We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Statistics Character and Complexity in Nonlinear Systems

Yagang Zhang and Zengping Wang

Key Laboratory of Power System Protection and Dynamic Security Monitoring and Control under Ministry of Education (North China Electric Power University), China

1. Introduction

Learning is the process of constructing a model from complex world. And machine learning is concerned with constructing computer programs that automatically improve with experience. Machine learning draws on concepts and results from many fields, including artificial intelligence, statistics, control theory, cognitive science, information theory, etc. Many successful machine learning applications have also been developed in recent years. Obviously, no matter what we adopt new analytical method or technical means, we must have a distinct recognition of system itself and its complexity, and increase continuously analysis, operation and control level.

In mathematics, nonlinear system represents a system whose behavior is not expressible as a linear function of its descriptors. Our world is inherently nonlinear in nature. Generally speaking, there have difficulties in solving nonlinear equations. Especially the nonlinear system may give rise to some interesting phenomena such as chaos, where simple changes in one part of the system will produce complex effects throughout.

It has been half century since the discovery of inherent randomicity in nonlinear systems (Ulam & Von Neumann, 1947). The study of chaotic symbolic sequences is gradually developing in theory. However, applied research of stochastic chaotic sequences has not been fully carried out, for most of studies focus on controlling or avoiding chaos. Chaos, nevertheless, affords inherent randomicity that can be calculated, which is an important applied domain. The stochastic symbolic sequences bear the following three features. First, computer can generate them iteratively. Second, like false stochastic numbers, they can set up a stochastic sequence simulation (in contradiction, they are based on corresponding symbolic spaces). Third, they can produce numerous symbolic spaces, which is not characteristic of common stochastic numbers. Therefore, the symbolic dynamics (Hao, 1989; Hao, 1991; Hao & Zheng, 1998; Collet & Eckmann, 1980; Alekseev & Yakobson, 1981; Xie, 1993; Xie, 1996; Peng & Luo, 1991; Zhou & Peng, 2000) developed by this means is supposed to be very useful.

Our researches are based on this kind of symbolic sequences, the generic iterative map in n symbolic map (Zhou & Cao, 2003) is:

$$x_{m+1} = \sum_{i=0}^{n} a_i x_m^i = a_0 x_m^0 + a_1 x_m^1 + \cdots + a_n x_m^n, (n \geq 1, m \geq 0)$$

For random n symbolic sequences, their corresponding symbolic spaces, symbolic expression and kneading sequences are listed in Table 1,

Symbolic Spaces	Symbolic Expression	Kneading Sequences
$\sum_2$	$L.R$	(L^∞, RL^∞)
$\sum_3$	$L.M.R$	(R^∞, L^∞) --Kneading plane
$\sum_4$	$L.M.N.R$	(L^∞, RL^∞) --Kneading space
$\vdots$	$\vdots$	$\vdots$

Table 1. The corresponding symbolic character in symbolic spaces

In this chapter, we will clarify the different kinds of statistic character and complexity in nonlinear systems. This chapter includes two parts, the fist part is about unimodal surjective map and Lorenz type maps nonlinear systems, which are two kinds of typical nonlinear systems. The distributions of frequency, inter-occurrence times, first passage time and visitation density in unimodal surjective map and Lorenz type maps are discussed carefully. These two kinds of nonlinear systems have same distributions, which will also be explained in theory, and the catholicity of the statistic character will be elicited. The second part is about the inherent randomicity in 4-symbolic dynamics. The distribution of frequency, inter-occurrence times and the alignment of two random sequences are amplified in detail. By using transfer probability of Markov chain (MC), we will obtain analytic expressions of generating functions in four probabilities stochastic wander model, which can be applied to all 4-symbolic systems. So, a perfect symbolic platform will be set up for our utilizing statistic character. The 4-symbolic sequences have natural relations with bioinformatics sequences, in the field of application, we hope to afford this kind of symbolic platform which satisfies these stochastic properties and study some properties of DNA sequences, 20 amino acids symbolic sequences of protein structure, and the time series that can be symbolic in finance market et al.

2. The statistic character in Unimodal surjective map

2.1 Symbolic dynamics of Unimodal surjective map
The generic iterative form in Unimodal surjective map is:

$$x_{n+1} = F(A, x_n) = 1 - 2x_n^2,$$

x_n is defined on interval $[-1,1]$.

Let us define an alphabet of 2 numbers, which is corresponding to the likely states of a random discrete nonlinear system, or all the likely outcomes of a random experiment:

$$\Omega = \{0,1\} = \{" \, faillure","success"\}$$

The forward sequence constitutes a space (or a set) composed of the generated outcomes:

$$\Omega^N = \left\{ \left(\xi_0, \xi_1, \xi_2, \cdots\right) : \xi_i \in \Omega, \forall i \in \{0,1,2,\ldots\} \right\}$$

These sequences themselves are iteratively generated (Collet & Eckmann, 1980; Peng & Luo, 1991), in fact it's a shift map $\sigma : \Omega^N \to \Omega^N$, which acting on the sequences by $\sigma(\xi_0, \xi_1, \xi_2, \ldots) = (\xi_1, \xi_2, \ldots)$. Another definition is μ, which is the product measure (Coelho & Collet, 1994; Coelho, 2000; Peng & Cao, 1996; Billingsley, 1986) on Ω^N generated by the measure $(1-p, p)$ on $\{0,1\}$, and will be denoted by $(1-p, p)^N$.

2.2 The distribution of frequency

Defining $f : \Omega^N \to \{0,1\}$ by $f\{\xi_0, \xi_1, \xi_2, \ldots\} = \xi_0$, it is coarse graining in theory, one can get:

$$X_i(\xi) = f(\sigma^i \xi), \text{ (for } i = 0,1,2,\ldots \text{)},$$

which are sequences of independent and identically distributed(i.i.d.) random variables defined on the probability space (Ω^N, μ) (all the following discussions are based on the random variables), that is, the random variables represented by $\xi_0, \xi_1, \xi_2, \ldots$ are i.i.d., and $\xi_i (i = 0,1,2,\ldots)$ is based on Ω,

$$\begin{aligned} Y_n &= X_0 + X_1 + \cdots + X_{n-1} \\ &= \sum_{i=0}^{n-1} f(\sigma^i x) \\ &= \xi_0 + \xi_1 + \cdots + \xi_{n-1} \end{aligned}$$

The stochastic symbolic sequences in Unimodal surjective map satisfy Binomial distribution:

$$\mu\{\xi \in \Omega^N : Y_n(x) = k\} = C_n^k p^k (1-p)^{n-k} \tag{1}$$

2.3 The inter-occurrence times in Unimodal surjective map

Now let us make a further study a given word's occurrence times in an independent repeated experiment, such as *success* in the alphabet of 2 numbers. Given outcomes of a random sequence

$$\xi = (\xi_0, \xi_1, \xi_2, \cdots) \in \Omega^N,$$

we are mainly interested in n such that $\xi_n = 1$, let

$$\tau(\xi) = \tau_1(\xi) = \inf\{n \geq 0 : \xi_n = 1\},$$

and accordingly, for $j \geq 2$,

$$\tau_j(\xi) = \inf\{n > \tau_{j-1}(\xi) : \xi_n = 1\},$$

then for all $k \geq 0$, the result is, for fixed $k > 0,$ and all $k_1 < \cdots < k_{j-1}$,

$$\begin{aligned}
&P(\tau_j - \tau_{j-1} = k \mid \tau_{j-1} = k_{j-1}, \cdots, \tau_1 = k_1) \\
&= P(\tau_j - \tau_{j-1} = k\} \\
&= p(1-p)^{k-1}
\end{aligned} \tag{2}$$

the inter-occurrence times $1 + \tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \cdots$ are i.i.d. with parameters p.

2.4 The first passage time in Unimodal surjective map

Using this method similar to study the distribution of first passage time T_y of one-dimensional simple random wander in stochastic processes, one gets τ_j satisfies Negative Binomial distribution $BN(j, p)$:

$$\begin{aligned}
P(\tau_j = k) = C_{k-1}^{r-1} p^r q^{k-r} \\
(k = r, r+1, r+2, \cdots, \quad 0 < p < 1, \quad q = 1-p)
\end{aligned} \tag{3}$$

3. The statistic character in Lorenz maps

3.1 Symbolic dynamics of Lorenz maps
Lorenz equation:

$$\begin{cases}
\dot{x} = \sigma(y - x), \\
\dot{y} = (r - z)x - y \\
\dot{z} = xy - bz
\end{cases}$$

On the Poincaré section, some geometrical structure of Lorenz flow may be reduced to a one-dimensional Lorenz map $f : [-\mu, v] \to [-\mu, v]$, $(\mu, v > 0, \lambda > 1)$

$$f(x) = \begin{cases} f_L(x) = v - \alpha |x|^\lambda + h.o.t, & x \le 0 \\ f_R(x) = -\mu + \beta x^\lambda + h.o.t, & x > 0 \end{cases}$$

Where λ is a constant greater than 1, "h.o.t" represents high-level term. Both of the branches f_L and f_R are monotone increasing. In order to get iterative sequences in the part of chaos, the Lorenz map used in this research is:

$$f(x) = \begin{cases} f_L = 1 - 2|x|^2, & x \le 0 \\ f_R = -1 + 2x^2, & x > 0 \end{cases} \tag{4}$$

The symbolic dynamics of Lorenz maps is also simple (Peng & Du, 1999). Following the kneading theory, the address $A(x)$ of any point x on the interval $[-1, 1]$ reads

$$A(x) = \begin{cases} R, & x \in [-1, 0) \\ L, & x \in [0, 1] \end{cases}$$

$x = 0$ is the turning (discontinuous) point, and one can define C and D as

$$C = \lim_{x \to 0^-} f_L(x),$$
$$D = \lim_{x \to 0^+} f_R(x).$$

Two infinite or finite symbolic sequences starting from C and D are kneading sequences which can be ordered lexicographically by $L \prec C, D \prec R$. For two kneading sequences, $\gamma_1 \gamma_2 \ldots \gamma_i \gamma_{i+1} \ldots$ and $\eta_1 \eta_2 \ldots \eta_i \eta_{i+1} \ldots$, with maximal common leading part:

$$\gamma_1 \gamma_2 \ldots \gamma_i = \eta_1 \eta_2 \ldots \eta_i ,$$

one has ,

$$\gamma_1 \gamma_2 \ldots \gamma_i \gamma_{i+1} \ldots \prec \eta_1 \eta_2 \ldots \eta_i \eta_{i+1} \ldots$$

if and only if $\gamma_{i+1} \prec \eta_{i+1}$.

The shift operator φ is defined as ,

$$\varphi^k(\xi) = \xi_{k+1}\xi_{k+2}\cdots \quad \text{for} \quad \xi = \xi_1\xi_2\cdots\xi_k\xi_{k+1}\cdots.$$

For any two sequences ,

$$\xi = \xi_1\xi_2\cdots\xi_i\xi_{i+1}\cdots \quad \text{and} \quad \zeta = \zeta_1\zeta_2\cdots\zeta_j\zeta_{j+1}\cdots,$$

$\xi_i, \zeta_j \in \{R, L\}$, if $\varphi^k(\xi) \prec \xi$ and $\zeta \prec \varphi^k(\zeta)$, for all $K \in \mathbb{Z}_+$,then ξ is called maximal, ζ minimal, and $S = (\xi, \zeta)$ is an extremal pair. Let the integers k_L and k_R be the order coordinates of a letter in the sequence such that $\varphi^{k_L-1}(\xi) = L\cdots$, and $\varphi^{k_R-1}(\xi) = R\cdots$, the set k_L and k_R describe successive sequences of L or R . Then, if the pair S further satisfies the following condition:

$$\varphi^{k_L}(\xi) \prec K^1, \ \varphi^{k_R}(\xi) \succ K^2, \ \{k_L\}\bigcup\{k_R\} = \{k\} \in \mathbb{Z}_+,$$
$$\varphi^{k_L'}(\zeta) \prec K^1, \ \varphi^{k_R'}(\zeta) \succ K^2, \ \{k_L'\}\bigcup\{k_R'\} = \{k'\} \in \mathbb{Z}_+.$$

S is admissible with respect to the kneading sequences K^1 and K^2 . All the admissible pairs form an admissible set K and fill up the whole kneading parameter plane of nonlinear systems of two letters.

3.2 The distribution of frequency
See expression (1).

3.3 The inter-occurrence times in Lorenz map
See expression (2).

3.4 The first passage time in Lorenz map
See expression (3).

4. Visitation density function of Unimodal surjective maps and Lorenz map

The orbital points' distribution of the Unimodal surjective maps and Lorenz map is,

$$\rho(x) = \frac{1}{\pi\sqrt{1-x^2}} \tag{5}$$

The concrete resolvent is using Frobenius-Perron operator (Lasota & Mackey, 1985; Yorke & Li, 1975; Ding & Li, 1991; Li, 1976). The general form of resolve visitation density problem by F-P operator P is,

$$Pf(x) = \frac{d}{dx} \int_{S^{-1}(\triangle)} f(u)du ,$$

here, $S = S(x)$ is a given map, $\triangle$ is an interval, $f(x)$ is a density function. In fact, it is an iterative process, the initial state is

$$\int_{\triangle} f_1(u)du = \int_{S^{-1}(\triangle)} f_0(u)du .$$

$f_0(x)$ is an arbitrary initial density and $f_1(x)$ is a new density transformed by map $S(x)$, that is,

$$f_1 = Pf_0 ,$$

until,

$$f_*(x) = P^n f(x) \quad \text{as} \quad n \to \infty .$$

Of course,

$$Pf_*(x) = f_*(x) ,$$

the unique limiting density is just the ultimate visitation density function.

It is mainly in numerical value meaning that getting visitation density functions of higher order maps, if the invariable density does exist. Figure 1 is the U-shaped probability density based on iterates, corresponding analytic form is just expression (5),
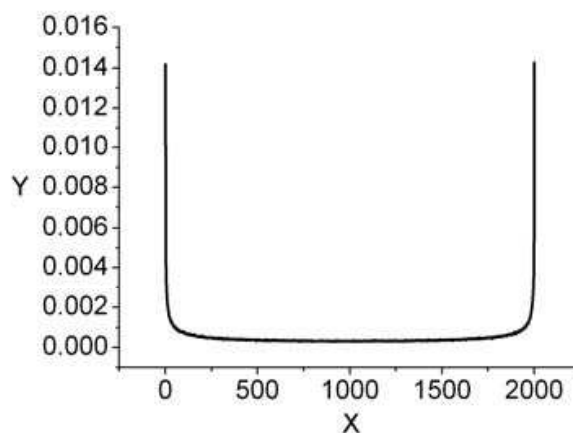
Fig. 1. The visitation density of Unimodal surjective map and Lorenz map based on 1000000 iterates, the interval $[-1,1]$ is divided into 2000 subintervals. X coordinate axis is corresponding interval, Y coordinate axis is the output proportion of each interval.
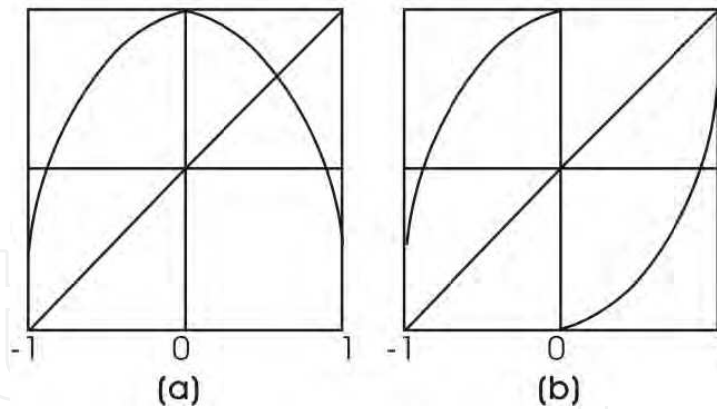
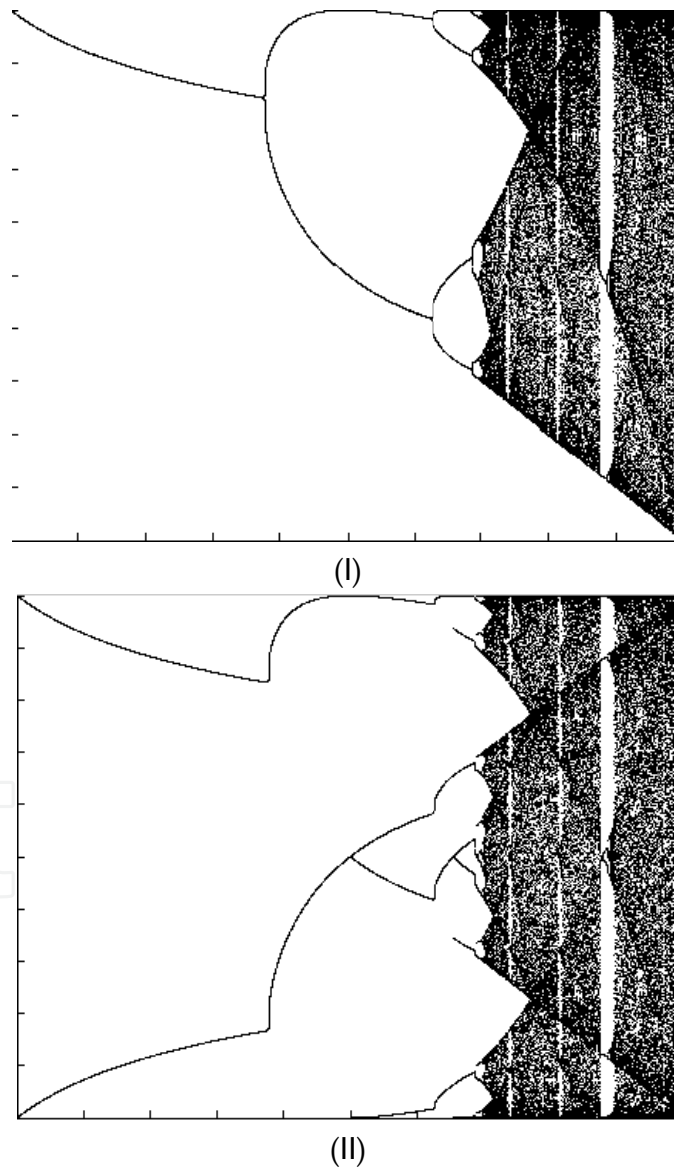Fig. 2. Unimodal surjective map (a) and Lorenz map (b)

Fig. 3. The bifurcation diagrams of Unimodal surjective map (I) and Lorenz map (II)

5. The comparability of statistic character in the Unimodal map and Lorenz map

The former statistic character in Lorenz map is similar entirely to that in Unimodal surjective map. This kind of comparability is determined by the relationship of Unimodal surjective map and Lorenz map. (See Figure 2)

The iterative form of Lorenz map is (4), and Unimodal surjective map is $y = 1 - 2x^2$. One can find this characteristic by Figure 2,

$$\begin{cases} f_a = -f_b, & x \geq 0 \\ f_a = f_b, & x < 0 \end{cases}.$$

A n-periods orbit of f_a corresponds to a couple of n-periods orbits of f_b. Both of them have the same topological entropy and marker behavior. The fixed point of f_a exhibits two-periods behavior of f_b, which can be found clearly by contrasting their bifurcation diagrams. (See Figure 3)

Compared the right branch of Lorenz map and Unimodal surjective map, the Lorenz map is only overturned by x coordinate axis. As these results reveal that this kind of overturn does not influence statistical properties of random sequences. Compared with Unimodal map, Lorenz map belongs to a more complex category, which presents more abundant dynamics actions. But as above study, these statistical results present regulation as a whole. These are randomicity in deterministic systems.

6. The stochastic properties in 4-letters maps

6.1 The distribution of frequency

Let us define an alphabet of four numbers, which is corresponding to the likely states of a random discrete dynamical system, or all the likely outcomes of a random experiment:

$$\Omega = \{0, 1, 2, 3\} = \{L, M, N, R\} = \{A, G, C, T\}$$
$$= \{\text{"Spring"}, \text{"Summer"}, \text{"Autumn"}, \text{"Winter"}\}$$

The forward sequence constitutes a space (or a set) composed of the generated outcomes:

$$\Omega^N = \left\{ (\xi_0, \xi_1, \xi_2, \cdots) : \xi_i \in \Omega, \forall i \in \{0, 1, 2, \ldots\} \right\}$$

These sequences themselves are iteratively generated, in fact it's a shift map $\sigma : \Omega^N \to \Omega^N$, which acting on the sequences by $\sigma(\xi_0, \xi_1, \xi_2, \ldots) = (\xi_1, \xi_2, \ldots)$. Another definition is μ, which is the product measure[6] on Ω^N generated by the measure

(p_1, p_2, p_3, p_4) on $\{0, 1, 2, 3\}$ ($p_1 + p_2 + p_3 + p_4 = 1$), and will be denoted by $(p_1, p_2, p_3, p_4)^N$. Defining $\varphi : \Omega^N \to \{0, 1, 2, 3\}$ by $\varphi\{\xi_0, \xi_1, \xi_2, \ldots\} = \xi_0$, it is also coarse graining in theory, one can get:

$$X_i(\xi) = \varphi(\sigma^i \xi), \text{ (for } i = 0, 1, 2, \ldots)$$

Which are sequences of independent and identically distributed (i.i.d.) random variables defined on the probability space (Ω^N, μ) (all the following discussions are based on the random variables), that is, the random variables represented by $\xi_0, \xi_1, \xi_2, \ldots$ are i.i.d., and $\xi_i (i = 0, 1, 2, \ldots)$ is based on Ω.

The numeric examinations reveal that the stochastic symbolic sequences in 4-letters maps satisfy multinomial distribution:

$$T(N_1 = n_1, N_2 = n_2, N_3 = n_3, N_4 = n_4) = \frac{n!}{n_1! n_2! n_3! n_4!} p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4}$$

$$(\sum_{i=1}^{4} p_i = 1, \quad \sum_{i=1}^{4} n_i = n)$$

The theoretic foundation of these results is that the topological entropy (Shi et al., 1996; Zhang et al., 1996; Cao et al., 1995; Chen et al. 1995; Peng et al., 1994; Chen & Zhou, 2003; Chen & Zhou, 2003; Liang & Jiang, 2002) in n letters surjective maps is $\ln(n)$, which is a deduction from chaotic symbolic sequences' Bernoulli property.

6.2 The inter-occurrence times of 4-letters maps

Now let us make a further study a given word's occurrence times in an independent repeated experiment, such as $"R", "T"$ or $"Winter"$ in the alphabet of four numbers. Given outcomes of a random sequence

$$\xi = (\xi_0, \xi_1, \xi_2, \cdots) \in \Omega^N,$$

we are mainly interested in n such that $\xi_n = T$, let

$$\tau^\lambda(\xi) = \tau_1^\lambda(\xi) = \inf\{n \geq 0 : \xi_n = T\}, \quad \lambda \in \Omega$$

and accordingly, for $j \geq 2$, λ corresponds to T,

$$\tau_j^\lambda(\xi) = \inf\{n > \tau_{j-1}^\lambda(\xi) : \xi_n = T\},$$

then for all $k \geq 0$, the result is, for fixed $k > 0$ and all $k_1 < \cdots < k_{j-1}$,

$$
\begin{aligned}
&\mathrm{P}(\tau_j^\lambda - \tau_{j-1}^\lambda = k \mid \tau_{j-1}^\lambda = k_{j-1}, \cdots, \tau_1^\lambda = k_1) \\
&= \mathrm{P}(\tau_j^\lambda - \tau_{j-1}^\lambda = k\} \\
&= p_1^a p_2^b p_3^c p_4
\end{aligned}
\tag{6}
$$

$$
(\sum_{i=1}^{4} p_i = 1, a,b,c \in \{0,1,2,\cdots\cdots\}, a+b+c = k-1)
\tag{7}
$$

the inter-occurrence times $1+\tau_1^\lambda, \tau_2^\lambda - \tau_1^\lambda, \tau_3^\lambda - \tau_2^\lambda, \cdots$ are i.i.d. with parameters p_1, p_2, p_3, p_4.

Correspondingly, for $j \geq 2$,

If λ corresponds to A, and

$$
\tau_j^\lambda(\xi) = \inf\{n > \tau_{j-1}^\lambda(\xi) : \xi_n = A\}
$$

Then

$$
\mathrm{P}(\tau_j^\lambda - \tau_{j-1}^\lambda = k\} = p_2^a p_3^b p_4^c p_1
\tag{8}
$$

If λ corresponds to G, and

$$
\tau_j^\lambda(\xi) = \inf\{n > \tau_{j-1}^\lambda(\xi) : \xi_n = G\}
$$

Then

$$
\mathrm{P}(\tau_j^\lambda - \tau_{j-1}^\lambda = k\} = p_1^a p_3^b p_4^c p_2
\tag{9}
$$

If λ corresponds to C, and

$$
\tau_j^\lambda(\xi) = \inf\{n > \tau_{j-1}^\lambda(\xi) : \xi_n = C\}
$$

then

$$
\mathrm{P}(\tau_j^\lambda - \tau_{j-1}^\lambda = k\} = p_1^a p_2^b p_4^c p_3
\tag{10}
$$

the conditions of expression (8)–(10) are also expression (7).

6.3 Exponential distribution of 4-letters maps

Suppose there are two random sequences of outcomes corresponding to the repetition of an experiment with four likely results. Let $\xi = (\xi_0, \xi_1, \xi_2, \cdots)$ and $\zeta = (\zeta_0, \zeta_1, \zeta_2, \cdots)$ denote the sequence of outcomes (ξ independent of ζ). There is an alignment at time n if $\xi_n = \zeta_n$. The alignment at time n as a success and no alignment is failure. Then note that, for all $n > 0$, $0 < p < 1$ and $q = 1 - p$,

$$P(\xi_n = \zeta_n) = P(\xi_0 = \zeta_0) = p$$

Now consider having a successive sequence of alignments. Define $\phi : \Omega^N * \Omega^N \to N$ by $\phi(\xi, \zeta) = k$ if $\xi_0 = \zeta_0, \xi_1 = \zeta_1, \ldots, \xi_{k-1} = \zeta_{k-1}, \xi_k \neq \zeta_k$, namely,

$$\phi(\xi, \zeta) = k, \text{ if } \xi_0 = \zeta_0, \xi_1 = \zeta_1, \ldots, \xi_{k-1} = \zeta_{k-1}, \xi_k \neq \zeta_k,$$

introduced shift arithmetic operators σ,

$$\sigma(\xi) = (\xi_1, \xi_2, \xi_3 \cdots),$$
$$\sigma(\zeta) = (\zeta_1, \zeta_2, \zeta_3 \cdots),$$

then,
for $i = 1, 2, 3, \cdots$,

$$\phi(\sigma(\xi), \sigma(\zeta)) = k, \text{ if } \xi_1 = \zeta_1, \ldots, \xi_k = \zeta_k, \xi_{k+1} \neq \zeta_{k+1};$$
$$\phi(\sigma^2(\xi), \sigma^2(\zeta)) = k, \text{ if } \xi_2 = \zeta_2, \ldots, \xi_{k+1} = \zeta_{k+1}, \xi_{k+2} \neq \zeta_{k+2};$$
$$\vdots \qquad\qquad\qquad \vdots$$
$$\phi(\sigma^i(\xi), \sigma^i(\zeta)) = k, \text{ if } \xi_i = \zeta_i, \ldots, \xi_{k-1+i} = \zeta_{k-1+i}, \xi_{k+i} \neq \zeta_{k+i}$$

in fact, alignment from i term is just keeping the former k terms success and the $k + i$ term failure.

Denote random variables

$$Z_n = \phi(\sigma^n \xi, \sigma^n \zeta), \ \gamma_k = \inf\{n \geq 0 : Z_n \geq k\}.$$

So,

$$\because \ \sigma^{n-1}(\xi) = (\xi_{n-1}, \xi_n, \xi_{n+1}, \cdots),$$
$$\sigma^{n-1}(\zeta) = (\zeta_{n-1}, \zeta_n, \zeta_{n+1}, \cdots)$$

and

$$\xi_{n-1} = \zeta_{n-1}, \xi_n = \zeta_n, \ldots\ldots, \xi_{n+k-2} = \zeta_{n+k-2}, \xi_{n+k-1} \neq \zeta_{n+k-1},$$

therefore,

$$Z_{n-1} = \phi(\sigma^{n-1}\xi, \sigma^{n-1}\zeta) = k.$$

The same way,

$$\because \quad \sigma^n(\xi) = (\xi_n, \xi_{n+1}, \xi_{n+2}, \cdots),$$
$$\sigma^n(\zeta) = (\zeta_n, \zeta_{n+1}, \zeta_{n+2}, \cdots)$$

and

$$\xi_n = \zeta_n, \xi_{n+1} = \zeta_{n+1}, \ldots\ldots, \xi_{n+k-2} = \zeta_{n+k-2}, \xi_{n+k-1} \neq \zeta_{n+k-1}$$

therefore,

$$Z_n = \phi(\sigma^n\xi, \sigma^n\zeta) = k-1$$

so,

$$P(Z_n = k-1 \mid Z_{n-1} = k) = 1, \text{ if } k \geq 1$$

for every $t > 0$, one gets the asymptotic exponential distribution of γ_k:

$$\lim_{k \to \infty} P(\gamma_k > tE(\gamma_k) \mid Z_0 = 0) = e^{-t},$$

$E(\gamma_k)$ represents mathematical expectation of γ_k, t is a time coordinate. Furthermore, if

$$Z_n = \phi(\sigma^n\xi, \sigma^n\zeta) = k-1$$

and $\xi_{n+k-1} = A$, then,

$$\lim_{k \to \infty} P(\gamma_k > t_A E(\gamma_k)) = e^{-t_A}$$

accordingly, one also gets,

$$\lim_{k \to \infty} P(\alpha_k > t_G E(\alpha_k)) = e^{-t_G},$$
$$\lim_{k \to \infty} P(\beta_k > t_C E(\beta_k)) = e^{-t_C},$$

$$\lim_{k \to \infty} P(\omega_k > t_T E(\omega_k)) = e^{-t_T} .$$

$\alpha_k , \beta_k , \omega_k$ correspond to γ_k when $\xi_{n+k-1} = G$, $\xi_{n+k-1} = C$, $\xi_{n+k-1} = T$.
We also get,

$$\lim_{k \to \infty} P(\gamma_k > t_A E(\gamma_k), \alpha_k > t_G E(\alpha_k), \beta_k > t_C E(\beta_k), \omega_k > t_T E(\omega_k))$$
$$= e^{-(t_A + t_G + t_C + t_T)}$$

Fig. 4 represents the alignment of two random sequences.

6.4 Transfer probability of Markov chain (MC) in 4-letters maps

We choose one of these transfer models (Figure. 5), such as Figure. 6.

If n fixation, the transfer probability of (i, j) is defined $p_{ij}^{(n)}$. The generating function is,

$$G^n(z) = (qz^{-1} + r + pz + vz^2)^n , \qquad (11)$$

which is determined by Figure. 3.

$$G^n(z) = \sum_{k \in R} p_{1k}^{(n)} z^k = \sum_{j-i \in R} p_{ij}^{(n)} z^{j-i} . \qquad (12)$$

Hereinto, $p_{1k}^{(n)} = P(X_n = k \mid X_0 = 1)$, $p_{ij}^{(n)} = P(X_n = j \mid X_0 = i)$, $k = j - i$,

$R = (-n, \cdots, 0, \cdots, 2n)$, R represents all coordinates a particle could attain during n steps transfer if n fixation. By using (11) and (12), one gets:

under restrictive condition of $\delta_{(2n_3 + n_2 - n_0), k}$:

$$p_{1k}^{(n)} = \frac{n!}{n_0 ! n_1 ! n_2 ! n_3 !} q^{n_0} r^{n_1} p^{n_2} v^{n_3} ,$$

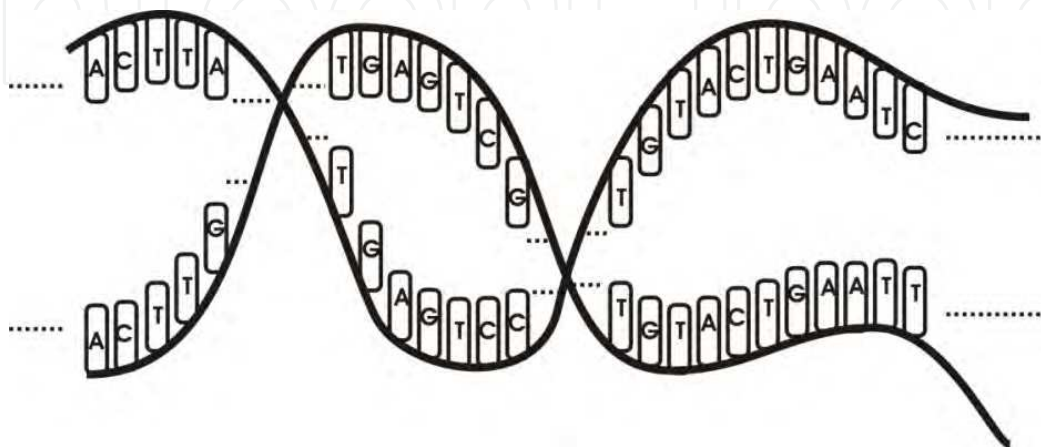$$(\delta : k = 2n_3 + n_2 - n_0, \quad n = n_0 + n_1 + n_2 + n_3)$$
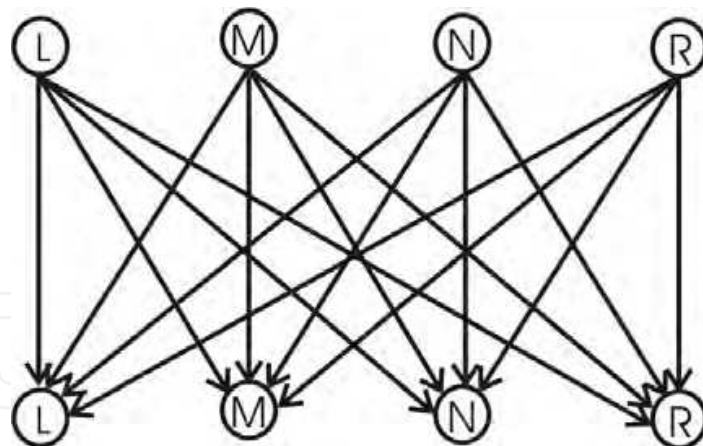

Fig. 4. The alignment of two random sequences

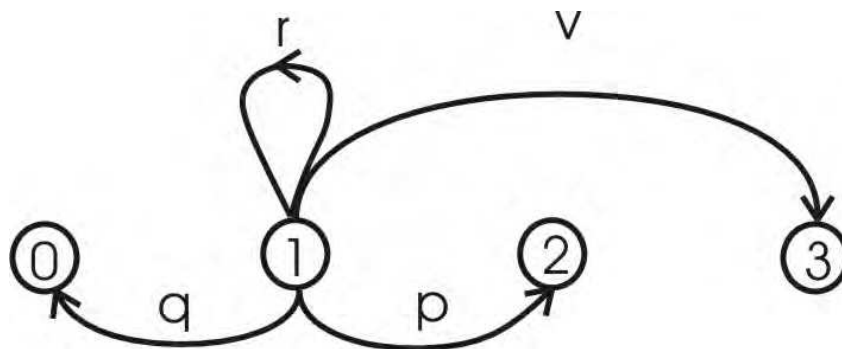Fig. 5. Four probabilities stochastic wander model

Fig. 6. One of these transfer models

Let,

$$n - n_1 + n_3 - k = 2l, \ n_0 = l, \ n_1 = m, \ n_2 = l + k - 2\lambda, \ n_3 = \lambda \,,$$

then,

$$p_{1k}^{(n)} = \frac{(2l + m + k - \lambda)!}{l!m!(l+k-2\lambda)!\lambda!} q^l r^m p^{(l+k-2\lambda)} v^\lambda, \quad k \in [-n, 2n].$$

The generating function $P_{1k}(s)$ if k fixation. During different n steps transfer, all the probabilities $p_{1k}^{(n)}$ are introduced into a generating function $P_{1k}(s)$, which is defined as

$$P_{1k}(s) = \sum_{n=0}^{\infty} p_{1k}^{(n)} s^n$$

$$= (ps)^k \sum_{\lambda=0}^{\infty} \sum_{l=0}^{\infty} C_{2l+k-\lambda}^l C_{l+k-\lambda}^\lambda (qs)^l (ps)^{(l-2\lambda)} (vs)^\lambda (1-rs)^{-(2l+k-\lambda+1)}$$

$$= (\frac{ps}{1-rs})^k \frac{1}{1-rs} \sum_{\lambda=0}^{\infty} \sum_{l=0}^{\infty} C_{2l+k-\lambda}^l C_{l+k-\lambda}^\lambda (\frac{qs}{1-rs})^l (\frac{ps}{1-rs})^{(l-2\lambda)} (\frac{vs}{1-rs})^\lambda$$

let,

$$x_0 = \frac{qs}{1-rs}, \; x_2 = \frac{ps}{1-rs}, \; x_3 = \frac{vs}{1-rs};$$

$$x = x_0 x_2 = \frac{qps^2}{(1-rs)^2}, \; y = x_0^2 x_3 = \frac{q^2 vs^3}{(1-rs)^3};$$

then,

$$P_{1k}(s) = (\frac{qs}{1-rs})^{-k} \frac{1}{1-rs} \sum_{\lambda=0}^{\infty} \frac{y^\lambda}{\lambda!} \frac{d^\lambda}{dx^\lambda} (\sum_{l=0}^{\infty} C_{2l+k+\lambda}^{l+k} x^{l+k}).$$

So, for $k = 1$,

$$P_{11}(s) = \frac{1}{1-rs} \sum_{\lambda=0}^{\infty} \frac{y^\lambda}{\lambda!} \frac{d^\lambda}{dx^\lambda} (\sum_{l=0}^{\infty} C_{2l+\lambda}^{l} x^l)$$

for $k = 2$,

$$P_{12}(s) = \frac{1}{qs} \sum_{\lambda=0}^{\infty} \frac{y^\lambda}{\lambda!} \frac{d^\lambda}{dx^\lambda} (\sum_{l=0}^{\infty} C_{2l+\lambda+1}^{l+1} x^{l+1})$$

for $k = 3$,

$$P_{13}(s) = \frac{1-rs}{q^2 s^2} \sum_{\lambda=0}^{\infty} \frac{y^\lambda}{\lambda!} \frac{d^\lambda}{dx^\lambda} (\sum_{l=0}^{\infty} C_{2l+\lambda+2}^{l+2} x^{l+2})$$

7. Conclusion and discussion

The statistic character and complexity in nonlinear systems have been clarified in this chapter. These stochastic symbolic sequences bear three characters. In two kinds of typical nonlinear systems-unimodal surjective map and Lorenz type maps nonlinear systems, the distributions of frequency, inter-occurrence times, first passage time and visitation density in unimodal surjective map and Lorenz type maps are discussed carefully. These two kinds of nonlinear systems have same distributions, which have also been explained in theory, and the catholicity of the statistic character has been elicited. In the 4-symbolic dynamics, the distribution of frequency, inter-occurrence times and the alignment of two random sequences have been amplified in detail. By using transfer probability of Markov chain (MC), we have obtained analytic expressions of generating functions in four probabilities stochastic wander model, which can be applied to all 4-symbolic systems. So, a perfect symbolic platform has been set up for our utilizing statistic character, in fact, it is a stochastic signal platform of symbolic simulation. The 4-symbolic sequences have natural relations with bioinformatics sequences, in the field of application, we hope to afford a symbolic platform which satisfies these statistic character and study some properties of DNA sequences (Hao, 2000; Hao et al., 2000; Bershadskii, 2001; Grimm & Rupprecht, 1997; Allegrini et al., 1996; Natalia & Avy, 2005; Elena et al., 2005), 20 amino acids symbolic sequences of protein structure, and the time series that can be symbolic in finance market et al, which are part of our future work. The symbolic platform provides a set of effective

methods to approach problems of this kind. The establishment of this symbolic platform will open up a vast vista.

8. Acknowledgements

9. References

Alekseev, V.M. & Yakobson, M.V. (1981). Symbolic dynamics and hyperbolic dynamics systems. *Physics Reports*, Vol. 75, 287-325, ISSN: 0370-1573

Allegrini, P., Grigolini, P. & West B.J. (1996). A dynamical approach to DNA sequences. *Physics Letters A*, Vol. 211, 217-222, ISSN: 0375-9601

Bershadskii, A. (2001). Multifractal and probabilistic properties of DNA sequences. *Physics Letters A*, Vol. 284, 136–140, ISSN: 0375-9601

Billingsley, P. (1986). *Probability and Measure*, John Wiley & Sons, ISBN: 0-471-80478-9, New York

Cao, K.F., Chen, Z.X. & Peng, S.L. (1995). Global metric regularity of the devil's staircase of topological entropy. *Physical Review E*, Vol. 51, 1989-1995, ISSN: 1539-3755

Chen, Z.-X., Cao, K.-F. & Peng, S.-L. (1995). Symbolic dynamics analysis of topological entropy and its multifractal structure. *Physical Review E*, Vol. 51, 1983-1988, ISSN: 1539-3755

Chen, Z.X. & Zhou, Z. (2003). Entropy invariants: I . The universal order relation of order-preserving star products. *Chaos, Solitons & Fractals*, Vol. 15, 713-727, ISSN: 0960-0779

Chen, Z.X. & Zhou, Z. (2003). Entropy invariants: II . The block structure of Stefan matrices. *Chaos, Solitons & Fractals*, Vol. 15, 729-742, ISSN: 0960-0779

Coelho, Z. (2000). On discrete stochastic processes generated by deterministic sequences and multiplication machines. *Indagationes Mathematicae*, Vol. 11, 359−378, ISSN: 0019-3577

Coelho, Z. & Collet, P. (1994). Asymptotic limit law for the close approach of two trajectories in expanding maps of the circle. *Journal of Probability Theory and related fields*, Vol. 99, 237-250, ISSN: 0178-8051

Collet, P. & Eckmann, J.P. (1980). *Iterated Maps on the Interval as Dynamical Systems*, Birkhauser, ISBN-13: 978-0817-63-026-3, Boston

Ding, J. & Li, T.Y. (1991). Markov finite approximation of Frobenius–Perron operator. *Nonlinear Analysis: Theory, Methods & Applications*, Vol. 17, 759–772, ISSN: 0362-546X

Elena, I., Andrei, P., Sergey V. & Yegor S. (2005). Mapping long-range chromatin organization within the chicken-globin gene domain using oligonucleotide DNA arrays. *Genomics*, Vol. 85, 143–151, ISSN: 0888-7543

Grimm, H. & Rupprecht, A. (1997). Low frequency dynamics of DNA. *Physica B*, Vol. 234-236, 183-187, ISSN: 0921-4526

Hao, B.L. (1989). *Elementary Symbolic Dynamics and Chaos in Dissipative Systems,* World Scientific, ISBN: 978-9971-50-682-7, Singapore

Hao, B.L. (1991). Symbolic Dynamics and characterization of complexity. *Physica D*, Vol. 51, 161-176, ISSN: 0167-2789

Hao, B.L. (2000). Fractals from genomes-exact solutions of a biology-inspired problem. *Physica A*, Vol. 282, 225-246, ISSN: 0378-4371

Hao, B.L., Lee, H.C. & Zhang, S.Y. (2000). Fractals related to long DNA sequences and complete genomes. *Chaos, Solitons & Fractals*, Vol. 11, 825-836, ISSN: 0960-0779

Hao, B.L. & Zheng, W.M. (1998). *Symbolic Dynamics and Chaos. Directions in Chaos* Vol. 7, World Scientific, ISBN: 978-9971-50-698-8, Singapore

Lasota, A. & Mackey, M.C. (1985). *Probabilistic properties of deterministic systems*, Cambridge University Press, ISBN: 0-521-30248-X, Cambridge

Li, T.Y. (1976). Finite approximation for the Frobenius–Perron operator, a solution to Ulam's conjecture. *Journal of Approximation Theory*, Vol. 17, 177–186, ISSN: 0021-9045

Liang, X. & Jiang, J.F. (2002). On the topological entropy, nonwandering set and chaos of monotone and competitive dynamical systems. *Chaos, Solitons & Fractals*, Vol. 14, 689–696, ISSN: 0960-0779

Natalia, L. & Avy, S. (2005). Nonlinear waves in double-stranded DNA. *Bulletin of Mathematical Biology*, Vol. 67, 701–718, ISSN: 0092-8240

Peng, S.L. & Cao, K.F. (1996). Global scaling behaviors and chaotic measure characterized by the convergent rates of period-p-tupling bifurcations. *Physical Review E*, Vol. 54, 3211-3220, ISSN: 1539-3755

Peng, S.L., Cao, K.F. & Chen, Z.X. (1994). Devil's staircase of topological entropy and global metric   regularity. *Physics Letters A*, Vol. 193, 437-443, ISSN: 0375-9601

Peng, S.L. & Du, L.M. (1999). Dual star products and symbolic dynamics of Lorenz maps with the same entropy. *Physics Letters A*, Vol.261, 63–73, ISSN: 0375-9601

Peng, S.L. & Luo, L.S. (1991). The ordering of critical periodic points in coordinate space by symbolic dynamics. *Physics Letters A*, Vol. 153, 345-352, ISSN: 0375-9601

Shi, J.X., Cao, K.F., Guo, T.L. & Peng, S.L. (1996). Metric universality for the devil's staircase of topological entropy. *Physics Letters A*, Vol. 211, 25-28, ISSN: 0375-9601

Ulam, S.M. & Von Neumann, J. (1947) . On combination of stochastic and deterministic processes. *Bulletin of the American Mathematical Society*, Vol. 53, 1120, ISSN: 0273-0979

Xie, H.M. (1993). On formal languages of one-dimensional dynamics systems. *Nonlinearity*, Vol. 6, 997-1007, ISSN: 0951-7715

Xie, H.M. (1996). Grammatical Complexity and one-dimensional dynamics systems. Directions in Chaos Vol. 6, World Scientific, ISBN 978-9810-22-398-4, Singapore

Yorke, J. & Li, T.Y. (1975). Period three implies chaos. *The American Mathematical Monthly*, Vol. 82, 985–992, ISSN: 0002-9890

Zhou, Z. & Cao, K.F. (2003). An effective numerical method of the word-lifting technique in one-dimensional multimodal maps. *Physics Letters A*, Vol. 310, 52-59, ISSN: 0375-9601

Zhou, Z. & Peng, S.L. (2000). Cyclic star products and universalities in symbolic dynamics of trimodal maps. *Physica D*, Vol. 140, 213–226, ISSN: 0167-2789

Zhang, X.S., Liu, X.D., Kwek, K.H. & Peng, S.L. (1996). Disorder versus order: Global multifractal relationship between topological entropies and universal convergence rates. *Physics Letters A*, Vol. 211, 148-154, ISSN: 0375-9601

Machine Learning

Edited by Yagang Zhang

Machine learning techniques have the potential of alleviating the complexity of knowledge acquisition. This book presents today's state and development tendencies of machine learning. It is a multi-author book. Taking into account the large amount of knowledge about machine learning and practice presented in the book, it is divided into three major parts: Introduction, Machine Learning Theory and Applications. Part I focuses on the introduction to machine learning. The author also attempts to promote a new design of thinking machines and development philosophy. Considering the growing complexity and serious difficulties of information processing in machine learning, in Part II of the book, the theoretical foundations of machine learning are considered, and they mainly include self-organizing maps (SOMs), clustering, artificial neural networks, nonlinear control, fuzzy system and knowledge-based system (KBS). Part III contains selected applications of various machine learning approaches, from flight delays, network intrusion, immune system, ship design to CT and RNA target prediction. The book will be of interest to industrial engineers and scientists as well as academics who wish to pursue machine learning. The book is intended for both graduate and postgraduate students in fields such as computer science, cybernetics, system sciences, engineering, statistics, and social sciences, and as a reference for software professionals and practitioners.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH
open science | open minds