots a_{km}) p(a_{km+1} / a_{k1} a_{k2} \cdots a_{km}) \log p(a_{km+1} / a_{k1} a_{k2} \cdots a_{km})$$

$$H_\infty = \sum_{i=1}^{r^m} \sum_{j\in J} p(S_i) p(S_j / S_i) \log p(S_j / S_i)$$

2-3 离散序列信源

从而得到m阶马尔科夫信源的**极限熵**，记为H_{m+1}：

$$H_\infty = \sum_{i=1}^{r^m}\sum_{j\in J} p(S_i)p(S_j/S_i)\log p(S_j/S_i)$$

$$= \sum_{i} p(S_i)H(X/S_i)$$

状态概率$p(S_i)$，$i=1,2,\cdots,r^m$应该是m阶M信源达到稳定时出现状态S_i的概率，将这种状态概率称为$\boxed{\text{状态极限概率}}$（稳态分布概率W_i）

状态极限概率可由下式计算出：

$$\mathbf{WP}=\mathbf{W} \qquad \sum_{i=1}^{r^m} W_i = 1$$

其中$\mathbf{W}=\left(W_1, W_2, \cdots W_{r^m}\right)$

2-3 离散序列信源

几点说明：

- 这里利用了m阶马尔可夫信源“有限记忆长度”的根本特性，使无限大参数N变为有限值m，把求极限熵的问题变成了一个求m阶条件熵的问题。
- 状态一步转移概率$p(S_j/S_i)$是给定或测定的。求解H_{m+1}条件熵的关键就是要得到$p(S_i)(i=1,2,\ldots,r^m)$。$p(S_i)$是马尔可夫信源稳定后(N→∞)各状态的极限概率。
- 有限状态的马尔可夫链：状态空间的状态是有限的。

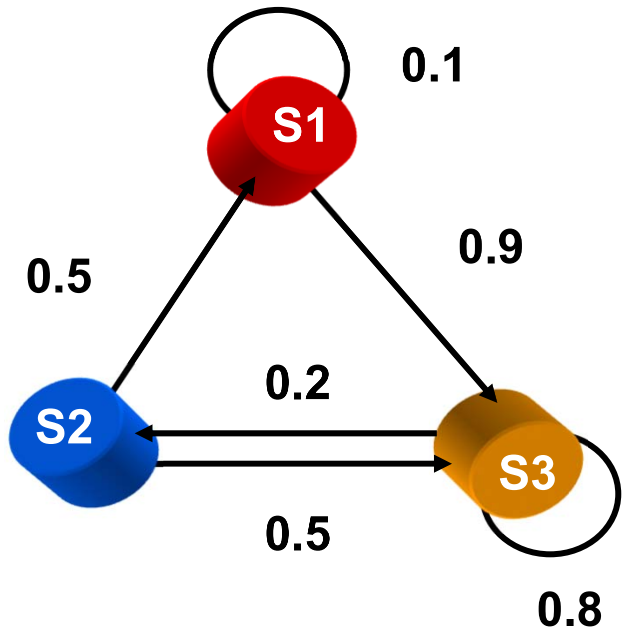
2-3 离散序列信源

例题2.11

如图所示的三状态马尔可夫信源，试计算该信源熵。

其状态一步转移矩阵为：

$$P=\begin{bmatrix} 0.1 & 0 & 0.9 \\ 0.5 & 0 & 0.5 \\ 0 & 0.2 & 0.8 \end{bmatrix}$$



2-3 离散序列信源

解：设信源稳态分布的概率向量为$\mathbf{W}=(W_1,W_2,W_3)$，则

$$\mathbf{WP}=\mathbf{W},\qquad \sum_{i=1}^{3}W_i=1,\qquad W_i\ge 0$$

$$(W_1,W_2,W_3)\begin{bmatrix}0.1 & 0 & 0.9\\ 0.5 & 0 & 0.5\\ 0 & 0.2 & 0.8\end{bmatrix}=(W_1,W_2,W_3)$$

解得：$W_1=5/59$，$W_2=9/59$，$W_3=45/59$

在S_i状态下信源输出一个符号的平均信息量为：

$$H(X/S_1)=-\sum_{j=1}^{3}p(S_j/S_1)\log p(S_j/S_1)$$

$$=\text{-}0.1\log 0.1\text{-}0.9\log 0.9=0.4690\ \ \text{(bit/symbol)}$$

$$H(X/S_2)=1\ \ \text{(bit/symbol)}$$

$$H(X/S_3)=0.7219\ \ \text{(bit/symbol)}$$

信源熵为：$H_\infty(X)=\sum_{i=1}^{3}W_iH(X/S_i)=0.7429\ \ \text{(bit/symbol)}$

2-3 离散序列信源

是不是所有的

Markov信源

都存在极限熵呢?

2-3 离散序列信源

(5) 有限齐次马尔可夫信源的各态遍历性

对于有限平稳的m阶M信源，由信源$X:\{a_1,a_2,\cdots,a_r\}$的m个符号组成的消息(状态)$S_i:\{a_{i1},a_{i2},\cdots,a_{im}\}$和消息(状态)$S_j:\{a_{j1},a_{j2},\cdots,a_{jm}\}$存在一个正整数$n\geq 1$，且经过$n$步，从状态$S_i$转移到状态$S_j$的$n$步转移概率$p_{ij}^{(n)}>0$，则这种$m$阶$M$信源具有各态遍历性。

2-3 离散序列信源

对于具有遍历性的m阶M信源，对于每个$j=1,2,\cdots,r^m$都存在不依赖于起始状态$S_i(i=1,2,\cdots,r^m)$的状态极限概率$\lim\limits_{n\to\infty} p_{ij}^{(n)}=p(S_j)$,而且,状态极限概率是在约束条件

$$p(S_j)>0(j=1,2,\cdots,r^m),\sum_{j=1}^{r^m}p(S_j)=1\text{的约束下,方程组}$$

$$p(S_j)=\sum_{i=1}^{r^m}p(S_i)p_{ij}\text{的唯一解}$$

说明：

■ 凡具有各态遍历性的m阶马尔可夫信源，其状态极限概率$p(S_i)$可由上式求出。有了$p(S_i)$和测定的$p(S_j/S_i)$，就可求出m阶马尔可夫信源的熵H_{m+1}。

2-3 离散序列信源

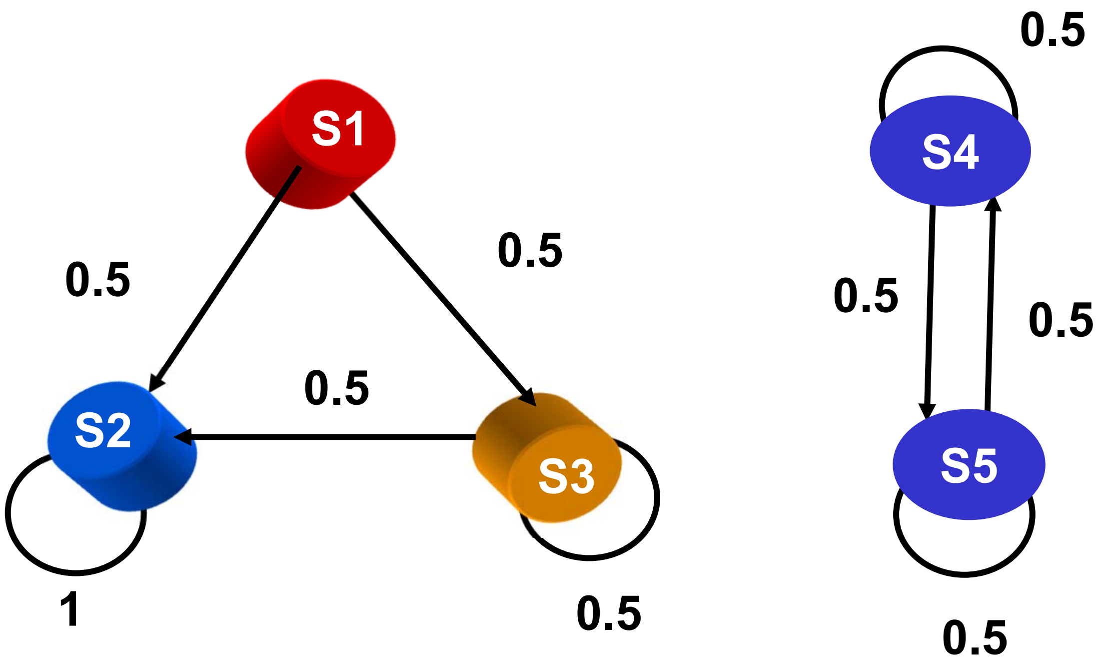
(a) 不可约性

如果状态空间$S:\{S_1,S_2,\cdots,S_{r^m}\}$中任意两个状态$S_i$，$S_j$都存在至少一个正整数$k$，使得$p_{ij}^{(k)}>0$,即从状态$S_i$开始，总有可能转移到另一状态$S_j$，状态空间中各个状态之间，都能相互到达，那么状态空间S组成的集合称为不可约闭集

2-3 离散序列信源

不可约闭集举例：

三个闭集$\{S_1,S_2,S_3,S_4,S_5\}$、$\{S_1,S_2,S_3\}$、$\{S_4,S_5\}$



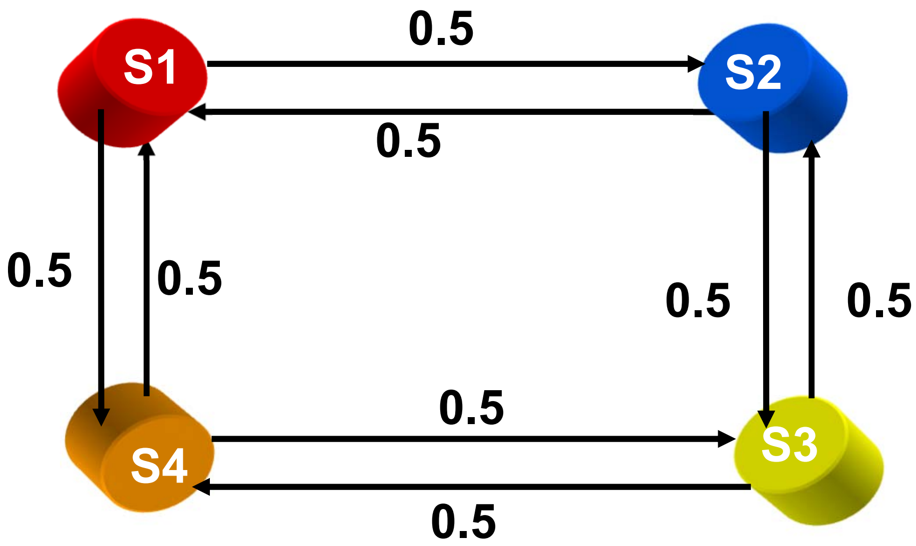
2-3 离散序列信源

(b) 非周期性

如果状态空间$S:\{S_1,S_2,\cdots,S_{r^m}\}$是一个不可约闭集，并且从每一状态出发，经过$k$步转移回到本状态的所有可能的步数$k$中，即使$p_{ii}^{(k)}>0$的所有$k$中，不存在大于 1 的公因子，则称不可约闭集$S:\{S_1,S_2,\cdots,S_{r^m}\}$具有非周期性。

2-3 离散序列信源

具有周期性不可约闭集举例：



周期为2，从任一状态出发，回到该状态的可能步数为2，4，6，…

2-3 离散序列信源

(c) 结论

在确定Markov信源极限熵时，必须首先判断该信源是否具有各态遍历性，方法如下：

(1)方法1：n步状态转移矩阵$\mathbf{P}^{[n]}$中，不存在"0"元素，则可判断具有各态遍历性；如果n为任意正整数的$\mathbf{P}^{[n]}$中，都有"0"元素，则不具有各态遍历性。

(2)方法2：状态转移图中的状态集合是不可约非周期闭集，则可判断具有各态遍历性；若不是不可约非周期闭集，则判断不具有各态遍历性。

2-3 离散序列信源

几点说明：

- m阶马尔可夫信源在起始的有限时间内，信源不是平稳和遍历/各态历经性的，状态的概率分布有一段起始渐变过程。经过足够长时间之后，信源处于什么状态已与初始状态无关，这时每种状态出现的概率已达到一种稳定分布。

- 一般马尔可夫信源并非是平稳信源。但当时齐、遍历的马尔可夫信源达到稳定后，这时就可以看成是平稳信源。

2-3 离散序列信源

例题2.12

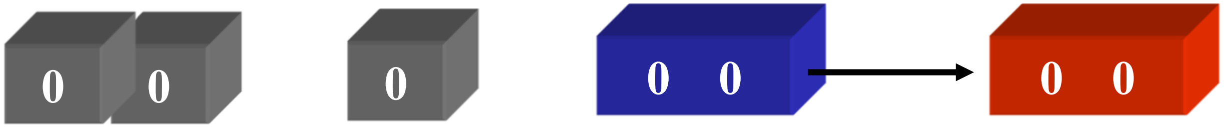
某m=2阶M信源X的符号集X:{0,1}，符号条件概率如表1，求解该信源的极限熵。

表1：符号条件概率

初始态	符号	
	0	1
S_1:00	1/2	1/2
S_2:01	1/3	2/3
S_3:10	1/4	3/4
S_4:11	1/5	4/5

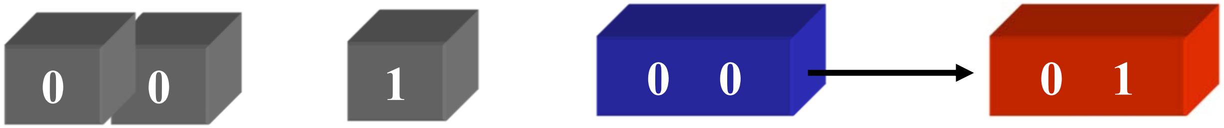
2-3 离散序列信源

解：步骤一、状态转移过程如下：



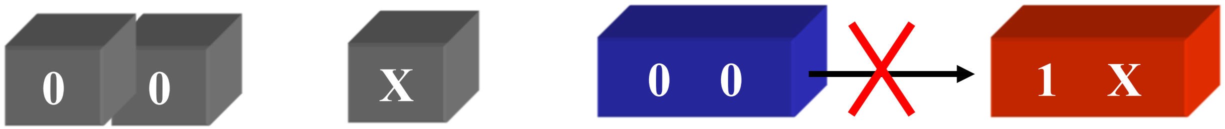
$S_1 \longrightarrow S_1$

$$p(0/00) = p(0/S_1) = p(S_1/S_1) = 0.5$$



$S_1 \longrightarrow S_2$

$$p(1/00) = p(1/S_1) = p(S_2/S_1) = 0.5$$



$S_1 \quad S_3$ or S_4

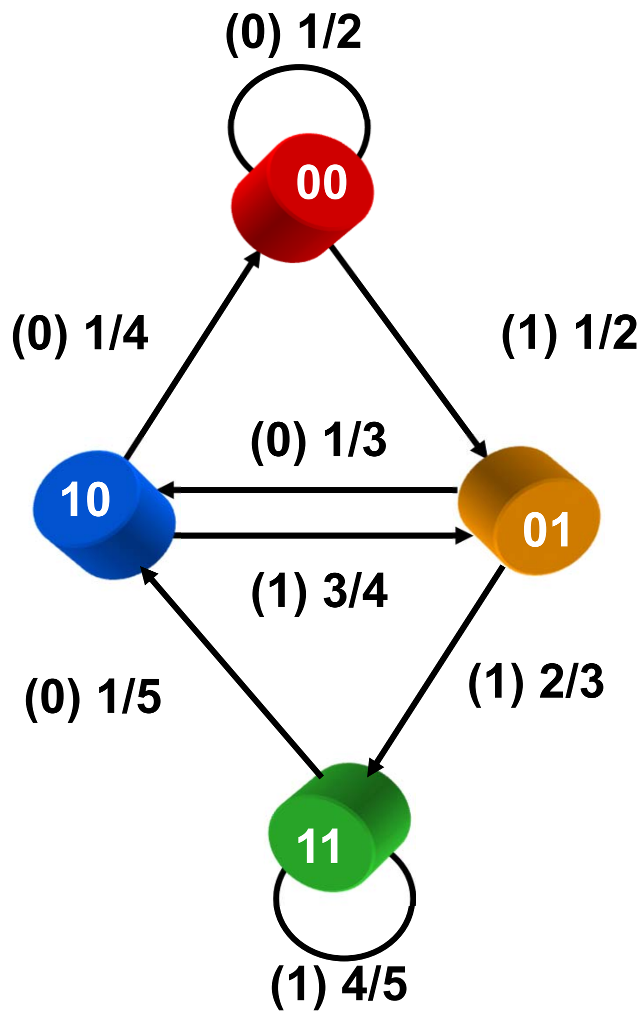
$$p(S_3/S_1) = 0 \qquad p(S_4/S_1) = 0$$

2-3 离散序列信源

步骤二、状态转移概率矩阵和状态转移图：

状态一步转移概率矩阵

$$\mathbf{P}=\begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 1/4 & 3/4 & 0 & 0 \\ 0 & 0 & 1/5 & 4/5 \end{bmatrix}$$



2-3 离散序列信源

步骤三、判断是否具有各态遍历性：

- 判断方法1：根据状态转移图判断是否是非周期不可约闭集，即不可约性和非周期性的判断

- 判断方法2：用状态一步转移概率判断：

$$\mathbf{P}^2 = \begin{bmatrix} 1/4 & 1/4 & 1/6 & 1/3 \\ 1/12 & 1/4 & 2/15 & 8/15 \\ 1/8 & 1/8 & 1/4 & 1/2 \\ 1/20 & 3/20 & 4/25 & 16/25 \end{bmatrix}$$

在$\mathbf{P}^2$中不存在0元素，故该信源具有各态遍历性

2-3 离散序列信源

步骤四、求信源的极限熵：

设各状态的稳态分布概率分别为W_1,W_2,W_3,W_4，则

根据$[W_1,W_2,W_3,W_4]=[W_1,W_2,W_3,W_4]\mathbf{P}$

$$[W_1,W_2,W_3,W_4]=[W_1,W_2,W_3,W_4]\begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 1/4 & 3/4 & 0 & 0 \\ 0 & 0 & 1/5 & 4/5 \end{bmatrix}$$

得到

$$\begin{cases} W_1=\dfrac{1}{2}W_1+\dfrac{1}{4}W_3, & W_2=\dfrac{1}{2}W_1+\dfrac{3}{4}W_3 \\ W_3=\dfrac{1}{3}W_2+\dfrac{1}{5}W_4, & W_4=\dfrac{2}{3}W_3+\dfrac{4}{5}W_4 \end{cases}$$

2-3 离散序列信源

又$\sum_{i=1}^{4} W_i = 1$　　即　　$W_1 + W_2 + W_3 + W_4 = 1$

解得：$W_1 = 3/35$，$W_2 = W_3 = 6/35$，$W_4 = 4/7$

在S_i状态下信源输出一个符号的平均信息量为：

$$H(X/S_1) = -\sum_{j=1}^{4} p(S_j/S_1)\log p(S_j/S_1)=1 \text{ (bit/symbol)},$$

$$H(X/S_2) = -\sum_{j=1}^{4} p(S_j/S_2)\log p(S_j/S_2)=0.9183 \text{ (bit/symbol)}$$

$$H(X/S_3) = -\sum_{j=1}^{4} p(S_j/S_3)\log p(S_j/S_3)=0.8113 \text{ (bit/symbol)}$$

$$H(X/S_4) = -\sum_{j=1}^{4} p(S_j/S_4)\log p(S_j/S_4)=0.7219 \text{ (bit/symbol)}$$

信源的极限熵为：$H_\infty(X) = H_{2+1}(X) = \sum_{i=1}^{4} W_i H(X/S_i)$

$$= \frac{3}{35}1 + \frac{6}{35}0.9183 + \frac{6}{35}0.8113 + \frac{4}{7}0.\ 7219 = 0.7947 \text{ (bit/symbol)}$$

2-3 离散序列信源

- **m阶M信源和消息长度为m的有记忆信源，其所含符号的依赖关系不同，对相应关系的数学描述不同，平均信息量的计算公式也不同。**

➢ m阶M信源的记忆长度虽为有限值，但符号之间的依赖关系延伸到无穷远，通常用**状态转移概率**来描述这种依赖关系。即M信源以转移概率发出每个信源符号，所以平均每发一个符号提供的信息量应为**极限熵**H_{m+1}

➢ 长度为m的有记忆信源，发出的是一组组符号序列，每m个符号构成一个符号序列组，代表一个消息。组与组之间也是相互统计独立的，因此符号之间的相互依赖关系仅限于每组之内的m个符号，一般用这m个符号的联合概率来描述符号间的依赖关系。平均每发一个符号提供的信息量，是m个符号的联合熵的m分之一，即**平均符号熵**$H_m(X)$

2-4信源的相关性和剩余度

试验：信源符号集如下15个字符

A C E G H I L M O Q S T R Z -

2-4信源的相关性和剩余度

试验： 信源符号集如下15个字符

A C E G H I L M O Q S T V Z -

T I L - G A Q S E M

2-4信源的相关性和剩余度

例题2.13：分析26个英文字母与一个空格符号组成的英文信源所提供的信息熵。

1、当信源发出的字母相互统计独立，互不相关且等概率分布时,得到该信源的最大熵值：

2、假设信源发出的字母相互统计独立，但不是等概率分布时,统计其出现的概率，得到：

2-4信源的相关性和剩余度

例题2.13：分析26个英文字母与一个空格符号组成的英文信源所提供的信息熵。

3、考虑到各个字母之间的相关性，将信源看成

m阶M信源处理,得到m阶M信源的极限熵：

1阶M信源极限熵：$H_{1+1}=H_2=3.32$比特/符号

2阶M信源极限熵：$H_{2+1}=H_3=3.1$比特/符号

……

$H_\infty(X)=1.4$比特/符号

2-4信源的相关性和剩余度

- 讨论信源的最主要目的是为了得到高效率的信源编码。
- 衡量信源编码效率的尺度是什么呢?
- 或者说能够使信源编码提高效率的根本原因是什么呢?

2-4信源的相关性和剩余度

信源符号的相关性与提供的平均信息量

- 把多符号离散信源都用马尔可夫信源来逼近，则记忆长度不同，熵值就不同，意味着平均每发一个符号就有不同的信息量。

$$\log_2 n = H_0 \geq H_1 \geq H_2 \geq \ldots \geq H_m \geq H_\infty$$

- 所以信源的记忆长度越长，熵值越小。当信源符号间彼此没有任何依赖关系且呈等概率分布时，信源熵达到最大值。即信源符号的相关性越强，提供的平均信息量越小。

2-4信源的相关性和剩余度

信息传输手段的浪费

- 一般离散平稳信源，H_∞就是实际信源熵。理论上只要有传送H_∞的手段，就能把信源包含的信息全部发送出去。但实际上确定H_∞非常困难，只好用H_m来代替。
- $H_{m+1}>H_\infty$，所以在传输手段上必然富裕，这样做很不经济，特别是有时只能得到H_1，甚至H_0，就更不经济。这种浪费是由信源符号的相关性引起的。

2-4信源的相关性和剩余度

信源符号组成的时间序列中，符号之间的依赖关系越强，信源每发一个符号提供的平均信息量就越小。

定义

为了衡量信源符号之间的依赖程度，将离散平稳有记忆信源的极限熵H_∞，与把这个信源当作离散平稳无记忆等概率信源达到的最大熵值$H_0=\log r$的比值，定义为这个离散平稳有记忆信源的相对熵率(信息效率)：$\eta=\dfrac{H_\infty}{H_0}$

2-4信源的相关性和剩余度

信源符号组成的时间序列中，符号之间的依赖关系越强，信源每发一个符号提供的平均信息量就越小。

定义

将1减去相对熵率η所得之差定义为离散平稳有记忆信源的剩余度（冗余度）：

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

2-4信源的相关性和剩余度

例题2.13：分析26个英文字母与一个空格符号组成的英文信源所提供的信息熵。

3、考虑到各个字母之间的相关性，将信源看成m阶M信源处理，得到m阶M信源的极限熵：

$$H_2 = 3.32\text{比特}/\text{符号}$$

$$H_3 = 3.1\text{比特}/\text{符号}$$

……

$$H_\infty(X) = 1.4\text{比特}/\text{符号}$$

熵的相对率： $\eta = \dfrac{H_\infty}{H_0} = 1.4/4.76 = 0.29$

信源剩余度： $\gamma = 1 - \eta = 0.71$

2-4信源的相关性和剩余度

不同文字在不同近似程度下的熵

文字	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（8000字计算）	近似13	9.41	8.1	7.7		4.1	0.315	0.685

2-4信源的相关性和剩余度

例题：2.14 计算汉字的剩余度（以10000个常用汉字统计）

类别	汉字个数	所占比例	每个汉字概率
1	140	0.5	0.5/140
2	485	0.35	0.35/485
3	1775	0.147	0.147/1775
4	7600	0.003	0.003/7600

解：为了计算方便，假设每类中汉字出现概率等概不考虑字符之间的相关性，信源的实际近似熵：

$$H(X) = -140 \times \frac{0.5}{140} \times \log\left(\frac{0.5}{140}\right) - 0.35 \times \log\left(\frac{0.35}{485}\right)$$

$$-0.147 \times \log\left(\frac{0.147}{1775}\right) - 0.003 \times \log\left(\frac{0.003}{7600}\right)$$

$= 9.773$比特/汉字

而 $H_0 = \log(10000) = 13.288$比特/汉字

则 $\gamma = 1 - \eta = 0.264$

2-4信源的相关性和剩余度

结论

- 写英语文章时，71%是由语言结构定好的，只有29%是写文字的人可以自由选择的。100页的书，大约只传输29页就可以了，其余71页可以压缩掉。信息的冗余度表示信源可压缩的程度。
- 从提高传输效率的观点出发总是希望减少或去掉冗余度。
- 冗余度大的消息抗干扰能力强。能通过前后字之间的关联纠正错误。
- 听母语广播和听外语广播的对比说明：听外语费劲是英语冗余度不够造成的。

2-4信源的相关性和剩余度

结论

- 初始信源的冗余度通常是很大的，这为信源的压缩编码提供了可能
- 压缩编码的目标就是寻找某种编码方法，使得编码后消息序列中的冗余度趋近于0
- 如果将这种编码包含在信源中，也可以说是寻找某种能够使信源冗余度趋近于0的编码方法
- 冗余度成为衡量信源编码效率的一个物理量，冗余度越低，编码效率就越高。

2-4 信源的相关性和剩余度