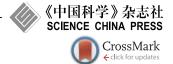
SCIENTIA SINICA Mathematica

综述



# 随机动力学: 内源与外源噪声的数学模型及其应用

献给钱敏教授 90 华诞

马易安,钱紘\*,叶晓峰

Department of Applied Mathematics, University of Washington, Seattle 98195-3925, USA E-mail: yianma.sjtu@gmail.com, hqian@u.washington.edu, yexf308@uw.edu

收稿日期: 2017-07-12; 接受日期: 2017-09-16; 网络出版日期: 2017-11-01; \* 通信作者

摘要 复杂系统与过程的数学建模需要用随机动力学 (stochastic dynamics) 的思想和方法. 随机动力学的理论有着两种不同的数学表述: 随机过程 (stochastic processes) 和随机动力系统 (random dynamical systems). 后者是比前者更为精细的数学模型, 它不但给出对应于每一个初值的随机过程, 还全面地描述不同初值的多条随机轨道如何同时随时间变化. 前者恰恰表述了有内在随机性的个体的运动, 而后者则反映了多个相同的确定性个体同时经历同一个随机环境. 本文称这两种情形为内源噪声和外源噪声. 两者都在化学和生物学中有广泛的应用. 近年来兴起的以图  $G(\mathcal{V},\mathcal{E})$  为基础的概率布尔网络正是一类以  $\{0,1\}^{\mathcal{V}}$  为状态空间的随机动力系统 (RDS). 本文介绍有关离散时间离散空间的RDS, 同时也给出一个它在统计推断隐 Markov 模型的收敛速率估算中的应用.

**关键词** Markov 链 随机映射 随机矩阵 概率布尔网络 随机梯度下降法 数学生物学 **MSC (2010) 主题分类** 60J10, 37H15, 34D08, 92C40, 65C05

# 1 引言

力学的基本单元是质点, 其运动完全由其质量、位置、速度和受的力决定. 与之相对应, 化学的基本单元是分子. 一个分子具有极其复杂的内在结构, 而且化学反应的机理和快慢与反应分子们的内在结构有着紧密的联系. 由于巨大的内在自由度和外在的复杂环境, 水溶液中一个化学反应的发生有很大的随机性. 事实上, 大多数水溶液中的化学反应可以用非时齐的 Poisson 过程来描述. Poisson 过程中的指数等待时间的速率, 化学家们称为反应速率. 因此从根本上来讲, 化学反应系统是一个随机动力学系统. 这样的认识在近年来单分子化学物理、单细胞的生物化学动力学和基因表达等建模中得以充分的发挥 (参见文献 [1,2]). 可以证明, 以常微分方程组表述的古典确定性宏观化学动力学是由上面的随机过程通过大数定律决定的 (参见文献 [3,4]). Avogadro 数在此起到了决定性的作用.

以上对分子的讨论也同样适用于其他的由个体组成的种群动力学,例如,活组织内不同表型的细胞,或者生态系统中众多的生物群体(参见文献[5,6]).因为所有这些应用的需要,随机动力学的思想、语言和方法正在应用数学中迅猛发展.在过去的半个世纪内,应用数学从以力学为主导,常微和偏微分

英文引用格式: Ma Y A, Qian H, Ye F X F. Stochastic dynamics: Models for intrinsic and extrinsic noises and their applications (in Chinese). Sci Sin Math, 2017, 47: 1693–1702, doi: 10.1360/N012017-00154

方程的解及其计算为主题扩展成由确定性数学和随机数学齐头并进,物理、生物和化学各有所取的庞大学科.

在纯数学研究领域内,有两个颇为不同的分支都为应用随机动力学提供理论基础,它们分别是随机过程 (stochastic processes) 的理论和随机动力系统 (random dynamical systems) 的理论. 前者是古典常微分方程的推广,后者则为近代动力系统的延深. 大概地说,对于应用数学来讲极为重要的是这两门数学理论对随机性的来源有着不同的假设,前者注重于每个独立的个体在其运动中的内在随机性;而后者强调的是每个独立的个体的运动遵循着同一个运动规律,但是这个统一的运动规律在随机的变化. 这两种想法恰好对应了生物随机动力学建模中的内源噪声 (intrinsic noise) 和外源噪声 (extrinsic noise) 的区分.

为了能够用初等的数学方法将随机动力学模型的基本思想介绍给广大的科学工作者们, 我们仅 考虑离散时间  $t \in \mathbb{N}$ 、离散状态空间  $\mathscr{S}$  上的随机运动. 在随机过程的理论框架内, 以一步转移概率  $p_{ij} = \Pr\{X_{t+1} = j \mid X_t = i\} \ (i, j \in \mathscr{S}, t \in \mathbb{Z}^*)$  定义的 Markov 链  $X_t$  是广为人知的. 本文平行地讨论随机动力系统, 或者更准确地讲, 随机映射 [7-9].

# 2 有限状态空间上随机映射及其定义的随机过程

# 2.1 有限状态空间上的确定性运动

给定一个离散空间上离散时间的动力系统  $(\mathcal{S}, M)$ : 对任意  $i \in \mathcal{S}$ , 存在唯一的  $j \in \mathcal{S}$ , j = M(i). 映射 M 可以用一个线性变换, 具体为矩阵 T, 来表示 (参见文献 [10]),

$$T = \{T_{ij} \mid i, j \in \mathcal{S}, T_{ij} = \delta_{i,M(i)}\}. \tag{2.1}$$

我们称这样的 T 为确定性转移矩阵,它是映射 M 的线性算子表示,是退化的 Markov 链转移矩阵.矩阵左乘一个向量在连续空间动力系统理论中对应的是 Perron-Frobenius-Ruelle 算子,而矩阵右乘一个向量在连续空间动力系统理论中对应的是 Koopman 算子 (参见文献 [11]).显而易见,一对一映射的 T 为置换矩阵,是可逆的. 但是多对一映射的 T 不可逆.

### 2.2 独立同分布的随机映射族及其定义的 Markov 链

一个独立同分布的随机动力系统 (RDS) 是由  $(\mathcal{S}, \Gamma, Q)$  决定的, 其中集合  $\mathcal{S}$  为状态空间,  $\Gamma$  为一族  $\mathcal{S} \to \mathcal{S}$  的映射, Q 为  $\Gamma$  的  $\sigma$ - 域  $\mathcal{F}$  上的概率测度,  $(\Gamma, \mathcal{F}, Q)$  为概率空间.

如果状态空间是有限的  $n \equiv \|\mathcal{S}\|$ , 则其上总共有  $n^n$  个不同的映射. 如果赋予每个映射  $M_i$  一个概率  $\alpha_i$ , 那么由无穷长的同分布映射序列的迭代

$$\cdots \circ M_{t_k} \circ \cdots \circ M_{t_2} \circ M_{t_1}, \quad M_{t_k} \in \Gamma$$
 (2.2)

构成的集合定义了一个有限的子移位 (subshift of finite type). 这映射序列也对应着一个矩阵序列

$$T_{t_1}T_{t_2}\cdots T_{t_k}\cdots, (2.3)$$

它是一个随机动力系统的表示. 如果初始状态  $x \in \mathcal{S}$  已经给定, 那么我们可以用一个标准正交基  $\vec{e}_x$  来表示此状态, 则

$$\vec{e}_x T_{t_1} T_{t_2} \cdots T_{t_k} \cdots \tag{2.4}$$

定义了一条随机轨道. 如果  $T_{t_k}$  是独立同分布抽样, 那么 (2.4) 中的轨道是 Markov 的. 此 Markov 链 的转移矩阵是

$$M = \sum_{i=1}^{n^n} \alpha_i T_i \equiv \mathbb{E}^Q[T]. \tag{2.5}$$

Perron-Frobenius-Ruelle 算子的期望满足 Kolmogorov 前进方程, Koopman 算子的期望满足 Kolmogorov 后退方程. 在 RDS 的理论框架里, 随机过程的半群应作为某种算子的数学期望来理解.

如果 (2.5) 中的每个  $T_i$  都是可逆的, 那么  $M = \{p_{ij}\}$  是双随机矩阵:

$$\sum_{i \in \mathscr{S}} p_{ij} = \sum_{j \in \mathscr{S}} p_{ij} = 1.$$

进一步, 如果 (2.5) 中的每个  $T_i$  和它的逆  $T_i^{-1} \equiv T_{i^*}$  有同样的  $\alpha_i = \alpha_{i^*}$ , 那么由 M 生成的平稳 Markov 链是可逆 Markov 过程 (参见文献 [12]).

# 2.3 Markov 链的随机映射表示

既然 Markov 链的转移矩阵是一种数学期望,那么一定有许多不同的 RDS 对应同一 Markov 链. 因此我们说前者是较后者更为精细的数学描述.换用优化的语言, $n \times n$  Markov 转移矩阵的集合是一凸集,确定性转移矩阵是其顶点.此观点对应 von Neumann-Birkhoff 定理在一般 Markov 链上的推广.一些有趣的但尚未完全解决的数学问题自然出现.例如,给定一个 Markov 矩阵 M,其对应的 RDS 至少需要几个 T,记为  $\kappa_*(M)$ ?可以证明, $\kappa_*(M) \le n^2 - n + 1$  (参见文献 [10]). 类似这样的问题的答案在应用建模中将会是非常有用的.

熵和熵产生是随机系统里的两个重要概念. 它们起源于热力学, 而热力学的基础是分子和原子的随机运动. 所以, 这两个概念自然而然在随机动力学的研究中出现. 香农的信息熵也与化学动力学里的熵力有着紧密的关系 (参见文献 [13]).

一个以  $\{p_{ij}\}$  为转移概率的 Markov 链的度规熵 (metric entropy) 是

$$h = -\sum_{i,j\in\mathscr{S}} \pi_i p_{ij} \ln p_{ij}, \tag{2.6}$$

其中  $\{\pi_i\}$  是此 Markov 链的平稳概率分布. 同一个 Markov 过程的拓扑熵可以如下得到. 首先引进矩阵 A: 当  $p_{ij}=0$  时,  $A_{ij}=0$ , 如果  $p_{ij}\neq 0$ , 则  $A_{ij}=1$ . 此 Markov 过程的拓扑熵  $\eta=\ln\lambda_A$ , 这里  $\lambda_A$  是矩阵 A 的最大特征值. 可以证明  $h\leqslant \eta$ . 当 Markov 链有  $p_{ij}=\frac{1}{n},\,h=\eta=\ln n$ . 作为 Markov 链的特例, 独立同分布序列有  $p_{ij}=\alpha_j$ . 于是独立同分布序列的度规熵为  $h_{RDS}=-\sum_i\alpha_i\ln\alpha_i$ . 将这个结果用到方程 (2.2) 中的序列, h 是独立同分布 RDS 的度规熵.

有了 RDS 度规熵的概念, Kifer [7] 和 Ye 等[10] 研究了对于一个给定的 Markov 矩阵 M, 其对应 RDS 的最小度规熵的问题, 并证明了  $h_{\rm RDS} \geqslant h_{\rm MC}$ . 对于给定的 M, 前面的等号不一定取得到, 所以 这是一个较松的下界.

# 2.4 非周期不可约 Markov 链的最大熵随机映射表示

在另一方向,对于一个给定的非周期不可约 Markov 矩阵 M,其对应 RDS 的最大度规熵有一个很直观的结果 (参见文献 [10]). 在优化的理论中,度规熵是一个凸函数. 它在一个凸集中存在唯一的最大值.下面的定理给出了最大值达到的显式条件及表达式.

# 定理 1 对应非周期不可约 Markov 链的独立同分布 RDS 的度规熵为

$$h_{\text{RDS}} \leqslant -\sum_{j,k} M_{jk} \ln M_{jk}.$$

记确定性转移矩阵  $T \equiv \{T_{ij} = \delta_{i,t_i}, i, t_i \in \mathcal{S}\}$  为  $P_{t_1,t_2,...,t_n}$ , 则对应上面等号的 RDS 赋予  $P_{t_1,t_2,...,t_n}$  概率

$$a_{t_1,t_2,\dots,t_n} = \prod_{k=1}^n M_{kt_k}. (2.7)$$

证明 计算可得

$$-\sum_{j,k=1}^{n} M_{kj} \ln M_{kj} - h_{RDS} = -\sum_{j,k=1}^{n} \left( \sum_{i_{k}=j} a_{i_{k},i_{2},...,i_{n}} \right) \ln M_{kj} - h_{RDS}$$

$$= -\sum_{i_{1},i_{2},...,i_{n}}^{n} a_{i_{k},i_{2},...,i_{n}} \ln \left( \prod_{k=1}^{n} M_{ki_{k}} \right) - h_{RDS}$$

$$= \sum_{i_{1},i_{2},...,i_{n}}^{n} a_{i_{k},i_{2},...,i_{n}} \ln \left( \frac{a_{i_{k},i_{2},...,i_{n}}}{\prod_{k} M_{ki_{k}}} \right)$$

$$\geqslant 0.$$

证毕.

# 2.5 独立同分布的随机映射下 Markov 链的同步

RDS 理论框架下的一个新课题是不同初值多条随机轨道有"同步"(synchronize)的可能性. 例如,随机映射族含有一确定性转移矩阵为多对一,那么在这个映射的作用下,多条轨道会并成一条. 之后所有的轨道将永远保持在一起. 此现象称为同步. 如果随机映射族含有的每个确定性转移矩阵皆为可逆的置换矩阵,那么此 RDS 不会同步. 有关同步有如下定理,其证明可见文献 [10],在此省略.

**定理 2** 对应 Markov 链 M 的最大度规熵 RDS 会同步的充分必要条件是 M 有唯一的吸收互通类,而且此互通类非周期.

同步的概念起源于非自治的动力系统理论<sup>[14]</sup>. 近年里, 它被 Lin 等<sup>[15]</sup> 应用到神经网络动态涌现 (emergent) 行为的描述. 最大度规熵 RDS 的同步也与 Markov 链的 Doeblin 偶合方法有着密切的关系. 前者为后者提供了一个直观的理解.

# 3 随机映射应用的两个例子

这里仅讨论有限状态的随机映射的应用, 有关连续状态的随机动力系统的应用, 参见专著 [14,16].

### 3.1 概率布尔网络 (probabilistic Boolean network) 的动力学

确定性的布尔网络动力学 (Boolean network dynamics) 正是有限状态确定性动力系统的一个极好的例子. 它是一类特殊的离散时间有限状态空间  $\mathcal{S} = \{0,1\}^{\mathcal{V}}$  上的运动, 其中  $\mathcal{V}$  是图  $G(\mathcal{V},\mathcal{E})$  上的节点的集合. 这类数学模型在研究神经网络<sup>[17]</sup>、细胞信号调控<sup>[18]</sup> 和系统生物学<sup>[19]</sup> 中被大量地运用. 带有随机性的布尔网络则在近代生物学中有着更加广泛的应用. Kauffman <sup>[20]</sup> 首先在 20 世纪 70 年

代就研究了以随机图为基础的 random Boolean networks. 在机器学习中被广为用到的 Boltzmann 机是确定性布尔神经网络的一个概率版本 (参见文献 [21,22]). 同样的思想也被运用到了细胞环调控网络的动力学 (参见文献 [23,24]).

上面提到的随机模型都是以随机过程为基础的. 而确定性布尔基因调控网络的另外一种随机推广,称作概率布尔网络 (probabilistic Boolean networks, PBN) [25], 是一个由细胞的随机环境决定 (context-sensitive) 的随机动力系统: 在给定的环境下, 不同细胞个体内的基因的调控取决于同一个确定性的布尔网络. 不同的细胞的运动为同一个随机动力系统中不同的随机轨道. PBN 同时有一个新的特征: 它的每一个确定性环境持续一段随机的时间, 以几何分布来描述. 因此, PBN 在映射层面更类似一个准Markov 过程 (Q process) 而非独立同分布 Markov 链, 其定义的随机轨道一般不再有 Markov 性.

受这些应用领域里工作的启发, 我们引进布尔网络上的随机动力学 (stochsatic Boolean network dynamics, SBND). 一个确定性的映射为  $f = (f_1, f_2, \dots, f_K) : \{0, 1\}^{\mathscr{V}} \to \{0, 1\}^{\mathscr{V}}, K = \|\mathscr{V}\|$ :

$$s_i(t+1) = f_i(s_{j_1}(t), s_{j_2}(t), \dots, s_{j_{k_i}}(t)),$$
(3.1)

其中  $s_i$  ( $f_j \in \{0,1\}$ ,  $j_1,j_2,\ldots,j_{k_i}$ ) 是与图上 i 点相邻的点. 我们将引进符号  $\mathfrak{N}_i$  作为所有与点 i 相邻的点的集合:  $\mathfrak{N}_i \equiv (j_1,j_2,\ldots,j_{k_i}) \subset \mathscr{V}$ . 于是,一个 random Boolean network 的集合  $\mathfrak{N}_i$  可以是随机的,它表示一种图  $G(\mathscr{V},\mathscr{E})$  的结构的随机性,而 PBN 则是在一族不同的映射中随机地一个一个取. 它们的区别并非本质,只是先引进一个图的结构,然后再在选定的图上定义映射.

选定一个随机映射  $f = (f_1, f_2, \dots, f_K)$ ,以及一个随机时间段  $\tau \in \mathbb{N}$ , $\Pr\{\tau = k\} = (1 - \beta)\beta^{k-1}$ ,此非独立同分布 RDS 定义的随机迭代为

$$s(t) = \underbrace{(f^{(t)} \cdots \circ f^{(t)})}_{\tau_t} \circ \cdots \circ \underbrace{f^{(2)} \cdots \circ f^{(2)}}_{\tau_2} \circ \underbrace{f^{(1)} \circ f^{(1)} \cdots \circ f^{(1)}}_{\tau_1}) s(0). \tag{3.2}$$

独立同分布的 RDS 有  $\beta = 0$ , 于是  $\tau = 1$ .

(3.2) 中的非 Markov 轨道可以用准 Markov 过程的思想来理解. 考虑每一个映射  $f^{(i)}$  和概率  $\alpha_i$ . 引入准映射

$$f^{i\circ\tau} \equiv \underbrace{f^{(i)} \circ f^{(i)} \cdots \circ f^{(i)}}_{\tau}, \quad \tau = 1, 2, \dots, \tag{3.3}$$

的全体的集合:  $\{f^{i\circ\tau} \mid i=1,2,\ldots,K; \tau=1,2,\ldots\}$ . 赋予准映射  $f^{i\circ\tau}$  概率

$$p_i(\tau) = \alpha_i (1 - \beta_i) \beta_i^{\tau - 1}, \quad \beta_i \leqslant 1. \tag{3.4}$$

注意到状态空间  $\mathscr{S} = \{0,1\}^{\checkmark}$  是有限的, 因此所有的准映射的集合是有限的. 于是在概率 (3.4) 的意义下, 它们构成了一个准独立同分布 RDS. 若  $f^{(i)}$  用一个矩阵 T 来表示, 则  $f^{i\circ\tau}$  的表示为  $T^{\tau}$ . 由以上准独立同分布 RDS 定义的非 Markov 平稳随机过程的一步转移概率是 RDS 的矩阵的期望为

$$\sum_{i} \sum_{\tau=1}^{\infty} p_{i}(\tau) \mathbf{T}_{i}^{\tau} = \sum_{i} \alpha_{i} (1 - \beta_{i}) \mathbf{T}_{i} \sum_{\tau=1}^{\infty} (\beta_{i} \mathbf{T}_{i})^{\tau-1}$$
$$= \sum_{i} \alpha_{i} \mathbf{T}_{i} \left[ \frac{\mathbf{I} - \beta_{i} \mathbf{T}_{i}}{1 - \beta_{i}} \right]^{-1}.$$

**定义 1** 如果 n 状态空间  $\mathcal S$  上的映射 M 的 n 次迭代的矩阵表示为

$$T^{n} = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}, \tag{3.5}$$

我们称映射 M 有唯一的吸收态.

显而易见, M 有唯一吸收态当且仅当其确定性转移矩阵 T 有唯一非零特征值, 且此特征值为 1. 以下是两个例子:

$$\begin{pmatrix}
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}
\qquad$$

$$\pi \quad
\begin{pmatrix}
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix},$$

它们的吸收态是状态都为 5:  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$ ;  $1 \rightarrow 2 \rightarrow 5 \leftarrow 4 \leftarrow 3$ . 有唯一吸收态的映射的确定性转移矩阵可以表示成上三角形和对角线上唯一一个 1.

定理 3 准独立同分布 RDS, 若存在一映射有唯一吸收态而且其  $\tau \neq 1$ , 则此 RDS 会同步.

**证明** 设  $f^{(1)}$  为有唯一吸收态的确定性映射, 则当  $\tau > n$  时,  $f^{1\circ\tau}$  有矩阵表示, 具有形式 (3.5), 而且其概率非零. 因此准独立同分布 RDS 会同步.

### 3.2 隐 Markov 模型统计推断的收敛速率

动力学的思想与 Monte Carlo 抽样方法<sup>[26]</sup> 的结合引导出了现今统计学的一个新研究方向 (参见文献 [27]). 非梯度系统和非对称 Markov 链正在拓广传统的、具有细致平衡的 Metropolis-Hasting 算法<sup>[27,28]</sup>. 这与以不可逆 Markov 过程为基础的非平衡态统计热力学的发展遥相呼应 (参见文献 [12,29]). 近年来在大数据的推动下,随机梯度 Monte Carlo 抽样方法<sup>[30,31]</sup> 快速发展,并与不可逆 Monte Carlo 抽样方法结合<sup>[31–33]</sup> 产生了既高效、又能处理大量数据的 Bayes 推断方法. 这些方法需要假设数据具有独立同分布的属性. 这一属性在很大程度上限制了随机梯度 Monte Carlo 抽样方法可以应用的范围. 例如,在对时间序列的分析中,假设数据独立同分布也就是意味着数据之间没有时间顺序. 因此,如何将随机梯度 Monte Carlo 抽样方法推广到具有相关性的数据中成为了目前研究的重点 (参见文献 [34]). 我们从隐 Markov 模型 (HMM) 入手,考察如何推广随机梯度 Monte Carlo 抽样方法对大规模时间序列数据进行机器学习.

隐 Markov 模型及其变体在连续的和离散的时间序列数据分析中有着广泛的应用. 它假设观察到的数据  $X_t$  概率依赖于隐含变量  $Y_t$ , 而离散隐含变量  $Y_t$  以转移概率  $p_{yy'}$  遵循 Markov 过程. 于是,  $(X_t,Y_t)$  是个二元 Markov 链, 其转移概率为

$$\Pr\{X_{t+1} = x', Y_{t+1} = y' \mid X_t = x, Y_t = y\} = g(x' \mid y') p_{uu'}, \tag{3.6}$$

其中  $p_{yy'}$  是 Markov 过程  $Y_t$  的转移概率矩阵, 而  $g(x \mid y)$  是当隐 Markov 状态为 y 时观测量为 x 的概率. 隐 Markov 模型是最简单的动态 Bayes 网络, 它在包括语音识别、计算生物、机器翻译、密码分析

和量化金融等诸多领域中都是强有力的分析工具. 通过推断模型参数和隐含变量, 我们可以进一步对数据作出分类和预测. 当数据量非常庞大时, 对隐 Markov 模型的统计推断不仅耗费大量的计算资源和时间, 还会产生计算误差, 甚至造成数值稳定性的问题. 因此, 我们考虑对依赖于全部数据集的统计量, 仅用一部分数据来计算它的统计估计. 如何选取这些互相之间不独立的数据取决于隐 Markov 模型对隐变量的边缘概率的收敛速率.

由 (3.6) 可得, 当观测得到序列  $x_1, x_2, \ldots, x_N$  时, 隐 Markov 状态为  $y_0, y_1, \ldots, y_N$  的条件概率是

$$\Pr\{Y_0 = y_0, \dots, Y_N = y_N \mid X_1 = x_1, \dots, X_N = x_N\}$$
$$= Z^{-1}(x_1, \dots, x_N) \prod_{k=1}^N g(x_k \mid y_k) \times \pi_{y_0} \prod_{k=1}^N p_{y_{k-1}y_k},$$

其中 元 为初分布,

$$Z(x_1, \dots, x_N) = \sum_{y_0, \dots, y_N} \pi_{y_0} \prod_{k=1}^N [g(x_k \mid y_k) p_{y_{k-1} y_k}].$$
 (3.7)

而现时的隐 Markov 状态为  $y_N$  的条件概率是

$$\rho_N = \Pr\{Y_N = y_N \mid X_1 = x_1, \dots, X_N = x_N\}$$

$$= Z^{-1}(x_1, \dots, x_N) \sum_{y_0, \dots, y_{N-1}} \prod_{k=1}^N g(x_k \mid y_k) \times \pi_{y_0} \prod_{k=1}^N p_{y_{k-1}y_k},$$
(3.8)

我们注意到计算 (3.7) 和 (3.8) 的关键是一串矩阵相乘. 如果没有  $g(x_k \mid y_k)$  项, 所有的矩阵都是相同的  $\mathbf{P} \equiv \{p_{yy'}\}$ . 但是, 如果考虑到  $g(x_k \mid y_k)$  项, 那么矩阵  $\mathbf{D}(x) \equiv \{g(x \mid y')p_{yy'}\}$  是观测量 x 的函数.  $\mathbf{D}(x)$  是非负对角矩阵  $\mathbf{G}(x) \equiv \{g(x \mid y')\delta_{yy'}\}$  和  $\mathbf{P}$  的乘积. (3.7) 和 (3.8) 含有一串不同的  $\mathbf{D}(x)$  矩阵的乘积, 称为非齐矩阵积 (non-homogeneous matrix product). 考虑非归一化的边缘概率分布  $\mathbf{p}_N$ ,

$$\mathbf{p}_N = \pi \mathbf{D}(x_1) \mathbf{D}(x_2) \cdots \mathbf{D}(x_N), \quad \boldsymbol{\rho}_N = \frac{\mathbf{p}_N}{\langle \mathbf{p}_N, \mathbf{1} \rangle}.$$
 (3.9)

当  $N \to \infty$  时, 如下有两个基本的数学问题是被广为关注的:

(1) 标量的极限,

$$\lim_{N \to \infty} \frac{1}{N} \ln \left\{ \pi \left( \prod_{k=1}^{N} \mathbf{D}(x_k) \right) \mathbf{1} \right\}, \tag{3.10}$$

其中  $\mathbf{1} = (1,1,\ldots,1)^{\mathrm{T}}$ . 此数学问题与一维自旋玻璃的配分函数<sup>[35]</sup> 或随机异聚高分子 (random heteropolymer) 的构象<sup>[36]</sup> 的计算异曲同工: 用最大 Lyapunov 指数来描述随机矩阵乘积的渐近行为. 自旋玻璃的平均场理论特别注重于计算当  $\{x_k \mid 1 \leq k \leq N\}$  为正态分布时  $\ln Z(x_1,\ldots,x_N)$  的数学期望<sup>[37]</sup>.

(2) 条件概率分布  $\rho_N$  当  $N \to \infty$  时的极限的收敛速率. 这是乘法遍历定理 (multiplicative ergodic theorem) 范畴内的课题 (参见文献 [16]). 始于两个不同的分布  $\rho_0$  和  $\rho_0'$ ,  $\|\rho_N - \rho_N'\| \to 0$  是指数收敛的. Atar 和 Zeitouni [38] 证明了收敛的速率的上界为序列  $\{p_k\}$  最大的两个 Lyapunov 指数  $\lambda_1$  和  $\lambda_2$  的差:

$$\limsup_{N \to \infty} \frac{1}{N} \| \boldsymbol{\rho}_N - \boldsymbol{\rho}_N' \| \leqslant \lambda_2 - \lambda_1. \tag{3.11}$$

这个上界是几乎处处可达到的. 有趣的是,  $\lambda_2 - \lambda_1$  可以用 RDS 的思想与方法来估算 (参见文献 [39]). 记  $\mathscr S$  上概率分布  $\rho$  的分量为  $(\rho_1, \rho_2, \dots, \rho_n)$ . 引进可逆变换, 它是保 Lyapunov 指数的,

$$\rho \leftrightarrow r \equiv \left( \ln \frac{\rho_1}{\rho_n}, \dots, \ln \frac{\rho_{n-1}}{\rho_n}, 0 \right).$$
(3.12)

 $\rho_k$  作为 k 的函数定义了一个  $\mathbb{R}^{n-1}$  上的随机映射  $r_k = d(y_k) + F(r_{k-1})$ , 其中  $d(y_k)$  是一个与  $r_{k-1}$  无关的随机移动, 取决于随机事件  $y_k$ , F 是一个确定性映射:

$$\boldsymbol{d}(y_k) = \left(\ln \frac{g(1\mid y_k)}{g(n\mid y_k)}, \dots, \ln \frac{g(n-1\mid y_k)}{g(n\mid y_k)}\right),\tag{3.13}$$

$$F(\mathbf{r}) = \left(\ln \frac{\sum_{j=1}^{n} e^{r_j} p_{j1}}{\sum_{j=1}^{n} e^{r_j} p_{jn}}, \dots, \ln \frac{\sum_{j=1}^{n} e^{r_j} p_{j,n-1}}{\sum_{j=1}^{n} e^{r_j} p_{jn}}\right).$$
(3.14)

这个 RDS 的最大 Lyapunov 指数  $\lambda_{r,\max}$  恰为  $\lambda_2 - \lambda_1$ . 而且  $\lambda_{r,\max}$  有表达式

$$\lambda_{r,\max} = \limsup_{N \to \infty} \frac{1}{N} \ln \| \boldsymbol{J}(\boldsymbol{r}^{N-1}) \cdots \boldsymbol{J}(\boldsymbol{r}^{0}) \|, \tag{3.15}$$

其中 J(r) 是确定性映射 F(r) 的 Jacobi 矩阵. 它与随机项  $y_k$  无关. 因此, (3.15) 的计算完全没有随机性. 它为有效地估算隐 Markov 模型统计推断的收敛速率提供了新的方法 (参见文献 [39]).

# 4 结论

动力系统不仅仅是数学与物理学里的研究领域. 非线性动力系统的思想与方法更是近代科学,包括化学、生物、甚至是经济、社会学理论思维的基础. 这些复杂系统的行为有别于古典力学系统;一个重要的特征是,无法忽略的内源的和外在的随机性. 近年来的研究表明,非线性动力系统的分枝现象与随机性相结合导致了凝聚态物理学里"相变"和"对称破缺"的概念(参见文献[40]). 这些思想正跟随着应用数学在进入细胞生物学(参见文献[6]),给出非基因表型差异更为严格的和定量的诠释. 从应用数学方法来讲,随机过程和随机动力系统都是随机动力学(stochastic dynamics)的一部分. 较之自 Newton 已被研究了三百多年的确定性动力学,随机动力学前景广阔(参见文献[41]).

致谢 作者感谢易英飞、沈文仙、刘培东和林揆宇 (亚利桑那大学) 几位教授多年来不断的指点和讨论.

#### 参考文献

- 1 Ge H, Qian H. Thermodynamic limit of a nonequilibrium steady-state: Maxwell-type construction for a bistable biochemical system. Phys Rev Lett, 2009, 103: 148103
- 2 Ge H, Qian H, Xie X S. Stochastic phenotype transition of a single cell in an intermediate region of gene state switching. Phys Rev Lett, 2015, 114: 078101
- 3 Kurtz T G. The relationship between stochastic and deterministic models for chemical reactions. J Chem Phys, 1972, 57: 2976–2978
- 4 Ethier S N, Kurtz T G. Markov Processes: Characterization and Convergence. New York: John Wiley & Sons, 2005
- 5 Qian H. Cellular biology in terms of stochastic nonlinear biochemical dynamics: Emergent properties, isogenetic variations and chemical system inheritability. J Stat Phys, 2010, 141: 990–1013
- 6 Huang S, Li F, Zhou J X, et al. Processes on the emergent landscapes of biochemical reaction networks and heterogeneous cell population dynamics-Differentiations in living matters. J R Soc Interfaces, 2017, 14: 20170097
- 7 Kifer Y. Ergodic Theory of Random Transformations. Basel: Birkhäuser, 1986

- 8 Diaconis P, Freedman D. Iterated random functions. SIAM Rev, 1999, 41: 45-76
- 9 Kifer Y, Liu P D. Random dynamics. In: Handbook of Dynamical Systems, vol. 1B. Amsterdam: Elsevier, 2006, 379–499
- 10 Ye F X F, Wang Y, Qian H. Stochastic dynamics: Markov chains and random transformations. Discrete Contin Dyn Syst Ser B, 2016, 21: 2337–2361
- 11 Lasota A, Mackey M C. Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics, 2nd ed. New York: Springer-Verlag, 1998
- 12 Jiang D Q, Qian M, Qian M P. Mathematical Theory of Noneqilibrium Steady States. Lecture Notes in Mathematics, vol. 1833. New York: Springer, 2004
- 13 钱紘. 信息与熵力: 生物细胞的物理描述—化学反应动力学和信息论. 中国科学: 生命科学, 2017, 47: 257-261
- 14 Kloeden P E, Pötzsche C. Nonautonomous Dynamical Systems in the Life Sciences. Lecture Notes in Mathematics, vol. 2102. New York: Springer, 2013
- 15 Lin K K, Shea-Brown E, Young L S. Spike-time reliability of layered neural oscillator networks. J Comput Neurosci, 2009, 27: 135–160
- 16 Arnold L, Crauel H. Lyapunov Exponents. Berlin: Springer, 1991
- 17 Hopfield J J. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA, 1982, 79: 2554–2558
- 18 Li F, Long T, Lu Y, et al. The yeast cell-cycle network is robustly designed. Proc Natl Acad Sci USA, 2004, 101: 4781–4786
- 19 Wang R S, Saadatpour A, Albert R. Boolean modeling in systems biology: An overview of methodology and applications. Phys Biol, 2012, 9: 055001
- 20 Kauffman S A. The large scale structure and dynamics of genetic control circuits: An ensemble approach. J Theoret Biol, 1974, 44: 167–190
- 21 Ackley D H, Hinton G E, Sejnowski T J. A learning algorithm for Boltzmann machines. Cognitive Sci, 1985, 9: 147–169
- 22 Qian M P, Gong G L, Clark J W. Relative entropy and learning rules. Phys Rev A (3), 1991, 43: 1061-1070
- 23 Zhang Y, Qian M P, Ouyang Q, et al. Stochastic model of yeast cell-cycle network. Phys D, 2006, 219: 35–39
- 24 Ge H, Qian H, Qian M. Synchronized dynamics and non-equilibrium steady states in a stochastic yeast cell-cycle network. Math Biosci, 2008, 211: 132–152
- 25 Shmulevich I, Dougherty E R. Probabilistic Boolean Networks: The Modeling and Control of Gene Regulatory Networks. Philadelphia: SIAM Publishing, 2010
- 26 Liu J S. Monte Carlo Strategies in Scientific Computing. New York: Springer, 2004
- 27 Ma Y A, Fox E B, Chen T, et al. Irreversible samplers from jump and continuous Markov processes. Stat Comput, submitted, 2017
- 28 Vucelja M. Lifting—A non-reversible Markov chain Monte Carlo algorithm. Amer J Phys, 2016, 84: 958–968
- 29 Qian H, Kjelstrup S, Kolomeisky A B, et al. Entropy production in mesoscopic stochastic thermodynamics: Nonequilibrium steady state cycles driven by chemical potentials, temperatures, and mechanical forces, J Phys Cond Matt, 2016, 28: 153004
- 30 Welling M, Teh Y W. Bayesian learning via stochastic gradient Langevin dynamics. In: Proceedings of the 28th International Conference on Machine Learning (ICML'11). Bellevue, 2011, 681–688
- 31 Ma Y A, Chen T, Fox E B. A complete recipe for stochastic gradient MCMC. Adv Neural Inf Process Syst, 2015, 28: 2899–2907
- 32 Ding N, Fang Y, Babbush R, et al. Bayesian sampling using stochastic gradient thermostats. Adv Neural Inf Process Syst, 2014, 27: 3203–3211
- 33 Shang X, Zhu Z, Leimkuhler B, et al. Covariance-controlled adaptive Langevin thermostat for large-scale Bayesian sampling. Adv Neural Inf Process Syst, 2015, 28: 37–45
- 34 Ma Y A, Foti N J, Fox E B. Stochastic Gradient MCMC Methods for Hidden Markov Models. In: Proceedings of the 34th International Conference on Machine Learning. Sydney, 2017, 2265–2274
- 35  $\,$  Edwards S F, Anderson P W. Theory of spin glasses. J Phys F Metal Phys, 1975, 5: 965–974
- 36  $\,$  Garel T, Orland H. Mean-field model for protein folding. Europhys Lett EPL, 1988, 6: 307–310
- 37 Talagrand M. Mean Field Models for Spin Glasses. New York: Springer, 2011
- 38 Atar R, Zeitouni O. Lyapunov exponents for finite state nonlinear filtering. SIAM J Control Optim, 1997, 35: 36–55
- 39 Ye F X F, Ma Y A, Qian H. Estimate exponential forgetting in HMM and its applications. ArXiv:1710.06078, 2017
- 40 Qian H, Ao P, Tu Y, et al. A framework towards understanding mesoscopic phenomena: Emergent unpredictability, symmetry breaking and dynamics across scales. Chem Phys Lett, 2016, 665: 153–161

41 葛颢, 钱紘. 数学动力学模型在生物物理和生物化学中的应用. 北京: 北京大学出版社, 2017

# Stochastic dynamics: Models for intrinsic and extrinsic noises and their applications

MA YiAn, QIAN Hong & YE Felix X.-F.

Abstract Mathematical modeling for complex systems and processes requires concepts from and techniques for stochastic dynamics. The theory of stochastic dynamics has two different mathematical representations: Stochastic processes and random dynamical systems. The latter is a more refined mathematical description of reality; it provides not only a stochastic trajectory following one initial condition, but also describes how the entire phase space, with all initial conditions, changes with time. The former represents the stochastic motion of individual systems with intrinsic noise while the latter describes many systems experiencing a common deterministic law of motion which is changing with time due to environmental fluctuations. We call these two situations with intrinsic and extrinsic noises; both have wide applications in chemistry and biology. The recently developed, graph  $G(\mathcal{V}, \mathcal{E})$  based probabilistic Boolean networks is precisely a class of random dynamical systems (RDS) with discrete state space  $\{0,1\}^{\mathcal{V}}$ . This paper introduces discrete-time RDS with discrete state space as well as discusses its applications in estimating a rate of convergence in hidden Markov model inference.

Keywords Markov chain, random transformation, random matrices, probabilistic Boolean network, stochastic gradient descent, mathematical biology

MSC(2010) 60J10, 37H15, 34D08, 92C40, 65C05

doi: 10.1360/N012017-00154