

## Chapter 2

# Stochastic Simulation of Biochemical Reaction Systems

This chapter presents the foundational theory of the *stochastic chemical kinetics* for modeling biochemical reaction networks, of which the discreteness in population of species and the randomness of reactions are treated as an intrinsic part. The dynamical behavior of the biochemical reactions, based on the fundamental premise of the stochastic chemical kinetics, is exactly described by the *chemical master equation* (CME). A class of Monte Carlo simulation techniques originating from the *stochastic simulation algorithm* (SSA) has been developed to realize the time evolution of the reaction networks. SSA outlines an exact computational procedure to sample the temporal evolution of biological systems consistently with CME and lays down the groundwork for developments in the next chapter.

The chapter is organized as follows. Section 2.1 presents the framework for stochastic modeling of biochemical reactions. Section 2.2 develops the mathematical basis for the stochastic simulation algorithm (practical implementations will be introduced in Chapter 3). The statistical techniques for analyzing the simulation results are presented in Section 2.3. Section 2.4 reports the conclusion remarks and Section 2.5 suggests further reading.

## 2.1 Stochastic Chemical Kinetics

This section first introduces biochemical reactions (Section 2.1.1), then defines the concept of reaction propensity (Section 2.1.2), and finally derives the chemical master equation (Section 2.1.3).

### 2.1.1 Biochemical Reactions

Biochemical reactions are the building blocks to model biological systems. They provide a unifying notation with sufficient level of details to represent complex bi-

ological processes. Biochemical reactions decorated with reaction kinetics can be simulated by a simulation algorithm to generate a realization of their dynamics.

Chemical species in a biological system move around and gain kinetic energy. Upon collisions with other species, they undergo reactions to modify and transform into different species. In order to make this concrete, consider the transformation of a molecule of species A to a molecule of species B. It is written schematically as



This reaction converts one A molecule on the left side of the arrow to a B molecule on the right side. Such a transforming reaction is called a *unimolecular reaction*. The special unimolecular reaction



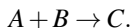
represents the degradation of species A. The species  $\emptyset$  denotes a special species that is not considered in the model (*e.g.*, because its population is large and does not change over time).

The reaction

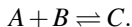


is called a *synthesis reaction* (or source reaction). The A molecules are introduced into the biological system from outside, *e.g.*, species reservoir. Synthesis reactions are often used to model the effects of outside environment on the system dynamics.

An A molecule can associate with a B molecule to produce a complex C through an association reaction



Such a reaction is called a *bimolecular reaction*. Often, the complexation process is reversible, *i.e.*, the complex C will disassociate into an A molecule and a B molecule. The association and disassociation reactions are written together as



A reversible reaction is only a convenient shorthand for writing two separated irreversible reactions. Changes caused by a reaction is only considered one direction at a time.

The special bimolecular reaction

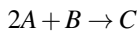


is called a *dimerization*, where two molecules of the same species A are consumed to produce a B molecule.

The four reaction types discussed above are called *elementary reactions* because they take one step to complete. A *non-elementary reaction*, which can be a higher order reaction or a multi-step reaction, can also be used to model biochemical reactions. For example, the *termolecular reaction*

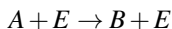


is used to represent the polymerization of three molecules of the same species A into a B molecule. Also, the termolecular reaction



is used to represent the combination of two A molecules with a B molecule to produce a complex C.

An example of multi-step reaction which is widely used is the *enzymatic reaction*



where E is the enzyme that catalyzes the rate of conversion of the species A into species B. The use of a specific reaction type in modeling depends on the knowledge of the biological system under study and the availability of the data.

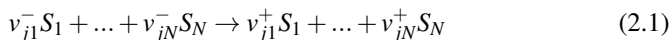
For a formal mathematical description, consider a biological system consisting of  $N$  chemical species  $S_1, \dots, S_N$ . The species are assumed to confine in a *well-mixed volume*  $V$  at thermal equilibrium (Definition 2.1). The legitimacy condition for the well-mixed volume is that nonreactive collisions, which do not lead to reactions, are much more frequent than reactive collisions, which lead to reactions. Chemical species under the well-mixed assumption at a thermal equilibrium are uniformly distributed in the volume  $V$  and their velocities are thermally randomized according to the Maxwell-Boltzmann distribution.

### Definition 2.1: Well-mixed reaction volume

The reaction volume in which all the molecular species are homogeneously distributed and spatially indistinguishable is called well-mixed. The biochemical reaction system with well-mixed volume thus satisfies the *spatial homogeneity* where spatial distribution of molecular species can be ignored.

The state of a spatially homogeneous biological system is determined by the population of each species, while the position and velocity of each individual molecule are ignored. Let  $X_i(t)$  be the population of species  $S_i$  at a particular time  $t$ . The  $N$ -vector  $X(t) = (X_1(t), \dots, X_N(t))$ , which determines the population of each species, constitutes the system state at the time  $t$ .

Chemical species can interact through  $M$  reactions  $R_1, \dots, R_M$ . A particular reaction  $R_j$  has a general scheme



in which a species on the left side of the arrow is called a *reactant*, while a species on the right side is called a *product*. The non-negative integers  $v_{ji}^-$  and  $v_{ji}^+$  are the *stoichiometric coefficients* which denote the number of molecules of a reactant that is consumed and the number of molecules of a product that is produced, respectively.

A reactant species that affects the speed of a reaction but is not consumed by the reaction, *i.e.*,  $v_{ji}^- = v_{ji}^+$ , is called a *catalyst*. The sum of stoichiometric coefficients of reactants of a reaction  $R_j$  is called *reaction order*. We note that the order of, for example, a multi-step reaction may not necessarily reflect through its reactant coefficients.

### Example 2.1: Reaction order

The orders of synthesis reaction, unimolecular, bimolecular and termolecular are 0, 1, 2 and 3, respectively.

For each reaction  $R_j$ , the net change in the population of species  $S_i$  involved in the reaction is equal to  $(v_{ji}^+ - v_{ji}^-)$ , which can be positive, negative or zero. The net changes by all reactions are described by a *stoichiometry matrix*  $\mathbf{v}$  with size  $M \times N$ . The  $j$ th row  $\mathbf{v}_j$  of the stoichiometry matrix expresses the changes caused by reaction  $R_j$  and it is called the *state change vector*.

### Example 2.2: Stoichiometry matrix of the Oscillator model

Consider the Oscillator model in Appendix A.3. It is an artificial model that implements the positive feedback motif. The model, shown in Fig. 2.1, consists of three species and three reactions.

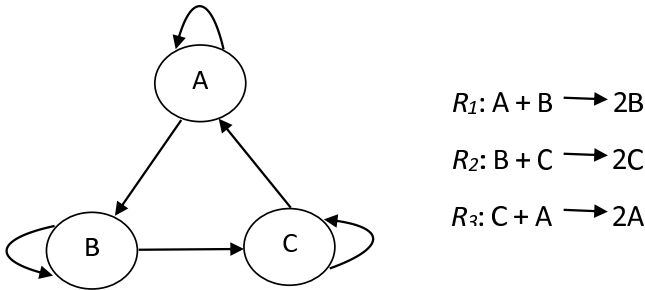


Fig. 2.1 Oscillator model.

Table 2.1 shows the stoichiometry matrix  $\mathbf{v}$  of the Oscillator model. The table has three rows in which each row expresses the net changes in the population of each species caused by a reaction. Specifically, consider the state change vector  $\mathbf{v}_1$  of the reaction  $R_1$  shown in the first row of the stoichiometry matrix  $\mathbf{v}$ . A firing of  $R_1$  consumes one A molecule and one B molecule and produces two B molecules, while the population of species C is unchanged. The net change in the population of species A, B and C by firing  $R_1$  is thus 1,  $-1$  and 0, respectively.

**Table 2.1** The stoichiometry matrix  $\mathbf{v}$  of the Oscillator model

Species	A	B	C
$\mathbf{v}_1$	-1	1	0
$\mathbf{v}_2$	0	-1	1
$\mathbf{v}_3$	1	0	-1

Suppose that at a time  $t$  the state is  $X(t)$ . It further assumes that the next reaction scheduled to fire at the next time  $t + \tau$  is  $R_\mu$ , which moves the system accordingly to a new state  $X(t + \tau)$ . Two important assumptions are imposed for the transition from the state  $X(t)$  to the new state  $X(t + \tau)$  by firing reaction  $R_\mu$ . First, no changes occur in the system state during the time interval  $[t, t + \tau)$  before the next reaction  $R_\mu$  fires at the time  $t + \tau$ . Second, the reaction occurs instantly after it is initiated. These assumptions are called the *Markov property*. The state transition by firing reaction  $R_\mu$  under the Markovian assumptions is expressed through the state change vector  $\mathbf{v}_\mu$  as

$$X(t + \tau) = X(t) + \mathbf{v}_\mu. \quad (2.2)$$

### 2.1.2 Reaction Propensity

Each reaction in the stochastic chemical kinetics is considered as a stochastic process where each of its occurrences is a random event with an assigned probability distribution. All reactions have chances to fire and move the system to new states. The system can be at each reachable state in a time interval. It is thus impossible to predict the progress of reactions deterministically, but only stochastically with a probability. To account for the uncertainty, each reaction  $R_j$  in the stochastic chemical kinetics is associated with a *propensity*  $a_j$  that expresses the probability per unit time of the occurrence of the reaction, given the current state  $X(t)$  at time  $t$ .

#### Definition 2.2: Reaction propensity

The propensity  $a_j$  of a reaction  $R_j$  is defined such that

$$a_j(\mathbf{x})dt = \text{probability that a reaction } R_j \text{ fires in the next infinitesimal time interval } [t, t + dt), \text{ given the state } X(t) = \mathbf{x} \text{ at time } t.$$

The propensity  $a_j(X(t))$  is a function of the state  $X(t)$ . It is important to note that although the propensity function  $a_j$  of a reaction depends on the time  $t$ , this happens implicitly through the state  $X(t)$  because the propensity  $a_j(X(t))$ , under the Markovian assumptions, changes only at the time the state  $X(t)$  changes due to a reaction firing. At a particular time  $t$ , the value of the propensity  $a_j(X(t))$  is a deterministic quantity. The propensity at a different time may have different values

depending on the state at that time. Therefore, the propensity value of a reaction in a state  $X(t)$  is often used as a measure of how fast the reaction proceeds to move to a new state.

Let  $\mathbb{P}\{R_j \text{ fires in } [t, t + dt)\}$  be the probability that reaction  $R_j$  fires in the next infinitesimal time interval  $[t, t + dt)$ , given the state  $X(t) = \mathbf{x}$  at time  $t$ . Definition 2.2 is equivalent with

$$\mathbb{P}\{R_j \text{ fires in } [t, t + dt)\} = a_j(\mathbf{x})dt + o(dt) \quad (2.3)$$

where the little-o term  $o(dt)$  is used to express that it asymptotically approaches zero faster than  $dt$ , i.e.,  $\lim_{dt \rightarrow 0} o(dt)/dt = 0$ . In other words, the probability that there are more than one firing of  $R_j$  in an infinitesimal time interval  $[t, t + dt)$  is in the order of  $o(dt)$  and thus it is negligible.

A precise formula of the propensity function  $a_j$  on the state  $X(t)$  is dependent on the kinetic theory and specific assumptions about how the reaction physically occurs. It is referred to as the *fundamental premise* of the stochastic chemical kinetics.

For the standard mass action kinetics, the propensity  $a_j$  of reaction  $R_j$  is proportional to a *stochastic reaction rate*  $c_j$  and the number of its reactants.

### Definition 2.3: Mass action propensity

For mass action kinetics, the propensity  $a_j$  of reaction  $R_j$  in Eq. (2.1), given the current state  $X(t)$  at time  $t$ , is

$$a_j(X(t)) = c_j h_j(X(t))$$

where  $c_j$  is the *stochastic reaction rate* and  $h_j(X(t))$  counts the number of distinct combinations of reactants,

$$h_j(X(t)) = \prod_i \binom{X_i(t)}{\mathbf{v}_{ji}^-} = \prod_i \frac{X_i(t)!}{\mathbf{v}_{ji}^-! (X_i(t) - \mathbf{v}_{ji}^-)!}.$$

The number of combinations  $h_j(X(t))$  of a synthesis reaction, where the stoichiometric coefficient of its reactants is zero, is set to  $h_j(X(t)) = 1$ .

The stochastic rate  $c_j$  denotes the average probability per unit time that a particular combination of reactant molecules of reaction  $R_j$  reacts in the volume  $V$  and it depends on the reaction type. For example, the stochastic rate of a unimolecular reaction is independent of the volume size, while the stochastic rate of a bimolecular reaction is inversely proportional to the volume  $V$  because for a pair of reactant molecules it is harder to find each other in a larger volume. The stochastic rate  $c_j$  of a reaction  $R_j$  is a constant provided that the volume  $V$  is constant, well-mixed and thermally homogeneous.

Hereafter, we write  $X_i$  in place of  $X_i(t)$ , when  $t$  is irrelevant or clear from the context.

### Example 2.3: Reaction propensity with mass action kinetics

Reaction propensity for reactions  $R_j$  with mass action kinetics

- Synthesis reaction ( $\emptyset \rightarrow$  products): the number of combinations  $h_j = 1$  and propensity  $a_j = c_j$
- Unimolecular reaction ( $S_i \rightarrow$  products): the number of combinations  $h_j = X_i$  and propensity  $a_j = c_j X_i$ .
- Bimolecular reaction ( $S_i + S_k \rightarrow$  products): the number of combinations  $h_j = X_i X_k$  and propensity  $a_j = c_j X_i X_k$ .
- Dimerization reaction ( $2S_i \rightarrow$  products): the number of combinations  $h_j = \frac{1}{2} X_i (X_i - 1)$  and propensity  $a_j = \frac{1}{2} c_j X_i (X_i - 1)$ .
- Polymerization reaction ( $3S_i \rightarrow$  products): the number of combinations  $h_j = \frac{1}{6} X_i (X_i - 1) (X_i - 2)$  and propensity  $a_j = \frac{1}{6} c_j X_i (X_i - 1) (X_i - 2)$ .
- Termolecular reaction ( $2S_i + S_k \rightarrow$  products): the number of combinations  $h_j = \frac{1}{2} X_i (X_i - 1) X_k$  and propensity  $a_j = \frac{1}{2} c_j X_i (X_i - 1) X_k$ .

Beyond the standard mass action kinetics, complex reaction kinetics can also be used. The propensity  $a_j$  of a reaction  $R_j$  in this setting often shows a complicated, nonlinear dependence on the chemical species, and may also contain more than one rate constant. The Michaelis-Menten kinetics, for instance, is commonly used to approximate the mechanism of enzymatic reactions (see also Section 4.8.1.1).

### Example 2.4: Reaction propensity with Michaelis-Menten kinetics

Consider an enzymatic reaction  $R_j$  with form  $S_i + S_k \rightarrow S_i + S_l$ , where  $S_i$  is the enzyme and  $S_k$  is the substrate. The reaction propensity according to the Michaelis-Menten kinetics is defined as  $a_j = \frac{V_{max}}{K_M + X_k} X_i X_k$ , where  $V_{max}$  is the maximum rate such that the substrate  $S_k$  is saturated and  $K_m$ , called the Michaelis constant, is the substrate concentration at which the reaction rate is half of  $V_{max}$ .

## 2.1.3 Chemical Master Equation

Suppose the biochemical reaction system starts with an initial state  $X(t_0) = \mathbf{x}_0$  at time  $t_0$ . Let  $t > t_0$  and the system at state  $X(t) = \mathbf{x}$ . The purpose of the stochastic chemical kinetics is to infer the probability  $\mathbb{P}\{\mathbf{x}, t | \mathbf{x}_0, t_0\}$ .

### Definition 2.4: Grand probability function

The probability function  $\mathbb{P}\{\mathbf{x}, t | \mathbf{x}_0, t_0\}$  is

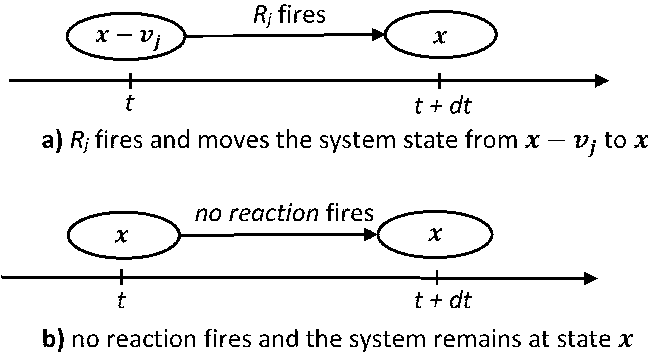
$\mathbb{P}\{\mathbf{x}, t | \mathbf{x}_0, t_0\}$  = probability that the system state is  $X(t) = \mathbf{x}$  at time  $t$ ,  
given the initial state  $X(t_0) = \mathbf{x}_0$  at time  $t_0$ .

The probability  $\mathbb{P}\{\mathbf{x}, t | \mathbf{x}_0, t_0\}$  is called the *grand probability* function because it gives the probabilities of all reachable states of the system at time  $t$ , given the initial state  $X(t_0) = \mathbf{x}_0$  at time  $t_0$ . Knowing  $\mathbb{P}\{\mathbf{x}, t | \mathbf{x}_0, t_0\}$ , all the statistical properties (e.g., mean, variance) can be calculated for every species at any time  $t > t_0$ .

To derive the time evolution for the grand probability, consider an infinitesimal time interval  $[t, t + dt)$  so that there is at most one reaction firing in this interval. Suppose that at time  $t + dt$  the system state is  $X(t + dt) = \mathbf{x}$ . There are two cases in order to reach the state  $\mathbf{x}$  in the next infinitesimal time  $t + dt$ , given the current time  $t$ . Either 1) be at state  $X(t) = \mathbf{x} - \mathbf{v}_j$  at time  $t$  and reaction  $R_j$  fires in the next time  $t + dt$  which leads to the next state  $X(t + dt) = \mathbf{x}$ , or 2) already be at state  $X(t) = \mathbf{x}$  at time  $t$  and no reaction fires in the next infinitesimal time interval  $[t, t + dt)$ . These two scenarios are depicted in Fig. 2.2. The grand probability  $\mathbb{P}\{\mathbf{x}, t + dt | \mathbf{x}_0, t_0\}$  is thus written as

$$\mathbb{P}\{\mathbf{x}, t + dt | \mathbf{x}_0, t_0\} = \sum_{j=1}^M \mathbb{P}\{R_j \text{ fires in } [t, t + dt)\} \mathbb{P}\{\mathbf{x} - \mathbf{v}_j, t | \mathbf{x}_0, t_0\} + \mathbb{P}\{\text{no reaction fires in } [t, t + dt)\} \mathbb{P}\{\mathbf{x}, t | \mathbf{x}_0, t_0\} \quad (2.4)$$

where  $\mathbb{P}\{\text{no reaction fires in } [t, t + dt)\}$  denotes the probability that no reaction fires in the infinitesimal time interval  $[t, t + dt)$ . Note that when the state vector  $\mathbf{x} - \mathbf{v}_j$  gives negative populations, the probability  $\mathbb{P}\{\mathbf{x} - \mathbf{v}_j, t | \mathbf{x}_0, t_0\}$  in Eq. (2.4) is zero because the populations of species must be positive.



**Fig. 2.2** Consider an infinitesimal time interval  $[t, t + dt)$ . At the current time  $t$ , there are two scenarios to reach the state  $\mathbf{x}$  in the next infinitesimal time  $t + dt$ . a) The state is  $\mathbf{x} - \mathbf{v}_j$  at time  $t$  and reaction  $R_j$  fires in the next infinitesimal time  $t + dt$  leading to state  $\mathbf{x}$ . b) The state is  $\mathbf{x}$  at time  $t$  and still remains in this state because there is no reaction firing in the next infinitesimal time interval  $[t, t + dt)$ .



The probability that no reaction fires in the infinitesimal time interval  $[t, t + dt)$  can be computed as:

$$\begin{aligned}
 \mathbb{P}\{\text{no reaction fires in } [t, t + dt)\} &= \prod_{j=1}^M (1 - \mathbb{P}\{R_j \text{ fires in } [t, t + dt)\}) \\
 &= \prod_{j=1}^M (1 - a_j(\mathbf{x})dt + o(dt)) \\
 &= 1 - \sum_{j=1}^M a_j(\mathbf{x})dt + o(dt) \tag{2.5}
 \end{aligned}$$

in which the first equality is derived from the complement rule and multiplication rule of probability, the second equality is obtained by applying Eq. (2.3), and the third equality is achieved by expanding and rearranging the product so that the terms with high orders of  $dt$  are collectively represented by  $o(dt)$  because they asymptotically approach zero faster than  $dt$ .

Substituting Eq. (2.3) and Eq. (2.5) into Eq. (2.4) gives

$$\begin{aligned}
 \mathbb{P}\{\mathbf{x}, t + dt | \mathbf{x}_0, t_0\} &= \sum_{j=1}^M \mathbb{P}\{\mathbf{x} - \mathbf{v}_j, t | \mathbf{x}_0, t_0\} (a_j(\mathbf{x} - \mathbf{v}_j)dt + o(dt)) + \\
 &\quad \mathbb{P}\{\mathbf{x}, t | \mathbf{x}_0, t_0\} (1 - \sum_{j=1}^M a_j(\mathbf{x})dt + o(dt)). \tag{2.6}
 \end{aligned}$$

Subtract  $\mathbb{P}\{\mathbf{x}, t | \mathbf{x}_0, t_0\}$  from both sides of Eq. (2.6), divide through by  $dt$  and finally consider the limit  $dt \rightarrow 0$  with a remark that  $\lim_{dt \rightarrow 0} o(dt)/dt = 0$ ; this results in

$$\frac{d\mathbb{P}\{\mathbf{x}, t | \mathbf{x}_0, t_0\}}{dt} = \sum_{j=1}^M (a_j(\mathbf{x} - \mathbf{v}_j) \mathbb{P}\{\mathbf{x} - \mathbf{v}_j, t | \mathbf{x}_0, t_0\}) - \mathbb{P}\{\mathbf{x}, t | \mathbf{x}_0, t_0\} \sum_{j=1}^M a_j(\mathbf{x}). \tag{2.7}$$

Eq. (2.7) is called the *chemical master equation* (CME). It is in fact a collection of differential equations in which each differential equation represents the probability of each possible state of the system at the time  $t$ . Thus, CME provides a complete description of the time evolution of the grand probability  $\mathbb{P}\{\mathbf{x}, t | \mathbf{x}_0, t_0\}$ .

#### Example 2.5: Solving CME for the Birth process

Consider the birth process model in Appendix A.1. The model contains a synthesis reaction that produces species S at rate  $c$ ,



Assume that at time  $t = 0$ , the number of S molecules is 0. Let  $n$  be number of molecules of species S produced at a particular time  $t > 0$ . Let  $\mathbb{P}\{n, t\}$  be the probability that there are  $n$  molecules of species S produced at time  $t$ , given zero S molecules at time 0 (the condition in the grand probability is removed to simplify the notation).

The collection of differential equations described by CME in Eq. (2.7) is explicitly written as:

$$\begin{cases} \frac{d\mathbb{P}\{i, t\}}{dt} = c\mathbb{P}\{i-1, t\} - c\mathbb{P}\{i, t\} & , \text{ for all } i = 1, \dots, n \\ \frac{d\mathbb{P}\{0, t\}}{dt} = -c\mathbb{P}\{0, t\} & , \text{ if } i = 0 \end{cases} \quad (2.8)$$

and the initial condition is  $\mathbb{P}\{0, 0\} = 1$ .

Eq. (2.8) has an analytical solution given by

$$\mathbb{P}\{n, t\} = \frac{(ct)^n}{n!} e^{-ct} \quad (2.9)$$

which denotes a Poisson distribution with parameter  $ct$ . Therefore, at a particular time  $t$  the expected number of S molecules is  $\mathbb{E}[n] = ct$  and the variance is  $\text{Var}[n] = ct$ .

The solution of CME gives the probabilities of all possible states at any time (see Example 2.5); however, directly solving CME poses a lot of computational challenges. An analytical and/or direct numerical approach to solve CME in general is non-trivial and difficult to find, except for rather simple cases such as Example 2.5. The challenge in solving CME is due to a huge number of differential equations required to specify probabilities of all possible states. Consider a simple model consisting of  $N$  species where the population of each species  $S_i$  has only two values, either 0 or 1 (*i.e.*,  $X_i = 0$  or 1). The system has total  $2^N$  possible states, hence CME needs  $2^N$  differential equations for describing the probabilities of all of these  $2^N$  possible states. The number of differential equations in the CME equation is thus exponentially increasing with the number of species  $N$ . Furthermore, the population of a species in a practical model may be very large and even infinite. The state space explosion problem (often referred to as the *curse of dimensionality*) prevents direct approaches in solving the CME equation.

## 2.2 Stochastic Simulation

Stochastic simulation is an alternative approach to solve CME by producing possible realizations of the grand probability function. It only explores possible states in the state space each time. Therefore, stochastic simulation can handle the biochemical reactions with very high dimensional state space. The mathematical basis of stochastic simulation is the *reaction probability density function* (pdf)  $p(\tau, \mu | \mathbf{x}, t)$ .

**Definition 2.5: Reaction probability density function**

The reaction probability density function  $p(\tau, \mu | \mathbf{x}, t)$  is defined such that

$p(\tau, \mu | \mathbf{x}, t) d\tau$  = probability that reaction  $R_\mu$  fires in the infinitesimal time interval  $[t + \tau, t + \tau + d\tau)$ , given the state  $X(t) = \mathbf{x}$  at time  $t$ .

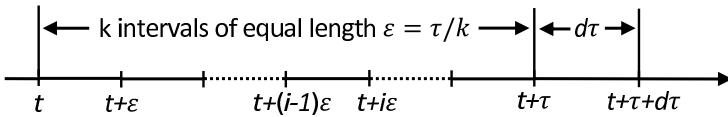
The pdf  $p(\tau, \mu | \mathbf{x}, t)$  is a joint distribution of two variables showing the index  $\mu$  of the reaction firing  $R_\mu$  and the time  $\tau$  to the firing, respectively, knowing that the system is at state  $X(t) = \mathbf{x}$  at time  $t$ . The domain of the reaction index  $\mu$  is an integer value from  $1 \leq \mu \leq M$ , while the domain of the next time  $\tau$  is a real value in  $0 \leq \tau < \infty$ .

The probability  $p(\tau, \mu | \mathbf{x}, t) d\tau$  in Definition 2.5 can be calculated as the product of two probabilities: 1) the probability that no reaction fires in the time interval  $[t, t + \tau)$  and 2) the probability that a reaction  $R_\mu$  fires in the next infinitesimal time interval  $[t + \tau, t + \tau + d\tau)$ . Let  $\mathbb{P}\{\text{no reaction fires in } [t, t + \tau)\}$  be the probability that no reaction fires in the time interval  $[t, t + \tau)$  and  $\mathbb{P}\{R_\mu \text{ fires in } [t + \tau, t + \tau + d\tau)\}$  be the probability that reaction  $R_\mu$  fires in the next infinitesimal time interval  $[t + \tau, t + \tau + d\tau)$ . Then,

$$p(\tau, \mu | \mathbf{x}, t) d\tau = \mathbb{P}\{\text{no reaction fires in } [t, t + \tau)\} \mathbb{P}\{R_\mu \text{ fires in } [t + \tau, t + \tau + d\tau)\}. \quad (2.10)$$

To calculate the first probability  $\mathbb{P}\{\text{no reaction fires in } [t, t + \tau)\}$ , divide the time interval  $[t, t + \tau)$  into  $k$  non-overlapping sub-intervals with equal length  $\varepsilon = \tau/k$  as shown in Fig. 2.3. The probability that no reaction fires in the  $i$ th interval  $[t + (i - 1)\varepsilon, t + i\varepsilon)$ , for  $i = 1, \dots, k$ , is (see Eq. (2.5))

$$\mathbb{P}\{\text{no reaction fires in } [t + (i - 1)\varepsilon, t + i\varepsilon)\} = 1 - \sum_{j=1}^M a_j(\mathbf{x}) \varepsilon + o(\varepsilon).$$



**Fig. 2.3** The  $k$  non-overlapping sub-intervals with equal length  $\varepsilon = \tau/k$  in which the  $i$ th time interval is  $[t + (i - 1)\varepsilon, t + i\varepsilon)$  constitute the time interval  $[t, t + \tau)$  and the last interval is  $[t + \tau - \varepsilon, t + \tau)$ .

Hence, by the multiplication rule of probability, the probability that no reaction fires in the time interval  $[t, t + \tau)$  is the product of the probabilities that no reaction fires in  $k$  non-overlapping intervals. Formally,

$$\begin{aligned}
\mathbb{P}\{\text{no reaction fires in } [t, t + \tau]\} &= \prod_{i=1}^k \mathbb{P}\{\text{no reaction fires in } [t + (i-1)\varepsilon, t + i\varepsilon]\} \\
&= \prod_{i=1}^k \left(1 - \sum_{j=1}^M a_j(\mathbf{x})\varepsilon + o(\varepsilon)\right) \\
&= \left(1 - \sum_{j=1}^M a_j(\mathbf{x})\varepsilon + o(\varepsilon)\right)^k \\
&= (1 - a_0(\mathbf{x})\varepsilon + o(\varepsilon))^k
\end{aligned} \tag{2.11}$$

where  $a_0(\mathbf{x})$  is the *total propensity* that is defined as

$$a_0(\mathbf{x}) = \sum_{j=1}^M a_j(\mathbf{x}). \tag{2.12}$$

Eq. (2.11) is valid for any integer  $k > 1$ , so it is valid for the limit case  $k \rightarrow \infty$ , and Eq. (2.11) becomes

$$\begin{aligned}
\mathbb{P}\{\text{no reaction fires in } [t, t + \tau]\} &= \lim_{k \rightarrow \infty} (1 - a_0(\mathbf{x})\varepsilon + o(\varepsilon))^k \\
&= \lim_{k \rightarrow \infty} \left(1 - \frac{a_0(\mathbf{x})k\varepsilon + ko(\varepsilon)}{k}\right)^k \\
&= \lim_{k \rightarrow \infty} \left(1 - \frac{a_0(\mathbf{x})\tau + \tau(o(\varepsilon)/\varepsilon)}{k}\right)^k \\
&= e^{-a_0(\mathbf{x})\tau}
\end{aligned} \tag{2.13}$$

in which the third equality is obtained by using the equality  $\varepsilon = \tau/k$ . The last equality is derived because 1)  $o(\varepsilon)/\varepsilon \rightarrow 0$  when  $k \rightarrow \infty$ , and 2)  $\lim_{k \rightarrow \infty} (1 - \frac{a_0(\mathbf{x})\tau}{k})^k = e^{-a_0(\mathbf{x})\tau}$ .

The second probability  $\mathbb{P}\{R_\mu \text{ fires in } [t + \tau, t + \tau + d\tau]\}$  is calculated by

$$\mathbb{P}\{R_\mu \text{ fires in } [t + \tau, t + \tau + d\tau]\} = a_\mu(\mathbf{x})d\tau + o(d\tau) \tag{2.14}$$

by Definition 2.2 of the reaction propensity.

Plugging Eqs. (2.13) and (2.14) into Eq. (2.10) gives

$$p(\tau, \mu | \mathbf{x}, t) d\tau = e^{-a_0(\mathbf{x})\tau} (a_\mu(\mathbf{x})d\tau + o(d\tau)). \tag{2.15}$$

Dividing both sides of Eq. (2.15) by  $d\tau$  and finally taking the limit  $d\tau \rightarrow 0$  with a remark that  $o(d\tau)/d\tau \rightarrow 0$ , the pdf  $p(\tau, \mu | \mathbf{x}, t)$  has the concrete formula

$$p(\tau, \mu | \mathbf{x}, t) = a_\mu(\mathbf{x})e^{-a_0(\mathbf{x})\tau}. \tag{2.16}$$

The pdf  $p(\tau, \mu | \mathbf{x}, t)$  in Eq. (2.16) is indeed the joint probability density function of the next reaction index  $\mu$  and the next firing time  $\tau$  over their domains. It can be

verified as

$$\int_0^\infty d\tau \sum_{\mu=1}^M p(\tau, \mu | \mathbf{x}, t) = \sum_{\mu=1}^M a_\mu(\mathbf{x}) \int_0^\infty d\tau e^{-a_0(\mathbf{x})\tau} = 1.$$

Furthermore, Eq. (2.16) shows that the pdf  $p(\tau, \mu | \mathbf{x}, t)$  depends on propensities of all reactions (not just only on the propensity  $a_\mu$ ) through the total propensity  $a_0$  in the exponential as well as on all species (not just only on the reactants of  $R_\mu$ ) through the current state  $\mathbf{x}$ .

The pdf  $p(\tau, \mu | \mathbf{x}, t)$  given in Eq. (2.16) is the mathematical framework for a class of exact Monte Carlo simulation techniques originating from the *stochastic simulation algorithm* (SSA). SSA is a *discrete event simulation* in which the state is updated by a random selected reaction  $R_\mu$  with index  $\mu$  at a discrete time  $\tau$  sampled from the pdf  $p(\tau, \mu | \mathbf{x}, t)$ . SSA is an exact simulation procedure because it exactly generates the reaction index  $\mu$  of the reaction firing  $R_\mu$  and the firing time  $\tau$  without introducing approximation in sampling  $p(\tau, \mu | \mathbf{x}, t)$ . In the following, a brief introduction to the general structure of SSA simulation is presented. The actual implementation of the Monte Carlo step to realize the pdf  $p(\tau, \mu | \mathbf{x}, t)$  will be discussed in detail in Chapter 3.

A general sketch of the SSA procedure is outlined in Algorithm 1. The input of SSA is a reaction network of  $M$  reactions in which each reaction  $R_j$ ,  $j = 1, \dots, M$ , is characterized by two quantities that are the state change vector  $\mathbf{v}_j$  and the propensity function  $a_j$ . The initial state  $\mathbf{x}_0$  denotes the initial population of each species  $S_i$ ,  $i = 1, \dots, N$ , at time  $t = 0$ . A specified time  $T_{max}$  is the ending time to stop the simulation. The population of each species at time  $t \leq T_{max}$  is stored in the state vector  $X$ .

---

#### Algorithm 1 Stochastic Simulation Algorithm (SSA) - General Sketch

---

**Input:** a biochemical reaction network of  $M$  reactions in which each reaction  $R_j$ ,  $j = 1, \dots, M$ , is accompanied with the state change vector  $\mathbf{v}_j$  and the propensity  $a_j$ , the initial state  $\mathbf{x}_0$  at time 0 and the simulation ending time  $T_{max}$

**Output:** a trajectory of the biochemical reaction network which is a collection of states  $X(t)$  for time  $0 \leq t \leq T_{max}$ .

```

1: initialize time  $t = 0$  and state  $X = \mathbf{x}_0$ 
2: while ( $t < T_{max}$ ) do
3:   set  $a_0 = 0$ 
4:   for all (reaction  $R_j$ ) do
5:     compute  $a_j$ 
6:     update  $a_0 = a_0 + a_j$ 
7:   end for
8:   sample reaction  $R_\mu$  and firing time  $\tau$  from pdf  $p(\tau, \mu | \mathbf{x}, t)$  in Eq. (2.16)
9:   update state  $X = X + \mathbf{v}_\mu$ 
10:  set  $t = t + \tau$ 
11: end while

```

---

SSA begins by assigning the initial state  $\mathbf{x}_0$  to the state  $X$ . It then goes into the main simulation loop in lines 2 - 11. For each iteration, the algorithm computes the propensity  $a_j$  of each reaction  $R_j$  for  $j = 1, \dots, M$  and the total propensity  $a_0 = \sum_{j=1}^M a_j$  (lines 4 - 7). The heart of the SSA algorithm is the line 8 where the next reaction  $R_\mu$  and its firing time  $\tau$  are sampled from the pdf  $p(\tau, \mu | \mathbf{x}, t)$ . This sampling step may need the generation of uniformly distributed random numbers. A brief recall on techniques for generating random numbers can be found in Appendix B. Lines 9 - 10 update the state to a new state  $X = X + \mathbf{v}_\mu$  and advance the time to a new time  $t = t + \tau$ . The simulation loop is repeated until the time  $t$  is greater than the time  $T_{max}$ . Note that the propensities of reactions in Algorithm 1 are updated at each simulation iteration to reflect changes in the populations of species caused by reaction firings, but this can be skipped by employing an appropriate sampling technique that is discussed in the next chapter.

The result of a SSA run is a trajectory, which shows the evolution of the biological system over time. The trajectory is a collection of states  $X(t)$  that denotes the state of the system at any time  $0 \leq t \leq T_{max}$ . It should be emphasized that because SSA is a discrete event simulation algorithm, the state changes only at discrete time instants when reactions fire. The state between two reaction firings is a constant.

## 2.3 Simulation Output Analysis

SSA is developed from the fundamental premise of the reaction propensity in Definition 2.2, so the trajectory obtained by a SSA run represents a *possible* realization of the grand probability  $\mathbb{P}\{\mathbf{x}, t | \mathbf{x}_0, t_0\}$ . In order to have a reasonable statistical estimation of the grand probability, many independent runs, in which each run starts with the same initial conditions, should be performed.

In this section, we present two techniques, the confidence interval estimation (Section 2.3.1) and the probability distribution estimation (Section 2.3.2), for analyzing statistical properties of simulation trajectories produced by SSA runs.

### 2.3.1 Confidence Interval Estimation

Let  $K$  be the number of simulations and let  $X^r$  with  $r = 1, \dots, K$  be a realization of the state  $X$  obtained at time  $t$  by the  $r$ th independent run of SSA under the same simulation conditions. The statistical properties (*e.g.*, mean and variance) can be derived from the ensemble of  $K$  trajectories and these properties are ensured to approach the exact solution of CME as  $K \rightarrow \infty$ .

Let  $\langle X \rangle$  be the sample mean and  $s^2$  be the (unbiased) sample variance of state  $X$  based on an ensemble of  $K$  independent simulations. They are computed as:

$$\langle X \rangle = \frac{\sum_{r=1}^K X^r}{K}$$

and

$$s^2 = \frac{\sum_{r=1}^K (X^r - \langle X \rangle)^2}{K - 1}.$$

By the law of large numbers, the sample mean and variance will asymptotically approach the mean  $\mathbb{E}[X]$  and variance  $\text{Var}[X]$  of the random variable  $X$  when  $K$  tends to infinity:

$$\begin{aligned}\mathbb{E}[X] &= \lim_{K \rightarrow \infty} \langle X \rangle, \\ \text{Var}[X] &= \lim_{K \rightarrow \infty} s^2.\end{aligned}$$

The number of simulation runs  $K$ , however, is often limited in practice. Thus, the convergence of the estimation is measured by the size of the confidence interval

$$d = \frac{zs}{\sqrt{K}} \quad (2.17)$$

where  $z$  is a specified *confidence level*, denoting the percentage of the range of estimated values that can be expected to include the true value. If the confidence level  $z$  is fixed, the probability that the mean  $\mathbb{E}[X]$  lies in the interval  $[\langle X \rangle - d, \langle X \rangle + d]$  is  $2\Phi(z) - 1$  where  $\Phi$  is the cumulative distribution function (cdf) of the standard normal distribution  $\mathcal{N}(0, 1)$ .

#### Example 2.6: Calculating confidence interval

Suppose choosing  $z = 1.96$ , the confidence level is  $2\Phi(z) - 1 \approx 0.95$ . Therefore, the probability that the mean falls in  $[\langle X \rangle - 1.96s/\sqrt{K}, \langle X \rangle + 1.96s/\sqrt{K}]$  is 95%.

Eq. (2.17) also suggests an estimation for the required number of simulation runs  $K$  to achieve a specified confidence interval size  $d$ . In particular, it can be computed as

$$K = \frac{z^2 s^2}{d^2}. \quad (2.18)$$

The number of simulation runs  $K$  in Eq. (2.18) shows two important facts. First,  $K$  reciprocally depends on the square of the confidence interval size  $d$ . In other words, to reduce the confidence interval size by a half, the number of simulation runs must be increased four times. Second,  $K$  depends on the sample variance  $s^2$ , which is unknown. Therefore, Eq. (2.18) cannot be implemented directly. One approach to circumvent the difficulty is first performing a small number of trial runs to estimate  $s_{trial}^2$ . Then, this value is applied to compute the number of simulation runs by

$$K = \frac{z^2 s_{trial}^2}{d^2}.$$

### 2.3.2 Probability Distribution Estimation

For biochemical reaction networks that expose *bistability*, the simplest form of multistability where two separated stable equilibrium points are separated by unstable equilibrium, the average population of species might not provide enough information for their dynamical behavior. In this case, the probability distribution must be used to quantitatively analyze the simulation results. The probability distribution can be estimated by using the *histogram* (or empirical distribution function) of the samples. The histogram is ensured to converge to the exact probability distribution given a large number of simulation runs  $K$ .

The calculation of the histogram in the following derivation assumes the state to be a scalar value, but it could be extended for the general case. To calculate the histogram, the state  $X$  at time  $t$  obtained by  $K$  simulation runs of SSA is supposed to be bounded into an interval  $[X_{min}, X_{max}]$ . Note that the interval  $[X_{min}, X_{max}]$  can be chosen arbitrarily. The interval then is divided into  $B$  bins in which the  $I_i$  bin,  $i = 1, \dots, B$ , defines as a subinterval  $[X_{min} + \frac{(i-1)L}{B}, X_{min} + \frac{iL}{B}]$  where  $L = X_{max} - X_{min}$ .

The histogram  $h_X$  of the state  $X$  is defined as

$$h_X(I_i) = \frac{B}{KL} \sum_{r=1}^K \chi(X^r, I_i)$$

where  $X^r$ ,  $r = 1, \dots, K$ , is the realization of  $X$  by  $r$ th simulation and the function  $\chi(X^r, I_i)$  is defined as

$$\chi(X^r, I_i) = \begin{cases} 1, & \text{if } X^r \in I_i \\ 0, & \text{otherwise.} \end{cases}$$

The histogram  $h_X(I_i)$  therefore gives the average probability of  $X$  in interval  $I_i$ .

Let  $p_X$  be the probability distribution of the state  $X$ . In the limit case when the number of simulation runs  $K \rightarrow \infty$  and the number of bins  $B \rightarrow \infty$ ,  $I_i$  reduces to a point and hence  $h_X$  converges to the probability distribution  $p_X$  at this point. Formally, it gives

$$p_X = \lim_{K, B \rightarrow \infty} h_X.$$

### 2.3.3 Illustrative Examples

The examples here are used to highlight the importance of stochasticity and to demonstrate the ability of stochastic simulation. Example 2.7 experimentally verifies the consistency of simulation trajectories obtained by SSA in comparison with the analytical solution of CME in Example 2.5. Example 2.8 shows that the behavior of a biological system can be drastically changed due to stochasticity. Finally, Example 2.9 demonstrates the ability of SSA to produce the bistability of the Schlögl model.

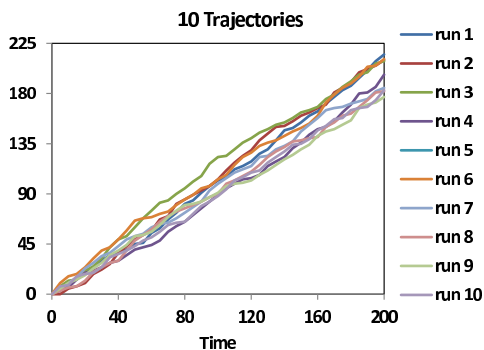


**Example 2.7: Simulation of the Birth process**

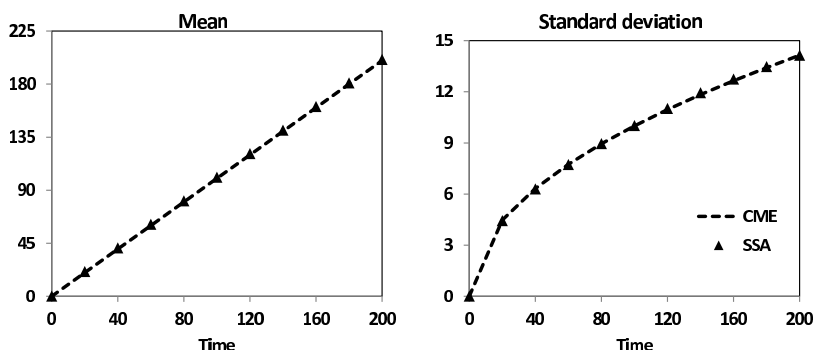
This example continues Example 2.5. It applies SSA to simulate the birth process in Appendix A.1 and compares its simulation results with the exact solution in Example 2.5. In order to simulate with SSA, the stochastic rate constant of the synthesis reaction is set to  $c = 1$  and the simulation time is set to  $T_{max} = 200$ .

Fig. 2.4 shows the trajectories of 10 SSA simulation runs. The figure shows that each simulation run produces a possible realization of the birth process model. Due to the stochastic nature of SSA, the population of S species at a particular time fluctuates in an interval rather than being a fixed value. In particular, as shown in Fig. 2.4, the population of species S at time  $t = 200$  spans from 175 to 217.

**Fig. 2.4** Trajectories of 10 SSA simulation runs of the birth process model with stochastic rate  $c = 1$  and simulation time  $T_{max} = 200$ .



We then compute the confidence interval of the number of S molecules produced by the birth process by performing 10,000 independent simulation runs. The sample mean and sample variance of the population of S at time  $T_{max} = 200$  are 200.93 and 199.98, respectively. The confidence interval of the population of S at time  $T_{max} = 200$  with 95% confidence level is thus  $[200.93 \pm 0.277]$ . The simulation results of SSA are compared against CME. It is shown in Example 2.5 that the mean of population of S at a particular time  $t$  is  $\mathbb{E}[n] = t$  and its standard deviation is  $\sigma = \sqrt{\text{Var}[n]} = \sqrt{t}$ . Fig. 2.5 depicts the mean and standard deviation, which is the square root of variance, by SSA and CME. The results depicted in Fig. 2.5 show a strong agreement between SSA and CME.

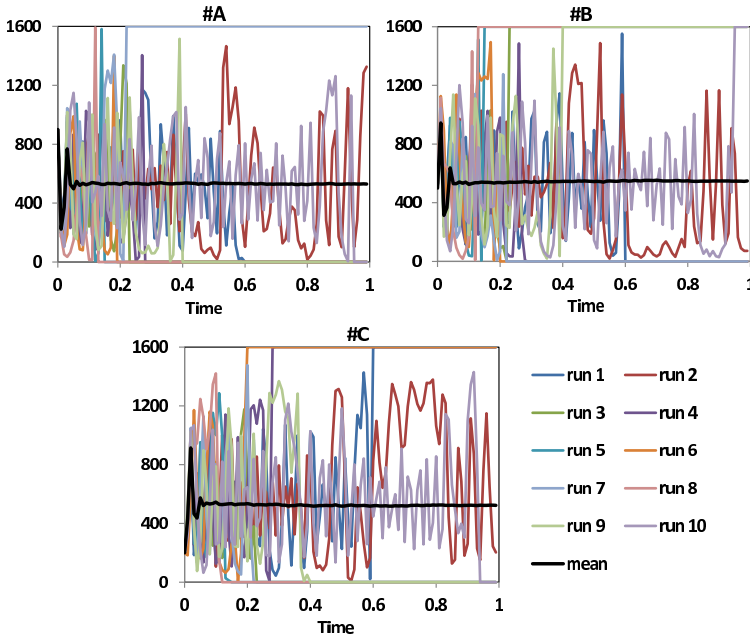


**Fig. 2.5** Mean and standard deviation of the number of S molecules produced by the birth process by CME and SSA. The sample mean and standard deviation by SSA is estimated by 10,000 simulation runs.

### Example 2.8: Simulation of the Oscillator model

This example applies SSA to simulate the Oscillator model described in Appendix A.3. We performed 10,000 independent runs of SSA, each with  $T_{max} = 1$ . Fig. 2.6 shows the population of each species through 10 simulation runs and its sample mean estimated by 10,000 independent simulation runs. The stochastic simulation of the Oscillator model obviously exhibits much more realistic behavior than the mean population of each species.

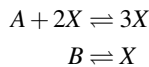
The stochastic change in the population of each species shown in Fig. 2.6 is significantly different from its average value. The average population of each species becomes stable, after a short fluctuation at beginning, while the population of each species significantly changes for each individual simulation run. For example, in the top left of Fig. 2.6 the average population of species A after the short transient time is kept around 530. The average population of species A, however, is significantly different from its population obtained from the stochastic simulations. Because of the inherent randomness in the SSA simulation, the population of species A by each individual simulation run may reach its maximum or degrade to zero. Note that the Oscillator model is closed; the total number of molecules during the simulation is conserved, *i.e.*,  $\#A(t) + \#B(t) + \#C(t) = \text{constant}$  for all time  $0 \leq t \leq T_{max}$ . The total number of molecules in the example is 1,600, hence the maximum number of molecules of species A is 1,600. The fluctuation interval for the population of species A is between  $0 \leq \#A(t) \leq 1,600$  for all time  $0 \leq t \leq T_{max}$ . The simulation stops if the population of a species reaches zero. The time at which the population of a species becomes zero is called the *extinction time*, which is a key issue in understanding the persistence and viability of the species. Because the mean population of species A is significantly larger than zero, the mean population may lead to misleading conclusion of the extinction time. In contrast, by considering probabilistic changes in the population of a species, SSA is able to quantify the distribution of extinction time of the species. This is one of the features that highlights the usefulness of SSA.



**Fig. 2.6** SSA simulation of the Oscillator model with simulation time  $T_{max} = 1$ . The black line shows the sample mean estimated over 10,000 independent runs.

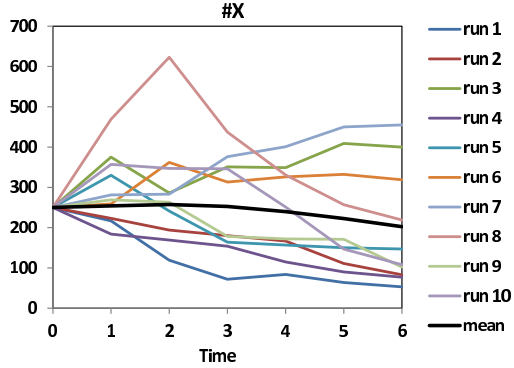
### Example 2.9: Simulation of the Schlögl model

This example demonstrates the ability of SSA to reproduce the *bistability* of the Schlögl model described in Appendix A.4. The model contains four reactions:

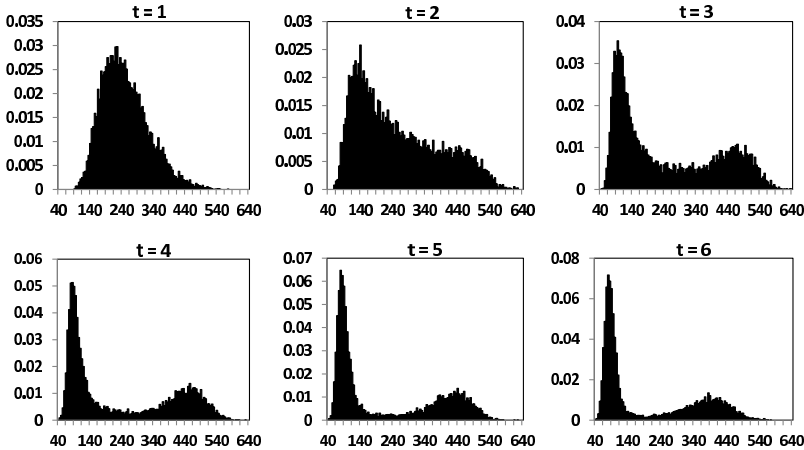


We performed 10,000 independent runs of SSA, each with  $T_{max} = 6$ , on the Schlögl model. Fig. 2.7 shows the trajectories of 10 SSA simulation runs. We only plot the population of species X because the populations of species A and B are large and are assumed to remain essentially constant over the simulation time. As shown qualitatively in the figure, the population of species X is roughly separated into two equilibrium parts. The mean population of X does not provide enough information to represent the behavior of the Schlögl model.

**Fig. 2.7** The population of species X by 10 independent SSA runs of the Schlögl model with simulation time  $T_{max} = 6$ . The black line shows the mean population of X through 10,000 independent simulation runs.



To quantitatively analyze the model, we approximate the probability distribution of species X by computing its histogram. The histogram is calculated by bounding the population of species X into the interval  $[40, 640]$  with the bin size of 5. Fig. 2.8 shows the histogram of population of species X at different times  $t$  by using  $K = 10,000$  SSA simulation runs. The figure quantitatively shows the Schlögl model exhibiting bistability. For example, at time  $t = 5$ , the model has two separated stable states. For the first stable state, the population of species X fluctuates around 70 and 100 with probability around 0.06. For the second stable state, the population of species X fluctuates around 380 and 460 with probability around 0.01.



**Fig. 2.8** Histogram of species X at different time calculated by 10,000 independent SSA runs on the Schlögl model. The x-axis is the interval of population of species X with bin size 5. The y-axis is the probability that the population of X is in a bin.

## 2.4 Summary

This chapter presented the theoretical foundations of stochastic chemical kinetics for modeling biochemical reaction systems. The occurrence of a reaction in the stochastic chemical kinetics is a random event where its probability is proportional to the reaction propensity. Based on this fundamental premise, the time evolution of reaction networks is exactly described by the chemical master equation. The chapter also presented the mathematical basis for the development of stochastic simulation algorithms for realizing temporal behavior of biochemical reactions. A class of exact simulation strategies originating from the stochastic simulation algorithm (SSA) is described. The algorithm is a discrete event simulator where a reaction is selected to fire according to a probability that is consistent with the chemical master equation. The chapter then introduced the concepts of confidence interval and histogram for statistically analyzing the outcome of simulation realizations. Finally, the section on numerical examples was used to introduce the nice features of stochastic simulation.

## 2.5 Further Reading

The seminal works on the stochastic modeling of biochemical reactions are investigated by McQuarrie [182] and subsequently by Gillespie [92, 93]. The rigorous derivation of the chemical master equation, also known as the forward Chapman-Kolmogorov equation in the context of a continuous-time Markov process, for the stochastic modeling of biochemical reactions from the mathematical principles of probability is developed by Gillespie in [96]. The mathematical background for stochastic modeling and simulation can be accessed from the books of Gillespie [95], Van Kampen [135], Wilkinson [276], Ullah and Wolkenhauer [266]. The book by Gardiner [89] presents a collection of methods for numerically solving the chemical master equation. Recent numerical methods for solving the chemical master equation have been developed that either analytically solve CME by limiting the biological network to unimolecular reactions [130] or numerically approximate CME by the linear noise approximation [75], the moment-closure approximation [91, 152], the finite state projection method [190, 191, 40, 244, 236], the sliding window method [114, 66, 277, 20] and the tensor approach [137, 67]. The stochastic simulation algorithm was first developed by Gillespie [92, 93], thus it is also called Gillespie's algorithm. Bortz *et al.* [36] also developed an algorithm, called *N*-fold way, that is similar to the Gillespie's algorithm but in the context of kinetic Monte Carlo. The application of the stochastic algorithm to understand the stochastic effects in the gene regulation is discussed in Arkin *et al.* [179, 21, 180]. Various applications of the stochastic simulation for dynamical systems are demonstrated in [100, 101, 132, 107]. Additional materials for stochastic simulation and analysis are in Bower and Boilouri [37], Szallasi *et al.* [245], Cao and Petzold [55], Asmussen and Glynn [23], Stumpf *et al.* [241], Anderson and Kurtz [16], and Priami and Morine [206].

Stochastic chemical kinetics have been widely adopted for modeling and simulating biochemical reactions where species are present at relative low copy numbers. Their assumptions are, however, often restricted to living cells. For example, molecular species in living cells are often localized, which is referred to as *molecular crowding*, to enhance species availability and to speed up cellular operations. The high concentration of macromolecular species like proteins is the key to explain the *excluded volume effect* in the cell. Validity conditions for stochastic simulation in this case are discussed in [104]. Inhomogeneous SSA (ISSA) is an extension of SSA to cope with the case when the well-mixed volume assumption is violated. ISSA divides the cell volume into well-mixed subvolumes and the diffusion of a species is explicitly modeled as a unimolecular reaction. ISSA can be accessed through the work of Stundzia and Lumsden [242], Bernstein [32], Elf [74, 76], and others [115, 68, 69, 116, 141]. It is also worth mentioning Smoldyn and Green's function reaction dynamics (GFRD), which are alternatives for spatial stochastic simulation. The Smoldyn approach is developed by Andrews and Bray [19] and further extended by Andrews *et al.* [17, 18]. The Green's function reaction dynamics approach is proposed in [269, 268, 196, 248, 88]. Another assumption of stochastic chemical kinetics that could be considered restrictive is that reaction firing is assumed to be instantaneous even if it requires a certain amount of time in living cells. Time delays in this case could be explicitly taken into account in order to accurately describe the system dynamics. The delayed SSA (DSSA) has been introduced to cope with delays by updating the state and the propensities as soon as a delayed reaction is scheduled to finish. Delayed stochastic simulation can be obtained from the work of Cai [43] and others [38, 28, 9, 26, 220, 252, 259].

Simulation Algorithms for Computational Systems  
Biology

Marchetti, L.; Priami, C.; Thanh, V.H.

2017, XI, 238 p. 52 illus., 23 illus. in color., Hardcover

ISBN: 978-3-319-63111-0