

Stochastic Description of Biochemical Networks

João P. Hespanha^{*a} and Mustafa Khammash^{**b}

^aCenter for Control, Dynamical Systems and Computation, University of California, Santa Barbara, CA, USA

^bDepartment of Biosystems Science and Engineering, Swiss Federal Institute of Technology at Zurich (ETHZ), Basel, Switzerland

Abstract

Conventional deterministic chemical kinetics often breaks down in the small volume of a living cell where cellular species (e.g., genes, mRNAs, etc.) exist in discrete, low copy numbers and react through reaction channels whose timing and order is random. In such an environment, a stochastic chemical kinetics framework that models species abundances as discrete random variables is more suitable. The resulting models consist of continue-time discrete-state Markov chains. Here we describe how such models can be formulated and numerically simulated, and we present some of the key analysis techniques for studying such reactions.

Keywords Stochastic biochemical reactions • Stochastic models • Gillespie algorithm • Chemical master equation • Moment dynamics

Introduction

The time evolution of a spatially homogeneous mixture of chemically reacting molecules is often modeled using a stochastic formulation, which takes into account the inherent randomness of thermal molecular motion. This formulation is important when modeling complex reactions inside living cells, where small populations of key reactants can set the stage for significant stochastic effects. In this entry, we review the basic stochastic model of chemical reactions and discuss the most common techniques used to simulate and analyze this model.

Stochastic Models of Chemical Reactions

We start by considering a set of N molecular species (reactants) $\mathcal{S}_1, \dots, \mathcal{S}_N$ that are confined to a fixed volume Ω . These species react through M possible reactions R_1, \dots, R_M . In this formulation of chemical kinetics, we shall assume that the system is in thermal equilibrium and is well mixed. Thus, the reacting molecules move due to their thermal energy. The population of the different reactants is described by a random process $X(t) = (X_1(t) \dots X_N(t))^T$, where $X_i(t)$ is a random variable that models the abundance (in terms of the number of copies) of molecules of species \mathcal{S}_i

*E-mail: hespanha@ece.ucsb.edu

**E-mail: mustafa.khammash@bsse.ethz.ch

Table 1 Propensity functions for elementary reactions. The constants c , c' , and c'' are related to k , k' , and k'' , the reaction rate constants from *deterministic* mass-action kinetics. Indeed it can be shown that $c = k$, $c' = k'/\Omega$, and $c'' = 2k''/\Omega$

Reaction type	Propensity function
$S_i \rightarrow \text{Products}$	$c\mathbf{x}_i$
$S_i + S_j \rightarrow \text{Products} \quad (i \neq j)$	$c'\mathbf{x}_i\mathbf{x}_j$
$S_i + S_i \rightarrow \text{Products}$	$c''\mathbf{x}_i(\mathbf{x}_i - 1)/2$

in the system at time t . For the allowable reactions, we shall only consider elementary reactions. These could either be monomolecular, $S_i \rightarrow \text{products}$, or bimolecular, $S_i + S_j \rightarrow \text{products}$. Upon the firing of reaction R_k , a transition occurs from some state $X = \mathbf{x}_i$ right before the reaction fires to some other state $X = \mathbf{x}_i + \mathbf{s}_k$, which reflects the change in the population immediately after the reaction has fired. \mathbf{s}_k is referred to as the *stoichiometric vector*. The set of allowable M reactions defines the so-called stoichiometry matrix:

$$S = [\mathbf{s}_1 \cdots \mathbf{s}_M].$$

To each reaction R_k , we associate a *propensity function*, $w_k(\mathbf{x})$ that describes the rate of that reaction. More precisely, $w_k(\mathbf{x})h$ is the probability that, given the system is in state \mathbf{x} at time t , R_k fires once in the time interval $[t, t + h)$. The propensity functions for elementary reactions is given in Table 1.

Limiting to the Deterministic Regime

There is an important connection between the stochastic process $X(t)$, as represented by the continuous-time discrete-state Markov chain described above, and the solution of a related deterministic reaction rate equations obtained from mass-action kinetics. To see this, let $\Phi(t) = [\Phi_1(t), \dots, \Phi_N(t)]^T$ be the vector concentrations of species S_1, \dots, S_N . According to mass-action kinetics, $\Phi(\cdot)$ satisfies the ordinary differential equation:

$$\dot{\Phi} = Sf(\Phi(t)), \quad \Phi(0) = \Phi_0.$$

In order to compare the $\Phi(t)$ with $X(t)$, which represents molecular counts, we divide $X(t)$ by the reaction volume to get $X^\Omega(t) = X(t)/\Omega$. It turns out that $X^\Omega(t)$ *limits* to $\Phi(t)$: According to Kurtz (Ethier and Kurtz 1986), for every $t \geq 0$:

$$\lim_{\Omega \rightarrow \infty} \sup_{s \leq t} |X^\Omega(s) - \Phi(s)| = 0, \quad \text{almost surely.}$$

Hence, over any finite time interval, the stochastic model *converges* to the deterministic mass-action one in the thermodynamic limit. Note that this is only a large volume limit result. In practice, for a fixed volume, a stochastic description may differ considerably from the deterministic description.

Stochastic Simulations

Gillespie's stochastic simulation algorithm (SSA) constructs sample paths for the random process $X(t) = (X_1(t) \dots X_N(t))^T$ that are consistent with the stochastic model described above (Gillespie 1976). It consists of the following basic steps:

1. Initialize the state $X(0)$ and set $t = 0$.
2. Draw a random number $\tau \in (0, \infty)$ with exponential distribution and mean equal to $1 / \sum_k w_k(X(t))$.
3. Draw a random number $k \in \{1, 2, \dots, M\}$ such that the probability of $k = i \in \{1, 2, \dots, M\}$ is proportional to $w_i(X(t))$.
4. Set $X(t + \tau) = X(t) + s_k$ and $t = t + \tau$.
5. Repeat from (2) until t reaches the desired simulation time.

By running this algorithm multiple times with independent random draws, one can estimate the distribution and statistical moments of the random process $X(t)$.

The Chemical Master Equation (CME)

The *chemical master equation* (CME), also known as the forward Kolmogorov equation, describes the time evolution of the probability that the system is in a given state \mathbf{x} . The CME can be derived based on the Markov property of chemical reactions. Suppose the system is in state \mathbf{x} at time t . Within an error of order $\mathcal{O}(h^2)$, the following statements apply:

- The probability that an R_k reaction fires exactly once in the time interval $[t, t + h)$ is given by $w_k(\mathbf{x})h$.
- The probability that no reactions fire in the time interval $[t, t + h)$ is given by $1 - \sum_k w_k(\mathbf{x})h$.
- The probability that more than one reaction fires in the time interval $[t, t + h)$ is zero.

Let $P(\mathbf{x}, t)$, denote the probability that the system is in state \mathbf{x} at time t . We can express $P(\mathbf{x}, t + h)$ as follows:

$$P(\mathbf{x}, t + h) = P(\mathbf{x}, t) \left(1 - \sum_k w_k(\mathbf{x})h \right) + \sum_k P(\mathbf{x} - s_k, t) w_k(\mathbf{x} - s_k)h + \mathcal{O}(h^2).$$

The first term on the right-hand side is the probability that the system is already in state \mathbf{x} at time t , and no reactions occur in the next h . In the second term on the right-hand side, the k th term in the summation is the probability that the system at time t is an R_k reaction away from being at state \mathbf{x} and that an R_k reaction takes place in the next h .

Moving $P(\mathbf{x}, t)$ to the left-hand side, dividing by h , and taking the limit as h goes to zero yields the chemical master equation (CME):

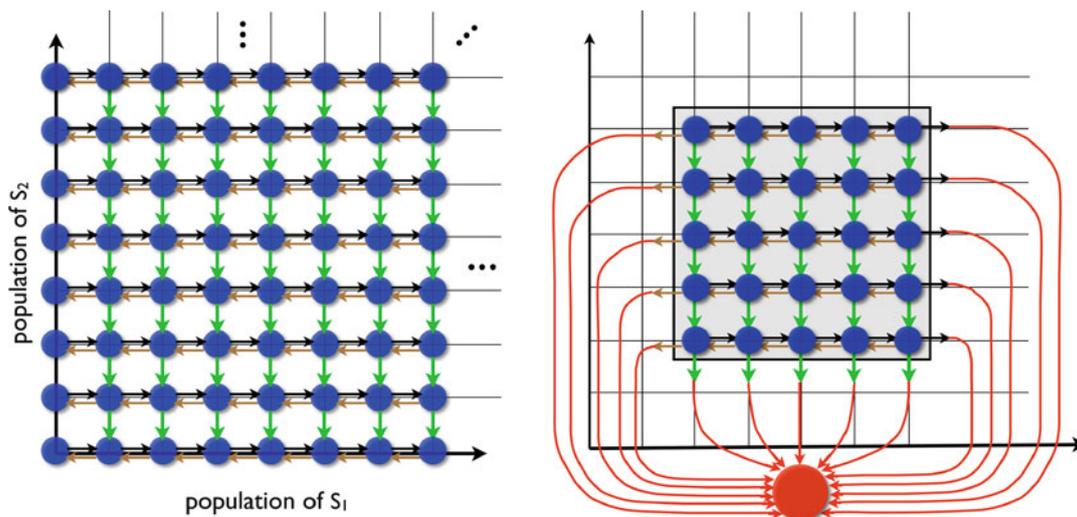


Fig. 1 The finite-state projection

$$\frac{dP(\mathbf{x}, t)}{dt} = \sum_{k=1}^M (w_k(\mathbf{x} - \mathbf{s}_k)P(\mathbf{x} - \mathbf{s}_k, t) - w_k(\mathbf{x})P(\mathbf{x}, t)). \quad (1)$$

The CME defines a linear dynamical system in the probabilities of the different states (each state is defined by a specific number of molecules of each of the species). However, there are generally an infinite number of states, and the resulting infinite linear system is not directly solvable. One approach to overcome this difficulty is to approximate the solution of the CME by truncating the states. A particular truncation procedure that gives error bounds is called the finite-state projection (FSP) (Munsky and Khammash 2006). The key idea behind the FSP approach is to keep those states that support the bulk of the probability distribution while projecting the remaining infinite states onto a single “absorbing” state. See Fig. 1.

The left panel in the figure shows the infinite states of a system with two species. The arrows indicate transitions among states caused by allowable chemical reactions. The underlying stochastic process is a continuous-time discrete-state Markov process. The right panel shows the projected (finite-state) system for a specific projection region (box). The projection is obtained as follows: transitions within the retained states are kept, while transitions that emanate from these states and end at states outside the box are channeled to a single new absorbing state. Transitions into the box are deleted. The resulting projected system is a finite-state Markov process. The probability of each of its finite states can be computed exactly. It can be shown that the truncation, as defined here, gives a lower bound for the probability for the original full system. The FSP algorithm provides a way for constructing an approximation of the CME that satisfies any prespecified accuracy requirement.

Moment Dynamics

While the probability distribution $P(\mathbf{x}, t)$ provides great detail on the state \mathbf{x} at time t , often statistical moments of the molecule copy numbers already provide important information about

their variability, which motivates the construction of mathematical models for the evolution of such models over time.

Given a vector of integers $m := (m_1, m_2, \dots, m_n)$, we use the notation $\mu^{(m)}$ to denote the following uncentered moment of X :

$$\mu^{(m)} := E[X_1^{m_1} X_2^{m_2} \cdots X_n^{m_n}].$$

Such moment is said to be of order $\sum_i m_i$. With N species, there are exactly N first-order moments $e[X_i]$, $\forall i \in \{1, 2, \dots, N\}$, which are just the means; $N(N-1)/2$ second-order moments $e[X_i^2]$, $\forall i$ and $e[X_i X_j]$, $\forall i \neq j$, which can be used to compute variances and covariance; $N(N-1)(N-2)/6$ third-order moments; and so on.

Using the CME (1), one can show that

$$\frac{d\mu^{(m)}}{dt} = E \left[\sum_k w_k(X) \left((X_1 + s_{1,k})^{m_1} (X_2 - s_{2,k})^{m_2} \cdots (X_N - s_{N,k})^{m_N} - X_1^{m_1} X_2^{m_2} \cdots X_N^{m_N} \right) \right],$$

and, because the propensity functions are all polynomials on x (cf. Table 1), the expected value in the right-hand side can actually be written as a linear combination of other uncentered moments of X . This means that if we construct a vector μ containing all the uncentered moments of x up to some order k , the evolution of μ is determined by a differential equation of the form

$$\frac{d\mu}{dt} = A\mu + B\bar{\mu}, \quad \mu \in \mathbb{R}^K, \quad \bar{\mu} \in \mathbb{R}^{\bar{K}} \quad (2)$$

where A and B are appropriately defined matrices and $\bar{\mu}$ is a vector containing moments of order larger than k . The equation (2) is exact, and we call it the (*exact*) k -order moment dynamics, and the integer k is called the *order of truncation*. Note that the dimension K of (2) is always larger than k since there are many moments of each order. In fact, in general, K is of order n^k .

When all chemical reactions have only one reactant, the term $B\bar{\mu}$ does not appear in (2), and we say that the exact moment dynamics are *closed*. However, when at least one chemical reaction has two or more reactants, then the term $B\bar{\mu}$ appears, and we say that the moment dynamics are *open* since (2) depends on the moments in $\bar{\mu}$, which are not part of the state μ . When all chemical reactions are elementary (i.e., with at most two reactants), then all moments in $\bar{\mu}$ are exactly of order $k+1$.

Moment closure is a procedure by which one approximates the exact (but open) moment dynamics (2) by an approximate (but now closed) equation of the form

$$\dot{v} = Av + B\varphi(v), \quad v \in \mathbb{R}^K \quad (3)$$

where $\varphi(v)$ is a column vector that approximates the moments in $\bar{\mu}$. The function $\varphi(v)$ is called the moment closure function, and (3) is called the *approximate k th-order moment dynamics*. The goal of any moment closure method is to construct $\varphi(v)$ so that the solution v to (3) is close to the solution μ to (2).

There are three main approaches to construct the moment closure function $\varphi(\cdot)$:

1. *Matching-based methods* directly attempt to match the solutions to (2) and (3) (e.g., Singh and Hespanha 2011).
2. *Distribution-based methods* construct $\varphi(\cdot)$ by making reasonable assumptions on the statistical distribution of the molecule counts vector x (e.g., Gomez-Uribe and Verghese 2007).
3. *Large volume methods* construct $\varphi(\cdot)$ by assuming that reactions take place on a large volume (e.g., Van Kampen 2001).

It is important to emphasize that this classification is about methods to *construct* moment closure. It turns out that sometimes different methods lead to the same moment closure function $\varphi(\cdot)$.

Conclusion and Outlook

We have introduced complementary approaches to study the evolution of biochemical networks that exhibit important stochastic effects.

Stochastic simulations permit the construction of sample paths for the molecule counts, which can be averaged to study the ensemble behavior of the system. This type of approach scales well with the number of molecular species, but can be computationally very intensive when the number of reactions is very large. This challenge has led to the development of approximate stochastic simulation algorithms that attempt to simulate multiple reactions in the same simulation step (e.g., Rathinam et al. 2003).

Solving the CME provides the most detailed and accurate approach to characterize the ensemble properties of the molecular counts, but for most biochemical systems such solution cannot be found in closed form, and numerical methods scale exponentially with the number of species. This challenge has led to the development of algorithms that compute approximate solutions to the CME, e.g., by aggregating states with low probability, while keeping track of the error (e.g., Munsky and Khammash 2006).

Moment dynamics is attractive in that the number of kt -order moments only scales polynomially with the number of chemical species, but one only obtains closed dynamics for very simple biochemical networks. This limitation has led to the development of moment closure techniques to approximate the open moment dynamics by a closed system of ordinary differential equations.

Bibliography

- Ethier SN, Kurtz TG (1986) Markov processes: characterization and convergence. Wiley series in probability and mathematical statistics: probability and mathematical statistics. Hoboken, New Jersey. See <http://books.google.com/books?id=zvE9RFouKoMC&printsec=frontcover>
- Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 22:403–434
- Gomez-Uribe CA, Verghese GC (2007) Mass fluctuation kinetics: capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. *J Chem Phys* 126(2):024109–024109–12
- Munsky B, Khammash M (2006) The finite state projection algorithm for the solution of the chemical master equation. *J Chem Phys* 124:044104

- Rathinam M, Petzold LR, Cao Y, Gillespie DT (2003) Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method. *J Chem Phys* 119(24):12784–12794
- Singh A, Hespanha JP (2011) Approximate moment dynamics for chemically reacting systems. *IEEE Trans Autom Control* 56(2):414–418
- Van Kampen NG (2001) *Stochastic processes in physics and chemistry*. Elsevier Science, Amsterdam