《信息论基础》

第2章 离散信源熵

吉小鹏

E-mail: 003163@nuist.edu.cn

南京信息工程大学 电子与信息工程学院 尚贤楼209




提纲

2.1 基本概念

2.2 离散信源熵的基本概念和性质

2.3 多符号离散平稳信源熵

2.1 基本概念

■ 信源的描述与分类

信源，是信息的来源，是产生消息或消息序列的源泉。

信息是抽象的，而消息是具体的。消息不是信息本身，但其包含和携带信息，因此可通过信息的表达者——消息来研究信源。

不研究信源的内部结构，不研究信源为什么产生和怎样产生各种不同的、可能的消息，而只研究信源的各种可能的输出，以及输出各种可能消息的不确定性。

在通信系统中，收信者在未受到消息以前，对信源发出什么消息是不确定的，是随机的。因此可以用一个样本空间及其概率测度（概率空间）来描述信源。

2.1 基本概念

在介绍信源描述前需要了解的条件概率、联合概率性质：

①非负性

$$0 \le p(x_i),\ p(y_j),\ p(y_j/x_i),\ p(x_i/y_j),\ p(x_iy_j) \le 1$$

②完备性

$$\sum_{i=1}^{n} p(x_i)=1, \sum_{j=1}^{m} p(y_j)=1, \sum_{i=1}^{n} p(x_i/y_j)=1,$$

$$\sum_{j=1}^{m} p(y_j/x_i)=1, \sum_{j=1}^{m}\sum_{i=1}^{n} p(x_iy_j)=1$$

$$\sum_{i=1}^{n} p(x_iy_j)=p(y_j), \sum_{j=1}^{m} p(x_iy_j)=p(x_i)$$

在介绍信源描述前需要了解的条件概率、联合概率性质：

③联合概率

$$p(x_i y_j) = p(x_i)p(y_j / x_i) = p(y_j)p(x_i / y_j)$$

当X与Y相互独立时，$p(x_i y_j) = p(x_i)p(y_j)$

$$p(y_j / x_i) = p(y_j),\ p(x_i / y_j) = p(x_i)$$

④贝叶斯公式

$$p(x_i / y_j) = \frac{p(x_i y_j)}{\sum_{i=1}^{n} p(x_i y_j)},\quad p(y_j / x_i) = \frac{p(x_i y_j)}{\sum_{j=1}^{m} p(x_i y_j)}$$

2.1 基本概念

- 信源的描述与分类

根据产生的消息的不同随机性质对信源进行分类。

1. 信源输出的消息用随机变量描述(意味着是单符号)

离散信源：信源输出的是单个符号的消息，且符号集的取值是有限的或者可数的，可用一维离散型随机变量X来描述信源的输出。符号集 $A:\{a_1,a_2,\cdots,a_q\}$

离散信源的数学模型（离散型概率空间）：

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

$$\sum_{i=1}^{q} P(a_i) = 1$$

2.1 基本概念

1. 信源输出的消息用**随机变量**描述

连续信源：信源输出的是单个符号的消息，但且其可能出现的消息数是不可数的无限值，即输出消息的符号集A的取值是连续的，可用一维连续型随机变量X来描述。

连续信源的数学模型（连续型概率空间）：

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

$$\int_a^b p(x)dx = 1$$

2.1 基本概念

2. 信源输出的消息用随机矢量描述（意味着多符号，即时间离散）

很多实际的信源输出的消息往往是由一系列符号序列所组成。如中文自然语言文字、离散化的平面灰度图像。

信源输出的消息是按一定概率选取的符号序列，因此可以把这种信源输出的消息看做时间或者空间上离散的一系列随机变量，即随机矢量。

可用N维随机矢量X=(X1 X2…XN)来描述，也称**随机序列**。N可为有限正整数或可数的无限值。

为了便于分析，假设信源输出的是平稳的随机序列，即序列的统计特性与时间的推移无关。

2.1 基本概念

2. 信源输出的消息用**随机矢量**描述

离散平稳信源：信源输出的随机序列 $X=(X_1X_2\cdots X_N)$ ，每个随机变量 $X_i(i=1,2,\cdots,N)$ 都是取值离散的离散型随机变量，且随机矢量X的各维概率分布都与时间起点无关。

如中文语言文字，离散化平面灰度图像等。

连续平稳信源：信源输出的随机序列 $X=(X_1X_2\cdots X_N)$ ，每个随机变量 $X_i(i=1,2,\cdots,N)$ 都是取值连续的连续型随机变量，且随机矢量X的各维概率密度函数都与时间起点无关。

如电压、电流信号在时间上取样后的信源。

2. 信源输出的消息用随机矢量描述

平稳信源又可以分为**无记忆信源**和**有记忆信源**。

在某些简单的离散平稳信源情况下，信源先后发出的一个个符号彼此是统计独立的，则N维随机矢量的联合概率分布满足：

$$P(X)=P(X_1X_2\cdots X_N)=P_1(X_1)P_2(X_2)\cdots P_N(X_N)$$

因为信源是平稳的，则个变量 X_i 的一维概率分布都相同，则

$$P(X)=P(X_1X_2\cdots X_N)=\prod_{i=1}^{N}P(X_i)$$

输出具有这种概率分布的信源，称为**离散无记忆信源**。

2.1 基本概念

2. 信源输出的消息用**随机矢量**描述

离散无记忆信源X所输出的随机矢量***X***所描述的信源称为**离散无记忆信源X的N次扩展信源。**

离散无记忆信源的N次扩展信源的数学模型是X信源空间的N重空间：

$$\begin{bmatrix} X^N \\ P(\alpha_i) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{q^N} \\ P(\alpha_1) & P(\alpha_2) & \cdots & P(\alpha_{q^N}) \end{bmatrix}$$

$$\alpha_i = (a_{i_1} a_{i_2} \cdots a_{i_N}) \qquad 0 \le P(\alpha_i) \le 1$$

$$P(\alpha_i) = P(a_{i_1} a_{i_2} \cdots a_{i_N}) = \prod_{k=1}^{N} P(a_{i_k})$$

$$\sum_{i=1}^{q^N} P(\alpha_i) = \sum_{i=1}^{q^N} \prod_{k=1}^{N} P(a_{i_k}) = 1$$

2. 信源输出的消息用随机矢量描述

一般情况下，信源在不同时刻发出的符号之间是相互依赖的。也就是信源输出的平稳随机序列X中，各随机变量Xi之间是有依赖的，这种信源称为**有记忆信源**。

需要在N维随机矢量的联合概率分布中引入条件概率分布来说明之间的关联。

当记忆长度为$m+1$时，称这种有记忆信源为**m阶马尔科夫信源**。也就是信源每次发出的符号只与前m个符号有关，与更前面的符号无关。

$$p(x_i \mid x_{i-1}x_{i-2}\cdots x_{i-m}\cdots) = p(x_i \mid x_{i-1}x_{i-2}\cdots x_{i-m})$$

2. 信源输出的消息用随机矢量描述

$$p(x_i \mid x_{i-1}x_{i-2}\cdots x_{i-m}\cdots) = p(x_i \mid x_{i-1}x_{i-2}\cdots x_{i-m})$$

当m=1时，可用简单的马尔科夫链描述，此时的条件概率就转化为**状态转移概率**。

$$P_{ji} = p(x_i = a_i \mid x_{i-1} = a_j)$$

能用马尔科夫链描述的信源称为**马尔科夫信源**。

3. 信源输出的消息用随机过程描述(时间连续，取值连续)

更一般地说，实际信源输出的消息常常是时间和取值都是连续的。同时，在某一固定时刻，他们的可能取值又是连续的和随机的。对于这种信源输出的消息，可用随机过程来描述。称这类信源为**随机波形信源**（也称**随机模拟信源**）。

分析一般随机波形信源比较复杂和困难。根据取样定理，只要是时间上或者频率上受限的随机过程，都可以把随机过程用一系列时间（频率）域上离散的随机序列来处理。

2.1 基本概念

随机变量 X 描述信源输出的消息：

- **离散信源**：可能输出的消息数是有限或可数的，每次只输出一个消息，即两两不相容

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} a_1, & a_2, & \cdots, & a_q \\ P(a_1), & P(a_2), & \cdots, & P(a_q) \end{bmatrix} \quad \sum_{i=1}^{q} P(a_i)=1 \text{ 且 } 0 \leqslant P(a_i) \leqslant 1 \ (i=1,2,\cdots q)$$

↑ 分层（量化）

- **连续信源**：可能输出的消息数是无限的或不可数的，每次只输出一个消息

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} (a,b) \\ p(x) \end{bmatrix} \text{或} \begin{bmatrix} \mathbf{R} \\ p(x) \end{bmatrix} \quad \int_a^b p(x)\mathrm{d}x=1 \text{ 或} \int_{\mathbf{R}} p(x)\mathrm{d}x=1$$

随机矢量 X 信源输出的消息是按一定概率选取的符号序列（时间或空间离散的随机序列）：

- **非平稳信源**：描述信源输出消息的随机序列 X 是非平稳随机序列——马尔可夫信源

  输出的随机序列 X 中各随机变量之间有依赖关系。但记忆长度有限，并满足马尔可夫链的条件式（2.8）

- **平稳信源**：描述信源输出消息的随机序列 X 是平稳的随机序列：
  - **离散平稳信源**：输出的随机序列 $X=(X_1X_2\cdots X_N)$ 中每个随机变量 $X_i(i=1,2,\cdots,N)$ 取值是离散的，并且随机矢量 X 的各维概率分布不随时间平移而改变：
    - **无记忆信源**

      输出的平稳随机序列 X 中各随机变量彼此统计独立（无依赖关系）。若每个随机变量 X_i 取值于同一离散概率空间 X，若满足式(2.5)则为离散无记忆 X 的 N 次扩展信源。若随机变量 X_i 满足式(2.9)则为连续平稳无记忆信源
    - **有记忆信源**

      输出的随机序列 X 中各随机变量之间有依赖关系，若记忆长度有限，则为有限记忆信源

    ↑ 量化 分层

  - **连续平稳信源**：输出的随机序列 $X=(X_1X_2\cdots X_N)$ 中每个随机变量 $X_i(i=1,2,\cdots,N)$ 取值是连续的，并且随机矢量 X 的各维概率密度函数不随时间平移而改变

↑ 取样定理

随机过程 $\{x(t)\}$：随机波形信源（随机模拟信源）——信源输出的消息是时间（或空间）上和取值上都是连续的函数

2.1 基本概念

举例：离散无记忆信源

例如扔骰子，每次试验结果必然是 1～6 点中的某一个面朝上。可以用一个离散型随机 变量 X 来描述这个信源输出的消息。

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} x_1, & x_2, & x_3, & x_4, & x_5, & x_6 \\ 1/6, & 1/6, & 1/6, & 1/6, & 1/6, & 1/6 \end{bmatrix}$$

并满足 $\sum_{i=1}^{6} P(x_i) = 1$

在实际情况中，存在着很多这样的信源、例如投硬币、书信文字、计算机的代码、电报符号、阿拉伯数字码等等。这些信源输出的都是单个符号（或代码）的消息，它们符号集的取值是有限的或可数的。

举例：连续无记忆信源

例如：随机取一干电池，测电压值作为输出符号,该信源每次输出一个符号,但符号的取值是在[0，1.5]之间的所有实数，每次测量值是随机的，可用连续型随机变最X 来描述。

在有些情况下，可将符号的连续幅度进行量化使其取值转换成有限的或可数的离散值．也就是把连续信源转换成离散信源来处理。

举例：有记忆信源

一般情况下，信源在不同时刻发出的符号之间是相互依赖的，也就是信源输出的平稳随机序列X 中，各随机变量Xi 之间是有依赖的。如在汉字序列中前后文字的出现是有依赖的，不能认为是彼此不相关的。

例如布袋中有100 个球，80 个红球，20 个白球。先取出一个球，记下颜色后不放回布袋，接着取另一个。而在取第二个球时布袋中的红球、白球概率已与取第一个球时不同，此时的概率分布与第1 个球的颜色有关：

若第1 个球为红色，则取第二个球时的概率p(a1)=0.79，p(a2)=0.2

若第1 个球为白色，则取第二个球时的概率p(a1)=0.8，p(a2)=0.19

总结：离散信源的统计特性

①离散消息是从有限个符号组成的符号集中选择排列组成的随机序列（组成消息的信息源的符号个数是有限的）。

一篇汉文，尽管文章优美，词汇丰富，一般所用的词都是从常用 10 000 个汉字里选出来的。一本英文书，不管它有多厚，总是从26 个英文字母选出来，按一定词汇结构，文法关系排列起来的。

②在形成消息时，从符号集中选择各个符号的概率不同。

对大量的由不同符号组成的消息进行统计，结果发现符号集中的每一个符号都是按一定的概率在消息中出现的。例如在英文中，每一个英文字母都是按照一定概率出现的，符号“e”出现最多，“z”出现最少。

总结：离散信源的统计特性

③组成消息的基本符号之间有一定的统计相关特性。

每一个基本符号在消息中，通常是按一定概率出现的。但在具体的条件下还应作具体的分析。如英文中出现字母h 的概率约为 0.04305，这是指对大量文章统计后所得到的字母h出现的可能性；但是，h 紧接在t 后面的单词特别多，紧接在y 后面的单词几乎没有。也就是说，在不同的具体条件下，同一个基本符号出现的概率是不同的。因此，做信源统计工作时，不仅需做每个独立的基本符号出现的概率，还须做前几个符号出现后下一个符号为某一基本符号的概率。

提纲

2.1 基本概念

2.2 离散信源熵的基本概念和性质

2.3 多符号离散平稳信源熵

- 单符号离散信源的数学模型

单符号离散信源：信源输出的是单个符号的消息，且符号集的取值是有限的或者可数的，可用一维离散型随机变量X来描述信源的输出。符号集$X:\{a_1,a_2,\cdots,a_q\}$

离散信源的数学模型（离散型概率空间）：

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

$$\sum_{i=1}^{q} P(a_i) = 1$$

2.2 离散信源熵的基本概念和性质

自信息与条件自信息

问题： 信源发出某一符号后，它提供了多少信息量?

在通信的一般情况下，接受者所获取的信息量，在数量上等于通信前后不确定性的消除的量。（最重要的结论之一）

A. 不确定度的定义：

随机事件的**不确定度**在**数量上**等于它的**自信息量。**

B. 自信息量的定义：

任意随机事件的自信息量定义为该事件发生概率的对数的负值。

设随机事件 x_i 发生的概率为 $p(x_i)$，那么它的自信息定义为：

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

2.2 离散信源熵的基本概念和性质

问题：不确定度和自信息量的区别

A. 两者的单位相同，但含义却不相同。

B. 具有某种概率分布的随机事件不管发生与否，都存在不确定度，不确定度表征了该事件的特性，而自信息量是在该事件发生后给予观察者的信息量。

C. 一个出现概率接近于 1 的随机事件，发生的可能性很大，所以它包含的不确定度就很小；反之，一个出现概率很小的随机事件，很难猜测在某个时刻它能否发生，所以它包含的不确定度就很大；若是确定性事件，出现概率为 1，则它包含的不确定度为 0。

2.2 离散信源熵的基本概念和性质

★**自信息量的单位**

自信息量的单位与所取的对数的底有关。

- 通常以2为底，信息量单位为比特(bit)【最常用】
- 若取自然对数(e), 单位为奈特(nat)

  $1nat=\log_2 e$(约为1.433bit)

- 若取10为底，单位为哈特(det)

  $1det=\log_2 10$(约为3.322bit)

★例: 自信息量计算

1. 一个以等概率出现的二进制码元（0，1）所包含的自信息量为：

$$I(0)=I(1)=-\log_2(1/2)=\log_2 2=1\text{bit}$$

2. 若是一个 m 位的二进制数，因为该数的每一位可从 0, 1 两个数字中任取一个，因此有 2^m 个等概率的可能组合。所以 $I=-\log_2(1/2^m)=m$ bit，就是需要 m 比特的信息来指明 这样的二进制数。

★例: 自信息量计算

设在甲袋中放入n个不同阻值的电阻，阻值分别为1……n，如果随机地取出一个，并对取出的电阻的阻值进行事先猜测，其猜测的困难程度相当于概率空间的不确定性。当被告知“取出的电阻阻值为i”所获信息量是多少?

概率空间为：$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}$

假设取出的电阻值是等概率的，即 $p(x_i) = \dfrac{1}{n}$

当被告知“取出的电阻阻值为i”所获信息量：

$$I(x_i) = -\log_2 p(x_i) = \log_2 n$$

2.2 离散信源熵的基本概念和性质

假如放入n(n+1)/2个电阻，其中阻值为1欧的放1个，阻值为n欧的放n个。当被告知“取出的电阻阻值为i”所获信息量是多少?

概率空间为：$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}$

取出电阻值是i的概率 $p(x_i) = \dfrac{i}{\frac{1}{2}n(n+1)}$

此时，当被告知“取出阻值为i的电阻”所获信息量：

$$I(x_i) = -\log_2 p(x_i) = \log_2 \frac{n(n+1)}{2i}$$

“取出1欧电阻”和“取出n欧电阻”，哪个的信息量大呢?

★联合自信息

定义：二维联合集 XY 上的元素 $(x_i y_i)$ 的联合自信息量定义为

$$I(x_i y_i) \triangleq -\log p(x_i y_i)$$

$x_i y_i$ 为积事件，

$p(x_i y_i)$ 为元素 $(x_i y_i)$ 的二维联合概率。

★条件自信息量

定义：联合集 XY 中，对事件 x_i 和 y_i，事件 x_i 在事件 y_i 给定的条件下的条件自信息量，定义为：

$$I(x_i|y_j) = log\frac{1}{p(x_i|y_j)}$$

条件自信息量也是非负值。

例： 在一个正方形棋盘上共有64个方格，如果甲将一粒棋子随意放在棋盘中的某方格，让乙猜棋子位置。

①将方格按顺序编号，猜棋子所在方格序号。

②将方格分别按行、列编号，甲将棋子所在行号（或列号）告诉乙，在让乙猜位置。

例：在一个正方形棋盘上共有64个方格，如果甲将一粒棋子随意放在棋盘中的某方格，让乙猜棋子位置。

①将方格按顺序编号，猜棋子所在方格序号。

棋子为二维等概率分布， $p(x_i)=\dfrac{1}{64}.$

$$\therefore I(x_i)=-\log p(x_i)=-\log_2\frac{1}{64}=6\ \ bit.$$

2.2 离散信源熵的基本概念和性质

例： 在一个正方形棋盘上共有64个方格，如果甲将一粒棋子随意放在棋盘中的某方格，让乙猜棋子位置。

②将方格分别按行、列编号，甲将棋子所在行号（或列号）告诉乙，在让乙猜位置。

元素 x_i 相对于 y_i 的条件自信息量：

$$I(x_i \mid y_i) = -\log_2 p(x_i \mid y_i) = -\log_2 \frac{p(x_i y_i)}{p(y_i)}$$

$$= -\log_2 \frac{1/64}{1/8} = 3 \ \ bit.$$

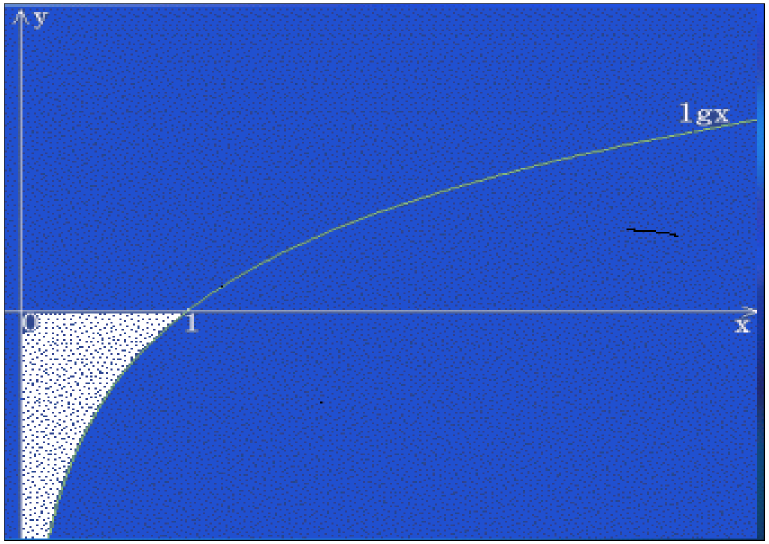
2.2 离散信源熵的基本概念和性质

★自信息量的性质

（1）自信息量为非负值。

当 $p(x_i)=1$ 时，$I(x_i)=0$；

当 $p(x_i)=0$ 时，$I(x_i)=\infty$。



（2）自信息量 $I(x_i)$ 是 $p(x_i)$ 的单调递减函数

若 $p(x_1)>p(x_2)\,,I(x_1)<I(x_2)$

2.2 离散信源熵的基本概念和性质

★**自信息量的性质**

$$I(a_ib_j)=-\log p(a_ib_j)=-\log p(a_i)p(b_j/a_i)=I(a_i)+I(b_j/a_i)$$
$$=-\log p(b_j)p(a_i/b_j)=I(b_j)+I(a_i/b_j)$$

（3）可加性

$$\boxed{I(x_i,y_j)}=\boxed{I(x_i/y_j)}+I(y_j)$$

两者相互独立时，有$I(x_i,y_j)=I(x_i)+I(y_j)$

- 联合自信息量：$I(x_i,y_j)$为事件x_i和y_j一起出现带来的信息量

$$I(x_i,y_j)=\log\frac{1}{p(x_i,y_j)}$$

- 条件自信息量：$I(x_i|y_j)$为在事件y_j已经出现的条件下事件x_i发生带来的信息量

$$I(x_i|y_j)=log\frac{1}{p(x_i|y_j)}$$

2.2 离散信源熵的基本概念和性质

■ 信息熵、条件熵、联合熵

问题：自信息(只是)表征了信源中各个符号的不确定度，但是一个信源总是包含多个符号消息，每个符号消息又有未必相同的先验概率分布，如何衡量一个信源总体的不确定度?

平均不确定度（平均自信息）：

$$H(X)=\sum p(x_i)I(x_i)$$

A. 平均不确定度*H(X)* 是**非负量**。

B. 平均不确定度*H(X)* 的定义公式与热力学中的熵表示形式相同，因此又把*H(X)* 称为**信源*X* 的熵(信息熵)**。

C. 熵是在**平均意义**上来表征**信源的总体特性**的，可以表征在信源输出前，信源的平均不确定度。

2.2 离散信源熵的基本概念和性质

信息熵、条件熵、联合熵

问题：自信息(只是)表征了信源中各个符号的不确定度，但是一个信源总是包含多个符号消息，每个符号消息又有未必相同的先验概率分布，如何衡量一个信源总体的不确定度?

平均不确定度（平均自信息）：

$$H(X)=\sum p(x_i)I(x_i)$$

A. 平均不确定度*H(X)*是**非负量**。

B. 平均不确定度*H(X)*的定义公式与热力学中的熵表示形式相同，因此又把*H(X)*称为**信源*X*的熵(信息熵)**。

C. 熵是在**平均意义**上来表征**信源的总体特性**的，可以表征在信源输出前，信源的平均不确定度。

2.2 离散信源熵的基本概念和性质

■ 信息熵、条件熵、联合熵

★离散信源熵（平均不确定度/平均信息量/平均自信息量）

定义：集合 X 上，随机变量 $I(x_i)$ 的数学期望定义为平均自信息量，又称作集合 X 的信息熵。

$$H(X)=\sum_i p(x_i)I(x_i)=-\sum_i p(x_i)\log p(x_i)$$

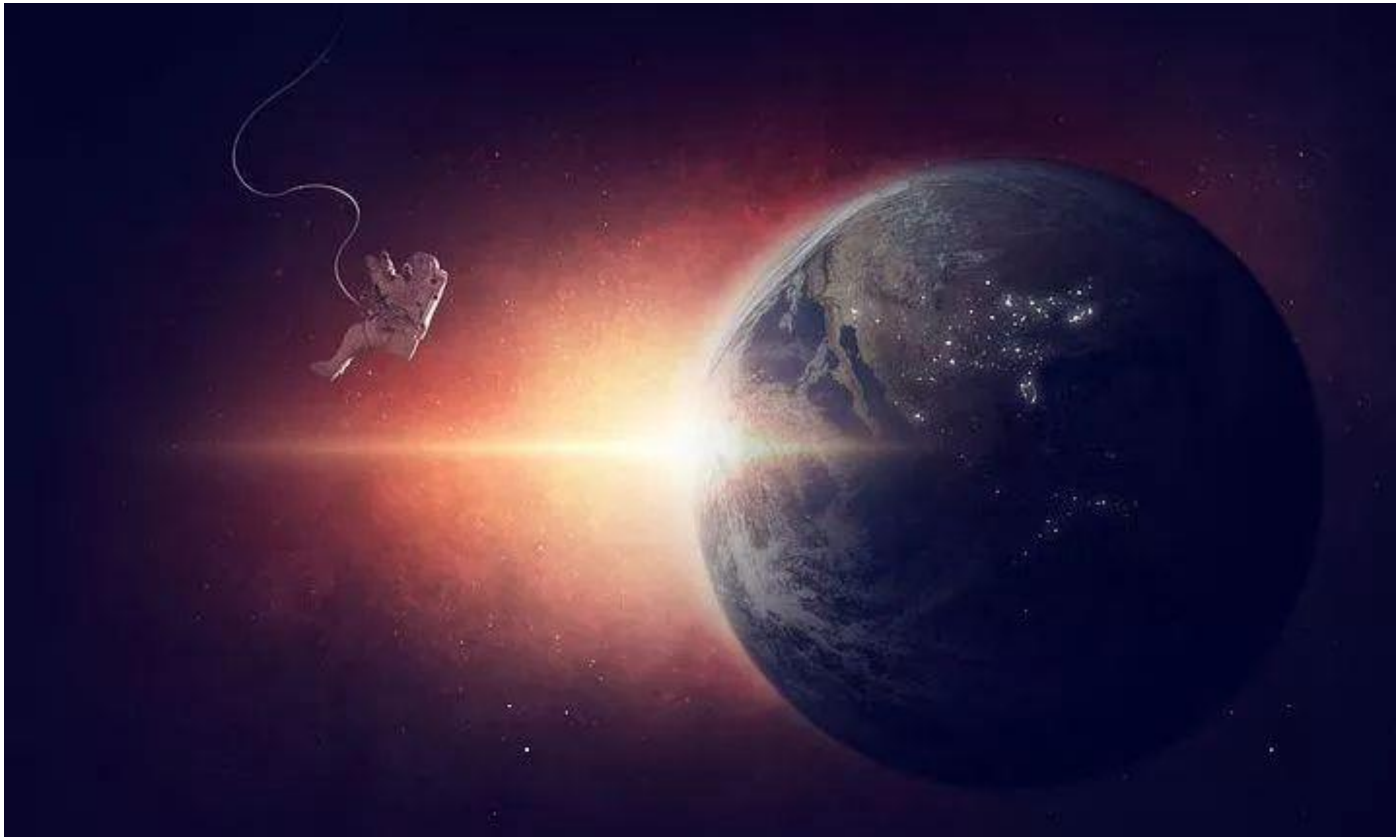
①集合X的平均自信息量表示集合X中事件出现的平均不确定性，每出现一个事件平均给出的信息量。

②平均自信息量的表示式和统计物理学中热熵的表示相似。

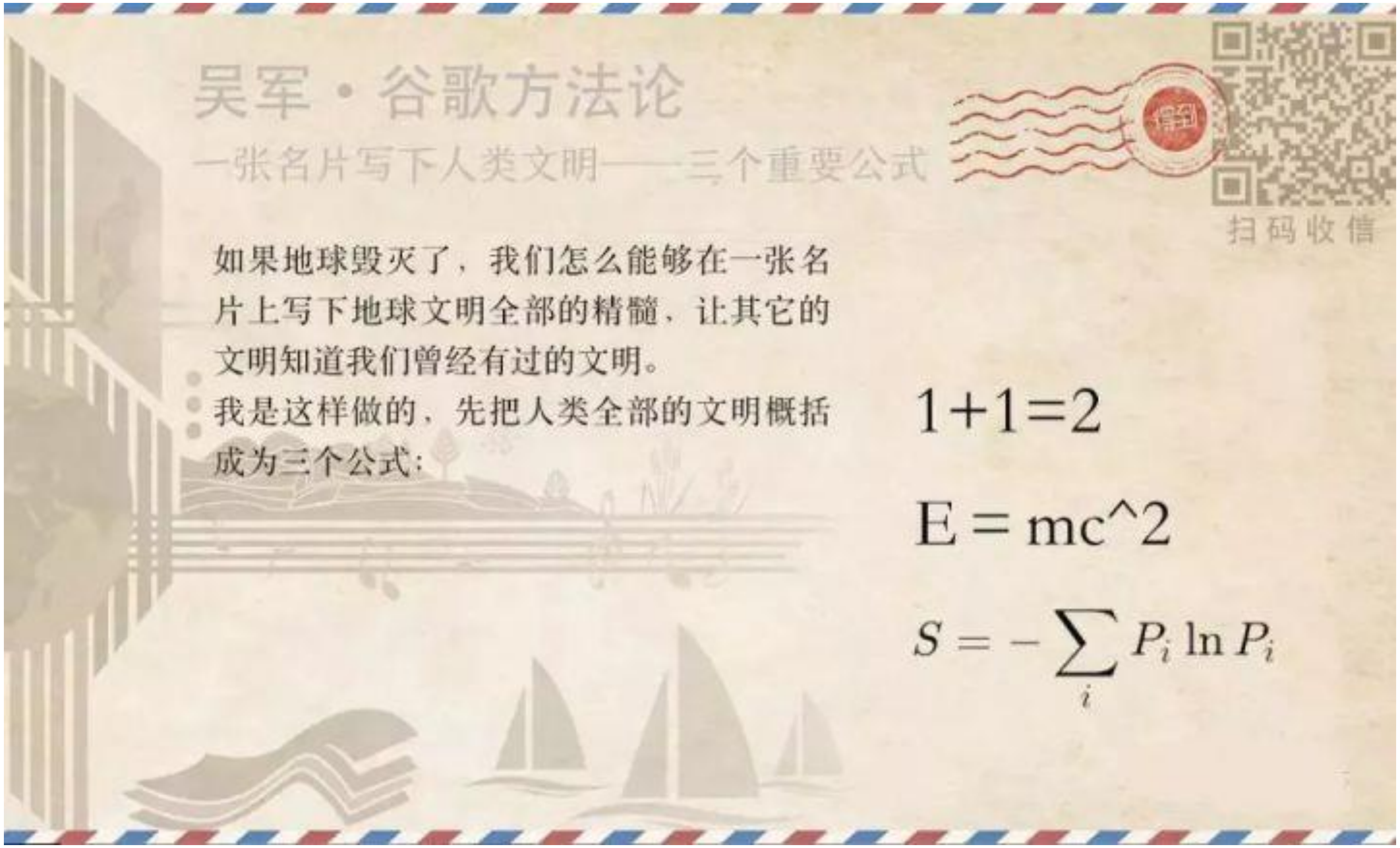
③若一个事件概率为0，无法提供任何信息，故定义0log0=0，即零概率事件的信息熵为0。信息熵的单位取决于对数所用的底。

2.2 离散信源熵的基本概念和性质

人活着就是在对抗熵增定律，生命以负熵为生。——薛定谔



2.2 离散信源熵的基本概念和性质



拓展片段：为什么熵增定律如此重要

因为它揭示了宇宙的终极演化规律。这个规律包括我们所有生命和非生命的演化规律，生命里又包含着个人和群体的演化规律。非生命，比如物质总是向着熵增演化，屋子不收拾会变乱，手机会越来越卡，耳机线会凌乱，热水会慢慢变凉，太阳会不断燃烧衰变……直到宇宙的尽头——热寂。

生命与个人，比如自律总是比懒散痛苦，放弃总是比坚持轻松，变坏总是比变好容易。少有人能做到自我管理，大多数人作息不规律，饮食不规律，学习不规律。生命与群体，比如大公司的组织架构会变得臃肿，员工会变得官僚化，整体效率和创新能力也会下降；封闭的国家会被世界淘汰。这些所有的现象都可以用一个定律来解释——熵增定律。

为什么懒散非常容易？
因为它符合熵增定律。

熵：系统中的无效能量，用以度量一个系统的“内在混乱程度”。

2.2 离散信源熵的基本概念和性质

- 信息熵、条件熵、联合熵

★离散信源熵（平均不确定度/平均信息量/平均自信息量）

设包含n个消息的集合 X，概率空间

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_0 & x_1 & \cdots & x_{n-1} \\ p(x_0) & p(x_1) & \cdots & p(x_{n-1}) \end{bmatrix}$$

每个消息的概率相等，均为1/n，选取对数底为n，由熵定义

$$H_n(X) = -\sum_{i=1}^{n} \frac{1}{n} \log_n \frac{1}{n} = 1$$

可以说此集合X包含了1个n进制单位的信息量。

（使用1个n进制数就可以表示此集合的消息。）

2.2 离散信源熵的基本概念和性质

- 信息熵、条件熵、联合熵

例：

电视屏幕上约有500*600=3*10^{5}个格点，按每点有10个不同的灰度等级考虑，则共能组10^{3*10^5}个不同画面。按等概率计算，平均每个画面可提供的信息量是多少?

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) = -\log_2 10^{-3\times10^5}$$

$$= 3\times10^5 \times 3.32 bit$$

2.2 离散信源熵的基本概念和性质

- 信息熵、条件熵、联合熵

例：

有一篇千字文章，假定每字可从万字表中任选，则共有不同的千字文 $N = 10000^{1000} = 10^{4000}$，仍按等概率计算，平均每篇千字文可提供的信息量为？

$$H(X) = \log_2 N = 4 \times 10^3 \times 3.32$$

$$\approx 1.3 \times 10^4 \, bit$$

“一个电视画面”平均提供的信息量远超“一篇千字文”提供的信息量。

2.2 离散信源熵的基本概念和性质

- 信息熵、条件熵、联合熵

例:

设信源符号集 $X=\{x_1,x_2,x_3\}$ ，每个符号发生的概率 $p(x_1)=1/2, p(x_2)=1/4, p(x_3)=1/4$ ，则信源熵为?

$$H(X)=\frac{1}{2}\log_2 2+\frac{1}{4}\log_2 4+\frac{1}{4}\log_2 4=1.5bit/\text{符号}。$$

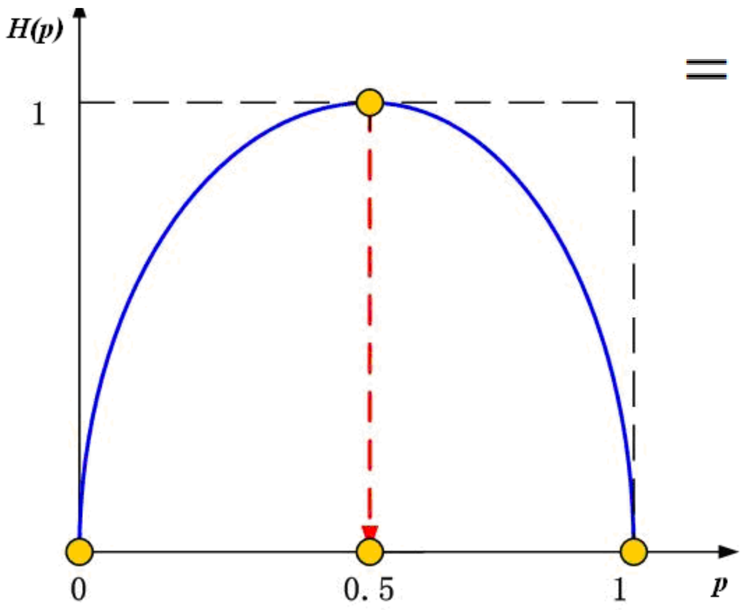
2.2 离散信源熵的基本概念和性质

■ 信息熵、条件熵、联合熵

该信源 X 输出符号只有两个，设为 0 和 1。输出符号发生的概率分别为 p 和 q，$p+q=1$。即信源的概率空间为 $\begin{pmatrix} X \\ P \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ p & q \end{pmatrix}$

则二元信源熵为：

$$H(X) = -p\log p - q\log q$$
$$= -p\log p - (1-p)\log(1-p) = H(p)$$



2.2 离散信源熵的基本概念和性质

■ 信息熵、条件熵、联合熵

熵函数的数学性质：

（1）对称性：概率矢量的每个分量可以任意变更位置，而熵值不变。

（2）非负性：$H(P)>=0$

（3）扩展性：增加一个概率近似于0的事件，熵值不变。

（4）可加性：当X、Y独立，$H(XY)=H(X)+H(Y)$

（5）强可加性： $H(XY)=H(X)+H(Y|X)$

（6）极值性：等概率场的平均不确定性为最大。$H(X)<=\log2(n)$

（7）确定性：只要有一个必然事件，则其余为不可能事件，熵为0

（8）上凸性：H是概率分布严格凸的。

■ 信息熵、条件熵、联合熵

定义：联合集 XY上，条件自信息量 $I(y|x)$ 的概率加权平均值定义为**条件熵**。其定义式为：

$$H(Y|X)=\sum_{XY}p(x_iy_j)I(y_j|x_i)$$

上式称为联合集XY中，集Y相对于集X的条件熵。

条件熵还可以写为：

$$H(Y|X)=-\sum_{XY}p(x_iy_j)\log p(y_j|x_i)$$

注意：

①式中求和的范围包括XY二维空间中的所有点。

②条件熵是用联合概率p(xy)，而不是用条件概率p(y|x)加权平均。

2.2 离散信源熵的基本概念和性质

- 信息熵、条件熵、联合熵

定义：联合集XY上，每对元素 $x_i y_j$ 的自信息量的概率加权平均值定义为**联合熵**。定义式为：

$$H(XY)=\sum_{XY}p(x_i y_j)I(x_i y_j)$$

联合熵又可以写为：

$$H(XY)=-\sum_{XY}p(x_i y_j)\log p(x_i y_j)$$

联合熵又称为共熵。

- 信息熵、条件熵、联合熵

联合熵与条件熵的关系：

1) $H(XY) = H(X) + H(Y|X)$

2) $H(XY) = H(Y) + H(X|Y)$

证明： 1） 由 $p(x_i y_j) = p(x_i)p(y_j \mid x_i) = p(y_j)p(x_i \mid y_j)$

得 $\sum_{i,j} p(x_i y_j) = \sum_i p(x_i) = \sum_j p(y_j)$

$$H(XY) = \sum_{i,j} p(x_i y_j) I(x_i y_j) = -\sum_{i,j} p(x_i y_j) \log p(x_i y_j)$$

$$= -\sum_{i,j} p(x_i y_j) \log p(x_i) p(y_j \mid x_i)$$

$$= -\sum_{i,j} p(x_i y_j) \log p(x_i) - \sum_{i,j} p(x_i y_j) \log p(y_j \mid x_i)$$

$$= -\sum_{i} p(x_i) \log p(x_i) - \sum_{i,j} p(x_i y_j) \log p(y_j \mid x_i)$$

$$= H(X) + H(Y \mid X)$$

2.2 离散信源熵的基本概念和性质

- 信息熵、条件熵、联合熵

联合熵与条件熵的关系：

3） $H(XY) \le H(X) + H(Y)$

注：

①若集合X与Y统计独立，则等号成立；

②当集合X和Y取自同一个符号集合Z时，H(X)=H(Y)=H(Z)，H(XY)<=2H(X)

③此性质可推广到N个概率空间的情况。

4） $H(Y \mid X) \le H(Y)$

等式成立的条件是当且仅当集合X和Y统计独立。

2.2 离散信源熵的基本概念和性质

■ 信息熵、条件熵、联合熵

例：设某系统的输入符号集为X=(x_1，x_2，x_3，x_4，x_5)，输出符号集为Y=(y_1，y_2，y_3，y_4)。输入符号与输出符号的联合分布为

$$
\begin{array}{c|cccc}
 & y_1 & y_2 & y_3 & y_4 \\
\hline
x_1 & 0.25 & 0 & 0 & 0 \\
x_2 & 0.10 & 0.30 & 0 & 0 \\
x_3 & 0 & 0.05 & 0.10 & 0 \\
x_4 & 0 & 0 & 0.05 & 0.10 \\
x_5 & 0 & 0 & 0.05 & 0
\end{array}
$$

计算H(XY), H(X), H(Y), H(Y|X), H(X|Y)。

提纲

2.1 基本概念

2.2 离散信源熵的基本概念和性质

2.3 多符号离散平稳信源熵

2.3 多符号离散平稳信源熵

■ 多符号离散平稳信源

多符号离散平稳信源，可用N 维随机矢量$X=(X_1\ X_2 \ldots X_N)$来描述，也称**随机序列**。N可为有限正整数或可数的无限值。

为了便于分析，假设信源输出的是平稳的随机序列，即序列的统计特性（各维联合概率）与时间的推移无关。

一维平稳：　$P(X_i)=P(X_j), i \neq j$

二维平稳：　$P(X_i X_{i+1})=P(X_j X_{j+1}), i \neq j$

… … …

N维平稳：　$P(X_i \cdots X_{i+N})=P(X_j \cdots X_{j+N}), i \neq j$

$N \to \infty$　这种信源称为（完全）离散平稳信源。

- 多符号离散平稳信源

离散信源的数学模型（离散型概率空间）：

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_q \\ P(a_1) & P(a_2) & \cdots & P(a_q) \end{bmatrix} \qquad \sum_{i=1}^{q} P(a_i) = 1$$

多符号离散信源的数学模型（离散型概率空间）：

$$\begin{bmatrix} \mathbf{X} \\ P(\alpha_i) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{q^N} \\ P(\alpha_1) & P(\alpha_2) & \cdots & P(\alpha_{q^N}) \end{bmatrix}$$

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

且 $X_i \in \{a_1, a_2, \cdots, a_q\}$

$$\alpha_i = (a_{i_1} a_{i_2} \cdots a_{i_N}) \qquad 0 \le P(\alpha_i) \le 1$$

$$P(\alpha_i) = P(a_{i_1} a_{i_2} \cdots a_{i_N})$$

$$\sum_{i=1}^{q^N} P(\alpha_i) = 1$$

2.3 多符号离散平稳信源熵

- 多符号离散平稳无记忆信源

多符号离散平稳无记忆信源：

输出的消息序列中各符号之间无相互依赖关系的信源。亦称为**单符号离散平稳无记忆信源X的*N*次扩展信源**。

序列的长度就是扩展次数。数学模型可表示为：

$$\begin{bmatrix} X^N \\ P(\alpha_i) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{q^N} \\ P(\alpha_1) & P(\alpha_2) & \cdots & P(\alpha_{q^N}) \end{bmatrix}$$

$$P(\alpha_i) = P(a_{i_1} a_{i_2} \cdots a_{i_N}) = \prod_{k=1}^{N} P(a_{i_k})$$

$$\sum_{i=1}^{q^N} P(\alpha_i) = \sum_{i=1}^{q^N} \prod_{k=1}^{N} P(a_{i_k}) = \sum_{i_1=1}^{q} P(a_{i_1}) \cdot \sum_{i_2=1}^{q} P(a_{i_2}) \cdots \sum_{i_N=1}^{q} P(a_{i_N}) = 1$$

离散无记忆信源X的N次扩展信源的概率空间也是完备集。

- 多符号离散平稳无记忆信源

N次扩展信源的熵：

N 次扩展信源 X^N 中共有 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_{q^N}$ 等 q^N 个消息（符号），每个符号 $\boldsymbol{\alpha}_i$ 都代表 X 的一 个 N 重序列（$a_{i_1},a_{i_2},\cdots,a_{i_N}$），其中（$i_1$，$i_2$，…，$i_N=1$，2，…，$q$）。

根据信息熵的定义，N次扩展信源的熵为：

$$H(\mathbf{X})=H(X^N)=-\sum_{X^N}P(\mathbf{X})\log P(\mathbf{X})$$

$$=-\sum_{X^N}P(\alpha_i)\log P(\alpha_i)$$

定理：**离散无记忆信源**X的N次扩展信源的熵等于离散信源X的熵的N倍，即

$$H(X^N)=NH(X).$$

2.3 多符号离散平稳信源熵

- **多符号离散平稳无记忆信源**

N 次扩展信源的熵：

例：假设有一离散无记忆信源X

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

① 求这个离散无记忆信源的二次扩展信源。（扩展信源的每个符号是信源X的输出长度为2的符号序列。）

② 求H(X)和H(X²)。

$$\begin{bmatrix} X^2 \\ P(\alpha_i) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 & \alpha_7 & \alpha_8 & \alpha_9 \\ a_1a_1 & a_1a_2 & a_1a_3 & a_2a_1 & a_2a_2 & a_2a_3 & a_3a_1 & a_3a_2 & a_3a_3 \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{16} & \frac{1}{16} & \frac{1}{8} & \frac{1}{16} & \frac{1}{16} \end{bmatrix}$$

- 多符号离散平稳有记忆信源

反映信源记忆特性的两种方法：

1. 用联合概率反映信源记忆特性
2. 用条件概率反映信源记忆特性

多符号离散平稳有记忆信源

二维信源（$N=2$）：

如何对二维离散平稳信源进行信息测度呢？

(1) 对其输出的随机序列每两个分组，此时新信源**X**=X_1X_2，其信源的概率空间为：

$$\begin{bmatrix} \mathbf{X} \\ P \end{bmatrix} = \begin{bmatrix} X_1 X_2 \\ P \end{bmatrix} = \begin{bmatrix} a_1a_1 & a_1a_2 & \cdots & a_qa_q \\ p(a_1a_1) & p(a_1a_2) & \cdots & p(a_qa_q) \end{bmatrix}$$

由熵定义，其联合熵为：

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

$H(X_1X_2)$称为X_1X_2的联合熵，表示原信源 **X** 输出的任意一对可能的消息的共熵。

- **多符号离散平稳有记忆信源**

二维信源（N=2）：

如何对二维离散平稳信源进行信息测度呢？

(2) 利用前后两个符号之间的依赖关系。在前一个符号X_1 = a_i时，信源下一个符号的平均不确定性为：

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

下一步对符号a_i求统计平均，即可得当前面一个符号已知时再输出后面一个符号的总的平均不确定性：

$$H(X_2 \mid X_1) = \sum_{i=1}^{q} P(a_i) H(X_2 \mid X_1 = a_i)$$

$$= -\sum_{i=1}^{q} \sum_{j=1}^{q} P(a_i) P(a_j \mid a_i) \log(a_j \mid a_i)$$

$$= -\sum_{i=1}^{q} \sum_{j=1}^{q} P(a_i a_j) \log(a_j \mid a_i)$$

■ 多符号离散平稳有记忆信源

二维信源（*N*=2）：

对于由两个符号组成的联合信源，有下列结论：

(1) $H(X_1X_2)=H(X_1)+H(X_2|X_1)=H(X_2)+H(X_1|X_2)$

(2) $H(X_1|X_2)\le H(X_1)\qquad H(X_2|X_1)\le H(X_2)$

信源联合熵等于，信源发出前一个符号的信息熵，加上前一个符号已知时信源发出下一个符号的条件熵。当前后符号无依存关系时，有下列推论：

$$H(X_1X_2)=H(X_1)+H(X_2)\qquad =2H(X)$$

$$H(X_1)=H(X_1|X_2)$$

$$H(X_2)=H(X_2|X_1)$$

- **多符号离散平稳有记忆信源**

二维信源（N=2）：

例：某二维离散平稳信源

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ \frac{11}{36} & \frac{4}{9} & \frac{1}{4} \end{bmatrix}$$

并设输出的符号只与前一个符号有关，即可用**联合概率**给出它们的关联关系如右表。

(1)假设无依赖，求信源 X 的信息熵；

(2)考虑符号依赖，求条件熵$H(X_2|X_1)$。

$$\begin{array}{cc} & X_2 \\ X_1 & \begin{bmatrix} & 0 & 1 & 2 \\ 0 & \frac{1}{4} & \frac{1}{18} & 0 \\ 1 & \frac{1}{18} & \frac{1}{3} & \frac{1}{18} \\ 2 & 0 & \frac{1}{18} & \frac{7}{36} \end{bmatrix} \end{array}$$

2.3 多符号离散平稳信源熵

- 多符号离散平稳**有记忆**信源

二维信源（N=2）：

例：某二维离散平稳信源

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ \frac{11}{36} & \frac{4}{9} & \frac{1}{4} \end{bmatrix}$$

并设输出的符号只与前一个符号有关，即可用**条件概率**给出它们的关联关系如右表。

(1)假设无依赖，求信源 X 的信息熵；

(2)考虑符号依赖，求条件熵$H(X_2|X_1)$；

$$\begin{array}{cc} & X_2 \\ X_1 & \begin{bmatrix} & 0 & 1 & 2 \\ 0 & \frac{9}{11} & \frac{2}{11} & 0 \\ 1 & \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ 2 & 0 & \frac{2}{9} & \frac{7}{9} \end{bmatrix} \end{array}$$

- **多符号离散平稳有记忆信源**

二维信源（$N=2$）：

例：某二维离散平稳信源

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ \frac{11}{36} & \frac{4}{9} & \frac{1}{4} \end{bmatrix}$$

$$\begin{array}{cc} & X_2 \\ X_1 & \begin{bmatrix} & 0 & 1 & 2 \\ 0 & \frac{1}{4} & \frac{1}{18} & 0 \\ 1 & \frac{1}{18} & \frac{1}{3} & \frac{1}{18} \\ 2 & 0 & \frac{1}{18} & \frac{7}{36} \end{bmatrix} \end{array}$$

联合概率

$$\begin{array}{cc} & X_2 \\ X_1 & \begin{bmatrix} & 0 & 1 & 2 \\ 0 & \frac{9}{11} & \frac{2}{11} & 0 \\ 1 & \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ 2 & 0 & \frac{2}{9} & \frac{7}{9} \end{bmatrix} \end{array}$$

条件概率

根据 $p(a_i,a_j) = p(a_i)\, p(a_j|\, a_i)$ ，

可由条件概率计算得联合概率$p(a_i\, a_j)$

信源 X 的信息熵为：$H(X) = -\sum_{i=0}^{2} p(a_i)\log p(a_i) = 1.543\ bit/$符号

当考虑符号之间有依赖时，计算得条件熵：

$$H(X_2|X_1) = -\sum_{i=0}^{2}\sum_{j=0}^{2} p(a_i a_j)\log p(a_j\,|\,a_i) = 0.872\ bit/\text{符号} \ < H(X)$$

2.3 多符号离散平稳信源熵

- 多符号离散平稳**有记忆**信源

二维信源（$N=2$）：

例：某二维离散平稳信源

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ \frac{11}{36} & \frac{4}{9} & \frac{1}{4} \end{bmatrix}$$

$$\begin{array}{cc} & X_2 \\ & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ X_1 \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{bmatrix} \frac{1}{4} & \frac{1}{18} & 0 \\ \frac{1}{18} & \frac{1}{3} & \frac{1}{18} \\ 0 & \frac{1}{18} & \frac{7}{36} \end{bmatrix} \end{array}$$

联合概率

$$\begin{array}{cc} & X_2 \\ & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ X_1 \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{bmatrix} \frac{9}{11} & \frac{2}{11} & 0 \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ 0 & \frac{2}{9} & \frac{7}{9} \end{bmatrix} \end{array}$$

条件概率

发二重符号序列的熵为：$H(X_1X_2) = -\sum_{i=0}^{2}\sum_{j=0}^{2} p(a_ia_j)\log p(a_ia_j) = 2.41\ bit/符号$

平均每个符号携带的信息量近似为：$H_2(X) = \frac{1}{2}H(X_1X_2) = 1.21\ bit/符号$

$$H_2(\mathrm{X}) < H_1(X)$$

2.3 多符号离散平稳信源熵

- 多符号离散平稳有记忆信源

N 维信源：

一般离散平稳有记忆信源中，符号的相互依赖关系往往不仅存在于相邻的两个符号之间，而且存在于更多的符号之间。

假设信源符号之间的依赖长度为N，则随机序列的概率可表示为：

$$P(\mathbf{X}=\mathbf{x}_i)=P(X_1=x_{i1},X_2=x_{i2},\cdots,X_N=x_{iN}) \qquad \mathbf{x}_i=(x_{i1},x_{i2},\cdots,x_{iN})$$

$$=P(X_1=x_{i1})P(X_2=x_{i2}\mid X_1=x_{i1})\cdots P(X_N=x_{iN}\mid X_1=x_{i1}X_2=x_{i2}\cdots X_{N-1}=x_{iN-1})$$

$$=P(x_{i1})P(x_{i2}\mid x_{i1})P(x_{i3}\mid x_{i1}^2)\cdots P(x_{iN}\mid x_{i1}^{N-1}) \qquad x_{i1}^k=x_{i1}x_{i2}\cdots x_{ik}$$

N 维信源熵：

$$H(\mathbf{X})=H(X_1\cdots X_N) \qquad =H(X_1\cdots X_{N-1})+H(X_N\mid X_1\cdots X_{N-1})$$

$$=H(X_1)+H(X_2\mid X_1)+H(X_3\mid X_1X_2)+\cdots+H(X_N\mid X_1X_2\cdots X_{N-1})$$

2.3 多符号离散平稳信源熵

- 多符号离散平稳信源

$$H_0(\mathbf{X}) \ge H_1(\mathbf{X}) \ge H_2(\mathbf{X}) \ge \cdots \ge H_\infty(\mathbf{X})$$

N 维信源：

平均符号熵：*N* 长的信源符号序列中，平均每个信源符号所携带的信息量，即

$$H_N(\mathbf{X}) = \frac{1}{N} H(X_1 X_2 \cdots X_N)$$

对于离散平稳信源，当$H_1(\mathbf{X})$小于无穷时，则具有以下几点性质：

① 条件熵 $H(X_N | X_1X_2\ldots X_{N-1})$ 随 N 的增加是非递增的；

$$H(X_N | X_1X_2\cdots X_{N-1}) \le H(X_{N-1} | X_1\cdots X_{N-2}) \le \cdots \le H(X_2 | X_1) \le H(X_1)$$

② N 给定时，平均符号熵大于等于条件熵；$H_N(\mathbf{X}) \ge H(X_N | X_1X_2\cdots X_{N-1})$

③ 平均符号熵随 N 的增加也是非递增的； $H_i(\mathbf{X}) \ge H_{i+1}(\mathbf{X})$

④ $H_\infty = \lim\limits_{N\to\infty} H_N(\mathbf{X})$ 存在，并且 $H_\infty = \lim\limits_{N\to\infty} H_N(\mathbf{X}) = \lim\limits_{N\to\infty} H(X_N | X_1X_2\cdots X_{N-1})$

称为离散平稳信源的**极限熵**或**极限信息量**。

2.3 多符号离散平稳信源熵

- 多符号离散平稳有记忆信源

N 维信源：

无记忆（可看做有记忆的特例）

平均符号熵： $H_N(\mathbf{X})=\frac{1}{N}H(X_1X_2\cdots X_N)$

$$=\frac{1}{N}NH(X)=H(X)$$ （单符号离散信源熵）

极限熵： $H_\infty=\lim_{N\to\infty}H_N(\mathbf{X})=H(X)$

2.3 多符号离散平稳信源熵

■ 马尔可夫信源

虽然马尔可夫信源是一个**非平稳信源**，但是当马尔可夫信源进入**稳定**状态后，可以看做一个平稳信源。

马尔可夫信源：若信源输出的符号序列和信源所处的状态满足以下两个条件：$X \in A = \{a_1, a_2, \cdots, a_q\}, S \in E = \{E_1, E_2, \cdots, E_J\}$

(1) 某一时刻信源符号的输出只与此刻信源所处的状态有关，而与以前的状态及以前的输出符号都无关，即

$$P(X_n = a_k \mid S_n = E_i, X_{n-1} = a_{k_1}, S_{n-1} = E_j, \cdots) = P(X_n = a_k \mid S_n = E_i) \xlongequal{\text{时齐}} P(a_k \mid E_i)$$

(2) 信源n时刻所处的状态由当前的输出符号和前一时刻（n-1）信源状态唯一决定，即

$$P(S_n = E_j \mid X_n = a_k, S_{n-1} = E_i) = \begin{cases} 0 & E_i, E_j \in E \\ 1 & a_k \in A \end{cases}$$

则此信源称为**马尔可夫信源**。

2.3 多符号离散平稳信源熵

■ 马尔可夫信源

m 阶马尔可夫信源：m 阶有记忆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，即

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1, a_2, a_3, \cdots, a_q \\ P(a_{k_{m+1}} \mid a_{k_1} a_{k_2} a_{k_3} \cdots a_{k_m}) \end{bmatrix} \quad (k_1, k_2, \cdots, k_m, k_{m+1} = 1, 2, \cdots, q)$$

并满足 $0 \le P(a_{k_{m+1}} \mid a_{k_1} a_{k_2} a_{k_3} \cdots a_{k_m}) \le 1$，且 $\sum_{k_{m+1}=1}^{q} P(a_{k_{m+1}} \mid a_{k_1} a_{k_2} a_{k_3} \cdots a_{k_m}) = 1$

则此信源 X 称为 **m 阶马尔可夫信源**。

当 m = 1 时，即任何时刻信源符号发生的概率只与前面一个符号有关，则称为**一阶马尔可夫信源**。

2.3 多符号离散平稳信源熵

■ 马尔可夫信源

一般马尔可夫信源的熵：平均符号熵的极限值，即

$$H_\infty = H_\infty(X) = \lim_{N\to\infty}\frac{1}{N}H(X_1X_2\cdots X_N)$$

信源处于某状态 E_i 时，输出一个信源符号所携带的平均信息量，即在状态 E_i 下输出一个符号的条件熵为

$$H(X\mid S=E_i) = -\sum_{k=1}^{q}P(a_k\mid E_i)\log P(a_k\mid E_i)$$

时齐、遍历的马尔可夫信源的熵：

$$H_\infty = \sum_{i=1}^{J}Q(E_i)H(X\mid E_i) = -\sum_{i=1}^{J}\sum_{k=1}^{q}Q(E_i)P(a_k\mid E_i)\log P(a_k\mid E_i)$$

$$\sum_{E_j\in E}Q(E_j)P(E_i\mid E_j) = Q(E_i) \qquad \sum_{E_i\in E}Q(E_i) = 1$$

■ 马尔可夫信源

时齐、遍历的 *m* 阶马尔可夫信源的熵：

$$H_\infty = H_{m+1}(X) = \sum_{i=1}^{J} Q(E_i) H(X \mid E_i) \qquad E_i = (a_{k_1} a_{k_2} \cdots a_{k_m})$$

$$\because H(X \mid E_i) = -\sum_{k=1}^{q} P(a_k \mid E_i) \log P(a_k \mid E_i)$$

$$= -\sum_{k_{m+1}=1}^{q} P(a_{k_{m+1}} \mid a_{k_1} a_{k_2} \cdots a_{k_m}) \log P(a_{k_{m+1}} \mid a_{k_1} a_{k_2} \cdots a_{k_m})$$

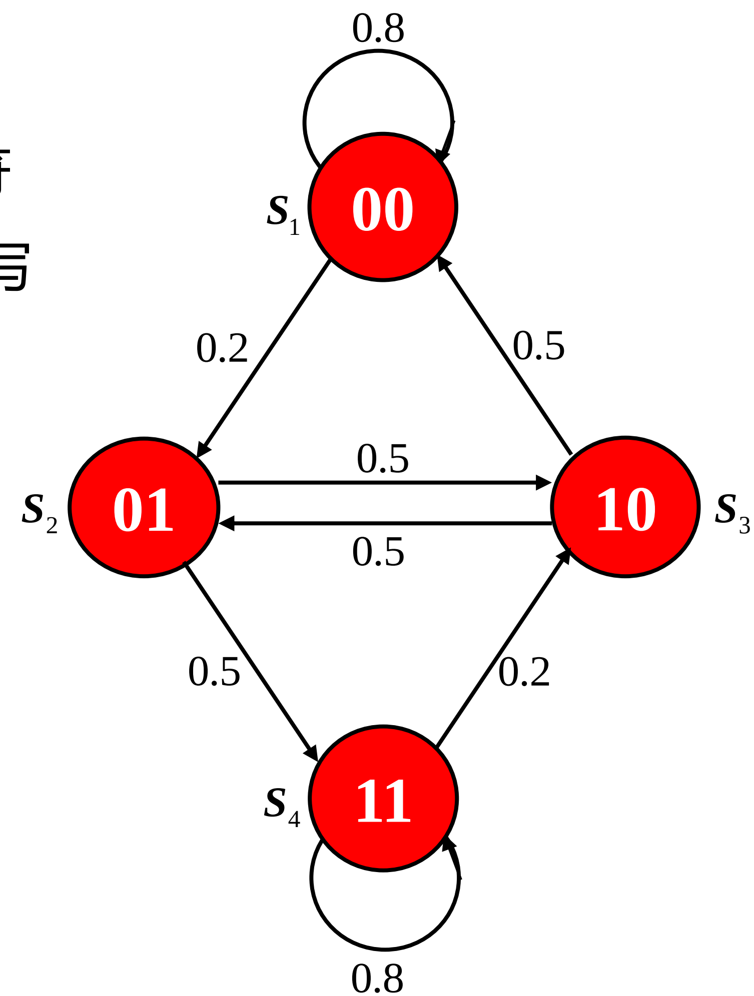
$$\therefore H_{m+1} = \sum_{i=1}^{J} Q(E_i) H(X \mid E_i)$$

$$= -\sum_{k_1=1}^{q} \sum_{k_2=1}^{q} \cdots \sum_{k_m=1}^{q} \sum_{k_{m+1}=1}^{q} Q(a_{k_1} a_{k_2} \cdots a_{k_m}) P(a_{k_{m+1}} \mid a_{k_1} a_{k_2} \cdots a_{k_m}) \log P(a_{k_{m+1}} \mid a_{k_1} a_{k_2} \cdots a_{k_m})$$

$$= H(X_{m+1} \mid X_1 X_2 \cdots X_m)$$

■ 马尔可夫信源

例：二元2 阶马尔可夫信源，信源符号集为{0,1}。信源状态图如右图所示，写出状态转移概率、输出符号的条件概率、条件熵、信源熵。



$$P(S_j\,|\,S_i)=\begin{bmatrix}0.8 & 0.2 & 0 & 0\\ 0 & 0 & 0.5 & 0.5\\ 0.5 & 0.5 & 0 & 0\\ 0 & 0 & 0.2 & 0.8\end{bmatrix}$$

$$P(0\,|\,S_1)=0.8, P(1\,|\,S_1)=0.2$$

$$P(0\,|\,S_2)=0.5, P(1\,|\,S_2)=0.5$$

$$P(0\,|\,S_3)=0.5, P(1\,|\,S_3)=0.5$$

$$P(0\,|\,S_4)=0.2, P(1\,|\,S_4)=0.8$$

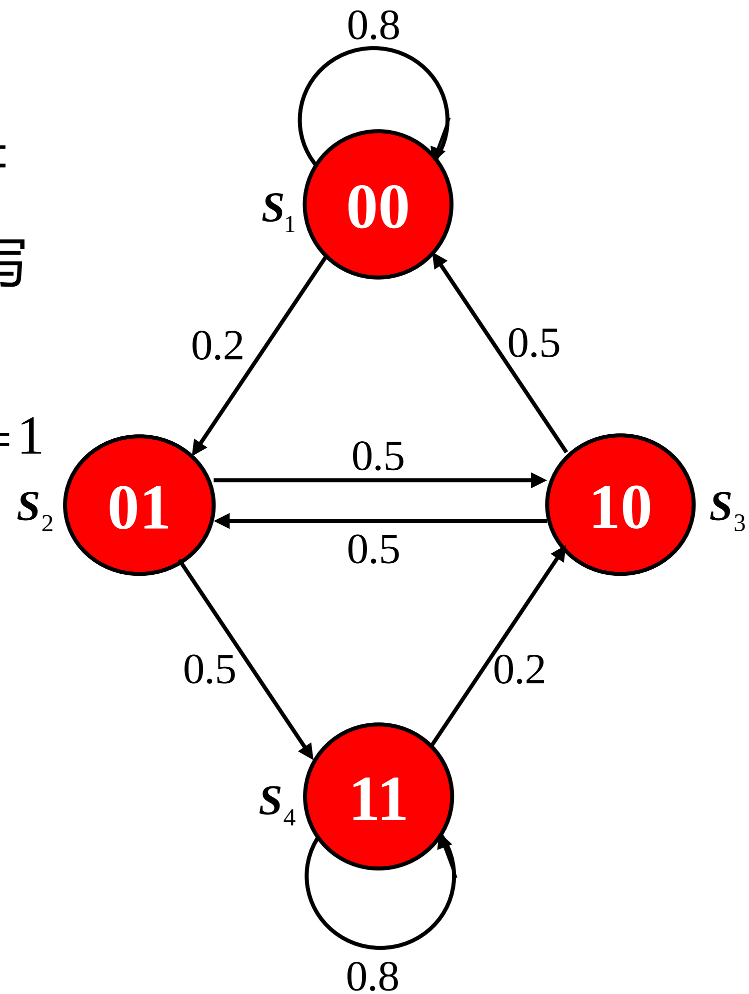
$$H(X\,|\,S_1)=-0.8\log 0.8-0.2\log 0.2=0.722\ bit/\text{符号} =H(X\,|\,S_4)$$

$$H(X\,|\,S_2)=H(X\,|\,S_3)=1\ bit/\text{符号}$$

2.3 多符号离散平稳信源熵

■ 马尔可夫信源

例：二元2 阶马尔可夫信源，信源符号集为{0,1}。信源状态图如右图所示，写出状态转移概率和输出符号的条件概率。



$$\sum_{E_j \in E} Q(E_j)P(E_i \mid E_j) = Q(E_i) \qquad \sum_{E_i \in E} Q(E_i) = 1$$

$$\begin{cases} 0.8Q_1 + 0.5Q_3 = Q_1 \\ 0.2Q_1 + 0.5Q_3 = Q_2 \\ 0.5Q_2 + 0.2Q_4 = Q_3 \\ 0.5Q_2 + 0.8Q_4 = Q_4 \end{cases}$$

$$Q_1 + Q_2 + Q_3 + Q_4 = 1$$

$$\Rightarrow Q_1 = \frac{5}{14}, Q_2 = \frac{1}{7}, Q_3 = \frac{1}{7}, Q_4 = \frac{5}{14}$$

$$H_{m+1} = \sum_i Q(S_i)H(X \mid S_i) = 0.801\ bit/\text{符号}$$

2.3 多符号离散平稳信源熵

信源相关性和冗余度

实际离散信源可能是非平稳的，其 H_∞ 不一定存在。

可以假定它是平稳的，用 H_∞ 来近似。然而，对于一般的平稳离散信源，求H_∞ 也是极为困难的。

可进一步假设它是 *m* 阶马尔可夫信源，用 *m* 阶马尔可夫信源的信息熵 H_{m+1} 来近似，近似程度取决于记忆长度 *m* 的大小。

若进一步简化，可假设信源是无记忆的，而信源符号有一定的概率分布，可用平均自信息 $H_1 = H(X)$ 来近似。

最后，可以假定是等概率分布的离散无记忆信源，用最大熵 $H_0 = \log q$ 来近似。

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

每个符号提供的平均自信息，随着符号间依赖关系长度的增大而减小。

2.3 多符号离散平稳信源熵

- 信源相关性和冗余度

熵的相对率：

一个信源实际的信息熵，与具有同样符号集的最大熵的比值，即

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

信源剩余度：等于1减去熵的相对率，即 $\gamma = 1 - \eta = 1 - \frac{H_\infty}{H_0}$

信息变差： $I_{0\infty} = H_0 - H_\infty$

信源剩余度的大小，能够反映离散信源输出的符号序列之间依赖关系的强弱。剩余度越大，表示信源的实际熵越小。

当剩余度等于0时，信源的信息熵就等于极大值 H_0 ，这表明符号之间不但统计独立无记忆，而且各个符号还是等概率分布。

2.3 多符号离散平稳信源熵

信源相关性和冗余度

举例：自然语言的剩余度

英文字母概率表：

字母	概率	字母	概率	字母	概率
空格	0.1859	I	0.0575	R	0.0484
A	0.0642	J	0.0008	S	0.0514
B	0.0127	K	0.0049	T	0.0796
C	0.0218	L	0.0321	U	0.0228
D	0.0317	M	0.0198	V	0.0083
E	0.1031	N	0.0574	W	0.0175
F	0.0208	O	0.0632	X	0.0013
G	0.0152	P	0.0152	Y	0.0164
H	0.0467	Q	0.0008	Z	0.0005

2.3 多符号离散平稳信源熵

- 信源相关性和冗余度

举例：自然语言的剩余度

对于实际英文字母组成的信源：

$$H_\infty = 1.4 \text{ (比特/符号)}$$

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

$$\gamma = 1 - 0.29 = 0.71$$

2.3 多符号离散平稳信源熵

总结

自信息：随机事件 a_i 的自信息 $I(a_i) = -\log P(a_i)$

信息熵：离散随机变量 $\boldsymbol{X}$ 的信息熵 $H(X) = -\sum_i P(a_i)\log P(a_i)$

熵函数的性质：对称性、确定性、非负性、扩展性、可加性、强可加性、递增性、极值性、上凸性。

离散无记忆信源的 N 次扩展信源的熵： $H(\mathbf{X}) = H(X^N) = NH(X)$

离散平稳信源的平均符号熵： $H_N(\mathbf{X}) = \dfrac{1}{N}H(X_1X_2\cdots X_N)$

离散平稳信源的条件熵： $H(X_N \mid X_{N-1}X_{N-2}\cdots X_1)$

离散平稳信源的极限熵： $H_\infty = \lim\limits_{N\to\infty} H_N(\mathbf{X}) = \lim\limits_{N\to\infty} H(X_N \mid X_{N-1}X_{N-2}\cdots X_1)$

■ 总结

离散平稳信源有：

$$H(X_N|X_1X_2\cdots X_{N-1})\le H(X_{N-1}|X_1\cdots X_{N-2})\le\cdots\le H(X_2|X_1)\le H(X_1)\le\log q=H_0$$

$$H(X_1X_2\cdots X_N)=\sum_{i=1}^{N}H(X_i|X_{i-1}\cdots X_1)$$

m 阶有记忆离散平稳信源的信息熵：

$$H_\infty=-\sum_{k_1=1}^{q}\sum_{k_2=1}^{q}\cdots\sum_{k_m=1}^{q}\sum_{k_{m+1}=1}^{q}P(a_{k_1}a_{k_2}\cdots a_{k_m})P(a_{k_{m+1}}|a_{k_1}a_{k_2}\cdots a_{k_m})\log P(a_{k_{m+1}}|a_{k_1}a_{k_2}\cdots a_{k_m})$$

总结

时齐、遍历马尔科夫信源的信息熵：

$$H_\infty = \sum_{i=1}^{J} Q(E_i) H(X \mid E_i) = -\sum_{i=1}^{J}\sum_{k=1}^{q} Q(E_i) P(a_k \mid E_i) \log P(a_k \mid E_i)$$

式中$Q(E_i)$满足： $\sum_{E_j \in E} Q(E_j) P(E_i \mid E_j) = Q(E_i)$ $\sum_{E_i \in E} Q(E_i) = 1$

时齐、遍历的 m 阶马尔科夫信源的信息熵：

$$H_{m+1} = \sum_{i=1}^{J} Q(E_i) H(X \mid E_i)$$

式中 $Q(E_i) = Q(a_{k_1} a_{k_2} \cdots a_{k_m})$ 满足：

$$\sum_{E_j \in E} Q(E_j) P(E_i \mid E_j) = Q(E_i) \qquad \sum_{E_i \in E} Q(E_i) = 1$$

信源的剩余度： $\gamma = 1 - \eta = 1 - \dfrac{H_\infty}{H_0}$