

6 时间序列的复杂度和熵

6	时间序列的复杂度和熵	1
6.1	引言	2
6.2	时间序列符号化方法	2
6.3	热力学熵—克劳修斯(Clausius)熵 ^[5, 6]	4
6.4	统计熵—玻尔兹曼(Boltzmann)熵 ^[5]	6
6.5	信息熵—先农(Shannon)熵 ^[5, 7]	6
6.6	Kolmogrov熵和K2 熵	8
6.7	非广延熵-Tsallis熵	9
6.8	近似熵—Approximate Entropy	10
6.9	样本熵—Sample Entropy	12
6.10	多尺度熵—Multiscale Entropy	14
6.11	Lempel-Ziv复杂度	16
6.12	相似指数分析 ^[24]	17
6.13	海杂波的复杂度和熵	18
6.14	本章小结	18
6.15	后记	19

6.1 引言

在对动力学结构进行动力学特性分析时，首先对结构进行非线性检测，并判断该系统的非线性因素是否可以忽略非常重要。一个系统的熵和复杂度隐含着整个系统运动规律，他们的物理意义直接和系统单一变量的性质相联系。在计算熵和复杂度时，一般同近年来由符号动力学理论发展出来的符号时间序列分析方法相联系。符号时间序列分析实质是结合混沌时间序列分析和信息理论的一种分析方法。其实质是先对序列值的符号化，符号化后编码，编码后计算熵或复杂度等特征。数据符号化的基本思想就是在几个可能值上对时间序列进行离散化，把许多可能值的数据序列变换为仅有几个互不相同值的符号序列。这是一个“粗粒化”(Coarse-grained)过程，这一过程能够捕获大尺度的特征，从而降低动力学噪声和测量噪声的影响。而且，复杂度和熵不仅适用于混沌时间序列，也适用于性质未知时间序列的特征提取和分析。

本文主要研究了时间序列的常用符号化过程，编码序列或原始时间序列的熵和复杂度的计算。本章中各节主要内容如下：6.2节介绍了时间序列的几种常用符号化方法；6.3节—6.10节介绍了各种熵的定义、计算方法和应用；6.11节为Lempel-Ziv复杂度的计算、应用和物理意义分析；6.12介绍了一种计算两时间序列相似度的方法；6.13节为海杂波的熵和复杂度及其应用；6.14节为本章小结。

6.2 时间序列符号化方法

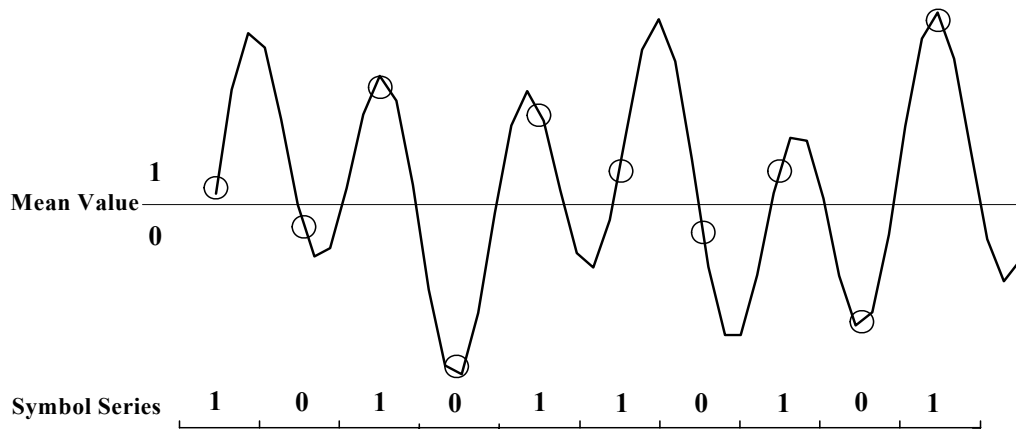
对于非线性时间序列复杂度的研究，时间序列粗粒化的方法有：均值法、一阶差分法、移动均值一阶差分法等，这些方法共同的特征是把时间序列简化成容易处理的符号序列^[1-4]。目前采用较多的是如图 1 所示的均值化方法，即对于某一时间序列

$x(i), i = 1, 2, \dots, n$ ，令：

$$X_{ave} = \frac{1}{n} \sum_{i=1}^n x(i) \quad (6.1)$$

X_{ave} 为这一时间序列的平均值， $S = (s_1 s_2 \dots s_n)$ 是一个与时间序列 $x(i), i = 1, 2, \dots, n$ 长度相同的空符号串。当时间序列中的某一元素 $x(i) > X_{ave}$ 时，则取符号 s_i 为“1”，否则为“0”，由此建立的符号序列为 S 。由于只考虑大于平均值和小于等于平均值这两种状态，所以在动力学结构分析中，这种划分方法对噪声不敏感。而且只能体现出序列的整体特性，不能体现出序列的局部特性。

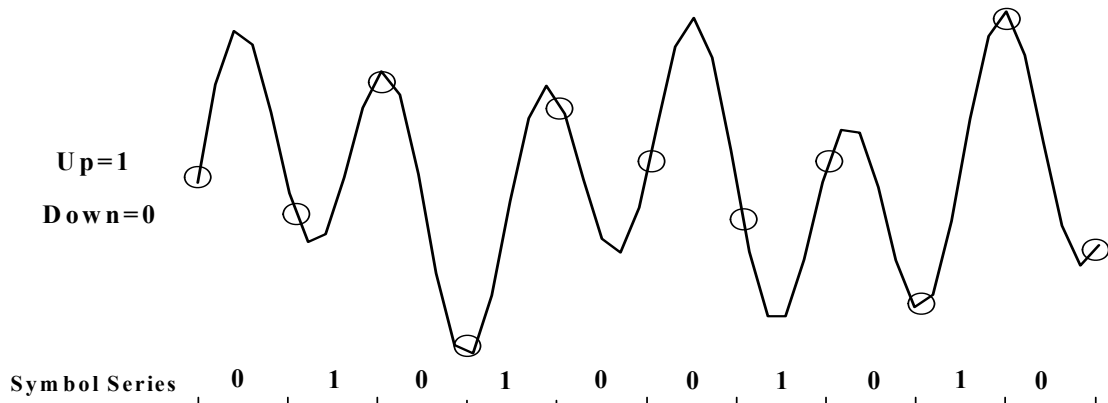
Static transformation (fixed partition)



图表 6-1: 连续模拟序列的静态均值符号序列化方法示意图

显然，以上取均值处理的方法存在明显缺陷，大量的细节信号被抹掉。为了体现时间序列的细节信息，可以使用一阶差分方法符号化时间序列。对一时间序列，任意点与其前一点进行比较，大于等于前者，记为 1，否则记为 0，得到的符号序列中的所有“01”或“10”对应一次局部极值。如图表 6-2所示，这种方法在很大程度上提取了信号中的细节成分。

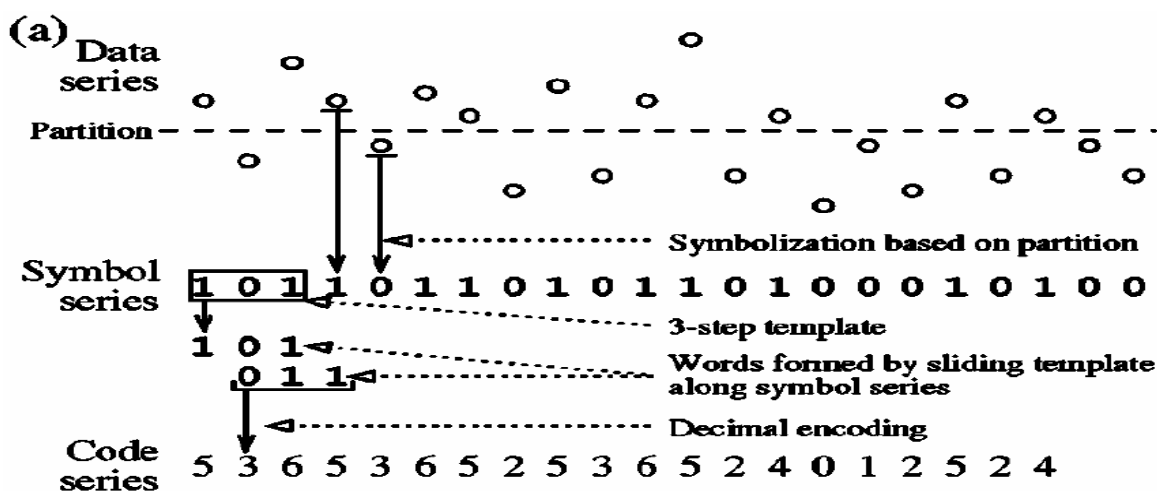
Dynamic transformation (first difference)



图表 6-2: 连续模拟序列的一阶差分符号序列化方法示意图

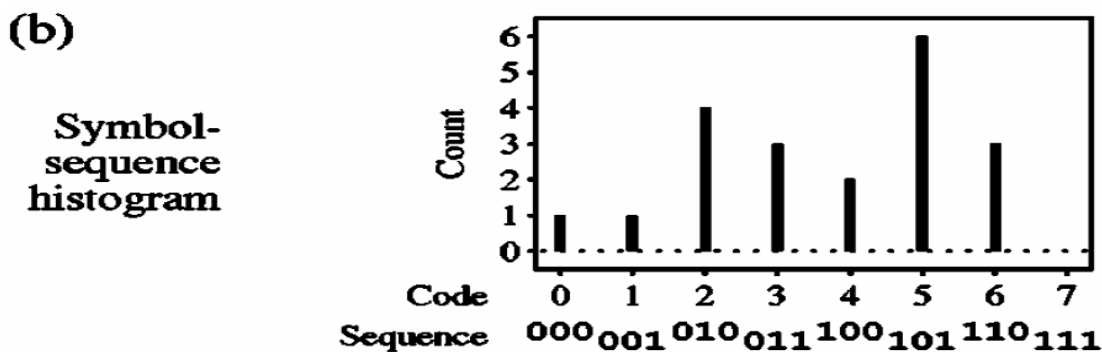
另外，为了综合上面两种方法的优点，也可动态符号化时间序列。动态符号化是指对长度为 N 的时间序列取长度为 L 的矩形窗，对窗内序列取均值，窗内大于均值的点置于 1，小于均值的点置为 0，这样将一长度为 N 的非“0, 1”序列变换为 $N-L+1$ 个长度为 L 的“0, 1”符号序列，然后对 $N-L+1$ 个序列分别计算其复杂度或熵，从而得到复杂度随时间的动态变化。使用这种方法简单方便，不仅考虑了全局特征也提取了局部细节信号。

图表 6-3给出了使用均值化方法根据原始数据序列产生符号序列和符号编码序列的示意图，它很好地解释了由原始时间序列到生成符号编码的整个过程。在原始数据被转化为符号序列形式后，就要提取其特征量。一个非常实用的方法就是选择一个标准长度的 L ， L 个连续的符号组成一个字，每一个字被编码为十进制数，形成一个新的序列。上述过程与时间延迟嵌入类似，即利用离散的符号代替连续的原始测量值。



图表 6-3: 根据原始数据序列产生符号序列和符号码序列示意图

在符号编码序列产生以后，可以将出现的字的频率作为时间序列分析的一个指标，如所示图表 6-4就是符号编码序列的直方图，它直观地描述了每一个字出现的概率，因此为后面各种熵和复杂度的计算提供了基础。



图表 6-4: 符号编码序列的直方图

6.3 热力学熵—克劳修斯 (Clausius) 熵^[5, 6]

1854 年克劳修斯 (Clausius) 发表了《力学的热理论的第二定律的另一种形式》的论文，给出了可逆循环过程中热力学第二定律的数学表示形式：

$$\oint \frac{dQ}{T} = 0 \quad (6.2)$$

引入了一个新的后来定名为熵的态参量 S 。1865 年他发表了《力学的热理论的主要方程

之便于应用的形式》的论文，把这一新的态参量正式定名为熵。并将上述积分推广到更一般的循环过程，得出了热力学第二定律的数学表示形式：

$$\oint \frac{dQ}{T} \leq 0 \quad (6.3)$$

其中等号对应于可逆过程，不等号对应于不可逆过程。在热力学过程中，为了描述不可逆过程的单向性，引入态函数熵 S 。对于只有热接触的封闭系统，熵变化 dS 为

$$dS \geq \frac{dQ}{T} \quad (6.4)$$

或

$$S_a - S_b \geq \int_a^b \frac{dQ}{T} \quad (6.5)$$

上式中的 a 、 b 表示始末两个状态， S_a 、 S_b 为始末两个状态的熵， dQ 为系统吸收的热量， T 为热源的温度，可逆过程中 T 也是系统的温度。

当系统经历绝热过程或系统是孤立的时候，系统与外界没有热量交换：

$$dQ = 0 \quad (6.6)$$

因此有

$$dS \geq 0 \quad (6.7)$$

或

$$S_a - S_b \geq 0 \quad (6.8)$$

即有“熵增加原理”：孤立系统或绝热过程是不可逆系统，熵总是增加的。

可逆过程只能沿熵增加的方向进行，最终达到平衡状态。由于系统的无序程度与熵值存在对应关系，所以由“熵增加原理”导致的孤立系统达到的平衡态是均匀无序状态。由此定义的熵称克劳修斯熵，或热力学熵。熵是一个态函数，是热力学宏观量。对绝热过程和孤立系统中所发生的过程，由熵函数的数值可判定过程进行的方向和限度。

对于非平衡的开放系统，Prigogine 等进一步推广熵的概念，将 dS 表示为

$$dS = d_e S + d_i S \quad (6.9)$$

其中 $d_e S$ 表示与外界交换能量与物质时，引起熵变换，称为熵交换； $d_i S$ 表示系统内部的不可逆过程产生的熵。根据式(6.9)，对于开放系统有

$$d_i S \geq 0 \quad (6.10)$$

$$\text{若 } d_e S < 0, \text{ 且有 } |d_e S| > |d_i S| \quad (6.11)$$

则由式(6.9)知，系统有

$$dS < 0 \quad (6.12)$$

即开放系统的熵的总变化是负的，故开放系统有可能形成有序结构。有序结构的形成，一方面需要外界条件，这是必要的；另一方面需要内部条件，这是充分的。由外界的能量或物质维持一种空间或时间上的有序结构，称为耗散系统。

Prigogine 等还证明, 相对于定态熵的二级偏差(超熵)总是负的, 即

$$\delta^2 S < 0 \quad (6.13)$$

超熵对时间的导数称为超熵产生, 其值可取不同的符号, 据稳定性理论有以下三种情形:

- (1) 当 $\frac{d}{dt}(\delta^2 S) < 0$ 时, 定态是渐近稳定的
- (2) 当 $\frac{d}{dt}(\delta^2 S) > 0$ 时, 定态是不稳定的
- (3) 当 $\frac{d}{dt}(\delta^2 S) = 0$ 时, 定态是临界稳定的

由系统的状态变量和动力学参量决定取三种定态情况中的哪一种。在接近平衡态区域, 已知定态应该是渐近稳定的; 当变量变化使状态远离平衡时, 系统可能越过(3)的情形进入(2)的情形。此时均匀无序的定态变成不稳定的, 从而通过分叉过程造成对称的破缺, 并出现有序结构(如极限环和混沌), 即有序结构只可能在远离平衡态的非线性区域出现。

6.4 统计熵—玻尔兹曼(Boltzmann)熵^[5]

1896年, 玻尔兹曼(Boltzmann)建立了熵 S 和系统宏观态所对应的可能微观态数目 W (即热力学概率) 的联系:

$$S \propto \ln W \quad (6.14)$$

1900年普朗克(Planck)引进了比例系数 k —称为玻尔兹曼常量, 写出了玻尔兹曼—普朗克公式:

$$S = k \ln W \quad (6.15)$$

所定义的熵称为玻尔兹曼熵, 或统计熵。由此玻尔兹曼表明了熵 S 是同热力学概率 W 相联系的, 揭示了宏观态与微观态之间的联系, 指出了热力学第二定律的统计本质: 熵增加原理所表示的孤立系统中热力学过程的方向性, 正相应于系统从热力学概率小的状态向热力学概率大的状态过渡, 平衡态热力学概率最大, 对应于 S 取极大值的状态; 熵自发地减小的过程不是绝对不可能的, 不过概率非常小而已。

6.5 信息熵—先农(Shannon)熵^[5, 7]

通常认为, 熵概念泛化的最早领域是信息论领域。Shannon 在 1948 年定义了信息熵作为某件事件不确定度的最度。信息是信息论中一个重要的基本概念, 从信息接收者的角度来看, “信息是能够减少或消除不确定性的一种客观存在和能动过程”。不确定性是一切客观事物的属性, 当接收者未收到某事物的信息之前, 对事物的认识是不确定的,

一旦收到信息，就了解了该事物，即可以说对它的认识上的不确定性减少或消除了。那么我们应该怎么样度量信息呢？关于信息的度量，首先是 Shannon 在通信领域进行研究的，他建立了由信源(信息的发送者)，信道与编、译码器，信宿(信息的接收者)组成的通信系统模型，而信息量就是解除信源不确定度所需的信息度量，或者说获得这样大的信息量后，信源不确定度就被解除。Shannon 以概率和数理统计为工具，提出度量信息量的数学公式，对于离散型信源，当它由若干随机事件所组成时，随机事件出现的不确定度用其出现的概率来描述。事件出现的可能性愈小，概率就愈小，而所含信息量却愈大;相反，事件出现可能性愈大，概率就愈大，而所含信息量却愈小，所以信息量可用事件出现概率的单调减函数来表示，若各事件概率分布不等，则信源提供的平均信息量 H 即为：

$$H = -\sum_{i=1}^n p_i \log p_i \quad (6.16)$$

上式就是信息熵，也称 Shannon 熵。其中， p_i 为第 i 个符号出现的概率，且有

$$\sum_{i=1}^n p_i = 1 \quad (6.17)$$

由符号编码序列直方图可得到反映符号序列总体特征的统计量—Shannon 熵。在二进制符号化规则下，可定义长度为 L 的 Shannon 熵为

$$H_s(L) = -\frac{1}{L \ln 2} \sum P_{s_1 s_2 \dots s_L} \ln(P_{s_1 s_2 \dots s_L}) \quad (6.18)$$

式中， $P_{s_1 s_2 \dots s_L}$ 为长度为 L 的短符号序列经二进制至十进制变换得到序列编码后，在直方图中出现的各个符号编码所对应的概率。

通过式(6.16)可以看出，信息熵是从平均意义上描述信源的总体特征的，它表示信源输出的平均信息量，它表征信源的平均不确定性。控制论的主要创立者Wiener曾说：“信息量是一个可以看做概率的量的对数的负数，它实质上是熵。”这实际上推广了原来热力学熵的概念。信息熵越大，说明信号源发出的平均信息量越大，即信源的不确定度大；而热力学概念的熵，是系统不确定性或无序度的量度。可见，信息熵与热力学概念的熵在性质上是等同的，所以我们可以说信息熵是热力学的自然推广。通过这种概念的扩展，我们可以看出：熵不仅不必与热力学过程相联系，而且也不必与微观分子运动相联系，它可以成为系统状态不确定程度的度量，这个状态可以是热力学的，也可以不是热力学的；可以是微观的，也可以是宏观的。

在熵的泛化当中，我们得到信息熵的表达式为式(6.16)，它把概率分布函数与熵联系起来，即每一个概率分布对应着唯一的一个信息熵值。也就是说，一个信息熵值表达

了一个概率分布的不确定性程度。进一步推广，如果用满足式(6.17)条件的非概率分布函数来代替上式中的概率分布函数，结果会如何？事实上，从数学上讲，只要是满足了式(6.17)的分布函数，无论它是否是概率分布函数，并没有这方面的要求。所以，由式(6.17)的各种不同形式，发展起来各种熵的定义形式。

6.6 Kolmogrov 熵和 K2 熵

1958年Kolmogrov定义了测度熵，称为Kolmogrov熵(又称信息维数)，用来度量系统运动的混乱或无序的程度。随后Sinai进行了改进，所以又称为Kolmogrov-Sinai熵，简称为K-S熵。考虑一个 n 维动力系统，将它的相空间分割成一个个边长为 ε 的 n 维立方体盒子，对于状态空间的一个吸引子和一条落在吸引域的轨道 $x(t)$ ，取时间间隔为一很小量 τ ，令 $P(i_0, i_1, \dots, i_d)$ 表示起始时刻轨道在第 i_0 个格子中， $t = \tau$ 时在第 i_d 个格子中的联合概率，则Kolmogorov熵定义为：

$$K = -\lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{d \rightarrow \infty} \frac{1}{d\tau} \sum_{i_0 \dots i_d} P(i_0, i_1, \dots, i_d) \ln P(i_0, i_1, \dots, i_d) \quad (6.19)$$

而 q 阶Renyi熵定义为

$$K_q = -\lim_{\tau \rightarrow 0} \lim_{l \rightarrow 0} \lim_{d \rightarrow \infty} \frac{1}{d\tau} \frac{1}{q-1} \log_2 \sum_{i_0 \dots i_d} P^q(i_0, i_1, \dots, i_d) \quad (6.20)$$

由于K-S熵的计算非常复杂，早期的定义主要是直接从定义入手，进行了一些研究，这些研究虽然很有意思，但不很成功。Crassbreger和Procaccia于1983年提出了相关维数的概念^[8]，并定义了相关熵(Correlation Entropy)的概念来逼近Kolmogorov熵^[9]，简写为 K_2 熵， K_2 熵和关联积分 $C_d^2(l)$ 存在如下关系：

$$K_2 = -\lim_{\tau \rightarrow 0} \lim_{r \rightarrow 0} \lim_{d \rightarrow \infty} \frac{1}{d\tau} \log_2 C_d^2(r) \quad (6.21)$$

对离散时间序列，固定延迟时间 τ ，式(6.21)为

$$K_2 = -\lim_{r \rightarrow 0} \lim_{d \rightarrow \infty} \frac{1}{d\tau} \log_2 C_d^2(r) \quad (6.22)$$

关联积分 $C_d^2(r)$ 在 $r \rightarrow 0$ 时与 r 存在如下关系：

$$\lim_{r \rightarrow 0} C_d^2(r) \propto r^{D_2} \quad (6.23)$$

其中， D_2 为关联维数， r 为尺度。结合式(6.22)和式(6.23)有

$$\lim_{r \rightarrow \infty} \lim_{d \rightarrow \infty} C_d^2(r) = r^{D_2} d^{-K_2 d \tau} \quad (6.24)$$

$$\text{即为} \quad \lim_{r \rightarrow \infty} \lim_{d \rightarrow \infty} \log_2 C_d^2(r) = D_2 \log_2 r - K_2 d \tau \quad (6.25)$$

若暂时不考虑式(6.25)左边的极限，重构分别为 d 和 $d+m$ 维下，由(6.25)式有 d 维情况下：

$$\log_2 C_d^2(r) = D_2 \log_2 r - K_2 d \tau \quad (6.26)$$

$$d+m \text{ 维情况下:} \quad \log_2 C_{d+m}^2(r) = D_2 \log_2 r - K_2 (d+m) \tau \quad (6.27)$$

将上面两式相减可以得到

$$K_2 = \lim_{l \rightarrow 0} \lim_{d \rightarrow \infty} \frac{1}{m \tau} \log_2 \frac{C_d^2(r)}{C_{d+m}^2(r)} \quad (6.28)$$

在一般情况下， K_2 是 K_1 的一个很好的估计，可以证明，在典型情况下 **K-S** 熵在数值上趋近于 K_2 熵。 K_2 熵的优点在于能直接地、容易地从信号中估计得到，具体的计算方法可以通过计算关联维来实现。根据前面的讨论，在嵌入维按等间隔 m 不断增加的情况下，在无标度区间内，由式(6.28)作等斜率线性回归，可同时得到关联维 D_2 和 Kolmogorov 熵的稳定估计^[10]。

6.7 非广延熵-Tsallis 熵

1988 年，为了解决非广延系统的问题 Tsallis 受多重分形的启示，提出非广延熵的概念^[11, 12]：

$$S_q = \frac{k(1 - \sum_{i=1}^m p_i^q)}{(q-1)}, \quad q \in R \quad (6.29)$$

其中， k 是 Boltzman 常数，通常取 1。以这种非广延熵为基础建立的统计力学称为非广延统计力学，或广义统计力学，而 Boltzmann-Gibbs 统计力学则作为非广延统计力学在 $q \rightarrow 1$ 时的极限情况被包括在内。 q 为非广延参数，它刻画了系统的非广延程度，反映在当系统有两个独立的子系统 A 和 B 组成时，系统的熵 $S_q(A+B)$ 满足下面的非广延可加性：

$$\frac{S_q(A+B)}{k} = \frac{S_q(A)}{k} + \frac{S_q(B)}{k} + (1-q) \frac{S_q(A)}{k} \frac{S_q(A)}{k} \quad (6.30)$$

当 $q \neq 1$ 时，包含两个独立子系统的系统熵就不等于两个系统熵的和。

- (1) $q < 1$ 时，系统对应着超广延性
- (2) $q = 1$ 时，系统对应着广延性
- (3) $q > 1$ 时，系统对应着次广延性

可以通过找到合适的通过非广延参数 q (q 不等于1时)来很好的描述复杂性既不是规则也不是完全混沌或随机的一类运动。事实上, 既不是规则也不是完全混沌或随机的运动是无处不在的。

Tsallis利用非广延熵函数极值的办法得出了广义的热力学统计的分布形式, 即Tsallis分布。Tsallis分布是在下面两个约束条件下获得的:

$$\begin{aligned} \int_{-\infty}^{+\infty} p(x)dx &= 1 \\ \int_{-\infty}^{+\infty} |x^2 - \delta^2| [p(x)]^q dx &= 0 \end{aligned} \quad (6.31)$$

根据最大熵原理引入lagrange乘数 α, β

$$\begin{aligned} S(y) &= \frac{1}{q-1} (1 - \int_{-\infty}^{+\infty} d(\frac{x}{\delta}) [\delta y]^q) + \alpha \int_{-\infty}^{+\infty} y dx + \beta \int_{-\infty}^{+\infty} (x^2 - \delta^2) y^q dx \\ &= \frac{1}{q-1} + \int_{-\infty}^{+\infty} [-\frac{\delta^{q-1}}{q-1} y^q + \alpha y + \beta(x^2 - \delta^2) y^q] dx \end{aligned} \quad (6.32)$$

为了简化, 表示成下式

$$L(x, y) = -\frac{\delta^{q-1}}{q-1} y^q + \alpha y + \beta(x^2 - \delta^2) y^q \quad (6.33)$$

因此
$$S(y + \delta y) - S(y) = \int_{-\infty}^{+\infty} \partial L / \partial y \delta y dx \quad (6.34)$$

为了去求 S 的最大值, 令 $\partial L / \partial y = 0$ 。可以得到分布

$$p(x) = \frac{1}{Z_q} [1 + \beta(q-1)x^2]^{1/(1-q)} \quad (6.35)$$

6.8 近似熵—Approximate Entropy

相对于其它非线性动力学参数(如关联维数、Lyapunov指数等)而言, 90年代初由Pincus提出的近似熵(Approximate Entropy, 简记为ApEn)更主要的是从衡量时间序列复杂性的角度来度量信号中产生新模式的概率大小, 产生新模式的概率越大, 序列的复杂性越大, 相应的近似熵也越大^[13]。设采集到的原始数据为 $\{x_i\}$, 预先给定模式维数 m 和相似容限 r , 则近似熵的值可通过以下步骤求得^[14]:

- 1) 将序列 $\{x_i\}$ 按顺序组成 m 维矢量, 即:

$$\mathbf{X}(i) = [x(i), x(i+1), \dots, x(i+m-1)], \quad \text{其中 } i = 1, 2, \dots, N-m+1 \quad (6.36)$$

- 2) 定义 $\mathbf{X}(i)$ 与 $\mathbf{X}(j)$ 间的距离 $d[X(i), X(j)]$ 为两者对应元素中差值最大的一个, 即

$$d[X(i), X(j)] = \max_{0 \sim m-1} |x(i+k) - x(j+k)| \quad (6.37)$$

此时 $\mathbf{X}(i)$ 和 $\mathbf{X}(j)$ 中其他对应元素间差值自然都小于 d ，并对每一个 i 值计算 $\mathbf{X}(i)$ 与其余矢量 $\mathbf{X}(j)$, ($j=1, 2, \dots, N-m+1$) 的距离 $d[X(i), X(j)]$ 。

- 3) 按照给定的阈值 $r (r > 0)$ ，对每一个 i 值统计 $d[X(i), X(j)] < r$ 的数目及此数目与总的矢量个数 $N-m+1$ 的比值，记作 $C_i^m(r)$ ：

$$C_i^m(r) = \frac{1}{N-m+1} \text{num}\{d[X(i), X(j)] < r\}, \quad i=1, 2, \dots, N-m+1 \quad (6.38)$$

- 4) 先将 $C_i^m(r)$ 取对数，再求其对所有 i 的平均值，记做 $\Phi^m(r)$ ，即：

$$\Phi^m(i) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \quad (6.39)$$

- 5) 再把维数加1，变成 $m+1$ ，重复1)–4)的过程，得到 $\Phi^{m+1}(r)$ 和 $C_i^{m+1}(r)$ ：

$$C_i^{m+1}(r) = \frac{1}{N-m} \text{num}\{d[X(i), X(j)] < r\}, \quad i=1, 2, \dots, N-m$$

$$\Phi^{m+1}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \ln C_i^{m+1}(r)$$

- 6) 此序列的近似熵为：

$$ApEn(m, r) = \lim_{N \rightarrow \infty} [\Phi^m(r) - \Phi^{m+1}(r)] \quad (6.40)$$

但在实际工作中 N 不可能为 ∞ ，当 N 取有限值时，估计：

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \quad (6.41)$$

$ApEn(m, r, N)$ 的值与参数 m ， r 和 N 的选取有关。Pincus 曾指出，当 $m=2$ ， $r=0.1 \sim 0.25SD_x$ (SD_x 为原始数据序列 $\{x_i\}$ 的标准差) 时， $ApEn(m, r, N)$ 的值对序列长度 N 的依赖程度最小，即：

$$ApEn(2, r, N) \approx ApEn(2, r) \quad (6.42)$$

因此实际的计算中，通常取 $m=2$ ， $r=0.1 \sim 0.25SD_x$ (SD_x 为原始数据序列 $\{x_i\}$ 的标准差) 时， $ApEn(m, r, N)$ 的值可以表示为：

$$\begin{aligned} ApEn(2, r, N) &= -[\Phi^{m+1}(r) - \Phi^m(r)] \\ &= \lim_{N \rightarrow \infty} \left[\frac{1}{N-m} \sum_{i=1}^{N-m} \ln C_i^{m+1}(r) - \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \right] \\ &= \lim_{N \rightarrow \infty} \left[\frac{1}{N-m+1} \sum_{i=1}^{N-m} \ln \frac{C_i^{m+1}(r)}{C_i^m(r)} \right] \end{aligned} \quad (6.43)$$

分析式(6.43)可以看出, 近似熵实际上是在衡量当维数变化时该时间序列中产生新模式的概率的大小, 产生新模式的概率越大, 序列就越复杂, 对应的近似熵也就越大, 因此从理论上讲, 近似熵能够用来表征信号的不规则性和复杂程度。近似熵大致相当于维数变化时新模式出现的对数条件概率的均值, 在衡量时间序列的复杂性方面具有一定意义。综合有关文献的论述, 可总结出近似熵的主要特点如下^[15]:

- ◆ 近似熵值的大小和采样序列复杂程度成正比关系。序列越复杂, 对应的近似熵值也就越大。
- ◆ 相对于K-S熵、 K_2 熵、E-R熵而言, 近似熵使用较短的数据就可以比较稳健的估计值, 分析出信号的特征, 适合工程应用。
- ◆ 它不仅仅是一个非线性动力学参数, 由于它大致相当于某种条件概率, 所以对随机过程和确定性过程都适用。
- ◆ 近似熵具有一定的抗噪、抗野点能力, 特别是对偶尔产生的瞬态强干扰有较好的承受能力。这一点从近似熵的定义可以看出, 若信号中噪声的幅度低于相似容限 r 该噪声将被抑制, 若时间序列中存在较大的瞬态干扰时, 干扰产生的数据(即所谓的“野点”)与相邻数据组成的矢量与 $\mathbf{X}(i)$ 的距离必定很大, 因而在阈值检测中将被去除。因此, 近似熵具有一定的抗噪、抗野点能力。

6.9 样本熵—Sample Entropy

熵定义为新信息产生率。 $ApEn(m, r, N)$ 是对数据长度为 N , 相似容限为 r , m 点数据段模式互相相似情况下 $m+1$ 点数据段模式互相相似的条件概率 CP 的负平均自然对数的近似值。关于参数 r 需要牢记的一点是, 它通常用数据的 SD (标准差)的百分比表示, 因此近似熵是一种标度不变的方法。我们定义 B 为维数 m 时序列自相似的概率, A 为维数 $m+1$ 时序列自相似的概率, 于是 $CP = A/B$ 。从近似熵的算法可以看出, 近似熵是以 $-\log(CP)$ 为计算模型, 并计算所有该模型的平均值。为避免 $\log(0)$ 的出现, 可知 A 和 B 都不应为0, 作为修正, CP 应定义为 $(A+1)/(B+1)$ 。为保证这一点, 在近似熵的定义中存在着对自身数据段的比较, 但这显然与新信息的观点不相容, 必然存在偏差。它包含两层含义: 一方面 $ApEn$ 的值是与数据长度有关的, 对短序列的 $ApEn$ 值明显比期望值小。另一方面是一致性差, 如一时间序列比另一时间序列有较高的 $ApEn$ 值的话, 那对于其它的 m 和 τ 值, 也应具有较高的 $ApEn$ 值, 但 $ApEn$ 不一定满足。这种偏差也是导致它对微小的复杂性变化不灵敏的原因。事实上, 即便如此, 当 A 、 B 过小, 匹配数目过少, 接近 $CP=1$, $ApEn=0$ 时便出现较大的误差。因此, 需要对近似熵的算法进行改进来克服这种偏差产生的影响^[14]。

样本熵(*SampEn*, Sample Entropy)是由Richman提出的一种新的时间序列复杂性测度方法^[16, 17]。它是*CP*严格的自然对数, 可以用*SampEn*(m, r, N)来表示。其中, N 为长度, r 为相似容限, 维数为 m 及 $m+1$ 。样本熵在算法上相对于近似熵算法的改进: 相对近似熵而言, 样本熵的计算则是和的对数, 优势在于包括更大的*A*、*B*值, 以及更加准确的*CP*估计。样本熵旨在降低近似熵的误差, 与已知的随机部分有更加紧密的一致性, 样本熵是一种与现在的近似熵类似但精度更好的方法。

样本熵的具体算法如下^[14]:

- 1) 将序列 $\{x_i\}$ 按顺序组成 m 维矢量, 即:

$$\mathbf{X}(i) = [x(i), x(i+1), \dots, x(i+m-1)], \quad \text{其中 } i = 1 \sim N-m+1 \quad (6.44)$$

- 2) 定义 $\mathbf{X}(i)$ 与 $\mathbf{X}(j)$ 间的距离 $d[X(i), X(j)]$ 为两者对应元素中差值最大的一个, 即

$$d[X(i), X(j)] = \max_{0 \sim m-1} |x(i+k) - x(j+k)| \quad (6.45)$$

此时 $\mathbf{X}(i)$ 和 $\mathbf{X}(j)$ 中其他对应元素间差值自然都小于 d , 并对每一个 i 值计算 $\mathbf{X}(i)$ 与其余矢量 $\mathbf{X}(j)$, ($j = 1, 2, \dots, N-m+1$)的距离 $d[X(i), X(j)]$ 。

- 3) 按照给定的阈值 r ($r > 0$), 对每一个 i 值统计 $d[X(i), X(j)] < r$ 的数目及此数目与总的矢量个数 $N-m$ 的比值, 记作 $B_i^m(r)$:

$$C_i^m(r) = \frac{1}{N-m} \text{num}\{d[X(i), X(j)] < r\}, \quad i = 1, 2, \dots, N-m+1, i \neq j \quad (6.46)$$

- 4) 再对所有 i 的平均值, 记做 $B^m(r)$, 即为:

$$B^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} B_i^m(r) \quad (6.47)$$

- 5) 再把维数加1, 变成 $m+1$, 重复1)–4)的过程, 得到 $B_i^{m+1}(r)$ 和 $B^{m+1}(r)$ 。

- 6) 因此理论上此序列的样本熵为:

$$\text{SampEn}(m, r) = \lim_{N \rightarrow \infty} \left\{ -\ln \left[B^{m+1}(r) / B^m(r) \right] \right\} \quad (6.48)$$

但在实际工作中 N 不可能为 ∞ , 当 N 取有限值时, 估计样本熵为:

$$\text{SampEn}(m, r, N) = -\ln \left[B^{m+1}(r) / B^m(r) \right] \quad (6.49)$$

$\text{SampEn}(m, r, N)$ 的值与参数 m , r 和 N 的选取有关。不同的嵌入维数 m 和相似容限 r 对应的样本熵也不同。在一般情况下 $m = 1$ 或 2 , $r = 0.1 \sim 0.25SD_x$ 计算得到的样本熵具有较为合理的统计特性。

样本熵在算法上相对于近似熵算法的改进, 具有以下很好的性质:

- ✧ 样本熵不包含自身数据段的比较,因此它是条件概率的负平均自然对数的精确值,因此样本熵的计算不依赖数据长度
- ✧ 样本熵具有更好的一致性,如一时间序列比另一时间序列有较高的 *SampEn* 值的话,那对于其他 m 和 r 值,也具有较高的 *SampEn* 值
- ✧ 样本熵对于丢失数据不敏感,即使数据丢失多达1/3,对 *SampEn* 计算值影响依然很小。

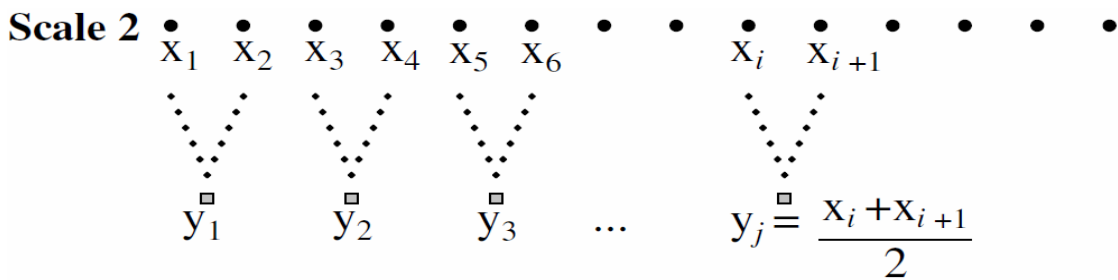
6.10 多尺度熵—Multiscale Entropy

多尺度熵方法首先是由Costa等人提出来的,已经广泛应用到生理时间序列的分析上^[18-21]。它基于样本熵,用于描述时间序列在不同时间尺度上的无规则程度。多尺度熵包含三个参数 τ , m 和 r , 其中 τ 是尺度因子, m 是嵌入维数, r 是阈值,也称相似系数,其计算方法如下:

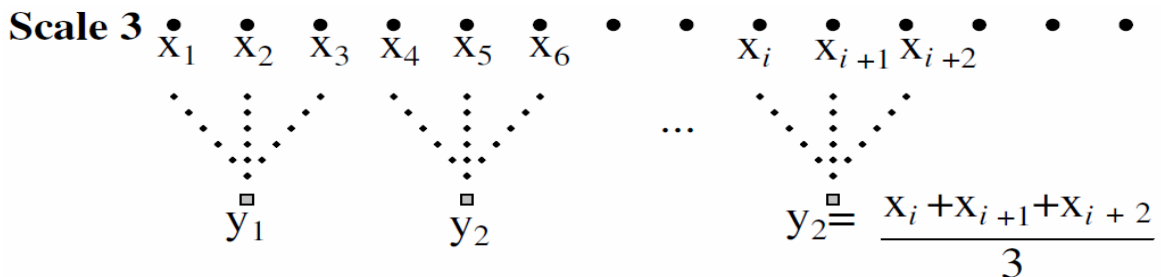
- (1) 设一离散时间序列 x_1, x_2, \dots, x_L , 共 L 个点: 对时间序列进行粗一断点 (coarse-graining) 变换, 得到新的时间序列:

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x(i) \quad (6.50)$$

其中 τ 为尺度因子, 每一个粗一断点时间序列的长度为 L/τ 。当尺度 $\tau=1$ 时, 粗一断点时间序列就是原始时间序列。图表 6-5和图表 6-6分别为 2 尺度和 3 尺度粗一断点时间序列的获取方法。



图表 6-5: 多尺度熵的 2 尺度时间序列获取方法



图表 6-6: 熵多尺度熵的 3 尺度时间序列获取方法

- (2) 根据尺度 τ 变化所得到的时间序列, 长度 $N = L/\tau$, 按连续序号组成一组 m 维矢

量：从 $Y^{(\tau)}(\mathbf{1})$ 到 $Y^{(\tau)}(\mathbf{N}-\mathbf{m}+1)$ ，其中，

$$Y^{(\tau)}(\mathbf{i}) = [y^{(\tau)}(i), y^{(\tau)}(i+1), \dots, y^{(\tau)}(i+m-1)], i=1 \sim N-m+1 \quad (6.51)$$

这些矢量代表了从第 i 个点开始的连续的 m 个 y 值。

(3) 定义 $d[Y^{(\tau)}(i), Y^{(\tau)}(j)]$ 为尺度 τ 上矢量 $Y^{(\tau)}(i)$ 和 $Y^{(\tau)}(j)$ 对应元素相减并取绝对值时最大的那个值，即

$$d[Y^{(\tau)}(i), Y^{(\tau)}(j)] = \max |y^{(\tau)}(i+k) - y^{(\tau)}(j+k)|, k=0 \sim m-1, i, j=1 \sim N-m+1, i \neq j \quad (6.52)$$

(此时 $Y^{(\tau)}(i)$ 和 $Y^{(\tau)}(j)$ 中对应元素间差的绝对值都小于 r)。并对每一个 i 值计算 $Y^{(\tau)}(i)$ 与其余矢量 $Y^{(\tau)}(j)$ 间的距离 $d[Y^{(\tau)}(i), Y^{(\tau)}(j)]$ 。

(4) 给定阈值 r ，对于每一个 $i \leq N-m+1$ 的值，统计 $d[Y^{(\tau)}(i), Y^{(\tau)}(j)]$ 小于 r 的数目以及数目与距离总数 $N-m$ 的比值，记作 $C_i^{\tau, m}(r)$ ，即

$$C_i^{\tau, m}(r) = \frac{1}{N-m} \left\{ \left[d[Y^{(\tau)}(i), Y^{(\tau)}(j)] < r \right] \text{的数目} \right\}, i, j=1 \sim N-m+1, i \neq j \quad (6.53)$$

上式表示在时间尺度 τ ，以 $Y(i)$ 为中心，在嵌入维数为 m ，阈值为 r 情形下，其余矢量 $Y^{(\tau)}(i)$ 与 $Y^{(\tau)}(j)$ 的距离 $d[Y^{(\tau)}(i), Y^{(\tau)}(j)]$ 小于 r 的概率，表示所有 $Y^{(\tau)}(j)(i \neq j)$ 与 $Y^{(\tau)}(i)$ 的关联程度，也就是表示矢量序列 $\{Y^{(\tau)}(j)\}$ 的规律性程度。

(5) 对所有的点求平均值，即

$$C^{\tau, m}(i) = (N-M+1)^{-1} \sum_{i=1}^{N-m+1} C_i^{\tau, m}(r) \quad (6.54)$$

$C^{\tau, m}(i)$ 表示矢量 $\{Y^{(\tau)}(j)\}$ 在尺度 τ 下的(也就是时一间序列 $y(i)$ 的平均自相关程度。

(6) 增加维数至 $m+1$ ，重复(2)–(5)步骤，从而得到尺度 τ ， $m+1$ 维数下的 $C_i^{\tau, m+1}(i)$ ，求其平均值，得到 $C^{\tau, m+1}(i)$ 。

理论上该时间序列在尺度 τ 下的样本熵值定义为

$$SampEn(m, r) = \lim_{N \rightarrow \infty} -\ln \left[C^{\tau, m+1}(r) / C^{\tau, m}(r) \right] \quad (6.55)$$

当 N 为有限值时，按上述步骤得出的是序列长度为 N ，尺度 τ 时样本熵估计值，将估计值记作：

$$SampEn(\tau, m, r) = -\ln \left[C^{\tau, m+1}(r) / C^{\tau, m}(r) \right] \quad (6.56)$$

多尺度熵定义为样本熵值在多个尺度下的集合，所以多尺度熵值为

$$MSE = \left\{ \tau \mid SampEn(\tau, m, r) = -\ln \left[C^{\tau, m+1}(r) / C^{\tau, m}(r) \right] \right\} \quad (6.57)$$

样本熵确定时间序列的在单一尺度上无规则程度，样本熵值越低，序列自相似性越高，样本熵值越大，序列越复杂。从多尺度熵的定义可以看出，它计算时间序列在多个尺度上的样本熵值，体现了时间序列在尺度上的无规则度。若熵值在尺度上单调递减，则序列在尺度上自相似性较低，结构简单，属于随机时间序列；若熵值在尺度上越大，序列自相似性大，复杂度大；若一个时间序列的熵值在绝大部分尺度上大于另一个时间序列的熵值，则说明前者要比后者复杂^[18, 19]。

6.11 Lempel-Ziv 复杂度

复杂性科学是一门最近发展起来的新兴交叉学科。20 世纪 60 年代，柯尔莫哥洛夫 (Kolmogorov) 等人对复杂性作了定义：一个系统的复杂程度与该系统行为(空间结构或时间序列所表示的变化)的最小描述有关，一般称为柯尔莫哥洛夫复杂性。其基本思想是认为：某事物的算法复杂度等于产生该事物的图形结构或符号序列的最短程序长度与该图形结构或符号序列本身大小之比的极限(当后者趋于极限时)^[22]。由于柯尔莫哥洛夫复杂性是不可计算的，1976 年 A. Lempel 和 J. Ziv 提出一种度量符号序列复杂性的简单算法，称为 Lempel-Ziv 复杂度^[23]。由于此法适用于符号序列，所以首先要对数据序列如 6.2 节进行符号化。

符号化时间序列以后，计算符号序列的 Lempel-Ziv 复杂度过程如下^[22]：

令 $c(n)$ 为某一给定的(一般为“0, 1”)符号序列 $S = (s_1 s_2 \cdots s_n)$ 的复杂性计数(复杂度)，其中 $s_i, i \in 1, 2, \dots, n$ 为一个字符)。令 S, Q 分别代表两个字符串， SQ 表示把 S, Q 两个字符串相加组成的总字符串， SQP 表示把 SQ 中最后一个字符删去所得的字符串(P 表示删去最后一个字符的操作)。令 $V(SQP)$ 表示 SQP 的所有不同的子串集合。 $c(n), S, Q$ 的初始化为 $c(n) = 1, S = s_1, Q = s_2$ ，因此 $SPQ = s_1$ 。现假定 $S = s_1 s_2 \cdots s_r, Q = s_{r+1}$ 。若 $Q \in V(SQP)$ ，则表示 s_{r+1} 是 $S = s_1 s_2 \cdots s_r$ 字符串的一子串，那么 S 不变，只将 Q 更新为 $Q = s_{r+1} s_{r+2}$ ，再判断 Q 是否属于 $V(SQP)$ (此时，因为 S 不变， Q 更新了，所以 SQP 也要更新)，如此反复进行，直到发现 $Q \notin V(SQP)$ 时为止。设此时 $Q = s_{r+1} s_{r+2} \cdots s_{r+i}$ ，即表明 $s_{r+1} s_{r+2} \cdots s_{r+i}$ 不是 $s_1 s_2 \cdots s_r s_{r+1} \cdots s_{r+i-1}$ 的子串。因此 $c(n)$ 值将加 1，然后将上述 Q 组合到 S 中，使 S 更新为 $S = s_1 s_2 \cdots s_r s_{r+1} \cdots s_{r+i}$ ，而取 Q 为 $Q = s_{r+i+1}$ 。

重复以上步骤，直到 Q 取到最后一位为止。这样就把 $s_1 s_2 \cdots s_n$ 分成了 $c(n)$ 个不同的子串。根据 A. Lempel 和 J. Ziv 的研究，对几乎所有属于 $[0, 1]$ 区间， x 对应的二进制分解所表示的序列都会趋向一个定值^[23]：

$$\lim_{n \rightarrow \infty} c(n) = b(n) = n / \log_2 n \quad (6.58)$$

其中 $b(n)$ 是随机序列的渐进行为，可以用它来使 $c(n)$ 归一化，称为“归一化复杂度”：

$$C_{LZN}(n) = c(n) / b(n) \quad (6.59)$$

通常就用这个函数来表达时间序列的复杂度的变化。从该算法可以看出，完全随机序列复杂度趋于 1，而规则序列则趋于 0， C_{LZN} 值越大，动力系统越复杂；相反越小，动力学系统的规律性就越明显。

例如，序列 $S = (10101010)$ 的复杂度可由以下步骤得到：

- 1) 先取 $S = s_1 = 1$ ， $Q = s_2 = 0$ ， $SQ = 10$ ， $SQP = 1$ ， $Q \notin V(SQP)$ ， Q 为插入， $SQ = 1 \cdot 0$
- 2) $S = s_1 s_2 = 10$ ， $Q = s_3 = 1$ ， $SQ = 101$ ， $SQP = 10$ ， $Q \in V(SQP)$ ，故第 2 步 Q 是复制， $SQ = 1 \cdot 0 \cdot 1$
- 3) $S = s_1 s_2 = 10$ ， $Q = s_3 s_4 = 10$ ， $SQ = 1010$ ， $SQP = 101$ ， $Q \in V(SQP)$ ，故第 2 步 Q 是复制， $SQ = 1 \cdot 0 \cdot 10$ 。

4) 以后几步都是重复 2) 和 3)，即都是复制。于是得： $S = 1 \cdot 0 \cdot 101010$ ，整个序列分为 3 段，故 $c(n) = 3$ 。

5) $b(8) = 8 \log_2 8 = 24$ ，故得归一化复杂度为： $C_{LZN} = c(8) / b(8) = 3 / 24 = 0.125$ ，结果表明此序列 S 的复杂度较小，原因在于此序列是一周期序列。

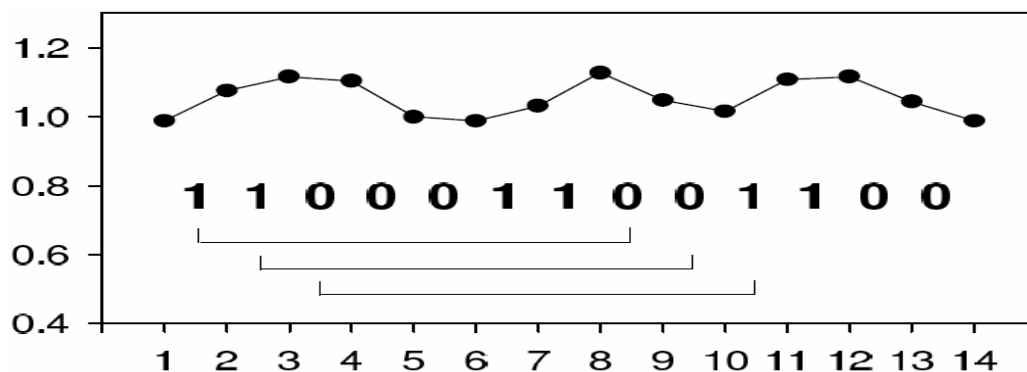
根据 Lempel-Ziv 复杂度算法可知，一个符号序列的复杂度越大，说明它的添加操作越多，新模式也越多，描述给定符号序列所需最少的、互不相同的“子串”个数越多，给定的符号序列周期性越弱，出现新模式的速率越快。反之，则复制操作越多，新模式越少，周期性越强，出现新模式的速率越慢。普遍认为复杂度的物理意义是它反映了一个时间序列随着序列长度的增加出现新模式的速率。复杂度越大，说明数据在窗口长度时期内随时间出现的新变化越多，发生新变化的速率越快，表明这一时期的数据变化是无序而复杂的。反之，复杂度越小，则说明发生新变化的速率越慢，数据变化是规则的，周期性越强。因此，计算一个时间序列的复杂度能够描述出系统的状态发生变化的情况，这是非常重要的性质。因此从这个意义上来说，可以用复杂度来描述动力学系统状态随时间变化的情况。不过要说明的一点：虽然这个描述是从数量(复杂度是一个具体数值)上进行分析，但目前还不能据此做定量描述，只能做较为清晰的定性描述^[22]。

6.12 相似指数分析^[24]

对于时间序列 $\{x_1, x_2, \dots, x_N\}$ ，我们检测 x_n 和它的临近值 x_{n-1} 的大小，根据如下规则生成一个新的二进制时间序列 I_n ：

$$I_n = \begin{cases} 0, & \text{if } x_n \leq x_{n-1} \\ 1, & \text{if } x_n > x_{n-1} \end{cases} \quad (6.60)$$

接着我们将二进制序列 I_n 每 m 个值当做一个bit组成一个字, 称为 w_k , 整个过程如图表 6-7所示:



图表 6-7: 在进行信息指数计算前的符号化和编码过程示意图

序列 S_1 和 S_2 的相似性指数可用如下公式进行计算:

$$D_m(S_1, S_2) = \frac{1}{2^m - 1} \sum_{k=1}^{2^m} |R_1(w_k) - R_2(w_k)| F(w_k) \quad (6.61)$$

其中

$$F(w_k) = \frac{1}{Z} [-p_1(w_k) \log p_1(w_k) - p_2(w_k) \log p_2(w_k)] \quad (6.62)$$

其中 $p_1(w_k)$ 和 $R_1(w_k)$ 表示时间序列 S_1 中的字 w_k 的出现的概率和秩; 同理 $p_2(w_k)$ 和 $R_2(w_k)$ 表示时间序列 S_2 中的字 w_k 的出现的概率和秩。 Z 为归一化因子, 它的表达式如下

$$Z = \sum_k [-p_1(w_k) \log p_1(w_k) - p_2(w_k) \log p_2(w_k)] \quad (6.63)$$

6.13 海杂波的复杂度和熵

内容 (略)

6.14 本章小结

本章介绍了表征时间序列复杂程度的熵和复杂度, 一般来讲, 复杂度和熵在数学上与原始时间序列的对应关系并不十分严格, 并且不一定要满足嵌入理论, 不要求时间序列一定为混沌信号, 因此非常适合于利用观测到的系统状态信号的有限时间序列对系统进行分析与评估。

6.15 后记

本版本为“时间序列的复杂度和熵”的初稿，写作的原因是由于对于多尺度熵概念的关注，走马观花的看了一下，熵的各种定义非常多，它的外延和内涵一直在发展，对于非线性时间序列而言，应用熵和复杂度去描述和识别动力系统隐含的规律性是很重要的方向。

你的关注是我前进的最大动力☺。

任何意见、建议、批评和讨论我都热烈欢迎。

我的信箱：xuxkboy@newmail.dlmu.edu.cn

欢迎大家到研学论坛(<http://bbs.matwav.com/>)混沌分形版进行讨论

本文版权目前归本人所有，引用格式如下：

许小可. 海杂波的非线性分析与建模:(博士学位论文). 大连:大连海事大学, 2007

参考文献:

- [1] Daw C. S., Finney C. E., Kennel M. B. Symbolic approach for measuring temporal "irreversibility". *Physical Review E*. 2000, 62(2): 1912-1921.
- [2] Finney C E A, Nguyen K, Daw C S, et al. Symbol-sequence statistics for monitoring fluidization. *International Mechanical Engineering Congress & Exposition, Anaheim, California, USA, 1998:15-20.*
- [3] 金宁德, 李伟波. 非线性时间序列的符号化分析方法研究. *动力学与控制学报*. 2004, 2(03): 54-59.
- [4] 张雨. 符号化时间序列分析. *湖南科技大学学报(自然科学版)*. 2004, 19(01): 75-79.
- [5] 李鹤龄. 信息熵、玻尔兹曼熵以及克劳修斯熵之间的关系——兼论玻尔兹曼熵和克劳修斯熵是否等价. *大学物理*. 2004, 23(12): 37-40.
- [6] 吕金虎, 陆君安, 陈士华. 混沌时间序列分析及其应用. 武汉: 武汉大学出版社, 2002.
- [7] 冯晓光. 近似熵在往复式压缩机故障诊断中的研究应用: 硕士学位论文. 大连: 大连理工大学, 2006.
- [8] Grassberger P, Procaccia I. Characterization of Strange Attractors. *Physical Review Letters*. 1983, 50(5): 346-349.
- [9] Grassberger P, Procaccia I. Estimation of the Kolmogorov entropy from a chaotic signal. *Physical Review A*. 1983, 28(4): 2591-2593.
- [10] 赵贵兵, 石炎福, 段文锋, 等. 从混沌时间序列同时计算关联维和Kolmogorov熵. *计算物理*. 1999, 15(03): 309-315.
- [11] Tsallis C., Plastino A. R., Zheng W. -m. Power-law sensitivity to initial conditions--New entropic representation. *Chaos, Solitons & Fractals*. 1997, 8(6): 885-891.
- [12] 曹克非, 王参军. Tsallis熵与非广延统计力学. *云南大学学报(自然科学版)*. 2005, 27(06): 514-520.
- [13] Pincus S M. Approximate entropy as a measure of system complexity. *Proc. Nati. Acad. Sci. USA*. 1991, 88(3): 2297-2301.
- [14] 白冬梅. 脑电信号的特性分析与特征提取: 硕士学位论文. 大连: 大连理工大学, 2006.
- [15] 杨福生, 廖旺才. 近似熵:一种适用于短数据的复杂性度量. *中国医疗器械杂志*. 1997, (05).
- [16] Richman J. S., Moorman J. R. Physiological time series analysis using approximate entropy and sample entropy. *Am J Physiol*. 2000, 278(6): 2039-2049.
- [17] Lake D E, Richman J S, Griffin M P, et al. Sample entropy analysis of neonatal heart rate variability. *Am J Physiol*. 2002, 283(3): 789-797.
- [18] Costa Madalena, Goldberger Ary L., Peng C. K. Multiscale entropy analysis of biological signals. *Physical Review E*. 2005, 71(2): 021906-1-18.
- [19] Costa Madalena, Goldberger Ary L., Peng C. -k. Multiscale Entropy Analysis of Complex Physiologic Time Series. *Physical Review Letters*. 2002, 89(6): 068102-1-4.
- [20] Thuraisingham R A, Gottwald G A. On multiscale entropy analysis for physiological data. *Physica A: Statistical Mechanics and its Applications*. 2006, 366: 323-332.
- [21] Costa M, Peng C -, L G A, et al. Multiscale entropy analysis of human gait dynamics. *Physica A: Statistical Mechanics and its Applications*. 2003, 330(1-2): 53-60.
- [22] 解幸幸, 李舒, 张春利, 等. Lempel-Ziv复杂度在非线形检测中的应用研究. *复杂系统与复杂性科学*. 2005, 2(03): 61-66.
- [23] Lempel A., Ziv J. On the Complexity of Finite Sequences. *IEEE Transactions on Information Theory*. 1976, 22(1): 75-81.
- [24] Yang A C C, Goldberger A L, Peng C K, et al. Information-Based Similarity Index.
<http://physionet.org/physiotools/ibs/doc/>.