

文章编号: 1000-5862(2015)01-0001-06

概率主方程的研究综述

周天寿

(中山大学数学与计算科学学院 广东 广州 510275)

摘要: 反应网络广泛存在于自然系统中. 原理上, 概率主方程对任何反应网络系统的随机行为提供了最为完整的数学模型. 然而, 分析和模拟这种类型的方程长期以来是一项挑战性任务. 至目前为止, 有关问题并没有得到根本解决. 相关研究仍在继续. 该文对概率主方程的研究进展进行了较为系统和全面的综述. 聚焦于若干常用的近似分析方法(如线性噪声逼近、普通矩封闭法、2项矩方法等)与常用的数值方法(如 Gillespie 随机模拟算法、有限状态映射法、矩封闭格式等). 特别地, 分析了概率主方程研究取得缓慢进展的主要原因, 讨论并提出了可能的解决方案.

关键词: 反应网络; 概率主方程; 矩封闭方法; 随机模拟

中图分类号: O 242; Q 332 **文献标志码:** A **DOI:** 10. 16357/j. cnki. issn1000-5862. 2015. 01. 01

0 引言

反应网络广泛存在于自然系统中, 如化学领域的 BZ(Zaikin-Zhabotinsky) 反应^[1]、生物学领域的基因调控网^[2-3]、生态学领域的微观种群模型^[4]等. 由于反应物种分子低的拷贝数, 因此这些系统的行为本质上是随机的. 确定性方程(即 ODE 方程)的描述常常无效, 需要发展随机格式(即需要检查系统中反应物种分子的联合概率分布). 假如所有涉及的反应事件是马氏的, 则概率主方程对任何反应网络系统的随机行为在原理上提供了最为完整的数学模型(实际仅提供了一种建模思路). 简单来说, 概率主方程描述出反应系统中所有反应物种分子的联合概率分布的时间演化^[5].

为帮助读者理解, 这里简要地描述这种类型的方程. 考虑一个耦合的反应网络, 假定它包含 M 个反应式, 涉及 N 个反应物种 $\{X_1, X_2, \dots, X_N\}$. 记网络系统的状态向量为 $\mathbf{n} = [n_1, n_2, \dots, n_N]^T$ (“T”表示转置), 这里 n_i 表示物种 X_i 的分子数目. 注意到: 构成该网络的 M 个反应完全由倾向函数向量 $\mathbf{a} = [a_1(\mathbf{n}), a_2(\mathbf{n}), \dots, a_M(\mathbf{n})]$ 和化学计量矩阵 $\mathbf{s} = [s_{ij}]$ 来决定, 这里 s_{ij} 表示物种 X_j 参加第 i 个反应的分子数目的变化. 例如, 对于简单的反应式: $X + Y \xrightarrow{k} Z$, 相应的反

应倾向函数为 $a(n_1, n_2, n_3) = kn_1n_2$, 化学计量矩阵为 $\mathbf{s} = [-1, -1, 1]^T$; 又例如, 对于简单的反应式: $X + X \xrightarrow{k} Y$, 相应的反应倾向函数为 $a(n_1, n_2) = kn_1(n_1 - 1)/2$, 化学计量矩阵为 $\mathbf{s} = [-2, 1]^T$.

记 $\mathbf{s}^i = [s_{1i}, \dots, s_{Mi}]^T$, 用 $P(\mathbf{n}; t)$ 表示物种 X_i 在时刻 t 有 n_i 个分子的概率(这里 $1 \leq i \leq N$), 并假定系统状态的变化是一个马氏过程(即当前的状态不依赖于其历史), 则概率主方程可表示为^[5]

$$\frac{\partial P(\mathbf{n}; t)}{\partial t} = \sum_{i=1}^M [a_i(\mathbf{n} - \mathbf{s}^i) P(\mathbf{n} - \mathbf{s}^i; t) - a_i(\mathbf{n}) P(\mathbf{n}; t)]. \quad (1)$$

这里, 求和括号中的第 1 项表示第 i 个反应式的流进概率, 第 2 项表示此反应的流出概率. 对于给定的反应网络, 为了写出相应的概率主方程, 关键是确定倾向函数向量及化学计量. 注意到: 格式 (1) 并没有考虑系统的体积或大小. 假如考虑系统的大小(记为 Ω), 则也容易导出相应的概率主方程. 例如,

考虑由简单反应式: $\sum_{i=1}^N r_i X_i \xrightleftharpoons[k_-]{k_+} \sum_{i=1}^N \tilde{r}_i X_i$ 组成的反应网络, 则相应的概率主方程为^[6]

$$\frac{\partial P(\mathbf{n}; t)}{\partial t} = k_+ \Omega \left(\prod_{i=1}^N E_i^{r_i - \tilde{r}_i} - I \right) \prod_{j=1}^N \left\{ \frac{\binom{n_j}{\tilde{r}_j}}{\Omega^{\tilde{r}_j}} \right\} P(\mathbf{n}; t) + k_- \Omega \left(\prod_{i=1}^N E_i^{\tilde{r}_i - r_i} - I \right) \prod_{j=1}^N \left\{ \frac{\binom{n_j}{r_j}}{\Omega^{r_j}} \right\} P(\mathbf{n}; t),$$

收稿日期: 2014-12-08

基金项目: 国家自然科学基金委/重大研究计划/重点支持(91230204)和科技部 973 项目子课题(2014CB964703)资助项目.

作者简介: 周天寿(1962-), 男, 江西新建人, 教授, 博士生导师, 目前主要从事系统生物学研究.

其中 $n_j(n_j - 1)(n_j - 2) \cdots (n_j - s_j + 1) = n_j! / (n_j - s_j)!$ 。类似地, 假如某个反应网络包含多个这种类型的反应式, 则也可给出其相应的概率主方程。事实上, 此时, 只要对各个反应式所对应的概率主方程进行求和即可。

为应用的方便, 这里给出概率主方程的另一种形式。让 $\mathbf{r} \rightarrow \tilde{\mathbf{r}}$ 代表反应网络中 1 个具有一般形式的

反应式: $\sum_{i=1}^N r_i X_i \rightarrow \sum_{i=1}^N \tilde{r}_i X_i$, 并记此反应的比率常数为 $c^{\tilde{\mathbf{r}}}$, 则概率方程(1)可改写为^[7]

$$\frac{\partial P(\mathbf{n}; t)}{\partial t} = \sum_{\mathbf{r} \rightarrow \tilde{\mathbf{r}}} c^{\tilde{\mathbf{r}}} [E^{-\tilde{\mathbf{r}}} E^{\mathbf{r}} - I] \Pi(\mathbf{n}; \mathbf{r}) P(\mathbf{n}; t), \quad (2)$$

其中 $\Pi(\mathbf{n}; \mathbf{v}) = \binom{\mathbf{n}}{\mathbf{v}} = \prod_{i=1}^K \binom{n_i}{v_i}$ 代表普通的二

项系数; E 是位移算子, E^{-1} 是 E 的逆算子, 它们的操作规则是: $E^{\alpha} f(\mathbf{n}) = f(\mathbf{n} + \alpha)$, 这里 f 是向量 \mathbf{n} 的函数, α 是 1 个由整数组成的向量; I 是单位算子。在方程(2)中, 必须对系统中所有可能的反应式 $\mathbf{r} \rightarrow \mathbf{s}$ (共有 M 个) 求和: 假如系统处在 $\mathbf{n} + \mathbf{r} - \tilde{\mathbf{r}}$ 状态, 则相应的反应式对 $P(\mathbf{n}; t)$ 有 1 个正的贡献; 假如系统处于 \mathbf{n} 状态, 则相应的反应式对 $P(\mathbf{n}; t)$ 有 1 个负的贡献。这蕴含着操作规则

$$E^{\mathbf{r} - \tilde{\mathbf{r}}} \Pi(\mathbf{n}; \mathbf{r}) = \Pi(\mathbf{n} + \mathbf{r} - \tilde{\mathbf{r}}; \mathbf{r}).$$

需要指出的是: 上述 2 种格式只是表示方式不同, 它们都描述同一个反应网络。

自从反应系统的第 1 个主方程^[8] 建立以来已经过去 70 余年, 但概率主方程的模拟与分析问题并没有得到很好地解决。主要原因是概率主方程关于反应物种数目是指数增长的。例如, 假如某个反应系统仅包含 1 个反应物种, 且其最大分子数目为 10^3 , 则相应的概率主方程相当于 10^3 个 ODE (ordinary differential equation) 方程; 假如某个反应系统包含 2 个反应物种, 且每个反应物种的最大分子数目均为 10^3 , 则相应的概率主方程相当于 $10^3 \times 10^3$ 个 ODE 方程。一般地, 一个真实的反应系统可能包含成千上万个反应物种, 可想而知其计算量是多么的大。这种指数增长性质极大地限制了现有随机分析方法与随机模拟方法的应用范围, 因此人们一直试图解决概率主方程的模拟与分析问题。尽管已经取得了某些进展, 但相关问题并没有得到根本解决, 研究仍在继续。特别是, 对生化反应网络, 有关计算问题目前仍是计算系统生物学的研究热点。

本文将从理论分析与数值模拟 2 方面对概率主方程进行较为系统和全面的综述。在理论分析方面,

主要介绍线性噪声逼近、普通矩封闭方法、2 项矩方法等; 在数值模拟方面, 主要介绍 Gillespie 随机模拟算法、有限状态映射法、矩封闭格式等。通过这种介绍或综述, 希望达到使读者对概率主方程的研究有一个较为全面地了解, 并能够运用概率主方程来灵活地处理反应网络随机行为的目的。

1 理论分析方法

目前, 分析反应网络的随机行为主要包括线性噪声逼近^[9-12]、矩封闭方法^[13-15]等, 且这些方法已得到了广泛应用。本质上, 前一种方法是后一种方法的特殊情形。在介绍时, 将用简单的例子加以说明, 以便更好地理解方法的精神实质。

1.1 线性噪声逼近

线性噪声逼近可分为静态线性噪声与动态线性噪声逼近^[16], 前者是在系统的静态处做线性近似, 而后者是在轨线处做线性近似。在线性噪声逼近方法中, 需要假设随机变量是高斯的, 即仅考虑随机变量的最初 2 阶矩: 1 阶矩和 2 阶矩, 其它高阶矩并不需要考虑。下面, 简单地导出静态线性噪声逼近法的矩阵方程。

1) 注意到: 一般地, 概率主方程(1)中物种数目的变化 s^i 相对于 \mathbf{n} 而言是很小的量, 因此可进行泰勒展开。特别是, 若展开到 2 阶, 则得下列 Fokker-Planck 方程

$$\frac{\partial P(\mathbf{n}; t)}{\partial t} = \sum_{i=1}^M \left[- \sum_{k=1}^N s_{ki} \frac{\partial}{\partial n_k} + \sum_{k,l=1}^N \frac{s_{ki} s_{li}}{2} \frac{\partial^2}{\partial n_k \partial n_l} \right] a_j(\mathbf{n}) P(\mathbf{n}; t). \quad (3)$$

用 $\langle n_i \rangle = \int n_i P(\mathbf{n}; t) d\mathbf{n}$ 表示物种 X_i 的平均数目 (即期望), 并用 $x_i x_j = (n_i - \langle n_i \rangle)(n_j - \langle n_j \rangle)$ 乘以方程(3)的 2 边并积分, 则可得下列矩阵方程

$$\frac{d\mathbf{C}}{dt} = \mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{A}^T + \mathbf{\Xi}, \quad (4)$$

其中协方差矩阵 \mathbf{C} 的元素为 $C_{ij} = \langle x_i x_j \rangle = \langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle$; 矩阵 \mathbf{A} 的元素为 $A_{ij} = \sum_{k=1}^M s_{ik} \partial a_k(\mathbf{n}) / \partial n_j$;

矩阵 $\mathbf{\Xi}$ 的元素为 $\gamma_{ij} = \sum_{k=1}^M s_{ik} s_{jk} a_k(\mathbf{n}_s)$, 这里 \mathbf{n}_s 是线性方程组 $d \langle n_i \rangle / dt = \sum_{k=1}^M s_{ik} \langle a_k(\mathbf{n}) \rangle$ ($1 \leq i \leq N$) 的静态解, 它近似地代表平均 $\langle \mathbf{n} \rangle$ 。

2) 更为实用的是考虑方程(4)的静态解。此时, 若记 $B_{ij} = A_{ij} \langle n_j \rangle / \langle n_i \rangle$, $\mathcal{D}_{ij} = \gamma_{ij} / (\langle n_i \rangle \langle n_j \rangle)$,

$\eta_{ij} = C_{ij}(\langle n_i \rangle \langle n_j \rangle)$ 则得 $B\eta + \eta B^T + D = 0$ 其中矩阵 η 代表定量化系统波动的协方差矩阵, 是未知变量. 特别是, 这一矩阵的对角线元素代表反应物种的噪声水平.

3) 由于假定随机变量为高斯型的, 因此在有了静态 1 阶矩 $\langle n \rangle \approx n_s$ 和静态 2 阶矩 (即协方差矩阵 Ξ) 之后, 并不困难地重构出相应的静态概率分布. 事实上, 重构的概率分布为

$$P(n) = \frac{1}{\sqrt{(2\pi)^N \det(\Xi)}} \exp\left(-\frac{n^T \Xi^{-1} n}{2}\right),$$

其中 $n = (n_1, \dots, n_N)^T$ 是列向量, $\det(\Xi)$ 代表矩阵 Ξ 的行列式.

下面, 考察一个简单的反应系统例子: 具有非线性降解的生灭过程, 相应的反应式为 $\emptyset \xrightarrow{k_1} X$,

$2X \xrightarrow{k_2} \emptyset$. 让 n 表示物种 X 的数目, 则概率主方程为

$$\frac{\partial P(n; t)}{\partial t} = k_1 [P(n-1; t) - P(n; t)] + \frac{k_2}{2} [(n+1)(n+2)P(n+2; t) - n(n-1)P(n; t)].$$

最初的 2 个中心矩的微分方程为

$$\frac{d\langle n \rangle}{dt} = k_1 + k_2 \langle n \rangle - k_2 \langle n \rangle^2 - k_2 \sigma,$$

$$\frac{d\sigma}{dt} = k_1 + 2k_2 \langle n \rangle - 6k_2 \langle n \rangle^2 + 4k_2 \langle n \rangle^3 +$$

$$4k_2(2\langle n \rangle - 1)\sigma + 2k_2 \langle n - \langle n \rangle \rangle^3,$$

其中 $\bar{\sigma} = \text{var}(n)$ 代表 n 的方差. 若按照线性噪声逼近法^[10], 则系统的静态 (记为 \bar{x}_s) 由代数方程 $k_1 + k_2 \bar{x}_s - k_2 \bar{x}_s^2 = 0$ 给出, 且 $\bar{x}_s = (1 + \sqrt{1 + 4c})/2$ 这里 $c = k_1/k_2$; 而在静态处 n 的噪声强度 (定义为方差与平均平方之比) 为 $\eta_{LNA}^2 = 10c/(1 + 4c + (1 + 2c)\sqrt{1 + 4c})$. 若采用近似 $\langle n - \langle n \rangle \rangle^3 \approx 0$, 则静态平均 (记为 x_s) 和静态方差 (记为 σ) 满足下列代数方程组

$$c + x_s - x_s^2 - \sigma = 0,$$

$$c + 2x_s - 6x_s^2 + 4x_s^3 + 4(2x_s - 1)\sigma = 0.$$

类似地, 可给出噪声强度的分析表示.

1.2 矩封闭方法

矩封闭方法^[13-15]的基本思路是: 首先由概率主方程导出矩的微分方程, 其中低阶矩的方程通常会依赖于高阶矩, 然后采用某种近似来导出矩方程的封闭格式.

定义随机变量 n 的矩生成函数 $m(\vartheta; n; t)$ 如下

$$m(\vartheta; n; t) = \sum_n e^{n\vartheta} P(n; t),$$

则概率分布 $P(n; t)$ 的矩函数可由生成函数 $m(\vartheta; n; t)$ 关于 ϑ 在 $\vartheta = 0$ 处的导数给出, 即 $\langle n^k \rangle = \partial^k m(\vartheta; n; t) / \partial \vartheta^k \big|_{\vartheta=0}$, 并不困难地导出下列微分方程^[17]

$$\frac{\partial m(\vartheta; n; t)}{\partial t} = \sum_l \left[(e^{\vartheta s_l} - 1) \sum_n e^{\vartheta n} a_l(n) P(n; t) \right].$$

若记 μ_i 为物种 X_i 的平均数目, $\mu = (\mu_1, \dots, \mu_N)^T$, 则不难导出下列微分方程

$$\frac{d\mu}{dt} = s \left[\sum_l \sum_{k_1=0}^{\infty} \dots \sum_{k_N=0}^{\infty} \frac{1}{k!} \frac{\partial^k a_l(n)}{\partial n^k} \bigg|_{n=\mu} M_{n^k} \right], \quad (5)$$

其中 $M_{n^k} = \langle (n_1 - \mu_1)^{k_1} \dots (n_N - \mu_N)^{k_N} \rangle$ 是高阶中心矩. 注意到 M_{n^k} 能够表示成

$$M_{n^k} = \sum_{i_1=0}^{k_1} \dots \sum_{i_N=0}^{k_N} \binom{k}{i} (-1)^{k-i} \underbrace{\mu}_{\alpha}^{k-i} \underbrace{\langle n^i \rangle}_{\beta}.$$

这样, 进一步能够导出下列微分方程

$$\frac{dM_{n^k}}{dt} = \sum_{i_1=0}^{k_1} \dots \sum_{i_N=0}^{k_N} \binom{k}{i} (-1)^{k-i} \left[\alpha \frac{d\beta}{dt} + \beta \frac{d\alpha}{dt} \right], \quad (6)$$

其中 α 和 β 满足下列微分方程

$$\frac{d\alpha}{dt} = \sum_{l=1}^N (k_l - i_l) \mu_l^{-1} \alpha \frac{d\mu_l}{dt},$$

$$\frac{d\beta}{dt} = \sum_{i_1=0}^{k_1} \dots \sum_{i_N=0}^{k_N} s^i \binom{k}{i} \frac{\langle n^{k-i} a(n) \rangle}{\langle F \rangle},$$

其中 $\langle F \rangle = \sum_{k_1=0}^{\infty} \dots \sum_{k_N=0}^{\infty} \frac{1}{k!} \frac{\partial^k F(n)}{\partial n^k} \bigg|_{n=\mu} M_{n^k}$, $k = (k_1, \dots, k_N)$.

在上面的矩方程 (5) 和 (6) 中, 低阶矩的方程通常依赖于高阶矩. 因此, 为了获得封闭的系统, 需要截取. 为此, 若把这些矩方程改写成下列形式

$$\frac{d\mu}{dt} = \mu_0 + A\mu + A'\mu', \quad (7)$$

其中 μ_0 为某一常数向量 (由概率的保守性条件或某些反应式中物种总数目是守恒的条件决定); μ 代表直到 m 阶的低阶矩 (既可以是原点矩也可以是中心矩), 其系数矩阵为 A ; μ' 代表高阶矩, 其系数矩阵为 A' , 则可以证明: 只要系数矩阵 A 的特征值的实部小于某个负常数, 则方程 (7) 右边的第 3 项可以舍去. 此时, 可获得一个封闭系统.

下面, 再介绍另一种截取矩的策略^[13]. 假定方程 (7) 中 μ 代表原点矩, 并限制讨论为 1 维情形. 假设单变量 x 的可能状态为 $\{x_1, x_2, \dots\}$, 状态 x_i 的概率为 $P(x_i)$. 引入熵函数 $H = - \sum_i P(x_i) \ln P(x_i)$. 假定前有限个矩是已知的 (例如由相应于方程 (7) 的静态解决定), 考虑下列最优化问题

$$\Lambda = H - \sum_{j=0}^L \left[\lambda_j \left(\langle x_j \rangle - \sum_i x_i^j P(x_i) \right) \right],$$

其中 λ_j 为拉格朗日乘子. 求解此最优化问题, 将导致 1 组代数方程; 再用牛顿法求解此代数方程, 由此确定这些拉格朗日乘子, 并不困难显示出: 满足最大熵的概率分布具有形式

$$P_H(\mathbf{x}) = \exp(-\lambda_0 - \lambda_1 x - \cdots - \lambda_L x^L),$$

这样, 基于最大熵的概率分布, 能够计算第 $(L+1)$ 阶矩 $\langle x^{L+1} \rangle_H = \sum_i x_i^{L+1} P_H(x_i)$. 类似地, 可讨论其它更高阶矩的计算.

上面的普通矩有一个明显的缺陷, 即当矩的阶趋于无穷时, 此矩并不能保证趋于零. 为了克服这种缺陷, 引入 2 项矩

$$b_k(t) = \sum_{N \geq k} z_0^{N-k} \Pi(N, k) P(N; t),$$

其中 $z_0 = (z_{10}, \cdots, z_{N0})$ 为 1 个参数向量, 它的选取需保证 $\lim_{|k| \rightarrow \infty} b_k(t) = 0$. 对于大多数反应网络, 能够选取 $z_0 = \mathbf{I}$. 注意到: 2 项矩 $b_k(t)$ 实际上为概率分布 $P(N; t)$ 对应的母函数 $G(z; t)$ 在 $z = z_0$ 处泰勒展开式中的系数. 确切地说 $b_k(t) = (1/k!) G_z^{(k)}(z_0; t)$. 根据概率主方程(2), 并不困难地导出下列 2 项矩方程^[7]

$$\frac{db_k(t)}{dt} = \sum_{r \rightarrow s} c_r \left[\sum_{i=0}^k \binom{s}{i} z_0^{s-i} - \binom{r}{i} z_0^{r-i} \right] \binom{r+k-i}{r} b_{r+k-i}(t),$$

求解这一方程获得 2 项矩; 反过来, 可用获得的 2 项矩来计算概率分布. 事实上, 有

$$P(N; t) = \sum_{k \geq N} (-1)^{k-N} \binom{k}{N} b_k(t). \quad (8)$$

因为当 $|k| \rightarrow \infty$ 时有 $b_k \rightarrow 0$, 因此存在一个大的正整数 \bar{M} , 使得当 $|k| > \bar{M}$ 时, 有 $b_k \approx 0$. 注意到:

阶为 $|k|$ 的不同 2 项矩的数目为 $\binom{|k|+N-1}{|k|}$, 这样

截取到 \bar{M} 阶的 2 项矩的数目为 $N_B = \sum_{|k|=1}^{\bar{M}} \binom{|k|+N-1}{|k|}$, 这表明矩方程的个数关于物种

数目 N 是以多项式的方式增长. 然而, 对于概率主方程, 由于 $\lim_{|n| \rightarrow \infty} P(n; t) = 0$, 因此, 假如当 $n_k > C_k$ 时, 有 $P(N; t) \approx 0$, 则概率主方程截取后的方程数目为

$N_M = \prod_{k=1}^N (1 + C_k)$, 它关于物种数目 N 是指数增长的, 这种性质严重地限制了主方程的应用范围. 从这种简单的分析, 可看出 2 项矩格式比概率主方程具有更大的优势.

为了帮助读者理解上面的 2 项矩及其方程, 仍然考虑前面研究过的生灭过程例子. 相应的 2 项矩方程(可选取 $z_0 = 1$) 为 $db_k/dt = k_1 b_{k-1} - k_2 k(k-1) b_k/2 - k_2 k(k+1) b_{k+1}$, 这里 $b_0 = 1, k = 1, 2, \cdots$. 假如截取到 2 阶矩, 即让 $b_3 \approx 0$, 则求得静态 1 阶矩和静态 2 阶矩为 $b_1 = 1/2, b_2 = c/2$. 因此, 噪声强度为 $\eta_{BM}^2 = (2b_2 + b_1 - b_1^2)/b_1^2 = 4c + 1$, 它比由线性噪声逼近法获得的噪声强度, 即 $\eta_{LNA}^2 = 10c/(1 + 4c + (1 + 2c)\sqrt{1 + 4c})$ 要大, 蕴含着线性噪声逼近低估了噪声. 假如截取到 3 阶, 即 $b_4 \approx 0$, 则求得前 3 个静态矩: $b_1 = (1/2) + c, b_2 = c/2, b_3 = c^2/6$. 因此, 噪声强度为 $B_M^2 = [2 - (2c - 1)^2]/(2c + 1)^2$, 这里 c 需要满足条件: $0 < c < (\sqrt{2} + 1)/2$.

2 数值计算方法

关于概率主方程的数值求解, 最广泛使用的方法是 Gillespie 随机模拟算法^[5] 及其后的改进版本^[18-19]. 目前, 有限状态映射法^[20] 也被广泛应用, 但仅能处理小规模的反应系统. 此外, 上面介绍的矩封闭方法也能够用于数值计算. 这里, 将简要介绍前种数值方法. 在介绍时, 力求阐明数值方法的精神实质.

2.1 Gillespie 随机模拟算法

注意到: $P(n_1, n_2, \cdots, n_N; t)$ 表示在时刻 t 、体积 Ω 内, 反应物种 X_k 有 n_k 个分子的概率 ($k = 1, 2, \cdots, N$). 概率主方程仅是函数 $P(n_1, n_2, \cdots, n_N; t)$ 的时间演化形式, 它可以通过应用概率理论的加法和乘法法则来写 $P(n_1, n_2, \cdots, n_N; t + dt)$ 为系统在时刻 $t + dt$ 达到状态 (n_1, n_2, \cdots, n_N) 的 $1 + M$ 个不同方式的概率之和. 这可表示为下列形式

$$P(n_1, n_2, \cdots, n_N; t + dt) = P(n_1, n_2, \cdots, n_N; t) \cdot \left(1 - \sum_{\mu=1}^M a_{\mu} dt\right) + \sum_{\mu=1}^M B_{\mu} dt, \quad (9)$$

其中

$a_{\mu} dt = c_{\mu} dt \times \{\text{在状态}(n_1, n_2, \cdots, n_N) \text{ 处不同 } R_{\mu} \text{ 分子组合的数目}\} = \text{给定系统在时刻 } t \text{ 处在状态}(n_1, n_2, \cdots, n_N) \text{ 的 } R_{\mu} \text{ 反应在时间间隔}(t, t + dt) \text{、体积 } \Omega \text{ 内发生的概率}.$ (10)

方程(9) 中的第 1 项代表系统首先在 t 时刻处在 (n_1, n_2, \cdots, n_N) 状态, 然后在 $(t, t + dt)$ 内仍然保持处在这一状态(即不经历新的反应)的概率; 而量 $B_{\mu} dt$ 给出这样一个概率: 系统首先在时刻 t 从状态

(n_1, n_2, \dots, n_N) 离开某个 R_μ 反应式, 然后在 $t + dt$ 内经历另外某些反应. 从 (9) 式可获得概率主方程

$$\frac{\partial}{\partial t} P(n_1, n_2, \dots, n_N; t) = \sum_{\mu=1}^M [B_\mu - a_\mu P(n_1, n_2, \dots, n_N; t)], \quad (11)$$

它与前面的方程 (1) 是等价的.

尽管这种描述是精确的, 但一般很难求解. 下面来讨论方程 (11) 的数值计算问题. 假设在时刻 t , 系统处在状态 (n_1, n_2, \dots, n_N) . 需要解决 2 个关键问题: 1) 何时发生下一个反应; 2) 哪个反应将会发生. 由于反应的随机性, 因此这 2 个问题只能指望在概率意义下才有答案. 为此, 引入 2 个变量的函数

$P(\tau, \mu) d\tau$ = 在时刻 t 给定状态 (n_1, n_2, \dots, n_N) 在时间间隔 $(t + \tau, t + \tau + d\tau)$ 、体积 Ω 内, 下一个反应将是 R_μ 的概率.

$P(\tau, \mu)$ 叫做该反应的概率密度函数, 它实际是连续变量 $\tau (0 \leq \tau < \infty)$ 和离散变量 $\mu (\mu = 1, 2, \dots, M)$ 的联合概率. 现在, 对每个反应式 R_μ , 定义 1 个数量函数 h_μ , h_μ = 在状态 (n_1, n_2, \dots, n_N) 处 R_μ 中不同反应分子的组合数目, 其中 $\mu = 1, 2, \dots, M$. 例如, 假如反应 R_μ 具有形式: $X_1 + X_2 \rightarrow \text{任何}$, 则 $h_\mu = n_1 n_2$; 假如 R_μ 具有形式: $2X_1 \rightarrow \text{任何}$, 则 $h_\mu = n_1(n_1 - 1)/2!$; 一般地, h_μ 是变量 n_1, n_2, \dots, n_N 的某些组合函数. 这样, (10) 式蕴含着: $a_\mu dt = c_\mu h_\mu dt$. 注意到: $P(\tau, \mu) d\tau$ 可以看成以下 2 项的乘积

$$P(\tau, \mu) d\tau = P_0(\tau) \cdot a_\mu d\tau, \quad (12)$$

其中 $P_0(\tau)$ 表示在时刻 t 处给定状态 (X_1, \dots, X_N) , 在时间区间 $(t, t + \tau)$ 内没有反应发生的概率; $a_\mu d\tau$ 表示某个反应 R_μ 将在 $(t + \tau, t + \tau + d\tau)$ 内发生的概率. 为了给出 $P_0(\tau)$ 的分析表示, 假如注意到 $[1 - \sum_{\nu=1}^M a_\nu d\tau']$ 表示在时间 $d\tau'$ 内没有反应发生的概率, 则有

$$P_0(\tau' + d\tau') = P_0(\tau') \cdot (1 - \sum_{\nu=1}^M a_\nu d\tau'). \quad (13)$$

从 (13) 式, 容易获得

$$P_0(\tau) = \exp(-\sum_{\nu=1}^M a_\nu \tau). \quad (14)$$

进一步, 从 (12) ~ (14) 式可知

$$P(\tau, \mu) = \begin{cases} a_\mu \exp(-a_0 \tau) & 0 \leq \tau < \infty, \mu = 1, 2, \dots, M, \\ 0, & \text{其它.} \end{cases}$$

这里 $a_\mu = h_\mu c_\mu$, $\mu = 1, 2, \dots, M$, $a_0 = \sum_{\nu=1}^M a_\nu =$

$\sum_{\nu=1}^M h_\nu c_\nu$. 根据上面的分析, 不难给出相应的数值算法. 关于细节, 参见文献 [16].

2.2 有限状态映射法

2006 年, B. Munsky 等^[20] 提出求解概率主方程的一种有效方法, 即有限状态映射法. 为了介绍此方法, 首先把概率方程 (1) 改写为下列形式

$$\frac{\partial}{\partial t} P(\mathbf{n}; t) = \left[- \sum_{l=1}^M a_l(\mathbf{n}) \rho_1(\mathbf{n} - \mathbf{s}^l) \rho_2(\mathbf{n} - \mathbf{s}^2) \dots \rho_M(\mathbf{n} - \mathbf{s}^N) \right] \begin{bmatrix} P(\mathbf{n}) \\ P(\mathbf{n} - \mathbf{s}^1) \\ \vdots \\ P(\mathbf{n} - \mathbf{s}^N) \end{bmatrix}, \quad (15)$$

然后, 让 $\mathbf{X} = [n_1, n_2, \dots]^T$ 表示系统的所有可能状态, $\mathbf{P}(\mathbf{X}) = [P(n_1), P(n_2), \dots]^T$ 表示在时刻 t 的完整概率密度状态向量, 则方程 (15) 又可改写为

$$\frac{\partial}{\partial t} \mathbf{P}(\mathbf{X}; t) = \mathbf{A} \cdot \mathbf{P}(\mathbf{X}; t). \quad (16)$$

它是一个常系数线性普通微分方程组, 其中矩阵 \mathbf{A} 称为状态反应矩阵, 且 \mathbf{A} 的元素为

$$A_{ij} = \begin{cases} - \sum_{l=1}^M a_l(\mathbf{n}_i), & \text{假如 } i = j, \\ a_l(\mathbf{n}_i), & \text{对所有的 } j: \mathbf{n}_j = \mathbf{n}_i + \mathbf{s}^l, \\ 0, & \text{其它.} \end{cases}$$

矩阵 \mathbf{A} 具有如下性质: (i) 独立于时间 t ; (ii) 所有的对角线元素非正; (iii) 所有的非对角线元素非负; (iv) 每一列元素之和为 0. 方程 (16) 开始于 $t = 0$, 结束于 $t = t_f$ 的解可表示为

$$\mathbf{P}(\mathbf{X}; t_f) = \Phi(0, t_f) \cdot \mathbf{P}(\mathbf{X}; 0),$$

其中, 算子 $\Phi(0, t_f)$. 特别是, 在只有有限个可达状态情形, 算子 $\Phi(0, t_f)$ 是 $\mathbf{A}t_f$ 的指数矩阵, 此时有关系:

$$\mathbf{P}(\mathbf{X}; t_f) = \exp(\mathbf{A}t_f) \cdot \mathbf{P}(\mathbf{X}; 0).$$

对于方程 (16), 有下列结论: 让 \mathbf{A}_K 是矩阵 \mathbf{A} 的 $K \times K$ 主子矩阵, $\mathbf{P}(\mathbf{X}_K; 0)$ 是 $\mathbf{P}(\mathbf{X}; 0)$ 相应元素的向量. 若对于 $\varepsilon > 0$, $t_f \geq 0$, $\mathbf{I}^T \exp(\mathbf{A}_K t_f) \mathbf{P}(\mathbf{X}_K; 0) \geq 1 - \varepsilon$, 则有误差估计:

$$\exp(\mathbf{A}_K t_f) \mathbf{P}(\mathbf{X}_K; 0) \leq \mathbf{P}(\mathbf{X}_K; t_f) \leq \exp(\mathbf{A}_K t_f) \mathbf{P}(\mathbf{X}_K; 0) + \varepsilon \mathbf{I},$$

其中 $\mathbf{I} = [1, 1, \dots, 1]^T$, $\mathbf{K} = [k_1, k_2, \dots, k_K]^T$.

由上面的分析, 可获得下列算法.

第 1 步 对系统中的所有的反应式, 定义倾向函数和化学计量, 选取初始的概率密度向量 $\mathbf{P}(\mathbf{X}; 0)$, 适当选取最后的时刻 t_f , 细化可接受的误差 $\varepsilon > 0$, 选取初始一套状态 \mathbf{X}_{K_0} , 初始化计数器 $i = 0$;

第 2 步 利用反应倾向函数和化学计量来形成 \mathbf{A}_{K_i} , 计算 $\Gamma_{K_i} = \mathbf{I}^T \exp(\mathbf{A}_{K_i} t_f) \mathbf{P}(\mathbf{X}_{K_i}; 0)$;

第3步 假如 $\Gamma_{K_i} \geq 1 - \varepsilon$, 则停止, 用 $\exp(A_{K_i} t_f) P(X_{K_i}; 0)$ 近似 $P(X_{K_i}; t_f)$ 到 ε 的误差范围内;

第4步 增加更多状态, 由此确定 $X_{K_{i+1}}$, 用 $i + 1$ 代替 i , 并返回到第2步.

3 讨论与展望

基于概率主方程的反应网络的模拟与分析是计算化学、计算生物学等领域的一个持续话题. 到目前为止, 相关问题并没有得到彻底解决. 一般地, 对于小规模的反应系统, Gillespie 随机模拟算法或有限状态映射法基本解决了计算问题, 但分析求解问题并没有得到解决, 仍然依赖于具体的反应网络. 对于大规模的反应系统, 现有的理论分析方法、数值计算方法通常无能为力. 本文综述了反应网络几种常用的分析与模拟方法, 并指出和讨论了它们的利弊.

关于反应网络随机行为的模拟, 主要解决2个问题: 1) 可计算问题, 即发展数值方法, 能够数值地给出反应网络中所有物种分子的联合概率分布; 2) 解决计算时的存储空间问题, 因为一个真实的反应网络一般涉及成千上万个反应物种. 现有的数值方法仅是部分地解决了有关计算问题, 需要发展更有效的、应用范围更广的数值算法.

相比较而言, 2项矩方法具有更大的优势和应用潜能, 因为它关于反应物种数目是以多项式的方式增长的. 然而, 在本文提议的2项矩格式中, 参数 z_0 的选取是1个需要讨论的问题: 对于许多反应系统, 选取 $z_0 = 1$ 就能保证2项矩在它的阶趋于无穷时能趋于零, 但是对某些反应系统, 必须细心地选取 z_0 ^[7], 且是1个计算量比较大的问题. 事实上, z_0 的选取涉及在多维空间中搜索满足特定条件(即保证2项矩在它的阶趋于无穷时能趋于零)的参数值, 工作量非常大. 此外, 当用2项矩来重构概率分布时(看方程(8))涉及阶乘的计算. 然而, 大正整数阶乘的计算必然导致大的误差. 这样, 这种重构需要克服阶乘的计算问题.

4 参考文献

[1] Zaikin A N, Zhabotinsky A M. Concentration wave propagation in two-dimensional liquid-phase self-oscillating system [J]. Nature, 1970, 225: 535-537.
[2] Thattai M, Van Oudenaarden A. Intrinsic noise in gene regulatory networks [J]. Proc Natl Acad Sci U S A, 2001, 98

(15): 8614-8619.

- [3] 胡长春, 周天寿. 三类基因振子和它们的基本动力学 [J]. 江西师范大学学报: 自然科学版, 2008, 32(1): 1-5.
[4] Lotka A J. Contribution to the theory of periodic reaction [J]. J Phys Chem, 1910, 14(3): 271-274.
[5] Gillespie D T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions [J]. J Comput Phys, 1976, 22(4): 403-434.
[6] 周天寿. 基因表达模型的研究进展: 概率分布 [J]. 江西师范大学学报: 自然科学版, 2012, 36(3): 221-229.
[7] Zhang Jiajun, Huang Lifang, Zhou Tianshou. Comment on Binomial moment equations for chemical reaction networks [J]. Phys Rev Lett, 2014, 112: 088901.
[8] Delbrück M. Statistical fluctuations in autocatalytic reactions [J]. J Chem Phys, 1940, 8(1): 120-140.
[9] Van Kampen N G. Stochastic process in physics and chemistry [M]. Amsterdam: North-Holland, 1992.
[10] Elf J, Ehrenberg M. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation [J]. Gen Res, 2013, 13(11): 2475-2484.
[11] Paulsson J. Summing up the noise in gene networks [J]. Nature, 2004, 427: 415-418.
[12] Zhang Jiajun, Yuan Zhanjiang, Zhou Tianshou. Physical limits of feedback noise-suppression in biological networks [J]. Phys Biol, 2009, 6(4): 046009.
[13] Smadbeck P, Kaznessis Y N. A closure scheme for chemical master equations [J]. Proc Natl Acad Sci U S A, 2013, 110(35): 14261-14265.
[14] Zechner C, Ruess J, Krenn P, et al. Moment-based inference predicts bimodality in transient gene expression [J]. Proc Natl Acad Sci U S A, 2012, 109(21): 8340-8345.
[15] Grima R. A study of the accuracy of moment-closure approximations for stochastic chemical kinetics [J]. J Chem Phys, 2012, 136(15): 154105.
[16] 周天寿. 生物系统的随机动力学 [M]. 北京: 科学出版社, 2009.
[17] Ale A, Kirk P, Stumpf M P H. A general moment expansion method for stochastic kinetic models [J]. J Chem Phys, 2013, 138(17): 174101.
[18] Gibson M A, Bruck J. Efficient exact stochastic simulation of chemical systems with many species and many channels [J]. J Phys Chem A, 2000, 104(9): 1876-1889.
[19] Gillespie D T. Approximate accelerated stochastic simulation of chemically reacting systems [J]. J Chem Phys, 2001, 115(4): 1716-1733.
[20] Munsky B, Khammash M. The finite state projection algorithm for the solution of the chemical master equation [J]. J Chem Phys, 2006, 124(4): 044104.

(下转第14页)

- [30] Jin Suoqin ,Zou Xiufen. Construction of the influenza a virus infection-induced cell-specific inflammatory regulatory network based on mutual information and optimization [J]. BMC Syst Biol 2013 ,7(1) : 105.
- [31] Chen Yu ,Xie Weicheng ,Zou Xiufen. A binary differential evolution algorithm learning from explored solutions [J]. Neurocomputing 2015 ,149(B) : 1038-1047.
- [32] Xie Weicheng ,Yu Wei ,Zou Xiufen. Diversity-maintained differential evolution embedded with gradient-based local search [J]. Soft Comput 2013 ,17(8) : 1511-1535.
- [33] Zou Xiufen ,Xiang Xueshuang ,Chen Yan ,et al. Understanding inhibition of viral proteins on type I IFN signaling pathways with modeling and optimization [J]. J Theor Biol 2010 ,265(4) : 691-703.

The Mathematical Modeling and Analysis for S1PR1-Mediated Cytokine Signaling Pathway

ZOU Xiufen ,NIU Lili ,JIN Suoqin

(School of Mathematics and Statistics ,Wuhan University ,Wuhan Hubei 430072 ,China)

Abstract: Based on the high-throughput data of H3N2 influenza virus ,a nonlinear ordinary differential equations (ODEs) model for S1PR1-mediated cytokines release signaling pathway is built. The parameters in the model are identified using the gradient-based differential evolution (DE) algorithm (DMGBDE) . Numerical simulation results show that the constructed mathematical model can fit the experimental data very well. Moreover ,computational analyses demonstrate that cytokines exhibit significantly different dynamical processes and steady state levels between the asymptomatic and symptomatic people who were infected by influenza virus. In addition ,parameters sensitivity analysis reveals the important factors whether the virus-infected people have severe clinical symptoms. These results can provide theoretical direction for clarifying the molecular pathogenic mechanisms of influenza virus infections.

Key words: cytokines; signaling pathway; mathematical modeling; nonlinear optimization; sensitivity analysis

(责任编辑:王金莲)

(上接第 6 页)

The Review on Study of Probability Master Equations

ZHOU Tianshou

(School of Mathematics and Computational Science ,Sun Yet-Sen University ,Guangzhou Guangdong 510275 ,China)

Abstract: Reaction networks exist extensively in natural systems. In principle ,probability master equations provide the most complete models of probabilistic behavior for any reaction network systems. However analysis and simulation of these equations have been a challenging task for a long time; these problems have not been thoroughly unsolved until now and relevant studies are still continuing. This article presents a systematic and comprehensive review on study of probability master equations focusing on common theoretical analysis methods such as linear noise approximation ,common moment-closure methods and binomial moment approach ,and common numerical approaches such as Gillespie stochastic simulation approach ,finite state mapping method and moment-closure formulations. In particular ,some reasons why slow progress is made in study of probability master equations are analyzed , and possible schemes for solving probability master equations are discussed and suggested.

Key words: reaction the network; probability master equation; moment-closure approach; stochastic simulation

(责任编辑:王金莲)