Stony Brook University

MOMENT PROPAGATION METHODS FOR STOCHASTIC SIMULATION OF COMPLEX BIOCHEMICAL SYSTEMS

A Dissertation Presented
by

Vibha Mane

to
The Graduate School
in Partial Fulfillment of the
Requirements
for the Degree of

Doctor of Philosophy
in
Electrical Engineering
Stony Brook University

May 2011

Stony Brook University
The Graduate School

Vibha Mane

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Petar M. Djurić, Dissertation Advisor
Professor, Department of Electrical & Computer Engineering

Murali Subbarao, Chairperson of Defense
Professor, Department of Electrical & Computer Engineering

Mónica F. Bugallo, Member of the Committee
Associate Professor, Department of Electrical & Computer Engineering

Sanford Simon, Outside Member
Professor, Department of Biochemistry and Cell Biology

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

MOMENT PROPAGATION METHODS FOR STOCHASTIC SIMULATION OF COMPLEX BIOCHEMICAL SYSTEMS

by

Vibha Mane

Doctor of Philosophy

in

Electrical Engineering

Stony Brook University

2011

We are interested in predicting the time dependent behavior of *biochemical networks* such as interaction between proteins. These networks are represented by a system of chemical reactions. In the *forward problem*, we have knowledge of all the reactions and the associated rate constants. We want to determine the joint probability density of the populations of all the molecular species at any time instant.

The given biochemical system is a discrete state, continuous time *Markov process*, and the time evolution of its probability density function is described by the *chemical master equation* (CME). We want to obtain the solution of the master equation in *complex biochemical systems* with a large number of species and reactions. *Analytical solutions* of the chemical master equation for first and second order reactions have only been obtained for select cases with a few species and reactions. Another approach to solving the problem is

to approximate the CME with the *Fokker-Planck* equations. But this would require solving partial differential equations with a large number of variables.

The present *state of the art* approaches for stochastic simulation of such systems are based on *Monte Carlo methods*. One such popular method is the Stochastic Simulation Algorithm (SSA) derived by Gillespie in 1976. Several authors have developed *accelerated* versions of the SSA such as the Next Reaction Method and the τ-leap methods, in order to reduce the computation time of SSA. The Monte Carlo methods provide approximations of the complete distribution, but they require simulations of many realizations of the Markov process and many time steps. Hence their computation times are prohibitively long for very complex systems.

Methods for modeling the biochemical networks based on *moment propagation* is a relatively unexplored area. We propose a new method for propagating the first two moments of the joint probability distribution of the number of molecules. In many systems, the distribution can be approximated as Gaussians and therefore computing the first two moments is sufficient. Simulation results show that our method yields accurate results for first order and second order reactions. Compared with the Monte Carlo methods, our method yields *significant savings in computation time*. Compared with other moment propagation methods, the *recursive expressions* in our method can be implemented by specifying rate constants and stoichiometries, without having to derive or solve any differential equations. Whereas other moment propagation methods with similar accuracy have been demonstrated for a few species, we demonstrate our method for complex biochemical systems with hundreds of species.

Contents

List of Figures

List of Tables

Acknowledgments

It has been my great privilege to work with my advisor Prof. Petar Djurić, an outstanding researcher and teacher. I learned from him how to develop new ideas. I hold him in highest esteem. He made it possible for me to enjoy many aspects of my stay at Stony Brook University including research, writing papers, learning topics outside of my research, teaching and an invigorating environment provided by my many students.

I am grateful to Prof. Sanford Simon for elucidating some interesting biology applications of this work. I want to express my deep gratitude to Prof. Mónica Bugallo for our discussions and her continued support. I would also like to thank Prof. Murali Subbarao, for serving on my advisory committee and for his helpful comments.

I have been very delighted to be a part of the COSINE lab with so many wonderful colleagues including Ting Lu, Akshay Athalye, Jae-Chan Lim, Mahesh Vemula, Yao Li, Zejie Zhang and Xueying Zhang who were greatly missed upon their graduation, Subodh Chiwate, Yunlong Wang, Shishir Das, Cagla Tasdemir, Bingxin Shen, Li Geng, Mingyi Hong, Pingyi Xiong and Kevin Mernick for contributing to an enriched atmosphere. This was further enhanced by visitors to our lab like Pau Closas. Our Journal Clubs provided a forum for intellectual stimulation and exchange of thoughts, made exuberant by the active participation of many members of our lab.

I want to express my love for my husband Sateesh Mane who has been very supportive throughout my doctoral endeavor, my son Rahul for his enthusiasm and interest in my work and my daughter Tarika who I hope to inspire with my work. My parents Sanval Kataria and Sheela Kataria have always encouraged me in my pursuit of knowledge.

Chapter 1

Introduction

1.1 Introduction

Molecular and systems biologists aim to develop a better understanding of complex interactions in biological systems. These include interactions between DNA, RNA and proteins in a gene regulatory system, enzyme-substrate interactions in metabolic pathways and interactions between proteins in cell signaling pathways. We start the modeling of a biological process by representing it with a system of chemical reactions. We next apply a computational model to predict the behavior of the system. In the *deterministic framework*, we have a set of reaction rate equations with the state variable being the concentration of molecular species. The deterministic approach adequately models many macroscopic systems. The underlying assumption in this approach is that the molecular populations are very large and can be approximated as continuously varying quantities. However, in intracellular biochemical reactions molecular populations of some species are very small, and the deterministic approach fails to predict the system behavior quantitatively or qualitatively. It is well known that gene expression is a random phenomenon due to fluctuations in transcription, translation, and degradation of proteins and messenger RNA. This has been experimentally observed by Novick et al. [56], and has been described by Arkin and many others [2], [52], [53], [60]. Stochasticity in gene expression can give rise to phenotypically distinct subgroups with genetically identical cells [42].

In the *stochastic framework*, the state variable is a random vector representing the number of molecules of each species. Given the initial state of the system, the reaction

model and the associated rate constants, we want to obtain the joint *probability mass function* (PMF) of the state of the system at any later time. The time evolution of this joint PMF is described by the *chemical master equation* (CME), a differential-difference equation [25], [76]. In the *analytical approach*, the probability generating function is employed to transform the master equation into a *partial differential equation* (PDE) [54], [55]. Another approach is to utilize the Laplace transform to transform the master equation into a system of linear equations [46]. The analytical solution of the master equation has been obtained for select cases, with very few species and this will be described in Chapter 2. In the continuous stochastic formulation, the evolution of the joint *probability density function* (PDF) of the state of the system is described by the *stochastic differential equation* (SDE) such as the Fokker-Planck equation or the Langevin equation [28], [76].

The *stochastic simulation algorithm* (SSA), which is a *Monte Carlo*-based method was first introduced by Doob in 1945 [12], and derived by Gillespie in 1976 for the biochemical reaction system [22]. In this method, we generate many realizations of the stochastic process using random sampling, and average over these realizations to obtain the PMF of the state of the system. However, the SSA simulates only one reaction per time step, and hence it is not scalable with the population of molecular species, number of species or the number of reactions. Several authors have developed accelerated versions of the SSA such as the τ-leap methods [9], [27], [58], [75]. The Monte Carlo methods will be discussed in Chapter 3.

Quite often, we are interested in the lower order moments of the PMF, such as the means and the covariance. In the *moment propagation method*, we trace a few lower order moments of the joint PMF of the state of the system. In 1995, Gillespie used the propagator function of a Markov process to derive moment evaluation equations [24]. In 2005, Hespanha et al. [38] and Singh et al. [67] presented a *Stochastic Hybrid Systems (SHS)* method to compute the first few moments of the state vector. In 2007, Gómez-Uribe et al. proposed a *Mass Fluctuation Kinetics (MFK)* method and obtained differential equations to estimate the means, variances and the covariances [32], [33]. These methods are evaluated in Chapter 6.

In the proposed method, a moment propagation method, our goal was to compute the evolution of the first two moments of the state vector for very large systems. We have utilized conditional expectation relations of random variables, Markovian property of the biochemical system, and expressions for simple reactions to build a general recipe for complex systems. We have demonstrated that the proposed method can be implemented

very easily by specifying rate constants and stoichiometries, without having to derive or solve any differential equations [50], [51]. The method is presented in Chapter 4, for first order reactions and Chapter 5 for second order reactions. Its comparison with the Monte Carlo methods and other moment propagation methods is provided in Chapter 6.

In the remainder of this chapter, we describe some biology examples where the stochastic simulation of molecular populations will be an additional tool to help comprehend such problems. Some of these examples are simulated in this research. We also discuss the deterministic method and introduce the stochastic approach.

1.2 Chemical Reactions

A chemical reaction such as

$$\nu_{11}S_1 \ + \ \nu_{12}S_2 \ \xrightarrow{k_1} \ \nu_{13}S_3 \ + \ \nu_{14}S_4 \tag{1.1}$$

states that ν_{11} molecules of type S_1 react with ν_{12} molecules of type S_2 to form ν_{13} molecules of type S_3 and ν_{14} molecules of type S_4 [71]. The coefficients ν_{11}, ν_{12}, ν_{13}, and ν_{14} are called *stoichiometric coefficients* of the reaction. In an *elementary reaction*, the above reaction occurs in a single step. In this report, a reaction written as in equation (1.1) will indicate an elementary reaction. The *reaction rate* tells us how fast a reaction takes place.

The *rate equation* relates the reaction rate to the concentration of reactants. For the chemical reaction in (1.1), the rate equation is

$$\mathsf{r} = k_1 \, [S_1]^{\nu_{11}} \, [S_2]^{\nu_{12}} \tag{1.2}$$

where r is the reaction rate, k_1 is temperature dependent *reaction coefficient*, and $[S_1]$ and $[S_2]$ are the concentrations of the reactants S_1 and S_2, respectively. The units of concentration are mole litre^{-1} or mole l^{-1}. We also refer to the reaction coefficient k_1 as *reaction constant* or *rate constant*, although it may be time dependent in the computational models described henceforth.

For elementary reactions, we define *reaction order* as the sum of the powers of the concentrations in the reaction rate. The reactions are categorized as zeroth, first or second order reactions. Reactions of order three or more are very rare and may be modeled as a

sequence of second order reactions [30].

In the zeroth order reaction

$$\emptyset \quad \xrightarrow{k_1} \quad S_1$$

the reaction rate is independent of the concentration of the reactants and therefore for modeling purposes there are no reactants. Its rate equation is given by

$$\mathsf{r} = k_1$$

and the units of the reaction constant k_1 are $\sec^{-1}$ mole litre^{-1}.

In the first order reaction

$$S_1 \quad \xrightarrow{k_2} \quad S_2$$

the reaction rate is proportional to the concentration of only one reactant, namely S_1, and the rate equation is

$$\mathsf{r} = k_2[S_1].$$

The units of the reaction constant k_2 are $\sec^{-1}$.

In the second order reaction

$$S_1 \; + \; S_2 \quad \xrightarrow{k_3} \quad S_3$$

the reaction rate is proportional to the product of the concentration of two reactants, namely S_1 and S_2 and the rate equation is

$$\mathsf{r} = k_3[S_1][S_2].$$

The units of the reaction constant k_3 are $\sec^{-1}/(\text{mole litre}^{-1})$. For a reaction of order n, the units of reaction constant are given by $\sec^{-1}/\left(\text{mole litre}^{-1}\right)^{n-1}$.

1.3 Applications

1.3.1 Example - Viral Kinetics

Srivastava et al. described a simplified model of intracellular viral infection, where a single virus infects its host cell [73]. The infection process includes transcription, translation, genome replication, assembly and virus release. It is characterized by the following six reactions:

$$R_1 : \ DNA \xrightarrow{c_1} mRNA \tag{1.3a}$$

$$R_2 : \ mRNA \xrightarrow{c_2} DNA + mRNA \tag{1.3b}$$

$$R_3 : \ mRNA \xrightarrow{c_3} Pr + mRNA \tag{1.3c}$$

$$R_4 : \ mRNA \xrightarrow{c_4} \emptyset \tag{1.3d}$$

$$R_5 : \ Pr \xrightarrow{c_5} \emptyset \tag{1.3e}$$

$$R_6 : \ DNA + Pr \xrightarrow{c_6} Vr. \tag{1.3f}$$

The authors studied the viral nucleic acids, namely DNA and $mRNA$ and the viral structure proteins (denoted by Pr in the above reactions). The DNA can either transcribe into an $mRNA$ (reaction R_1) or it may be packaged into structural proteins to form progeny virus (denoted by Vr in reaction R_6). The $mRNA$ molecule is used as a template to duplicate the DNA (reaction R_2) and also translate to protein (reaction R_3). In reactions R_4 and R_5, the $mRNA$ and the protein molecules are degraded, respectively.

1.3.2 Example - Gene Regulatory Network

Gene expression is a complex process consisting of transcription of a gene into an mRNA and its translation into a protein [1]. Each of the transcription and translation processes consist of a multitude of elementary reactions, and therefore cannot be modeled as instantaneous reactions. Roussel el al. and Ribeiro et al. modeled the gene expression process in a single step, using time delays [63], [65]

$$RNAP(t) \ + \ Prom(t) \ \longrightarrow \ RNAP(t+\tau_1) + Prom(t+\tau_2) + K \times Pr(t+\tau_3), \tag{1.4}$$

where $RNAP$ is the RNA polymerase enzyme which binds to the promotor site to initiate gene transcription, $Prom$ is the promotor site, Pr is the resultant protein created from the translation of $mRNA$, K is the number of copies of protein created in the process and τ_1, τ_2 and τ_3 are the times required for each of the resultant products to become available. A gene regulatory network consists of transcription factors and target genes which interact with each other to determine the level of $mRNA$ and proteins created. Kierzek et al. and Ramsey et al. have performed stochastic modeling of gene regulatory networks using the SSA [44], [59].

1.4 Computational Models

Various modeling approaches for biochemical networks are discussed in several texts [5], [15], [78] and review papers [11], [17], [31]. The modeling approach we choose depends on the level of detail required, time scales, molecular populations, the complexity of the reaction system and so forth [5]. The deterministic Boolean models require synchronous updating and are not suitable for reactions with considerably different timescales [17]. The stochastic Petri nets are a graphical representation of the biochemical networks, and allow qualitative analysis of the reaction model [34], [61]. The most pervasive approaches are the Monte Carlo-based approaches such as the SSA and the τ-leap methods to reduce the computation times of the SSA.

We will now describe the deterministic approach, the stochastic approach using the chemical master equation and the stochastic approximation such as the Fokker-Planck and the Langevin equation.

1.4.1 Deterministic Approach

In the deterministic approach, the state variable is the concentration of all the molecular species. These concentrations are continuous variables. From the reaction system, we obtain a system of coupled *ordinary differential equations* (ODEs), with one equation per molecular species. We will also refer to these as *reaction rate equations* (RREs).

Example - A First Order Reaction

For the first order reaction

$$S_1 \xrightarrow{k_2} S_2 \tag{1.5}$$

we have the following system of equations:

$$\frac{d[S_1]}{dt} = -k_2[S_1]$$

$$\frac{d[S_2]}{dt} = k_2[S_1].$$

Example - A Second Order Reaction

For the second order reaction

$$S_1 + S_2 \xrightarrow{k_3} S_3 \tag{1.6}$$

we have the following system of equations:

$$\frac{d[S_1]}{dt} = -k_3[S_1][S_2]$$

$$\frac{d[S_2]}{dt} = -k_3[S_1][S_2]$$

$$\frac{d[S_3]}{dt} = k_3[S_1][S_2].$$

Example - A Reversible First Order Reaction

For the reversible reaction

$$S_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} S_2$$

we have the following system of equations:

$$\frac{d[S_1]}{dt} = -k_1[S_1] + k_2[S_2]$$

$$\frac{d[S_2]}{dt} = k_1[S_1] - k_2[S_2].$$

In complex systems, the above reaction rate equations, together with the initial conditions, can be solved numerically using the Euler method or the Runge-Kutta method.

In the deterministic framework we assume that the number of molecules are sufficiently high so that the discrete changes of molecular populations may be approximated by continuous changes in the concentration. However, as discussed in Section 1.1, under certain conditions, the deterministic approach will produce incorrect values for the molecular concentrations. We now give two examples comparing the deterministic method with the SSA. As mentioned earlier, SSA is a Monte Carlo-based method and it will be described in detail in Chapter 3. The stochastic parameters used in these examples are defined in Section 1.4.2. In the first example, the stochastic averages agree with the deterministic values. In the second example, the stochastic method produces a pronounced oscillatory behavior which is not predicted by the deterministic method.

Example - Deterministic Competing Reactions

Consider the following reaction system:

$$R_1 : \; S_1 \; + \; S_2 \quad \xrightarrow{k_1} \quad S_3 \tag{1.7a}$$

$$R_2 : \; S_1 \; + \; S_4 \quad \xrightarrow{k_2} \quad S_5. \tag{1.7b}$$

The RREs were solved using the Euler method. The initial concentrations were $[S_1] = 1000\,\text{molecules l}^{-1}$, $[S_2] = 1000\,\text{molecules l}^{-1}$ and $[S_4] = 2000\,\text{molecules l}^{-1}$. The reaction constants were $k_1 = 0.001\,\text{sec}^{-1}/(\text{molecules l}^{-1})$, $k_2 = 0.001\,\text{sec}^{-1}/(\text{molecules l}^{-1})$ and plots were generated until time $t = 1.0\,\text{sec}$. We set the volume $\mathsf{V} = 1\,\text{l}$. The initial number of molecules were $X_1(t=0) = 1000$, $X_2(t=0) = 1000$ and $X_4(t=0) = 2000$. The stochastic rate constants, to be defined in Section 1.4.2, were $c_1 = 0.001\,\text{sec}^{-1}$, $c_2 = 0.001\,\text{sec}^{-1}$. In the deterministic method, $\Delta t = 0.01\,\text{sec}$ and in the SSA method the results were averaged over 5 realizations. Figure 1.1 shows that the simulation results using the deterministic method compare well with the averages obtained using the SSA algorithm.

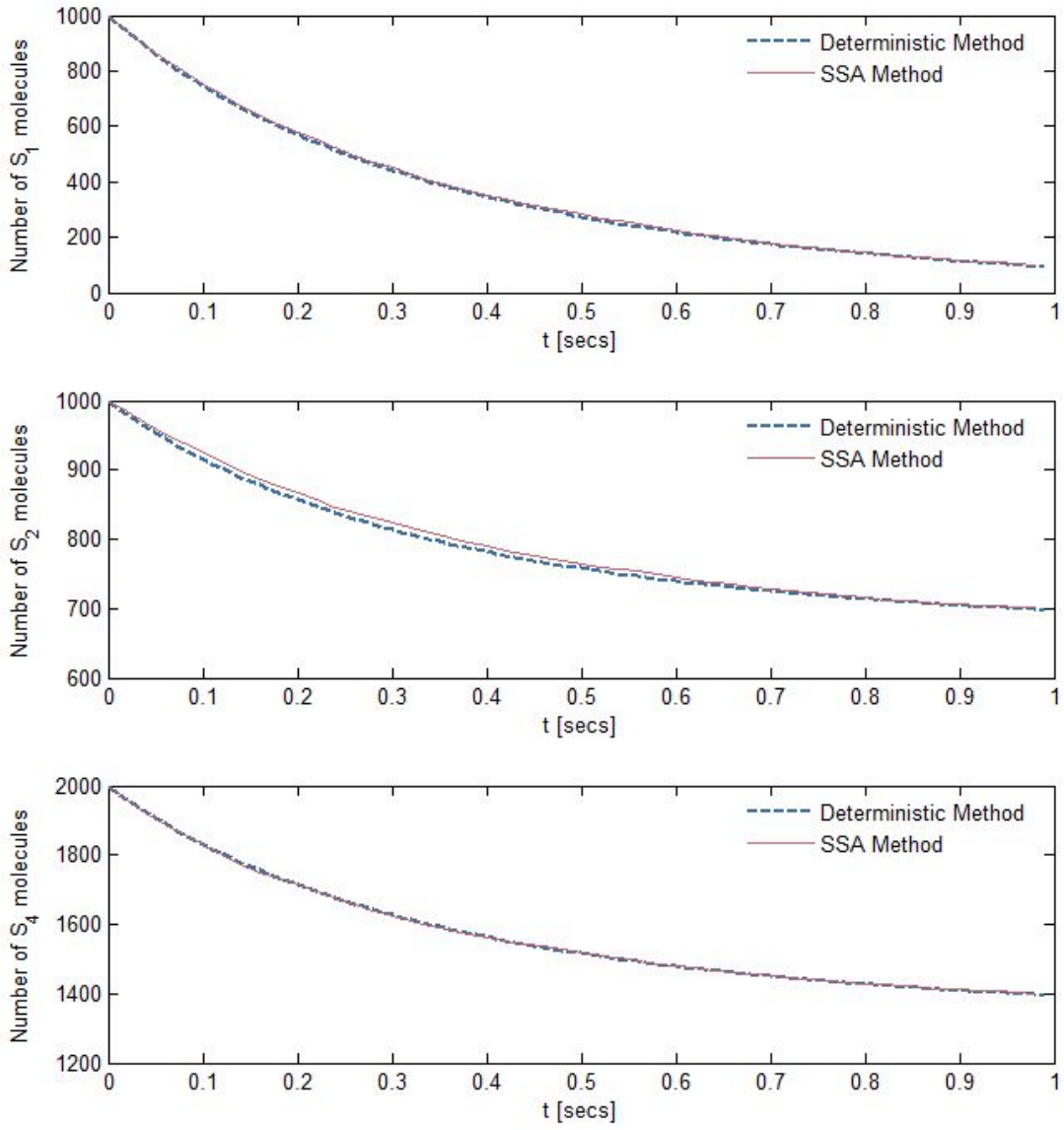

Figure 1.1: In the Competing Reaction system, simulation results using the deterministic method (solid line) compare well with the averages obtained using the SSA method (dashed line).

Example - Deterministic Lotka Reactions

In this example, we show that the results using the deterministic method do not agree with the averages obtained using the stochastic method. The Lotka reactions represent a predator-prey model proposed by Lotka in 1925 [48] and Volterra in 1930 [77] to study population dynamics

$$R_1 : \; S_1 \xrightarrow{k_1} 2S_1 \tag{1.8a}$$

$$R_2 : \; S_1 + S_2 \xrightarrow{k_2} 2S_2 \tag{1.8b}$$

$$R_3 : \; S_2 \xrightarrow{k_3} S_3. \tag{1.8c}$$

In this model, the prey S_1 feeds on food with constant supply and reproduces (reaction R_1), and the predator S_2 feeds on the prey S_1 and reproduces (reaction R_2). Reaction R_3 indicates the decline of the predator due to natural causes. The reaction rate equations are

$$\frac{d[S_1]}{dt} = k_1[S_1] - k_2[S_1][S_2] \tag{1.9a}$$

$$\frac{d[S_2]}{dt} = k_2[S_1][S_2] - k_3[S_2]. \tag{1.9b}$$

The steady-state solution obtained by setting

$$\frac{d[S_1]}{dt} = 0 \tag{1.10a}$$

$$\frac{d[S_2]}{dt} = 0 \tag{1.10b}$$

gives $S_{1ss} = k_3/k_2$ and $S_{2ss} = k_1/k_2$. That is, if we start with the initial values $S_1 = S_{1ss}$ and $S_2 = S_{2ss}$, then these values will remain constant. However, the simulations in Figure 1.2 using the SSA method show that the populations of S_1 and S_2 exhibit oscillatory behavior, with a rise in prey population followed by a rise in predator population [23]. The parameters used for these simulations were $[S_1] = 1000\,\text{molecules}\,\text{l}^{-1}$, $[S_2] = 1000\,\text{molecules}\,\text{l}^{-1}$, $k_1 = 10\,\text{sec}^{-1}/(\text{molecules}\,\text{l}^{-1})$, $k_2 = 0.01\,\text{sec}^{-1}/(\text{molecules}\,\text{l}^{-1})$ and $k_3 = 10\,\text{sec}^{-1}/(\text{molecules}\,\text{l}^{-1})$.

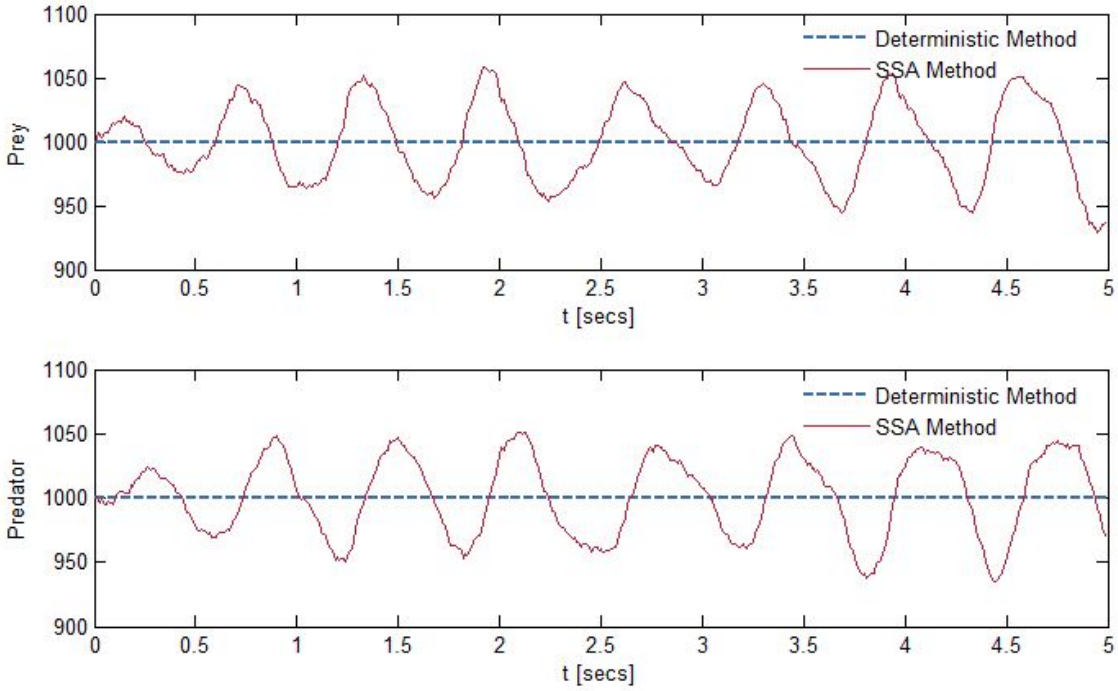

Figure 1.2: In Lotka reactions, the SSA method produces oscillatory behavior for the predator and the prey populations, whereas the deterministic values remain constant with time.

1.4.2 Stochastic Approach

In the stochastic approach, we want to determine the probability that the number of molecules of the species is in a given range. We define our biochemical system as a spatially homogeneous mixture of N molecular species S_1, S_2, ..., S_N inside a fixed volume V. The species interact through M reaction channels R_1, R_2, ... R_M. The state of the system is specified by the random vector $\boldsymbol{X}(t) = \begin{pmatrix} X_1(t) & X_2(t) & \dots & X_N(t) \end{pmatrix}^T$, where T indicates the vector transpose and $X_i(t)$, $i = 1, 2 \dots N$, is the number of molecules of species S_i at time instant t. Let $\boldsymbol{x}(t)$ be an observation of the random vector $\boldsymbol{X}(t)$, where $\boldsymbol{x}(t) = \begin{pmatrix} x_1(t) & x_2(t) & \dots & x_N(t) \end{pmatrix}^T$. Define l successive time indices $t_0 < t_1 < \dots t_l$, and let $P(\boldsymbol{x}(t_l))$ denote the joint probability mass function (PMF) of $\boldsymbol{X}(t)$ at any time t_l. The following argument also holds for continuous-state Markov process with $P(\boldsymbol{x}(t_l))$ being the joint probability density function (PDF). Let $P_{l|(0,1,\dots,l-1)}(\cdot)$ denote the conditional PMF at time t_l given its PMF at times $t_0, t_1, \dots, t_{l-1}$ and $P_{l|l-1}(\cdot)$ denote the conditional PMF at time t_l given its PMF at time t_{l-1}. Then $\boldsymbol{X}(t)$ is a discrete-state continuous-time stochastic process with the Markov property

$$\begin{aligned} &P_{l|(0,1,\dots,l-1)}(\boldsymbol{X}(t_l) = \boldsymbol{x}(t_l) | \boldsymbol{X}(t_0) = \boldsymbol{x}(t_0), \boldsymbol{X}(t_1) = \boldsymbol{x}(t_1) \dots, \boldsymbol{X}(t_{l-1}) = \boldsymbol{x}(t_{l-1})) \\ &= P_{l|l-1}(\boldsymbol{X}(t_l) = \boldsymbol{x}(t_l) | \boldsymbol{X}(t_{l-1}) = \boldsymbol{x}(t_{l-1})). \end{aligned} \tag{1.11}$$

That is, conditional on the present state, the future state is independent of the past states. The aforementioned Markov process is completely determined by an initial state $P_0(\boldsymbol{x}(t_0))$ and a one-step transition probability $P_{l|l-1}(\boldsymbol{X}(t_l) = \boldsymbol{x}(t_l) | \boldsymbol{X}(t_{l-1}) = \boldsymbol{x}(t_{l-1}))$. We want to determine the time evolution of the joint probability mass function $P(\boldsymbol{x}(t))$ of $\boldsymbol{X}(t)$. (We will also denote this probability function with $P(\boldsymbol{x}, t)$). This is described by the *master equation* and its general form is given by [26], [76]:

$$\frac{dP(\boldsymbol{x}, t)}{dt} = \sum_{\boldsymbol{x}'} \left[U_{\boldsymbol{x}'\boldsymbol{x}} P(\boldsymbol{x}', t) - U_{\boldsymbol{x}\boldsymbol{x}'} P(\boldsymbol{x}, t) \right], \tag{1.12}$$

where $U_{\boldsymbol{x}'\boldsymbol{x}}$ is the transition probability from state $\boldsymbol{x}'$ to state $\boldsymbol{x}$ and $U_{\boldsymbol{x}\boldsymbol{x}'}$ is the transition probability from state $\boldsymbol{x}$ to $\boldsymbol{x}'$. It is a gain-loss equation for the probability of the state $\boldsymbol{X} = \boldsymbol{x}$.

Next, we will define reaction parameters which will give a characterization of a reaction system. Consider the following reactions:

$$R_1 : \ \emptyset \ \longrightarrow \ S_1 \tag{1.13a}$$

$$R_2 : \ S_1 \ \longrightarrow \ S_2 \tag{1.13b}$$

$$R_3 : \ 2S_1 \ \longrightarrow \ S_2 \tag{1.13c}$$

$$R_4 : \ S_1 \ + \ S_2 \ \longrightarrow \ S_3. \tag{1.13d}$$

Let c_m be the stochastic rate constant for the reaction R_m. In the deterministic framework, the unknown quantities are the *molecular concentrations* and in the stochastic framework, the state variables are the *number of molecules*. For the four reactions in (1.13), the stochastic and the deterministic rate constants are related by:

$$c_1 \ = k_1 n_A \mathsf{V} \tag{1.14a}$$

$$c_2 \ = k_2 \tag{1.14b}$$

$$c_3 \ = k_3/(n_A \mathsf{V}/2) \tag{1.14c}$$

$$c_4 \ = k_4/(n_A \mathsf{V}), \tag{1.14d}$$

where n_A is the Avogadro number. The units of c_m are $\sec^{-1}$ for the four reactions. The reaction probabilities are constant for zeroth order reaction, proportional to the number of molecules for first order reactions and proportional to the product of the number of molecules for the second order reactions. Let $h_m(\boldsymbol{x}, t)$ be the number of distinct molecular combinations for reaction R_m and define *reaction propensity* to be $a_m(\boldsymbol{x}, t) \equiv c_m h_m(\boldsymbol{x}, t)$. Then the probability that an R_m reaction will take place in an infinitesimal time interval dt is given by $a_m(\boldsymbol{x}, t)dt$, for the system in state $\boldsymbol{x}$. For the four reactions in (1.13), $h_m(\boldsymbol{x}, t)$ is given by:

$$h_1(\boldsymbol{x}, t) = 1 \tag{1.15a}$$

$$h_2(\boldsymbol{x}, t) = x_1(t) \tag{1.15b}$$

$$h_3(\boldsymbol{x}, t) = x_1(t)(x_1(t) - 1)/2 \tag{1.15c}$$

$$h_4(\boldsymbol{x}, t) = x_1(t)x_2(t). \tag{1.15d}$$

We also define ν_{mi} to be the stoichiometry of species S_i in reaction R_m. Therefore, $\boldsymbol{\nu}_m = \begin{pmatrix}\nu_{m1} & \nu_{m2} & \dots & \nu_{mN}\end{pmatrix}$ is the state-change vector for reaction R_m, and for the four reactions in (1.13), the stoichiometries are given by:

$$\nu_{11} = 1 \tag{1.16a}$$

$$\nu_{21} = -1 \quad \nu_{22} = 1 \tag{1.16b}$$

$$\nu_{31} = -2 \quad \nu_{32} = 1 \tag{1.16c}$$

$$\nu_{41} = -1 \quad \nu_{42} = -1 \quad \nu_{43} = 1. \tag{1.16d}$$

Using the above description, the general form of the *chemical master equation (CME)* is expressed as:

$$\frac{dP(\boldsymbol{x},t)}{dt} = \sum_{m=1}^{M} \left[a_m(\boldsymbol{x}-\boldsymbol{\nu}_m,t)P(\boldsymbol{x}-\boldsymbol{\nu}_m,t) - a_m(\boldsymbol{x},t)P(\boldsymbol{x},t)\right], \tag{1.17}$$

where M is the total number of reactions in the system.

Example - Master Equation, Competing Reactions

A derivation of the master equation is given in [25] and [76]. We will derive the master equation for the system of competing second order reactions (1.7) which we reproduce here:

$$R_1 : \; S_1 \; + \; S_2 \; \xrightarrow{c_1} \; S_3 \tag{1.18a}$$

$$R_2 : \; S_1 \; + \; S_4 \; \xrightarrow{c_2} \; S_5. \tag{1.18b}$$

If we track all five species, $N = 5$, $\boldsymbol{X}(t) = \begin{pmatrix}X_1(t) & X_2(t) & X_3(t) & X_4(t) & X_5(t)\end{pmatrix}^T$, $P(\boldsymbol{x},t)$ is its joint PMF, and the number of reaction channels $M = 2$. Consider the time interval $(t, t+\Delta t)$ and let $o(\Delta t)$ represent terms that go to zero with Δt faster than Δt. We use the following axioms:

- The probability of exactly one reaction $S_1 \; + \; S_2 \longrightarrow S_3$ in the time interval $(t, t+\Delta t)$ is given by $c_1 x_1 x_2 \Delta t + o(\Delta t)$. The probability of exactly one reaction $S_1 \; + \; S_4 \longrightarrow S_5$ in the time interval $(t, t+\Delta t)$ is given by $c_2 x_1 x_4 \Delta t + o(\Delta t)$.

- The probability that no reaction will occur in the system in the time interval $(t, t+\Delta t)$ is $(1 - c_1 x_1 x_2 \Delta t - c_2 x_1 x_4 \Delta t) + o(\Delta t)$.
- The probability of more than one reaction in the time interval $(t, t + \Delta t)$ is $o(\Delta t)$.

Hence we obtain the following balance equation:

$$\begin{aligned} P(x_1, x_2, x_3, x_4, x_5, t + \Delta t) &= (1 - c_1 x_1 x_2 \Delta t - c_2 x_1 x_4 \Delta t) P(x_1, x_2, x_3, x_4, x_5, t) + \\ & c_1 (x_1 + 1)(x_2 + 1) \Delta t P(x_1 + 1, x_2 + 1, x_3 - 1, x_4, x_5, t) + \\ & c_2 (x_1 + 1)(x_4 + 1) \Delta t P(x_1 + 1, x_2, x_3, x_4 + 1, x_5 - 1, t) + \\ & o(\Delta t). \end{aligned} \tag{1.19}$$

We rearrange the terms, divide by Δt, and take the limit as $\Delta t \longrightarrow 0$. All the terms of the form $o(\Delta t)/dt$ vanish and we obtain the following differential-difference equation:

$$\begin{aligned} \frac{dP(x_1, x_2, x_3, x_4, x_5, t)}{dt} &= -c_1 x_1 x_2 P(x_1, x_2, x_3, x_4, x_5, t) - c_2 x_1 x_4 P(x_1, x_2, x_3, x_4, x_5, t) + \\ & c_1 (x_1 + 1)(x_2 + 1) P(x_1 + 1, x_2 + 1, x_3 - 1, x_4, x_5, t) + \\ & c_2 (x_1 + 1)(x_4 + 1) P(x_1 + 1, x_2, x_3, x_4 + 1, x_5 - 1, t). \end{aligned} \tag{1.20}$$

At time $t = 0$, let the number of S_1, S_2 and S_4 molecules be α_1, α_2 and α_4, respectively. We can also describe this system with three variables, for example (X_1, X_2, X_4), and rewrite the above equation

$$\begin{aligned} \frac{dP(x_1, x_2, x_4, t)}{dt} &= -c_1 x_1 x_2 P(x_1, x_2, x_4, t) - c_2 x_1 x_4 P(x_1, x_2, x_4, t) + \\ & c_1 (x_1 + 1)(x_2 + 1) P(x_1 + 1, x_2 + 1, x_4, t) + \\ & c_2 (x_1 + 1)(x_4 + 1) P(x_1 + 1, x_2, x_4 + 1, t). \end{aligned} \tag{1.21}$$

In Chapter 2, we will discuss various techniques for solving the master equation analytically.

1.4.3 Stochastic Approximation

The time evolution of a continuous stochastic process $X(t)$ is described by the integral equation [57].

$$X(t) = X(0) + \int_0^t b(X(s), s)ds + \int_0^t u(X(s), s)d\mathcal{W}(s), \tag{1.22}$$

where $b(\cdot)$ is a drift coefficient, $u(\cdot)$ is a diffusion coefficient and $\mathcal{W}(\cdot)$ is a Wiener process, which is a continuous stochastic process with independent increments such that

- $E\left(\mathcal{W}(t) = 0\right)$.
- For $0 < t_1 < t_2$, $\mathcal{W}(t_2) - \mathcal{W}(t_1)$ has the distribution $\mathcal{N}(0, t_2 - t_1)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes a normal random variable with mean μ and variance σ^2.

The ds-integral in (1.22) is a Reimann integral and the $d\mathcal{W}(s)$-integral is an Itō integral [45], [57]. The corresponding differential equation

$$dX(t) = b(\boldsymbol{X}(t), t)dt + u(\boldsymbol{X}(t), t)d\mathcal{W}(t), \tag{1.23}$$

with the initial condition $X(t{=}0) = X(0)$, is known as *stochastic differential equation* (SDE). The Langevin equation and the Fokker-Planck equation are forms of SDE.

Langevin Equation

The Langevin equation is a stochastic differential equation for the state of the system $\boldsymbol{X}(t)$. For the biochemical system described above, the *chemical Langevin equation (CLE)* is derived by Gillespie [26], [28]

$$dX_i(t) = \sum_{m=1}^{M} \nu_{mi} a_m(\boldsymbol{X}, t)dt + \sum_{m=1}^{M} \nu_{mi} \sqrt{a_m(\boldsymbol{X}, t)} d\mathcal{W}_{mi}(t) \qquad i = 1 \ldots N, \tag{1.24}$$

where we have set $a_m(\boldsymbol{x}, t) \equiv a_m(\boldsymbol{X}, t)$. For the above CLE to be valid, the macroscopically infinitesimal time increment dt must satisfy the following two conditions:

1. dt be small enough that none of the propensity functions $a_m(\cdot)$ changes appreciably in this time

2. dt be large enough that the number of times each R_m reaction takes place is much greater than one.

One method of solution of the Langevin equation is to use the Euler approximation and generate realizations of the stochastic process in (1.24). Several other methods of solution are discussed in Kloeden and Platen [45].

Fokker-Planck Equation

For the Langevin equation in (1.24), there is an equivalent *Fokker-Planck equation (FPE)* for the PDF $P(\boldsymbol{x},t)$ of $\boldsymbol{X}(t)$ [28]

$$\begin{aligned}\frac{\partial P(\boldsymbol{x},t)}{\partial t} &= -\sum_{i=1}^{N}\frac{\partial}{\partial x_i}\left[\sum_{m=1}^{M}\nu_{mi}a_m(\boldsymbol{x},t)P(\boldsymbol{x},t)\right]+\\ &\quad\frac{1}{2}\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2}\left[\sum_{m=1}^{M}\nu_{mi}^2a_m(\boldsymbol{x},t)P(\boldsymbol{x},t)\right]+\\ &\quad\sum_{\substack{i,j=1\\ i<j}}^{N}\frac{\partial^2}{\partial x_i\partial x_j}\left[\sum_{m=1}^{M}\nu_{mi}\nu_{mj}a_m(\boldsymbol{x},t)P(\boldsymbol{x},t)\right].\end{aligned} \tag{1.25}$$

Example - Fokker-Planck Equation, Competing Reactions

We track three species, S_1 S_2 and S_4, $N=3$ and the number of reactions $M=2$, with reaction propensities (omitting the t's),

$$a_1(\boldsymbol{x}) = c_1x_1x_2 \tag{1.26a}$$

$$a_2(\boldsymbol{x}) = c_2x_1x_4. \tag{1.26b}$$

The FPE is

$$\begin{aligned}\frac{\partial P(\boldsymbol{x},t)}{\partial t} &= -\frac{\partial}{\partial x_1}\left[(-1)c_1x_1x_2P(\boldsymbol{x},t)+(-1)c_2x_1x_4P(\boldsymbol{x},t)\right] \\ &\quad -\frac{\partial}{\partial x_2}\left[(-1)c_1x_1x_2P(\boldsymbol{x},t)\right] - \frac{\partial}{\partial x_4}\left[(-1)c_2x_1x_4P(\boldsymbol{x},t)\right] \\ &\quad +\frac{\partial^2}{\partial x_1\partial x_2}\left[c_1x_1x_2P(\boldsymbol{x},t)\right] + \frac{\partial^2}{\partial x_1\partial x_4}\left[c_2x_1x_4P(\boldsymbol{x},t)\right]. \qquad (1.27)\end{aligned}$$

The Fokker-Planck equation can also be derived as an approximation of the master equation [76]. The FPE is a parabolic partial differential equation (PDE), and it is an N-dimensional equation for N molecular species. A numerical solution of the master equation is very difficult as the state space grows exponentially with the number of variables. Elf and Sjöberg present numerical solutions of the CME and the FPE using finite difference method [14], [70]. For the master equation, the authors use an N-dimensional grid of size x_{max} in each dimension, with grid points at $x = 0, 1 \ldots x_{max}$ and step size $\Delta x = 1$. The total number of grid points is approximately x_{max}^N. The Fokker-Planck equation is solved on a coarser grid with step size $\Delta x > 1$ and the number of grid points less than x_{max}^N. The authors compare the efficiency of the numerical solution of the Fokker-Planck equation with the SSA, and conclude that for $N > 4$ the SSA would be more efficient.

Chapter 2

Master Equation Approach

2.1 Introduction

In the stochastic approach for modeling the biochemical reactions, we obtain the PMF (or PDF) of the state of the system. The problem may be addressed in several ways, including analytical solution of the master equation, Monte Carlo methods and the moment propagation methods. We give a brief review of the analytical approach.

In 1954 Rényi presented a first treatment of the stochastic analysis of the second order reaction $S_1 + S_2 \longrightarrow S_3$. He gave a recursive expression for the Laplace transform of the distribution [15], [62]. In 1958, Bartholomay derived a stochastic model of the first order reaction $S_1 \longrightarrow S_2$, obtained the master equation and the distributions of the molecular populations with time, and also the limiting forms of the distributions and the standard deviation [3]. In his 1963 and 1967 papers, McQuarrie made use of the probability generating function to transform the master equation, a differential-difference equation, into a PDE, and obtained closed form expressions for a few first order reactions and the irreversible second order reaction $S_1 + S_2 \longrightarrow S_3$ [54], [55]. In 1964, Ishida studied some second order reactions and obtained expressions for the probability generating functions, which were expressed as Jacobi polynomials for the reaction $S_1 + S_2 \longrightarrow S_3$ and Legendre polynomials for the reaction $2\,S_1 \longrightarrow S_2$ [40]. In 2000 Laurenzi employed the Laplace transform to transform the master equation into a system of linear equations and solved the reversible second order reaction $S_1 + S_2 \longleftrightarrow S_3$ [46]. In 2005, Zhang et al. used Laplace transform to solve some first order reaction systems and obtained the molecular distributions as convolution of multinomials

[79]. In 2007, Jahnke et al. obtained solution of the CME for first order reactions. These solutions are expressed as convolution of multinomial and Poisson distributions, with time-dependent parameters obtained as solutions of the reaction rate equations [41]. In this chapter, we present the results of McQuarrie and Zhang et al.

2.2 Method of Probability Generating Function

We review McQuarrie's results for the case of irreversible first order reactions, and give the expressions for a few other cases. We will utilize some of these results in the proposed method.

2.2.1 Irreversible First Order Reactions

For the first order reaction

$$S_1 \xrightarrow{c_1} S_2, \tag{2.1}$$

the master equation is

$$\frac{dP(x_1,t)}{dt} = c_1(x_1+1)P(x_1+1,t) - c_1 x_1 P(x_1,t), \tag{2.2}$$

where $X_1(t)$ is the number of molecules of S_1 at time t. Let α_1 be the initial number of S_1 molecules. We use the probability generating function of $X_1(t)$

$$\mathcal{G}(s,t) = \sum_{x_1=0}^{\infty} P(x_1,t)s^{x_1}, \qquad x_1 \in \mathbb{N}_0, \tag{2.3}$$

where $\mathbb{N}_0$ is the set of natural numbers with zero. The above sum is defined for those values of s for which the sum converges. Utilizing (2.3) in (2.2) yields the PDE

$$\frac{\partial \mathcal{G}(s,t)}{\partial t} = c_1(1-s)\frac{\partial \mathcal{G}(s,t)}{\partial s} \tag{2.4}$$

with the initial condition

$$\mathcal{G}(s,t=0) = s^{\alpha_1}. \tag{2.5}$$

We rewrite (2.4) as

$$\frac{\partial \mathcal{G}(s,t)}{\partial t} + c_1(s-1)\frac{\partial \mathcal{G}(s,t)}{\partial s} = 0. \tag{2.6}$$

This is an equation of the form

$$e_1(s,t)\frac{\partial \mathcal{G}(s,t)}{\partial t} + e_2(s,t)\frac{\partial \mathcal{G}(s,t)}{\partial s} = 0, \tag{2.7}$$

and its characteristic equation is [4]

$$\frac{ds}{dt} = \frac{e_2(s,t)}{e_1(s,t)}. \tag{2.8}$$

Its solution curves $e_3(s,t) = A$, where A is an arbitrary constant, give the characteristic curves of the PDE. We next make a transformation from the (s,t) coordinate system to the (v,w) coordinate system

$$v = e_3(s,t) \qquad w = s \tag{2.9}$$

so that w is constant on each curve and the PDE (2.6) becomes an ODE. The characteristic equation of (2.6) is

$$\frac{ds}{dt} = c_1(s-1). \tag{2.10}$$

Integrating, we obtain

$$(s-1) = Ae^{c_1 t}. \tag{2.11}$$

The characteristic curves are given by

$$(s-1)e^{-c_1 t} = A. \tag{2.12}$$

Applying the transformation of variables

$$v = (s-1)e^{-c_1 t} \qquad w = s \qquad \mathsf{u}(v,w) = \mathcal{G}(s,t) \tag{2.13}$$

to (2.6), we obtain

$$\frac{\partial \mathsf{u}}{\partial w} = 0. \tag{2.14}$$

That is, u is only a function of $v = (s-1)e^{-c_1 t}$. Applying the initial condition (2.5), gives the solution

$$\mathcal{G}(s,t) = (1 + (s-1)e^{-c_1 t})^{\alpha_1}. \tag{2.15}$$

Utilizing the expressions for the first and the second moments

$$E\left(X_1(t)\right) = \left.\frac{\partial \mathcal{G}}{\partial s}\right|_{s=1} \tag{2.16a}$$

$$E\left(X_1^2(t)\right) = \left.\frac{\partial^2 \mathcal{G}}{\partial s^2}\right|_{s=1} + \left.\frac{\partial \mathcal{G}}{\partial s}\right|_{s=1}, \tag{2.16b}$$

we have for the mean and the variance of $X_1(t)$

$$\mu_1(t) \equiv E\left(X_1(t)\right) = \alpha_1 e^{-c_1 t} \tag{2.17}$$

$$\sigma_1^2(t) \equiv E\left(X_1^2(t)\right) - E^2\left(X_1(t)\right) = \alpha_1 e^{-c_1 t}\left(1 - e^{-c_1 t}\right). \tag{2.18}$$

2.2.2 Parallel First Order Reactions

Consider the parallel first order reactions

$$R_1 : \; S_1 \xrightarrow{c_1} S_2 \tag{2.19a}$$

$$R_2 : \; S_1 \xrightarrow{c_2} S_3. \tag{2.19b}$$

Let $X_1(t)$ and $X_2(t)$ be the number of molecules of S_1 and S_2, respectively, and let α_1 and α_2 be the initial number of S_1 and S_2 molecules, respectively. For the probability generating function of $P(x_1, x_2, t)$,

$$\mathcal{G}(s_1, s_2, t) = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} P(x_1, x_2, t) s_1^{x_1} s_2^{x_2} \tag{2.20}$$

we have the following expression from [55]:

$$\mathcal{G}(s_1, s_2, t) = \left[\frac{c_1 s_2 + c_2 - (c_1 s_2 + c_2 - C_1 s_1)\, e^{-C_1 t}}{C_1}\right]^{\alpha_1} s_2^{\alpha_2}, \tag{2.21}$$

where $C_1 = c_1 + c_2$. For the mean and the variance of $X_1(t)$ and $X_2(t)$, we have

$$\begin{aligned}
\mu_1(t) &= \alpha_1 e^{-C_1 t} && (2.22)\\
\sigma_1^2(t) &= \alpha_1 e^{-C_1 t}\left(1 - e^{-C_1 t}\right) && (2.23)\\
\mu_2(t) &= \alpha_2 + \frac{c_1\alpha_1}{C_1}\left(1 - e^{-C_1 t}\right) && (2.24)\\
\sigma_2^2(t) &= \frac{c_1\alpha_1}{C_1}\left(1 - e^{-C_1 t}\right)\left[1 - \frac{c_1\left(1 - e^{-C_1 t}\right)}{C_1}\right]. && (2.25)
\end{aligned}$$

2.2.3 Irreversible Second Order Reactions

For the second order reaction

$$S_1 \; + \; S_2 \;\xrightarrow{c_1}\; S_3, \tag{2.26}$$

let $X_1(t)$ and $X_2(t)$ be the number of S_1 and S_2 molecules, respectively, at time t. Let $\alpha_1 = X_1(0)$ and $\beta_{12} = X_2(0) - X_1(0)$. We track one species and the master equation is

$$\frac{dP(x_1,t)}{dt} = c_1(x_1+1)(\beta_{12}+x_1+1)P(x_1+1,t) - c_1 x_1(\beta_{12}+x_1)P(x_1,t). \tag{2.27}$$

Using the probability generating function gives the PDE

$$\frac{\partial \mathcal{G}(s,t)}{\partial t} = c_1 s(1-s)\frac{\partial^2 \mathcal{G}(s,t)}{\partial s^2} + c_1(\beta_{12}+1)(1-s)\frac{\partial \mathcal{G}(s,t)}{\partial s}. \tag{2.28}$$

This equation is solved using separation of variables and we have the following expressions for the first and second factorial moments of $X_1(t)$ [55]:

$$E\left(X_1(t)\right) = \sum_{n=1}^{\alpha_1} \frac{(2n+\beta_{12})\,\Gamma\left(\alpha_1+1\right)\Gamma\left(\alpha_1+\beta_{12}+1\right)}{\Gamma\left(\alpha_1-n+1\right)\Gamma\left(\alpha_1+\beta_{12}+n+1\right)}\Upsilon_n(t) \tag{2.29}$$

$$E\left(X_1(t)(X_1(t)-1)\right) =$$

$$\sum_{n=1}^{\alpha_1} \frac{(n-1)(n+\beta_{12}+1)(2n+\beta_{12})\Gamma\left(\alpha_1+1\right)\Gamma\left(\alpha_1+\beta_{12}+1\right)}{\Gamma\left(\alpha_1-n+1\right)\Gamma\left(\alpha_1+\beta_{12}+n+1\right)}\Upsilon_n(t). \tag{2.30}$$

In the above equations, $\Gamma(\cdot)$ is the gamma function defined by

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx \tag{2.31}$$

and

$$\Upsilon_n(t) = \exp\{-n(n+\beta_{12})c_1 t\}. \tag{2.32}$$

2.3 Method of Laplace Transform

In this method, Laplace transforms are employed to transform the master equation into a set of linear equations which are easier to handle.

2.3.1 First Order Reactions

We now present the method of Zhang et al. for computing the distributions of first order reactions [79]. This approach makes use of the fact that in a first order reaction system each molecule reacts independently of the others. Starting with one molecule from one source, the transition probabilities to other states are derived. Subsequently, the authors arrived at the distributions arising from several molecules and several sources by summing up the contribution from each. Several examples, including a chain reaction with several species, were presented.

The system has N molecular species S_1, S_2, ..., S_N, M reaction channels R_1, R_2, ... R_M and its state is specified by the random vector $\boldsymbol{X}(t)$. For the reactions

$$R_m: \quad S_i \xrightarrow{c_m} S_j, \qquad m = 1, \ldots, M, \tag{2.33}$$

an $N \times N$ matrix D is defined as follows:

$$d_{ii} = -C_i \tag{2.34a}$$

$$d_{ij} = c_m, \qquad i \neq j, \tag{2.34b}$$

where C_i is the sum of all the rate constants in which species S_i is a source.

One Source

First consider *one source* S_i with one molecule. The injection of the molecule into the system is modeled by a source probability density function $f_i(t)$. For example, if the molecule is injected at time $t = 0$, the source probability density is $f_i(t) = \delta(t)$, where $\delta(t)$ is the Dirac delta function. At a later time t, the molecule will be in any one of the states S_j, $j = 1, 2 \ldots N$, with probability $p_{ij}(t)$ satisfying the condition

$$\sum_{j=1}^{N} p_{ij}(t) = 1. \tag{2.35}$$

The probabilities $p_{ij}(t)$ are determined by solving the following system of first-order differential equations:

$$\frac{d\mathsf{p}}{dt} = \mathsf{D}\mathsf{p} + \mathsf{f}, \tag{2.36}$$

where

$$\mathsf{p} = \begin{pmatrix} p_{i1}(t) \\ p_{i2}(t) \\ . \\ . \\ p_{iN}(t) \end{pmatrix}, \qquad \mathsf{f} = \begin{pmatrix} 0 \\ . \\ f_i(t) \\ . \\ 0 \end{pmatrix}. \tag{2.37}$$

The Laplace Transform of (2.36) gives

$$s\mathsf{P} = \mathsf{D}\mathsf{P} + \mathsf{F}, \tag{2.38}$$

where P and F denote the Laplace transforms of p and f, respectively. Therefore P is obtained by solving

$$\mathsf{P} = (sI - \mathsf{D})^{-1}\, \mathsf{F}. \tag{2.39}$$

The inverse Laplace transform of P gives the required probabilities, and the joint distribution of $\big(X_1(t) \;\; X_2(t) \;\; \ldots \;\; X_N(t)\big)$ is the categorical distribution

$$P(x_1, x_2 \ldots, x_N, t) \;=\; \prod_{j=1}^{N} \left(p_{ij}(t)\right)^{x_j}. \tag{2.40}$$

Next, suppose that α_i molecules of S_i are injected into the system. Since each molecule reacts independently of the others, we have α_i independent trials each with N possible outcomes $X_1(t)$, $X_2(t)$, ... and $X_N(t)$. The probabilities of these outcomes are $p_{i1}(t)$, $p_{i2}(t)$, ... and $p_{iN}(t)$, respectively, and the resulting distribution is multinomial

$$P(x_1, x_2, \ldots, x_N, t) \;=\; \frac{\alpha_i!}{x_1! x_2! \ldots x_N!} \left(p_{i1}(t)\right)^{x_1} \left(p_{i2}(t)\right)^{x_2} \ldots \left(p_{iN}(t)\right)^{x_N}, \tag{2.41}$$

where $x_1 + x_2 + \ldots x_N = \alpha_i$. The marginal of any one species is the binomial distribution

$$P(x_j, t) \;=\; \frac{\alpha_i!}{x_j!\,(\alpha_i - x_j)!} \left(p_{ij}(t)\right)^{x_j} \left(1 - p_{ij}(t)\right)^{\alpha_i - x_j}, \qquad j = 1, 2 \ldots N. \tag{2.42}$$

Several Sources

If there are *several independent sources*, then the marginal probability density of a species is given by a convolution of binomials from each source. So for example, if we have two sources S_i and S_k with initial number of molecules α_i and α_k, respectively, then the resulting distribution is

$$\begin{aligned} P(x_j, t) \;&=\; \frac{\alpha_i!}{(x_j)!\,(\alpha_i - x_j)!} \left(p_{ij}(t)\right)^{x_j} \left(1 - p_{ij}(t)\right)^{\alpha_i - x_j} \\ &\otimes\; \frac{\alpha_k!}{(x_j)!\,(\alpha_k - x_j)!} \left(p_{kj}(t)\right)^{x_j} \left(1 - p_{kj}(t)\right)^{\alpha_k - x_j} \\ & \qquad j = 1, 2 \ldots N, \end{aligned} \tag{2.43}$$

where $\sum_{j=1}^{N} p_{ij}(t) = 1$ and $\sum_{j=1}^{N} p_{kj}(t) = 1$. We now present some examples.

Example - Reversible First Order Reaction

Consider the reactions

$$R_1 :\; S_1 \;\xrightarrow{c_1}\; S_2 \tag{2.44a}$$

$$R_2 :\; S_2 \;\xrightarrow{c_2}\; S_1, \tag{2.44b}$$

with one source S_1 consisting of α_1 molecules. For each S_1 molecule, we have the following system of equations:

$$\begin{pmatrix} \dfrac{dp_{11}(t)}{dt} \\ \dfrac{dp_{12}(t)}{dt} \end{pmatrix} = \begin{pmatrix} -c_1 & c_2 \\ c_1 & -c_2 \end{pmatrix} \begin{pmatrix} p_{11}(t) \\ p_{12}(t) \end{pmatrix} + \begin{pmatrix} \delta(t) \\ 0 \end{pmatrix}. \tag{2.45}$$

Taking the Laplace transform of the above equation, and solving for P gives

$$\mathsf{P} = \begin{pmatrix} \dfrac{s+c_2}{s\left(s+c_1+c_2\right)} \\ \dfrac{c_1}{s\left(s+c_1+c_2\right)} \end{pmatrix}. \tag{2.46}$$

The inverse Laplace transform of the above expression yields the following one molecule probabilities:

$$p_{11}(t) = \frac{1}{c_1+c_2}\left(c_2 + c_1 e^{-(c_1+c_2)t}\right) \tag{2.47a}$$

$$p_{12}(t) = \frac{c_1}{c_1+c_2}\left(1 - e^{-(c_1+c_2)t}\right). \tag{2.47b}$$

With α_1 initial molecules of S_1, the distributions of $X_1(t)$ and $X_2(t)$ are given by

$$P(x_j,t) = \frac{\alpha_1!}{x_j!\,(\alpha_1-x_j)!}\left(p_{1j}(t)\right)^{x_j}\left(1-p_{1j}(t)\right)^{\alpha_1-x_j}, \qquad j=1,2. \tag{2.48}$$

Example - Three Species

Consider the reactions

$$R_1: \; S_1 \xrightarrow{c_1} S_2 \tag{2.49a}$$

$$R_2: \; S_2 \xrightarrow{c_2} S_3, \tag{2.49b}$$

with one source S_1 consisting of α_1 molecules. Once again, for each S_1 molecule, we have the following system of equations:

$$\begin{pmatrix} \frac{dp_{11}(t)}{dt} \\ \frac{dp_{12}(t)}{dt} \\ \frac{dp_{13}(t)}{dt} \end{pmatrix} = \begin{pmatrix} -c_1 & 0 & 0 \\ c_1 & -c_2 & 0 \\ 0 & c_2 & 0 \end{pmatrix} \begin{pmatrix} p_{11}(t) \\ p_{12}(t) \\ p_{13}(t) \end{pmatrix} + \begin{pmatrix} \delta(t) \\ 0 \\ 0 \end{pmatrix}. \tag{2.50}$$

Using the above method yields the following one molecule probabilities:
Case i $c_1 \neq c_2$

$$p_{11}(t) = e^{-c_1 t} \tag{2.51a}$$

$$p_{12}(t) = \frac{c_1}{c_2 - c_1}\left(e^{-c_1 t} - e^{-c_2 t}\right) \tag{2.51b}$$

$$p_{13}(t) = 1 - \frac{c_2}{c_2 - c_1}e^{-c_1 t} + \frac{c_1}{c_2 - c_1}e^{-c_2 t}. \tag{2.51c}$$

Case ii $c_1 = c_2$

$$p_{11}(t) = e^{-c_1 t} \tag{2.52a}$$

$$p_{12}(t) = c_1 t e^{-c_1 t} \tag{2.52b}$$

$$p_{13}(t) = 1 - e^{-c_1 t} - c_1 t e^{-c_1 t}. \tag{2.52c}$$

With α_1 initial molecules of S_1, the distributions of $X_1(t)$, $X_2(t)$ and $X_3(t)$ are given by (2.42), with $i = 1$ and $N = 3$.

Chapter 3

Monte Carlo Methods

3.1 Introduction

The SSA method generates complete realizations of the Markov process $\boldsymbol{X}(t)$, and it was formulated for biochemical reaction systems by Gillespie in 1976 [22], [23]. A form of this method was introduced by Doob in 1945 [12]. In this method, the system evolves one reaction at a time, and uses two samples, one each, to determine the time to next reaction and the index of next reaction, respectively. The marginal densities and other statistics are inferred from the generated realizations. However, simulating only one reaction per time step is a major drawback of the SSA algorithm. If some molecular population is very large, the time to next reaction is very small and this renders the algorithm too slow to be useful. There are two different formulations of the SSA, namely, the *direct method* and the *first reaction method*. Gibson et al. modified the first reaction method to produce the *next reaction method*. In this method, the time to next reaction is sampled separately for each reaction, and the next reaction to be executed is the one with the smallest such time [20]. All of the samples are utilized, and this method scales better than SSA with the number of reactions. Cao et al. proposed an optimization of the direct method and compare it with the next reaction method [7]. Both the SSA and the next reaction methods are exact, as the trajectories generated by these methods would be the same as those generated from the solution of the master equation.

Numerous approximate methods have been established to reduce the computation time. In the τ-leap algorithm introduced by Gillespie, one steps through time in intervals large

enough to contain several firings of each reaction channel [27]. Each reaction channel fires independently of the others, with the number of such firings exhibiting Poisson distribution. However, the range of Poisson random variable varies from zero to infinity, and can cause negative molecular populations in the τ-leap method. Tian et al. [75] and Chatterjee et al. [9] proposed τ-leap methods in which the reaction channels fire from a binomial distribution. Pettigrew et al. developed a multinomial τ-leap method, which is an extension of the binomial τ-leap method, and increases the speed by partitioning a large reaction network into smaller groups [58]. Slepoy et al. proposed a method to reduce the computation time for very large reaction networks [72]. They use a method called composition and rejection, for faster computation of the index to next reaction. Their algorithm assumes that the average number of coupling between reactions does not grow as the size of the reaction network grows. In this chapter, we will describe the SSA and the τ-leap algorithm proposed by Tian et al.

3.2 Stochastic Simulation Algorithm

In this section, we recount the stochastic simulation algorithm [22], [23]. We have a system of N molecular species S_1, S_2, ..., S_N distributed uniformly inside a fixed volume V. The number of non-reactive molecular collisions occur much more frequently than reactive molecular collisions. The mixture is in thermal equilibrium, so that the collisions occur randomly. The species interact through M reaction channels R_1, R_2, ... R_M. The state of the system is specified by the random vector $\boldsymbol{X}(t)$. As described in Section 1.4.2, $h_m(\boldsymbol{x}, t)$ is the number of distinct molecular combinations for reaction R_m and the reaction propensity is defined by $a_m(\boldsymbol{x}, t) \equiv c_m h_m(\boldsymbol{x}, t)$. Then the probability that an R_m reaction will take place in an infinitesimal time interval dt is given by $a_m(\boldsymbol{x}, t)dt$, for the system in state $\boldsymbol{x}$. The algorithm then generates samples of the following events:

- The time to next reaction.
- The index of next reaction.

Let $P(\tau, k)$ denote the *next reaction probability density*, where τ is the time to next reaction and k is the index of next reaction. Then $P(\tau, k)d\tau$ gives the probability that the

next reaction will occur in time $d\tau$, and it will be the reaction R_k. $P(\tau, k)$ takes the form

$$P(\tau, k) \;=\; a_k(\boldsymbol{x}, t)\, e^{-a_0(\boldsymbol{x},t)\,\tau}, \qquad \tau > 0, \;\; k = 1, 2, \ldots M \tag{3.1}$$

where

$$a_0(\boldsymbol{x}, t) = \sum_{m=1}^{M} a_m(\boldsymbol{x}, t). \tag{3.2}$$

Marginalizing $P(\tau, k)$, with respect to k, gives the PDF of τ

$$P_1(\tau) \;=\; a_0(\boldsymbol{x}, t)\, e^{-a_0(\boldsymbol{x},t)\,\tau}, \qquad \tau > 0, \tag{3.3}$$

an exponential distribution with parameter $a_0(\boldsymbol{x}, t)$. Again, marginalizing $P(\tau, k)$, with respect to τ, gives the index of next reaction with probability

$$\text{Prob}(k = m) \;=\; \frac{a_m(\boldsymbol{x}, t)}{a_0(\boldsymbol{x}, t)}. \tag{3.4}$$

Therefore τ is sampled from the exponential distribution (3.3) and the index k is sampled from the categorical distribution, with M possible outcomes, and probabilities a_1/a_0, a_2/a_0 …, a_M/a_0 .

3.2.1 SSA Implementation

Given the reaction stoichiometries, reaction constants, the initial number of molecules $\boldsymbol{X}(t=0)$, the stopping time T_s and the number of simulations, the algorithm is implemented as follows:
For each simulation, perform the following initialization and recursion:

Initialization

- Set the initial time $t=0$ and the molecular populations to the given initial values.
- Compute the values of $h_m(\boldsymbol{x}, t)$ and $a_m(\boldsymbol{x}, t)$, $m = 1 \ldots M$.
- Compute the value of $a_0(\boldsymbol{x}, t)$.

Recursion

Repeat the following steps until stop time $t = T_s$ is reached:

- Generate a random number τ from distribution $P_1(\tau)$.
- Generate a random number m from the multinomial distribution with probabilities a_1/a_0, a_2/a_0 ..., a_M/a_0. This gives the reaction R_m to be executed.
- Update the molecular populations of all the species affected by the reaction R_m.
- Update $a_m(\boldsymbol{x}, t)$ and $a_0(\boldsymbol{x}, t)$.
- Advance the time to $t + \tau$.

Each simulation of the above algorithm generates a *realization* of the biochemical process from time $t = 0$ to time $t = T_s$. The marginal densities $P_1(x_1, t)$, $P_2(x_2, t)$, ..., $P_N(x_N, t)$, moments of $\boldsymbol{X}(t)$ and other statistics can be obtained from these realizations.

3.3 Binomial τ-leap Method

In the τ-leap algorithms, the time is divided into contiguous time intervals, and the system leaps along the time axis, to allow for many reaction events to take place in these time intervals [27]. We will now recite the τ-leap algorithm proposed by Tian et al. [75]. Once again, we have a uniform mixture of N molecular species S_1, S_2, ..., S_N, which interact through M reaction channels R_1, R_2, ... R_M. Define a parameter K_m for the first and second order reactions as follows:

$$R_m : \; S_1 \;\longrightarrow\; S_2, \qquad K_m = x_1(t). \tag{3.5a}$$

$$R_m : \; 2S_1 \;\longrightarrow\; S_2 \;, \qquad K_m = \left\lfloor \frac{x_1(t)}{2} \right\rfloor . \tag{3.5b}$$

$$R_m : \; S_1 \;+\; S_2 \;\longrightarrow\; S_3, \qquad K_m = \min(x_1(t), x_2(t)), \tag{3.5c}$$

where $\lfloor x \rfloor$ denotes largest integer less than or equal to x.

A time step τ is selected, and the criteria for doing so will be described shortly. At each time step, and for each reaction channel R_m, $m = 1 \ldots M$, we draw a sample L_m from the

distribution

$$\text{bin}\left(K_m, \frac{a_m(\boldsymbol{x},t)\,\tau}{K_m}\right), \tag{3.6}$$

where $\text{bin}\,(K,p)$ denotes binomial distribution with K independent trials each of which is successful with probability p. After sampling from all the channels, the molecular populations are updated as follows:

$$\boldsymbol{x}(t+\tau) = \boldsymbol{x}(t) + \sum_{m=1}^{M} \boldsymbol{\nu}_m L_m. \tag{3.7}$$

The time step τ must satisfy the condition,

$$0 \leq \frac{a_m(\boldsymbol{x},t)\,\tau}{K_m} \leq 1. \tag{3.8}$$

In addition, τ should be consistent with the *leap condition* which states that no propensity function must change appreciably in this time [29]. The leap condition is mathematically formulated as

$$|a_m(\boldsymbol{x},t+\tau) - a_m(\boldsymbol{x},t)| \leq \epsilon a_0(\boldsymbol{x},t), \qquad m = 1,\ldots M, \tag{3.9}$$

where $a_0(\boldsymbol{x},t)$ is defined in (3.2) and ϵ is some specified error control parameter. Starting from the above leap condition, τ is found to be

$$\tau = \min_{m\in[1,M]} \left(\frac{\epsilon a_0(\boldsymbol{x},t)}{|\mu_{\tau_m}(\boldsymbol{x},t)|}, \frac{\epsilon^2 a_0^2(\boldsymbol{x},t)}{\sigma^2_{\tau_m}(\boldsymbol{x},t)}\right), \tag{3.10}$$

where

$$\mu_{\tau_m}(\boldsymbol{x},t) = \sum_{m'=1}^{M} \eta_{mm'}(\boldsymbol{x},t) a_{m'}(\boldsymbol{x},t) \qquad m = 1,\ldots,M \tag{3.11a}$$

$$\sigma^2_{\tau_m}(\boldsymbol{x},t) = \sum_{m'=1}^{M} \eta^2_{mm'}(\boldsymbol{x},t) a_{m'}(\boldsymbol{x},t) \qquad m = 1,\ldots,M \tag{3.11b}$$

$$\eta_{mm'}(\boldsymbol{x},t) = \sum_{i=1}^{N} \frac{\partial a_m(\boldsymbol{x},t)}{\partial x_i} v_{m'i} \qquad m,m' = 1,\ldots,M. \tag{3.11c}$$

3.3.1 Simultaneous Sampling

In the above method, if the same species S_i undergoes reactions in two reaction channels R_m and $R_{m'}$, then sample as follows:

- Determine K_m and $K_{m'}$ for the reactions R_m and $R_{m'}$, respectively, from equations (3.5).

- Generate a sample $L_{mm'}$ for the total number of reactions that S_i undergoes in channels R_m and $R_{m'}$ That is sample from the distribution

$$\text{bin}\left(K_i, \frac{(a_m(\boldsymbol{x},t) + a_{m'}(\boldsymbol{x},t))\ \tau}{K_i}\right) \tag{3.12}$$

  under the conditions

$$K_i = \min\left(K_m, K_{m'}\right) \neq 0 \tag{3.13}$$

  and

$$0 \leq \frac{(a_m(\boldsymbol{x},t) + a_{m'}(\boldsymbol{x},t))\ \tau}{K_i} \leq 1. \tag{3.14}$$

- Next generate a sample value L_m for the total number of reactions that S_i undergoes in channel R_m. That is sample from

$$\text{bin}\left(L_{mm'}, \frac{a_m(\boldsymbol{x},t)}{a_m(\boldsymbol{x},t) + a_{m'}(\boldsymbol{x},t)}\right). \tag{3.15}$$

- The number of reactions in channel $R_{m'}$ is $L_{m'} = L_{mm'} - L_m$.

The authors give an example with twenty two reactions, where the τ-leap method shows a gain in speed by a factor of fifteen, compared with the SSA for $\epsilon = 0.01$. Since the SSA is used as a benchmark by most authors in the accelerated Monte Carlo-based and other methods, we will compare our proposed method with the SSA for both accuracy and speed.

Chapter 4

Proposed Method - First Order Reactions

4.1 Introduction

We propose a method for computing recursively, with time, the first two moments of the state of the system, and we call it the *recursive moment* (RM) method [50], [51]. The derivation of our method is based on conditional expectation relations of the random variables, Markovian properties of the biochemical system and approximations for small time steps. Our goal is to derive general expressions for very complex systems consisting of a large number of species and reactions. A second goal is that the expressions be straightforward to implement.

General expressions for the first order reactions are presented in this chapter and for the second order reactions are presented in the next chapter. The recipe for the first order reactions is simpler and can be used for a system consisting of first order reactions and those second order reactions which may be approximated as first order. The recipe for second order reactions can be used for systems consisting of both the first and the second order reactions. We compare the RM method with the SSA for accuracy, and further comparisons are characterized in Chapter 6.

4.2 Recursive Moment Method

Recall that we have a spatially homogeneous mixture of N molecular species S_1, S_2, ..., S_N inside a fixed volume V. The species interact through M reaction channels R_1, R_2, ... R_M. The state of the system is specified by the random vector $\boldsymbol{X}(t) = \begin{pmatrix} X_1(t) & X_2(t) & \dots & X_N(t) \end{pmatrix}^T$, where $X_i(t)$, $i = 1, 2 \dots N$, is the number of molecules of species S_i at time instant t.

The joint non-central moments of $\boldsymbol{X}(t)$ are given by [8], [74]

$$\gamma^{\boldsymbol{q}}(t) \;\equiv\; \gamma^{(q_1\; q_2\; \cdots\; q_N)}(t) \;=\; E\left(X_1^{q_1}(t) X_2^{q_2}(t) \dots X_N^{q_N}(t)\right), \tag{4.1}$$

where $\boldsymbol{q} = (q_1\; q_2\; \dots\; q_N) \in \mathbb{N}^N$ and $\mathbb{N}$ is the set of positive integers. We will also denote $(X_1^{q_1}(t) X_2^{q_2}(t) \dots X_N^{q_N}(t))$ by $\boldsymbol{X}^{\boldsymbol{q}}(t)$. The joint central moments of $\boldsymbol{X}(t)$ are given by

$$\begin{aligned} &\gamma'^{\,\boldsymbol{q}}(t) \;\equiv\; \gamma'^{\,(q_1\; q_2\; \cdots\; q_N)}(t) \\ &=\; E\left(\left(X_1(t) - \mu_1(t)\right)^{q_1} \left(X_2(t) - \mu_2(t)\right)^{q_2} \dots \left(X_N(t) - \mu_N(t)\right)^{q_N}(t)\right), \end{aligned} \tag{4.2}$$

where

$$\mu_i(t) \;=\; E\left(X_i(t)\right), \quad i = 1, \dots N. \tag{4.3}$$

The order of the moments Q is defined by $Q = \sum_{n=1}^{N} q_n$.

Further, let $\boldsymbol{\mu}(t)$ denote the mean of $\boldsymbol{X}(t)$ with elements $\mu_i(t)$, $i = 1, 2 \dots N$, and $\boldsymbol{C}(t)$ denote the covariance matrix of $\boldsymbol{X}(t)$ with its elements given by

$$\sigma_{ij}(t) \;=\; E\left(\left(X_i(t) - \mu_i(t)\right)\left(X_j(t) - \mu_j(t)\right)\right), \quad i, j = 1, \dots N. \tag{4.4}$$

The diagonal elements of $\boldsymbol{C}(t)$ will be denoted by $\sigma_i^2(t)$. In the proposed method, we will compute the mean and the covariance matrix of $\boldsymbol{X}(t)$.

In this chapter, we will discuss reactions of the type

$$R_m: \quad S_i \;\xrightarrow{c_m}\; S_j, \qquad m = 1, \dots, M \tag{4.5}$$

consisting of one reactant and one product. For a system consisting of one reaction as in (4.5), (2.15) shows that the distribution of X_i is binomial, where the probability that a

molecule of species S_i converts to a molecule of species S_j in time Δt is $1 - e^{-c_m \Delta t}$, and the probability that it stays as a molecule of species S_i is $e^{-c_m \Delta t}$.

We choose a time step Δt between two recursions and the criterion for doing so is discussed in Section 4.3. An $N \times N$ matrix of transition probabilities $\boldsymbol{P}$ is defined such that its $ij-$th element, p_{ij}, represents the probability that a molecule of species S_i converts to a molecule of species S_j during Δt. These elements are given by

$$p_{ii} = e^{-C_i \Delta t} \tag{4.6a}$$

$$p_{ij} = \frac{c_m}{C_i}(1 - e^{-C_i \Delta t}), \qquad i \neq j, \tag{4.6b}$$

where C_i is the sum of all the rate constants in which species S_i is a source. These probabilities are assumed constant during Δt.

For each molecular species in the system, and at each time step, we obtain expressions for the first two moments conditioned on the previous time step. We employ the following relations for conditional expectations [8], [16]:

$$E_U(U) = E_Z\left(E_{U|Z}(U|Z)\right) \tag{4.7}$$

$$E_{U|Z}\left(U^2|Z\right) = \mathrm{var}(U|Z) + E^2{}_{U|Z}(U|Z) \tag{4.8}$$

$$E_{U,V|Z}(UV|Z) = \mathrm{cov}(U,V|Z) + E_{U|Z}(U|Z)\, E_{V|Z}(V|Z), \tag{4.9}$$

where U, V and Z are random variables, $E(\cdot)$ is the expectation operator, $\mathrm{var}(\cdot)$ is the variance of a random variable and $\mathrm{cov}(\cdot)$ indicates the covariance between two random variables. The subscript of E indicates that the sums (or integrals) are evaluated with respect to the probability mass function (or probability density function) of that random variable. These subscripts will be omitted whenever it is clear from the context how the integrals are evaluated. For example, $E_{U|Z}(U|Z)$ in the above expression is the conditional expectation of U given Z. Marginalizing (4.8) and (4.9) gives the required second moments

$$E_U\left(U^2\right) = E_Z\left(E_{U|Z}\left(U^2|Z\right)\right) \tag{4.10}$$

$$E_{U,V}(UV) = E_Z\left(E_{U,V|Z}(UV|Z)\right). \tag{4.11}$$

We also define covariance and conditional covariance of the random vectors $\boldsymbol{U}$, $\boldsymbol{V}$ and $\boldsymbol{Z}$ as follows:

$$\operatorname{cov}(\boldsymbol{U},\boldsymbol{V}) = E(\boldsymbol{U}\boldsymbol{V}) - E(\boldsymbol{U})\,E(\boldsymbol{V}) \tag{4.12}$$

$$\operatorname{cov}(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{Z}) = E(\boldsymbol{U}\boldsymbol{V}|\boldsymbol{Z}) - E(\boldsymbol{U}|\boldsymbol{Z})\,E(\boldsymbol{V}|\boldsymbol{Z})\,. \tag{4.13}$$

For the random vector $\boldsymbol{X}(t+\Delta t)$, we have from relations (4.7) - (4.9)

$$\begin{aligned}
&\boldsymbol{\mu}(t+\Delta t) \;\equiv\; E_{\boldsymbol{X}(t+\Delta t)}\left(\boldsymbol{X}(t+\Delta t)\right)\\
&=\; E_{\boldsymbol{X}(t)}\left(E_{\boldsymbol{X}(t+\Delta t)\,|\,\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\,|\,\boldsymbol{X}(t)\right)\right)
\end{aligned} \tag{4.14}$$

$$\begin{aligned}
&E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\,\boldsymbol{X}^T(t+\Delta t)\,|\,\boldsymbol{X}(t)\right)\\
&=\; \operatorname{cov}\left(\boldsymbol{X}(t+\Delta t),\boldsymbol{X}^T(t+\Delta t)\,|\,\boldsymbol{X}(t)\right)\\
&\quad+\; E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\,|\,\boldsymbol{X}(t)\right)E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}^T(t+\Delta t)\,|\,\boldsymbol{X}(t)\right).
\end{aligned} \tag{4.15}$$

To evaluate (4.14), we obtain the expression for $E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\,|\,\boldsymbol{X}(t)\right)$ by summing up the contributions from all the species. This yields

$$E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\,|\,\boldsymbol{X}(t)\right) \;=\; \boldsymbol{P}^T\,\boldsymbol{X}(t) \tag{4.16}$$

and after taking the expectation $E_{\boldsymbol{X}(t)}(\cdot)$ of the above expression, we obtain

$$\boldsymbol{\mu}(t+\Delta t) \;=\; \boldsymbol{P}^T\,\boldsymbol{\mu}(t). \tag{4.17}$$

We next evaluate (4.15). The second term is

$$\begin{aligned}
&E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\,|\,\boldsymbol{X}(t)\right)E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}^T(t+\Delta t)\,|\,\boldsymbol{X}(t)\right)\\
&=\; \boldsymbol{P}^T\,\boldsymbol{X}(t)\,\boldsymbol{X}^T(t)\,\boldsymbol{P},
\end{aligned} \tag{4.18}$$

where we have utilized (4.16). To evaluate the first term in (4.15), we set

$$\boldsymbol{R}(t) \equiv \operatorname{cov}\left(\boldsymbol{X}(t+\Delta t),\boldsymbol{X}^T(t+\Delta t)\,|\,\boldsymbol{X}(t)\right). \tag{4.19}$$

The elements of $\boldsymbol{R}(t)$ are obtained by utilizing the multinomial distribution approximation as follows. Given K_m molecules of species S_m at time t, let $p_{m1}, p_{m2}, \ldots, p_{mN}$ be the probabilities that species S_m will become species $S_1, S_2 \ldots, S_N$, respectively. Then the joint probability of the number of molecules $(k_{m1}, k_{m2}, \ldots, k_{mN})$ of species $S_1, S_2, \ldots S_N$, respectively, created by the K_m molecules of S_m during the interval $(t, t+\Delta t)$ is given by the multinomial distribution

$$f(k_{m1}, k_{m2}, \ldots, k_{mN}) \;=\; \frac{K_m!}{k_{m1}!k_{m2}! \ldots k_{mN}!} p_{m1}^{k_{m1}}\, p_{m2}^{k_{m2}}\, \ldots p_{mN}^{k_{mN}}, \tag{4.20}$$

where $\sum_{i=1}^{N} k_{mi} = K_m$ and $\sum_{i=1}^{N} p_{mi} = 1$. From the moment generating function of the multinomial distribution

$$\mathcal{M}_{\boldsymbol{X}(t)}(\omega_1, \omega_2, \ldots, \omega_{N-1}) \;=\; \left(\sum_{i=1}^{N-1} p_{mi} e^{\omega_i} + p_{mN} \right)^{K_m}, \tag{4.21}$$

we compute the expectations $E\left(k_{mi}^2\right)$ and $E\left(k_{mi} k_{mj}\right)$ and readily obtain

$$\mathrm{var}(k_{mi}) \;=\; K_m\, p_{mi}\, (1 - p_{mi}) \tag{4.22}$$

$$\mathrm{cov}(k_{mi}, k_{mj}) \;=\; -K_m\, p_{mi}\, p_{mj}. \tag{4.23}$$

We can also arrive at the above result by noting that the marginal of any random variable in (4.20) is binomial and the marginal of any two random variables in (4.20) is trinomial.

We assume that in the time interval $(t, t + \Delta t)$, the distribution of the population of each reactant species evolves as multinomial, and contributes to all of its product species. Using (4.22) and (4.23) and summing up the contributions from all the species yields the elements of $\boldsymbol{R}(t)$

$$r_{ii}(t) \;=\; \sum_{k=1}^{N} X_k(t)\; p_{ki}\, (1 - p_{ki}) \tag{4.24a}$$

$$r_{ij}(t) \;=\; -\sum_{k=1}^{N} X_k(t)\; p_{ki}\, p_{kj} \qquad\qquad i \neq j. \tag{4.24b}$$

We will also need the expectation $\boldsymbol{B}(t) \equiv E_{\boldsymbol{X}(t)}\left(\boldsymbol{R}(t)\right)$. Its elements are given by

$$b_{ii}(t) \;=\; \sum_{k=1}^{N} \mu_k(t)\, p_{ki}\,(1-p_{ki}) \tag{4.25a}$$

$$b_{ij}(t) \;=\; -\sum_{k=1}^{N} \mu_k(t)\, p_{ki}\, p_{kj} \qquad\qquad i \neq j, \tag{4.25b}$$

and $\mu_j(t) = E\left(X_j(t)\right)$. Substituting (4.18) and (4.19) in (4.15) yields

$$E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\boldsymbol{X}^T(t+\Delta t)\,|\,\boldsymbol{X}(t)\right) \;=\; \boldsymbol{R}(t) \;+\; \boldsymbol{P}^T\,\boldsymbol{X}(t)\,\boldsymbol{X}^T(t)\,\boldsymbol{P}. \tag{4.26}$$

Marginalizing the above expression gives

$$E\left(\boldsymbol{X}(t+\Delta t)\boldsymbol{X}^T(t+\Delta t)\right) \;=\; \boldsymbol{B}(t) \;+\; \boldsymbol{P}^T\,E_{\boldsymbol{X}(t)}\left(\boldsymbol{X}(t)\,\boldsymbol{X}^T(t)\right)\,\boldsymbol{P}. \tag{4.27}$$

The covariance matrix is given by

$$\begin{aligned}\boldsymbol{C}(t+\Delta t) &= E\left(\boldsymbol{X}(t+\Delta t)\boldsymbol{X}^T(t+\Delta t)\right) \;-\; \boldsymbol{\mu}(t+\Delta t)\boldsymbol{\mu}^T(t+\Delta t) \\ &= \boldsymbol{B}(t) \;+\; \boldsymbol{P}^T\,E_{\boldsymbol{X}(t)}\left(\boldsymbol{X}(t)\,\boldsymbol{X}^T(t)\right)\,\boldsymbol{P} \;-\; \boldsymbol{P}^T\,\boldsymbol{\mu}(t)\,\boldsymbol{\mu}^T(t)\,\boldsymbol{P},\end{aligned} \tag{4.28}$$

where we have utilized (4.17) and (4.27) in the last equality. Therefore, the recursive expression for the covariance matrix is

$$\boldsymbol{C}(t+\Delta t) \;=\; \boldsymbol{B}(t) + \boldsymbol{P}^T\,\boldsymbol{C}(t)\,\boldsymbol{P}. \tag{4.29}$$

4.2.1 Example - A Simple Reversible Reaction

Consider the following system of reactions:

$$R_1:\; S_1 \xrightarrow{c_1} S_2 \tag{4.30a}$$

$$R_2:\; S_2 \xrightarrow{c_2} S_1. \tag{4.30b}$$

For reaction R_1, the expressions for the mean and the variance of $X_1(t)$ due to McQuarrie are [55]

$$E\left(X_1(t)\right) = X_1(t{=}0)\, e^{-c_1 t} \tag{4.31}$$

$$\mathrm{var}\left(X_1(t)\right) = X_1(t{=}0)\, e^{-c_1 t}\left(1 - e^{-c_1 t}\right). \tag{4.32}$$

We have analogous expressions for the mean and the variance of $X_2(t)$. By summing up the contributions from reactions R_1 and R_2, in time interval Δt, we have the following expressions for the conditional expectations of $X_1(t+\Delta t)$ and $X_1^2(t+\Delta t)$:

$$E\left(X_1(t+\Delta t)\,|\,X_1(t), X_2(t)\right) \;=\; X_1(t)\, e^{-c_1 \Delta t} + \;\; X_2(t)\;\left(1 - e^{-c_2 \Delta t}\right) \tag{4.33}$$

$$\begin{aligned} E\left(X_1^2(t+\Delta t)\,|\,X_1(t), X_2(t)\right) \;=\;& X_1(t)\, e^{-c_1 \Delta t}\;\left(1 - e^{-c_1 \Delta t}\right) + \\ & X_2(t)\, e^{-c_2 \Delta t}\;\left(1 - e^{-c_2 \Delta t}\right) + \\ & E^2\left(X_1(t+\Delta t)\,|\,X_1(t), X_2(t)\right). \end{aligned} \tag{4.34}$$

After taking the expectations $E_{X_1(t),X_2(t)}\left(\cdot\right)$ of (4.33) and (4.34), we obtain the recursive expressions for the mean and the variance of the number of S_1 molecules

$$\mu_1(t+\Delta t) = \mu_1(t)\, e^{-c_1 \Delta t} + \;\; \mu_2(t)\;\left(1 - e^{-c_2 \Delta t}\right) \tag{4.35}$$

$$\begin{aligned} \sigma_1^2(t+\Delta t) &= E\left(X_1^2(t+\Delta t)\right) - \mu_1^2(t+\Delta t) \\ &= \mu_1(t)\, e^{-c_1 \Delta t}\;\left(1 - e^{-c_1 \Delta t}\right) + \;\; \mu_2(t)\, e^{-c_2 \Delta t}\;\left(1 - e^{-c_2 \Delta t}\right) \\ &\quad + \;\; \sigma_1^2(t)\;\left(e^{-c_1 \Delta t}\right)^2 + \;\; \sigma_2^2(t)\;\left(1 - e^{-c_2 \Delta t}\right)^2 \\ &\quad - \;\; 2\sigma_1(t)\sigma_2(t)\, e^{-c_1 \Delta t}\;\left(1 - e^{-c_2 \Delta t}\right), \end{aligned} \tag{4.36}$$

where $\mu_i(t), i = 1, 2$ and $\sigma_i^2(t), i = 1, 2$ are the means and the variances of species S_1 and S_2, respectively.

4.3 General Expressions and Implementation

In order to implement the RM method, we choose a time step Δt such that the probability of each reaction is less than a predefined value. That is

$$c_m \Delta t \leq \epsilon, \;\; \forall \, m = 1, \ldots M, \tag{4.37}$$

where ϵ is some specified value. Given the reaction stoichiometries, reaction constants, the initial number of molecules $\boldsymbol{X}(t=0)$ and the stop time T_s, the algorithm is implemented as follows:

Initialization

- Select the time step as described above.
- Compute the matrix of transition probabilities $\boldsymbol{P}$ as given by the expressions (4.6a) and (4.6b).
- Set time $t=0$,

  $\boldsymbol{\mu}(t=0) = \boldsymbol{X}(t=0)$,

  $\boldsymbol{C}(t=0) = \boldsymbol{0}$

Recursion

Repeat the following steps until stop time $t=T_s$ is reached:

- Compute the matrix $\boldsymbol{B}(t)$ by employing the expressions (4.25a) and (4.25b).
- Compute $\boldsymbol{\mu}(t + \Delta t)$ by employing expression (4.17).
- Compute $\boldsymbol{C}(t + \Delta t)$ by employing expression (4.29).
- Advance the time to $t + \Delta t$.

4.4 Simulation Results

As described in Section 3.2, the SSA method generates realizations from the exact distribution. In this section, we compare results from the RM method with the SSA method

for accuracy and computation time. We track the means and the covariance matrix with time. In some cases, at a given time t, a normal distribution $\mathcal{N}\left(\mu_i(t), \sigma_i^2(t)\right)$ for species S_i, using the mean and the variance computed from the RM method, is compared with the distribution generated by the SSA method. The timing measurements are done on a machine called *Fermat* which consists of an Intel Pentium(R) Dual 2.4 GHz CPU.

4.4.1 Example - Three Species

Consider the reaction system

$$R_1 : \; S_1 \xrightarrow{c_1} S_2 \tag{4.38a}$$

$$R_2 : \; S_2 \xrightarrow{c_2} S_3. \tag{4.38b}$$

We track all three species, and therefore $N = 3$ and the number of reactions $M = 2$. The state of the system is given by $\boldsymbol{X}(t) = \begin{pmatrix} X_1(t) & X_2(t) & X_3(t) \end{pmatrix}^T$. The model is specified by the initial number of molecules $\boldsymbol{X}(t{=}0) = \begin{pmatrix} 10000 & 500 & 0 \end{pmatrix}^T$ and the reaction constants $c_1 = 0.1$ sec^{-1} and $c_2 = 1.0$ sec^{-1}. The matrix of transition probabilities is defined by

$$\boldsymbol{P} = \begin{pmatrix} e^{-c_1 \Delta t} & 1 - e^{-c_1 \Delta t} & 0 \\ 0 & e^{-c_2 \Delta t} & 1 - e^{-c_2 \Delta t} \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.39}$$

$\boldsymbol{B}(t)$ is symmetric and its elements are given by

$$\begin{aligned}
b_{11}(t) &= \mu_1(t) p_{11}(1 - p_{11}) \\
b_{12}(t) &= \mu_1(t) p_{11} p_{12} \\
b_{13}(t) &= b_{31}(t) = 0 \\
b_{22}(t) &= \mu_1(t) p_{12}(1 - p_{12}) + \mu_2(t) p_{22}(1 - p_{22}) \\
b_{23}(t) &= b_{32}(t) = \mu_2(t) p_{22} p_{23} \\
b_{33}(t) &= \mu_2(t) p_{23}(1 - p_{23}).
\end{aligned}$$

By applying equations (4.17) and (4.29) repeatedly, we obtain the mean and the covariance of all the species from time $t{=}0$ sec to $t{=}10$ sec. In Figure 4.1, we show the tracking results

of some of the first two moments of the species. In implementing the methods, $\Delta t = 0.005$ sec for the RM method and the number of realizations is $5,000$ for the SSA method. The average number of time steps with the SSA method is $12,721$. As seen from the figure, the estimated moments obtained by the RM and SSA methods are in good agreement.

4.4.2 Complex Example

We now present a complex example with the number of species $N = 100$ and the number of reactions $M = 200$. The reactions are listed in Tables 4.1 and 4.2, and the reaction parameters are listed in Table 4.3. The initial population $X_1(t=0)$ of the number of S_1 molecules is $2,000$. The initial populations of all the other species are zero. The time step Δt in the RM method was selected to be $\Delta t = 0.001$ sec. The average number of time steps with the SSA method is $30,663$. We compared the results with the SSA for 500 realizations (Figures 4.2 and 4.3) and $2,000$ realizations (Figures 4.4 and 4.5). The SSA method produces much smoother results with $2,000$ realizations and the results from the two methods compare well.

Next, we compare the computation time of the RM method with the SSA method with varying molecular populations. We vary the initial population $X_1(t=0)$ of the number of S_1 molecules from $5,000$ to $30,000$. The initial populations of all the other species are zero. As seen from (4.37), the time step Δt of the RM method does not depend on the size of molecular population, and it is set to $\Delta t = 0.001$ sec as indicated above.

For the SSA method, the simulation time depends on the molecular population, as the time to next reaction has an exponential distribution with an expected value $1/a_0$ (expression (3.3)). Therefore, the number of time steps and hence the computation time increases with the molecular population. The number of realizations required to estimate the means and the variances also depend on the size of molecular populations. To achieve the same accuracy, we set the number of realizations equal to the molecular populations. Table 4.4 lists the computation times from both methods on the Fermat machine. With the SSA method, the CPU time is linear with the number of realizations. It was measured for a smaller number of realizations, and estimated for the number of realizations indicated in the table. The cumulative increase in the SSA CPU time is quadratic in the molecular populations. With the RM method the CPU time remains constant at 35 sec. We would like to remark that

the computation time employing the RM method will increase with the number of species N, as this requires multiplication of matrices of size N. Also for a given number of species, the computation time with the RM method remains constant as the number of reactions M increases, whereas it increases with the SSA method.

$S_1 \longrightarrow S_2$	$S_2 \longrightarrow S_3$	$S_3 \longrightarrow S_4$	$S_4 \longrightarrow S_5$	$S_5 \longrightarrow S_6$
$S_6 \longrightarrow S_7$	$S_7 \longrightarrow S_8$	$S_8 \longrightarrow S_9$	$S_9 \longrightarrow S_{10}$	$S_{10} \longrightarrow S_{11}$
$S_{11} \longrightarrow S_{12}$	$S_{12} \longrightarrow S_{13}$	$S_{13} \longrightarrow S_{14}$	$S_{14} \longrightarrow S_{15}$	$S_{15} \longrightarrow S_{16}$
$S_{16} \longrightarrow S_{17}$	$S_{17} \longrightarrow S_{18}$	$S_{18} \longrightarrow S_{19}$	$S_{19} \longrightarrow S_{20}$	$S_{20} \longrightarrow S_{21}$
$S_{21} \longrightarrow S_{22}$	$S_{22} \longrightarrow S_{23}$	$S_{23} \longrightarrow S_{24}$	$S_{24} \longrightarrow S_{25}$	$S_{25} \longrightarrow S_{26}$
$S_{26} \longrightarrow S_{27}$	$S_{27} \longrightarrow S_{28}$	$S_{28} \longrightarrow S_{29}$	$S_{29} \longrightarrow S_{30}$	$S_{30} \longrightarrow S_{31}$
$S_{31} \longrightarrow S_{32}$	$S_{32} \longrightarrow S_{33}$	$S_{33} \longrightarrow S_{34}$	$S_{34} \longrightarrow S_{35}$	$S_{35} \longrightarrow S_{36}$
$S_{36} \longrightarrow S_{37}$	$S_{37} \longrightarrow S_{38}$	$S_{38} \longrightarrow S_{39}$	$S_{39} \longrightarrow S_{40}$	$S_{40} \longrightarrow S_{41}$
$S_{41} \longrightarrow S_{42}$	$S_{42} \longrightarrow S_{43}$	$S_{43} \longrightarrow S_{44}$	$S_{44} \longrightarrow S_{45}$	$S_{45} \longrightarrow S_{46}$
$S_{46} \longrightarrow S_{47}$	$S_{47} \longrightarrow S_{48}$	$S_{48} \longrightarrow S_{49}$	$S_{49} \longrightarrow S_{50}$	$S_{50} \longrightarrow S_{51}$
$S_{51} \longrightarrow S_{52}$	$S_{52} \longrightarrow S_{53}$	$S_{53} \longrightarrow S_{54}$	$S_{54} \longrightarrow S_{55}$	$S_{55} \longrightarrow S_{56}$
$S_{56} \longrightarrow S_{57}$	$S_{57} \longrightarrow S_{58}$	$S_{58} \longrightarrow S_{59}$	$S_{59} \longrightarrow S_{60}$	$S_{60} \longrightarrow S_{61}$
$S_{61} \longrightarrow S_{62}$	$S_{62} \longrightarrow S_{63}$	$S_{63} \longrightarrow S_{64}$	$S_{64} \longrightarrow S_{65}$	$S_{65} \longrightarrow S_{66}$
$S_{66} \longrightarrow S_{67}$	$S_{67} \longrightarrow S_{68}$	$S_{68} \longrightarrow S_{69}$	$S_{69} \longrightarrow S_{70}$	$S_{70} \longrightarrow S_{71}$
$S_{71} \longrightarrow S_{72}$	$S_{72} \longrightarrow S_{73}$	$S_{73} \longrightarrow S_{74}$	$S_{74} \longrightarrow S_{75}$	$S_{75} \longrightarrow S_{76}$
$S_{76} \longrightarrow S_{77}$	$S_{77} \longrightarrow S_{78}$	$S_{78} \longrightarrow S_{79}$	$S_{79} \longrightarrow S_{80}$	$S_{80} \longrightarrow S_{81}$
$S_{81} \longrightarrow S_{82}$	$S_{82} \longrightarrow S_{83}$	$S_{83} \longrightarrow S_{84}$	$S_{84} \longrightarrow S_{85}$	$S_{85} \longrightarrow S_{86}$
$S_{86} \longrightarrow S_{87}$	$S_{87} \longrightarrow S_{88}$	$S_{88} \longrightarrow S_{89}$	$S_{89} \longrightarrow S_{90}$	$S_{90} \longrightarrow S_{91}$
$S_{91} \longrightarrow S_{92}$	$S_{92} \longrightarrow S_{93}$	$S_{93} \longrightarrow S_{94}$	$S_{94} \longrightarrow S_{95}$	$S_{95} \longrightarrow S_{96}$
$S_{96} \longrightarrow S_{97}$	$S_{97} \longrightarrow S_{98}$	$S_{98} \longrightarrow S_{99}$	$S_{99} \longrightarrow S_{100}$	$S_{100} \longrightarrow S_1$

Table 4.1: Complex Example, reactions R_1 through R_{100}, in left-to-right top-to-bottom order.

$S_1 \longrightarrow S_{11}$	$S_2 \longrightarrow S_{12}$	$S_3 \longrightarrow S_{13}$	$S_4 \longrightarrow S_{14}$	$S_5 \longrightarrow S_{15}$
$S_6 \longrightarrow S_{16}$	$S_7 \longrightarrow S_{17}$	$S_8 \longrightarrow S_{18}$	$S_9 \longrightarrow S_{19}$	$S_{10} \longrightarrow S_{20}$
$S_{11} \longrightarrow S_{21}$	$S_{12} \longrightarrow S_{22}$	$S_{13} \longrightarrow S_{23}$	$S_{14} \longrightarrow S_{24}$	$S_{15} \longrightarrow S_{25}$
$S_{16} \longrightarrow S_{26}$	$S_{17} \longrightarrow S_{27}$	$S_{18} \longrightarrow S_{28}$	$S_{19} \longrightarrow S_{29}$	$S_{20} \longrightarrow S_{30}$
$S_{21} \longrightarrow S_{31}$	$S_{22} \longrightarrow S_{32}$	$S_{23} \longrightarrow S_{33}$	$S_{24} \longrightarrow S_{34}$	$S_{25} \longrightarrow S_{35}$
$S_{26} \longrightarrow S_{36}$	$S_{27} \longrightarrow S_{37}$	$S_{28} \longrightarrow S_{38}$	$S_{29} \longrightarrow S_{39}$	$S_{30} \longrightarrow S_{40}$
$S_{31} \longrightarrow S_{41}$	$S_{32} \longrightarrow S_{42}$	$S_{33} \longrightarrow S_{43}$	$S_{34} \longrightarrow S_{44}$	$S_{35} \longrightarrow S_{45}$
$S_{36} \longrightarrow S_{46}$	$S_{37} \longrightarrow S_{47}$	$S_{38} \longrightarrow S_{48}$	$S_{39} \longrightarrow S_{49}$	$S_{40} \longrightarrow S_{50}$
$S_{41} \longrightarrow S_{51}$	$S_{42} \longrightarrow S_{52}$	$S_{43} \longrightarrow S_{53}$	$S_{44} \longrightarrow S_{54}$	$S_{45} \longrightarrow S_{55}$
$S_{46} \longrightarrow S_{56}$	$S_{47} \longrightarrow S_{57}$	$S_{48} \longrightarrow S_{58}$	$S_{49} \longrightarrow S_{59}$	$S_{50} \longrightarrow S_{60}$
$S_{51} \longrightarrow S_{61}$	$S_{52} \longrightarrow S_{62}$	$S_{53} \longrightarrow S_{63}$	$S_{54} \longrightarrow S_{64}$	$S_{55} \longrightarrow S_{65}$
$S_{56} \longrightarrow S_{66}$	$S_{57} \longrightarrow S_{67}$	$S_{58} \longrightarrow S_{68}$	$S_{59} \longrightarrow S_{69}$	$S_{60} \longrightarrow S_{70}$
$S_{61} \longrightarrow S_{71}$	$S_{62} \longrightarrow S_{72}$	$S_{63} \longrightarrow S_{73}$	$S_{64} \longrightarrow S_{74}$	$S_{65} \longrightarrow S_{75}$
$S_{66} \longrightarrow S_{76}$	$S_{67} \longrightarrow S_{77}$	$S_{68} \longrightarrow S_{78}$	$S_{69} \longrightarrow S_{79}$	$S_{70} \longrightarrow S_{80}$
$S_{71} \longrightarrow S_{81}$	$S_{72} \longrightarrow S_{82}$	$S_{73} \longrightarrow S_{83}$	$S_{74} \longrightarrow S_{84}$	$S_{75} \longrightarrow S_{85}$
$S_{76} \longrightarrow S_{86}$	$S_{77} \longrightarrow S_{87}$	$S_{78} \longrightarrow S_{88}$	$S_{79} \longrightarrow S_{89}$	$S_{80} \longrightarrow S_{90}$
$S_{81} \longrightarrow S_{91}$	$S_{82} \longrightarrow S_{92}$	$S_{83} \longrightarrow S_{93}$	$S_{84} \longrightarrow S_{94}$	$S_{85} \longrightarrow S_{95}$
$S_{86} \longrightarrow S_{96}$	$S_{87} \longrightarrow S_{97}$	$S_{88} \longrightarrow S_{98}$	$S_{89} \longrightarrow S_{99}$	$S_{90} \longrightarrow S_{100}$
$S_{91} \longrightarrow S_1$	$S_{92} \longrightarrow S_1$	$S_{93} \longrightarrow S_3$	$S_{94} \longrightarrow S_4$	$S_{95} \longrightarrow S_5$
$S_{96} \longrightarrow S_6$	$S_{97} \longrightarrow S_7$	$S_{98} \longrightarrow S_8$	$S_{99} \longrightarrow S_9$	$S_{100} \longrightarrow S_{10}$

Table 4.2: Complex Example, reactions R_{101} through R_{200}, in left-to-right top-to-bottom order.

0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03
1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04
1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06
1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07
1.08	1.08	1.08	1.08	1.08	1.08	1.08	1.08	1.08	1.08
1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09
2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0

Table 4.3: Complex Example, reaction constants for reactions R_1 through R_{200}, in units of sec^{-1}, in left-to-right top-to-bottom order.

Initial number of S_1 molecules	Number of realizations in SSA	CPU time	
		RM	SSA
5,000	5,000	35 sec	2.35 days
10,000	10,000	35 sec	8.9 days
15,000	15,000	35 sec	19.2 days
20,000	20,000	35 sec	36.6 days
25,000	25,000	35 sec	61.1 days
30,000	30,000	35 sec	85.1 days

Table 4.4: Computation times of the RM and the SSA methods, versus the molecular populations, for the Complex Example given in Section 4.4.2.

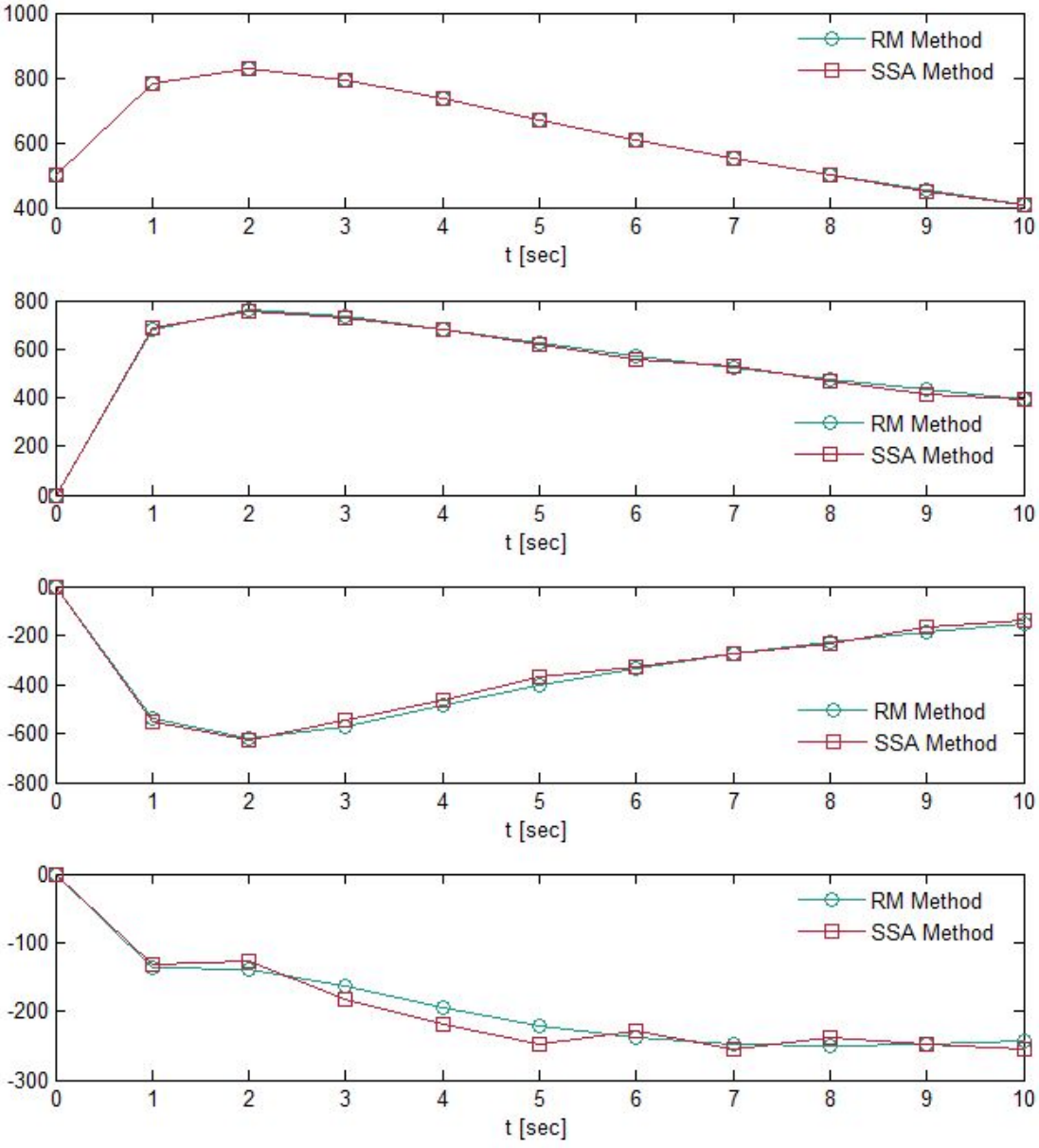

Figure 4.1: Comparison of the results from the RM and the SSA methods for the Three Species example given in Section 4.4.1. The curves represent the computed means $\mu_2(t)$ (top plot), the variances $\sigma_2^2(t)$ (second plot), and the covariances $\sigma_{21}(t)$ (third plot) and $\sigma_{23}(t)$ (bottom plot). Δt is .005 sec with the RM method and the number of realizations in the SSA method is $5,000$.

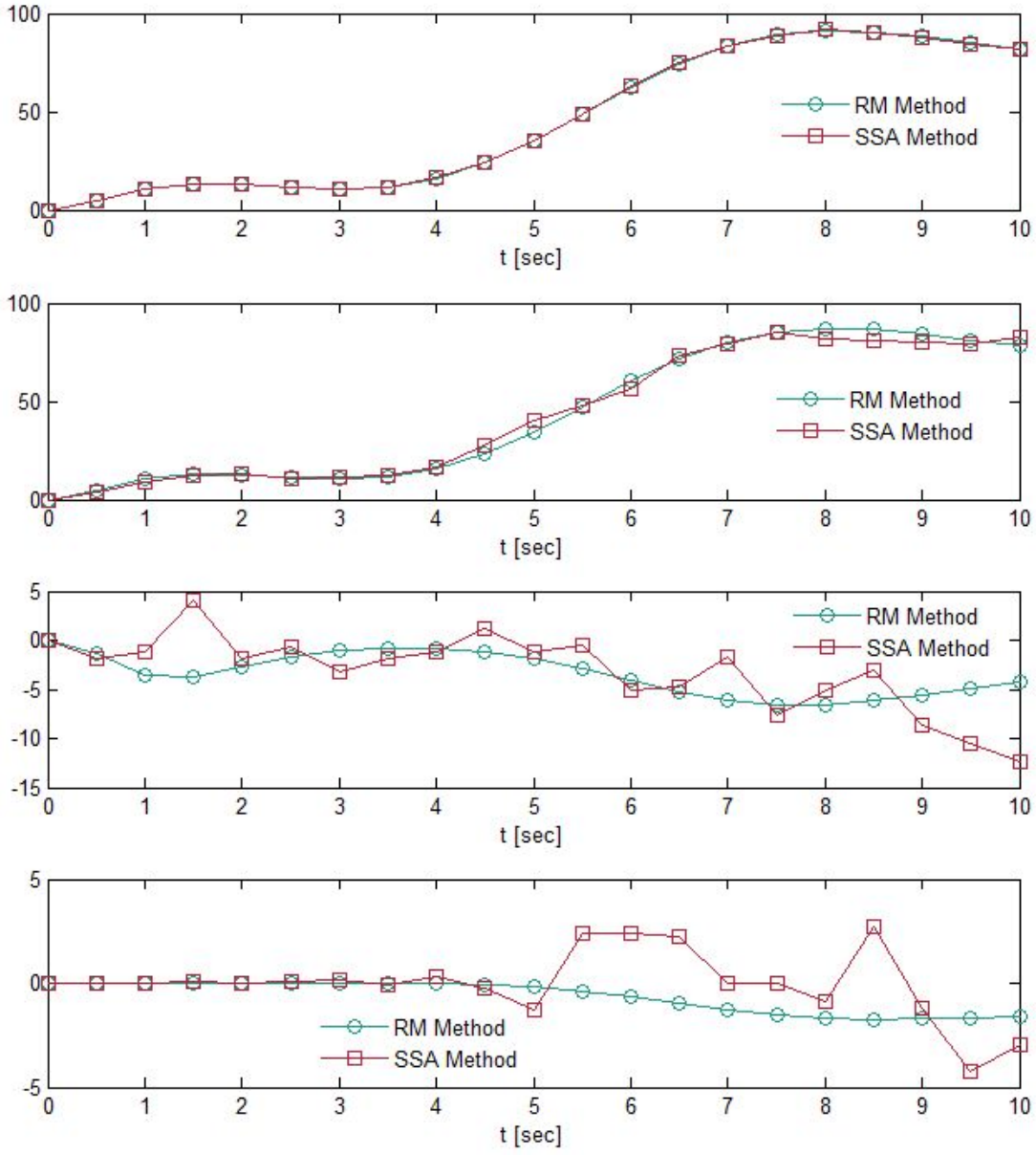

Figure 4.2: Comparison of the results from the RM and the SSA methods for the Complex Example given in Section 4.4.2. The curves represent the computed means $\mu_{11}(t)$ (top plot), the variances $\sigma^2_{11}(t)$ (second plot), and the covariances $\sigma_{11,10}(t)$ (third plot) and $\sigma_{11,12}(t)$ (bottom plot). Δt is .0005 sec with the RM method and the number of realizations in the SSA method is $1,000$.

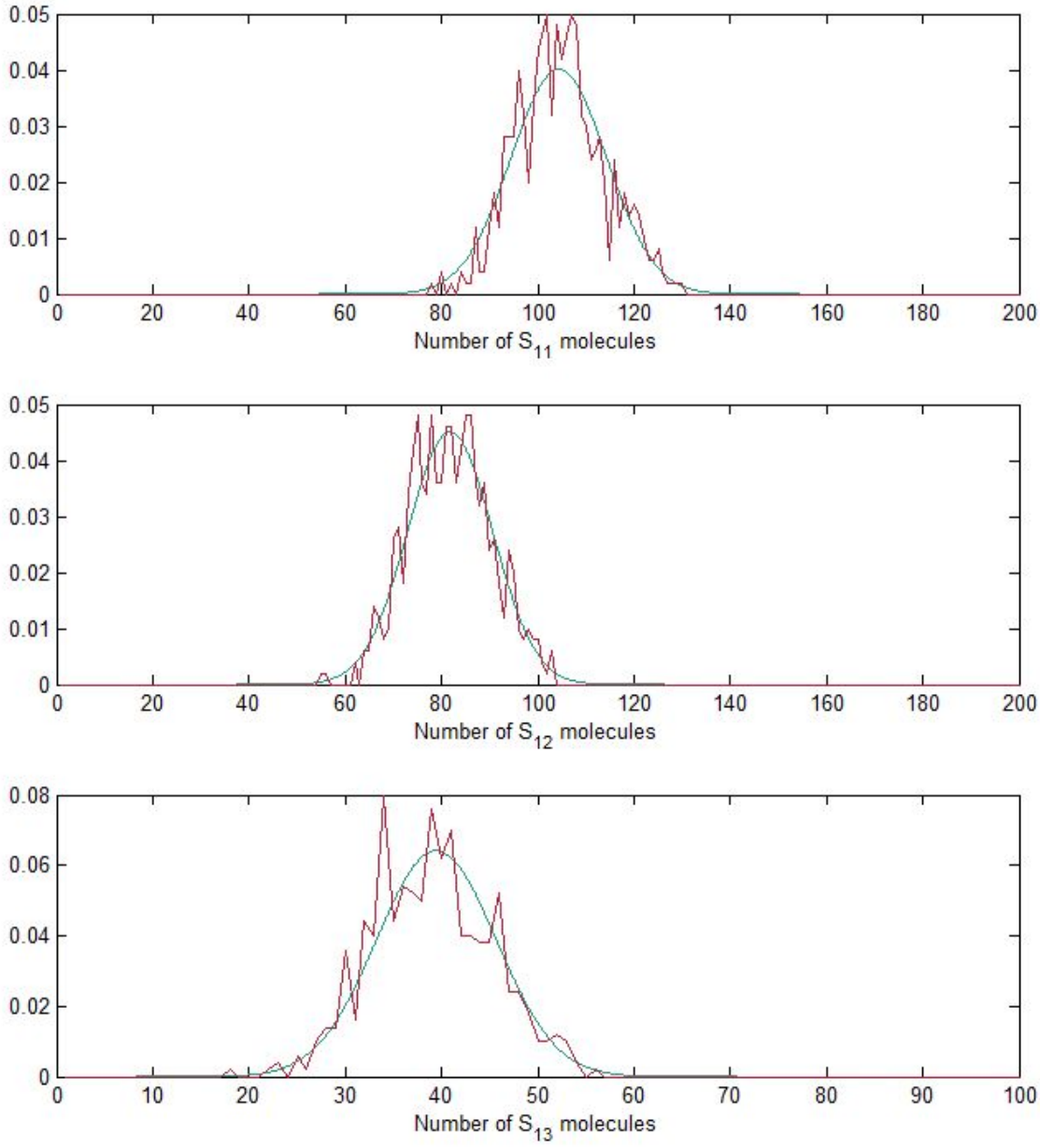

Figure 4.3: The normal distribution generated from the RM method (solid line) is compared with the distribution generated by the SSA method ("noisy" plot), at time $t = 10$ sec, for the Complex Example given in Section 4.4.2.

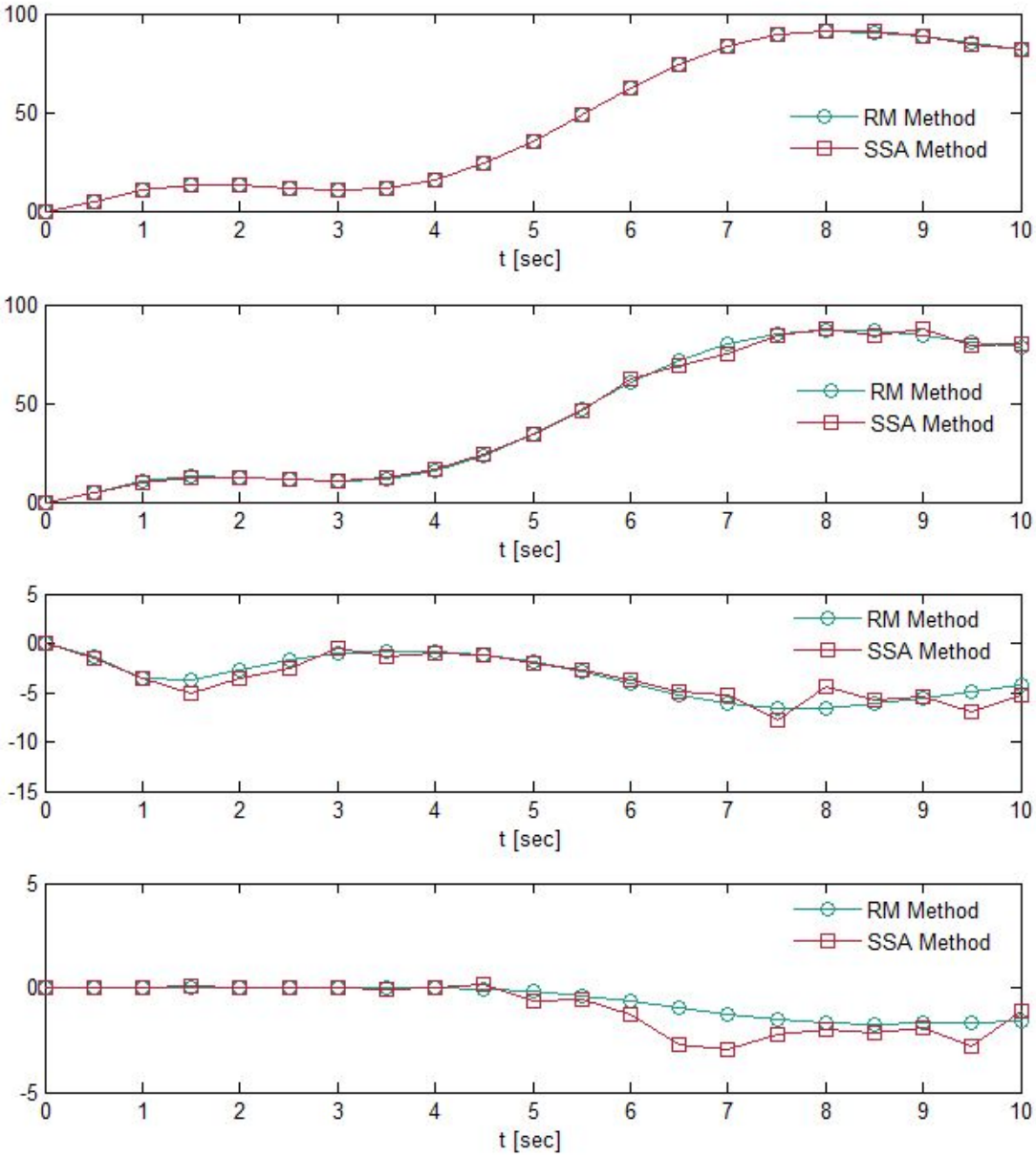

Figure 4.4: Same example and parameters as in Figure 4.2, except that the number of realizations in the SSA method is $2,000$.

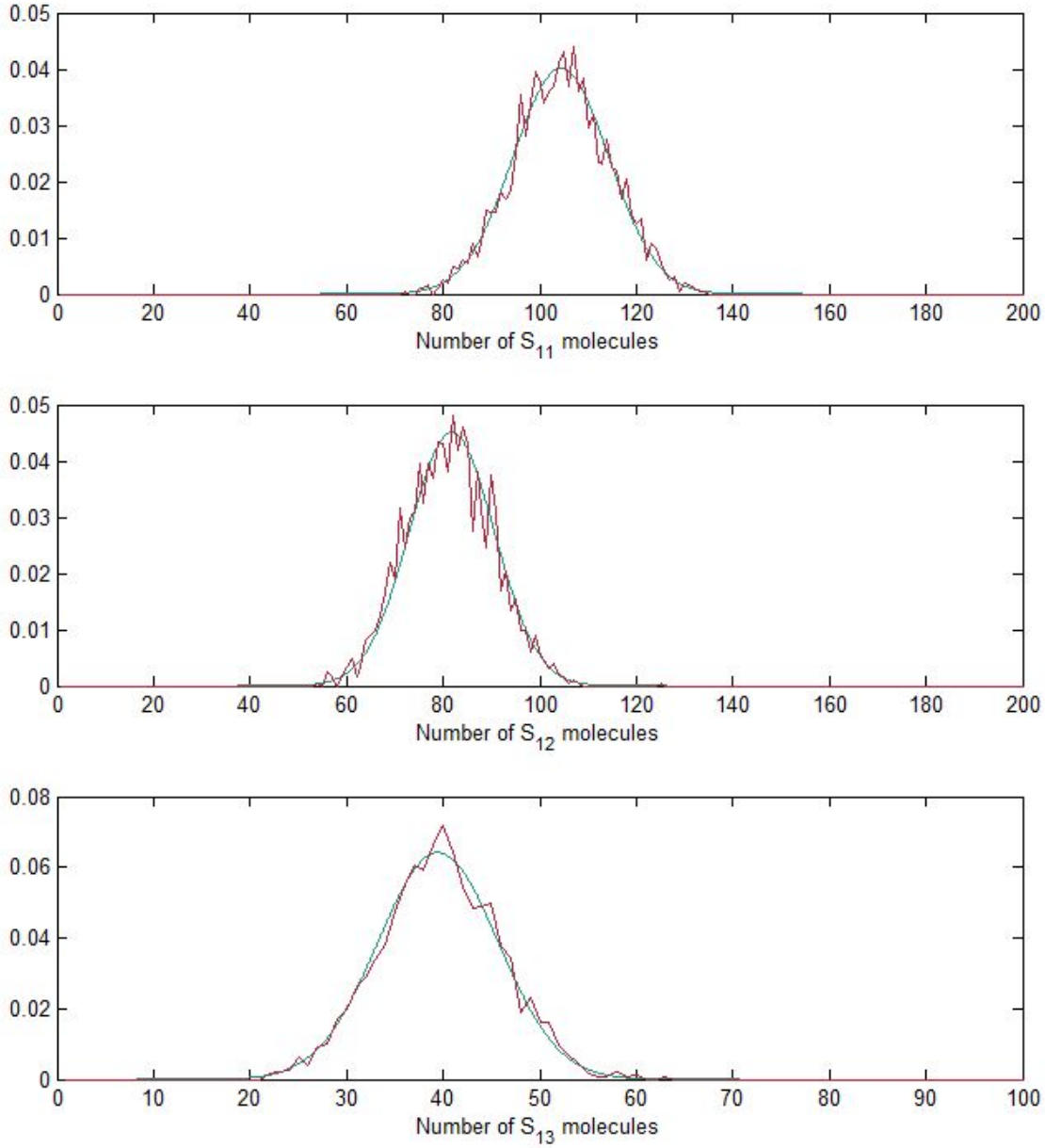

Figure 4.5: Same example and parameters as in Figure 4.3, except that the number of realizations in the SSA method is $2,000$.

Chapter 5

Proposed Method - Second Order Reactions

5.1 Introduction

In Chapter 4, we obtained recursive expressions for propagating the first two moments of the joint distribution of the molecular species. Our results were applied to a complex system consisting of a hundred species and two hundred reactions. We now want to extend our method to the more difficult problem of second order reactions. Some issues with the second order reactions are:

- In first order reactions, the reaction rates are proportional to the number of molecules of the reactant. In second order reactions, the reaction rates are proportional to the product of the number of molecules of the reactants. That is, the second order reactions are *nonlinear* in propensities.
- The evaluation of the covariance matrix requires computing expectations of third and higher order cross terms.
- For the second order reaction $S_i + S_j \longrightarrow S_k$, the expressions (2.29) for $E\left(X_1(t)\right)$ and (2.30) containing $E\left(X_1^2(t)\right)$ do not have a simple form as for the first order reaction $S_i \longrightarrow S_j$.

5.2 The Methodology

Once again, we use the conditional expectation relations as in (4.14) and (4.15)

$$\begin{aligned}
&\boldsymbol{\mu}(t+\Delta t) \;\equiv\; E_{\boldsymbol{X}(t+\Delta t)}\left(\boldsymbol{X}(t+\Delta t)\right) \\
&=\; E_{\boldsymbol{X}(t)}\left(E_{\boldsymbol{X}(t+\Delta t)\,|\,\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\,|\,\boldsymbol{X}(t)\right)\right)
\end{aligned} \tag{5.1}$$

$$\begin{aligned}
&E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\,\boldsymbol{X}^T(t+\Delta t)\,|\,\boldsymbol{X}(t)\right) \\
&=\; \mathrm{cov}\left(\boldsymbol{X}(t+\Delta t), \boldsymbol{X}^T(t+\Delta t)\,|\,\boldsymbol{X}(t)\right) \\
&\quad +\; E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\,|\,\boldsymbol{X}(t)\right)\, E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}^T(t+\Delta t)\,|\,\boldsymbol{X}(t)\right).
\end{aligned} \tag{5.2}$$

For the second order reaction

$$R_m: \;\; S_i \;+\; S_j \;\xrightarrow{c_m}\; S_k, \tag{5.3}$$

we do not have a simple expression for the transition probabilities of the above reaction. We use the approximation that in the time interval Δt, the probability that a molecule of species S_i (or S_j) changes into a molecule of species S_k is $c_m \Delta t$, and the probability that a molecule of species S_i (or S_j) remains a molecule of species S_i (or S_j) is $1 - c_m \Delta t$. We also have reactions of the form

$$2S_1 \;\longrightarrow\; S_2, \tag{5.4}$$

where the stoichiometry of species S_1 is -2 and

$$S_1 \;+\; S_2 \;\longrightarrow\; 2S_3, \tag{5.5}$$

where the stoichiometry of species S_3 is 2. Therefore, in the following derivations we take into account v_{mi}, the stoichiometry of species S_i in reaction R_m. We define a matrix $\boldsymbol{W}$ of size $M \times N$ with elements

$$w_{mi} = v_{mi}(c_m \Delta t), \qquad m = 1, 2 \ldots M, \qquad i = 1, 2 \ldots N. \tag{5.6}$$

We further define a vector $\boldsymbol{H}(t)$ of size M such that its mth element $h_m(t)$ is the number

of distinct molecular combinations for reaction R_m as described in Section 1.4.2. We obtain an expression for $E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\,|\,\boldsymbol{X}(t)\right)$ by summing up the contribution for each species from all the reactions. This gives

$$E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\,|\,\boldsymbol{X}(t)\right) \;=\; \boldsymbol{X}(t) + \boldsymbol{W}^T\,\boldsymbol{H}(t). \tag{5.7}$$

Taking expectation $E_{\boldsymbol{X}(t)}\left(\cdot\right)$ of the above expression and setting $\boldsymbol{G}(t) \equiv E_{\boldsymbol{X}(t)}\left(\boldsymbol{H}(t)\right)$, we obtain

$$\boldsymbol{\mu}(t+\Delta t) \;=\; \boldsymbol{\mu}(t) + \boldsymbol{W}^T\,\boldsymbol{G}(t). \tag{5.8}$$

We next evaluate the recursive expression for the covariance matrix. Utilizing (5.7) yields the second term in (5.2)

$$\begin{aligned}
&E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}(t+\Delta t)\,|\,\boldsymbol{X}(t)\right)\,E_{\boldsymbol{X}(t+\Delta t)|\boldsymbol{X}(t)}\left(\boldsymbol{X}^T(t+\Delta t)\,|\,\boldsymbol{X}(t)\right)\\
&=\; \boldsymbol{X}(t)\,\boldsymbol{X}^T(t) \;+\; \boldsymbol{X}(t)\,\boldsymbol{H}^T(t)\,\boldsymbol{W}\\
&\quad+\; \boldsymbol{W}^T\,\boldsymbol{H}(t)\,\boldsymbol{X}^T(t) \;+\; \boldsymbol{W}^T\,\boldsymbol{H}(t)\boldsymbol{H}^T(t)\boldsymbol{W}.
\end{aligned} \tag{5.9}$$

To evaluate the first term in (5.2) we set

$$\boldsymbol{R}(t) \equiv \mathrm{cov}\left(\boldsymbol{X}(t+\Delta t), \boldsymbol{X}^T(t+\Delta t)\,|\,\boldsymbol{X}(t)\right), \tag{5.10}$$

where $\boldsymbol{R}(t)$ is a matrix of size N and its elements are obtained using the multinomial distribution approximation, as in the case of first order reactions. That is, in the time interval $(t, t+\Delta t)$, the reactant species in the reaction channel R_m either remain as reactants or change into its product species, as specified by the stoichiometries. Further, the reaction propensities are proportional to $h_m(t)$ for second order reactions, and therefore we replace K_m by $h_m(t)$ in expressions (4.22) and (4.23). This yields

$$r_{ii}(t) \;=\; \sum_{m=1}^{M} v_{mi}^2\,h_m(t)\,p_{mi}\,(1-p_{mi}) \tag{5.11a}$$

$$r_{ij}(t) \;=\; \sum_{m=1}^{M} v_{mi}\, v_{mj}\, h_m(t)\, p_{mi}\, p_{mj} \qquad\qquad i \neq j, \tag{5.11b}$$

where $p_{mi} = 1 - c_m \Delta t$, if species S_i is a reactant and $p_{mi} = c_m \Delta t$ if S_i is a product in reaction R_m. Marginalizing the above expression gives the elements of $\boldsymbol{B}(t) = E_{\boldsymbol{X}(t)}\left(\boldsymbol{R}(t)\right)$

$$b_{ii}(t) \;=\; \sum_{m=1}^{M} v_{mi}^2\, g_m(t)\, p_{mi}\, (1 - p_{mi}) \tag{5.12a}$$

$$b_{ij}(t) \;=\; \sum_{m=1}^{M} v_{mi}\, v_{mj}\, g_m(t)\, p_{mi}\, p_{mj} \qquad\qquad i \neq j, \tag{5.12b}$$

where $g_m(t)$ are the elements of $\boldsymbol{G}(t)$.

Substituting (5.9) and (5.10) in (5.2), and taking its expectation $E_{\boldsymbol{X}(t)}(\cdot)$ gives the second moments

$$\begin{aligned} &E\left(\boldsymbol{X}(t+\Delta t)\, \boldsymbol{X}^T(t+\Delta t)\right) \\ &\qquad = \; E_{\boldsymbol{X}(t)}\left(\boldsymbol{X}(t)\, \boldsymbol{X}^T(t)\right) + \boldsymbol{B}(t) \;+\; E_{\boldsymbol{X}(t)}\left(\boldsymbol{X}(t)\, \boldsymbol{H}^T(t)\right) \boldsymbol{W} \\ &\qquad\quad + \boldsymbol{W}^T E_{\boldsymbol{X}(t)}\left(\boldsymbol{H}(t)\, \boldsymbol{X}^T(t)\right) \;+\; \boldsymbol{W}^T E_{\boldsymbol{X}(t)}\left(\boldsymbol{H}(t) \boldsymbol{H}^T(t)\right) \boldsymbol{W}. \end{aligned} \tag{5.13}$$

From (5.8) and (5.13), we obtain the following recursive expression for the covariance matrix

$$\begin{aligned} \boldsymbol{C}(t+\Delta t) \;&=\; \boldsymbol{C}(t) + \boldsymbol{B}(t) + \left(E_{\boldsymbol{X}(t)}\left(\boldsymbol{X}(t)\, \boldsymbol{H}^T(t)\right) - \boldsymbol{\mu}(t)\, \boldsymbol{G}^T(t)\right) \boldsymbol{W} \\ &\quad + \boldsymbol{W}^T \left(E_{\boldsymbol{X}(t)}\left(\boldsymbol{H}(t)\, \boldsymbol{X}^T(t)\right) \;- \boldsymbol{G}(t)\, \boldsymbol{\mu}^T(t)\right) \\ &\quad + \boldsymbol{W}^T \left(E_{\boldsymbol{X}(t)}\left(\boldsymbol{H}(t) \boldsymbol{H}^T(t)\right) - \boldsymbol{G}(t) \boldsymbol{G}^T(t)\right) \boldsymbol{W}, \end{aligned} \tag{5.14}$$

where the fourth term is the transpose of the third term.

5.2.1 Example - Simple Reversible Reaction

Consider the reactions

$$R_1 : \; S_1 \; + \; S_2 \; \xrightarrow{c_1} \; S_3 \tag{5.15a}$$

$$R_2 : \; S_3 \; \xrightarrow{c_2} \; S_1 \; + \; S_2. \tag{5.15b}$$

If we know the initial number of molecules of all the species, then it is sufficient to track only one species. We will track species, S_1, and therefore we have $N = 1$, $M = 2$ and $\boldsymbol{X}(t) = X_1(t)$. The stoichiometries of species S_1 in reactions R_1 and R_2 are $\nu_{11} = -1$ and $\nu_{21} = 1$, respectively. The matrices $\boldsymbol{H}(t)$ and $\boldsymbol{W}$ are given by

$$\boldsymbol{H}(t) \; = \; \begin{pmatrix} X_1(t)X_2(t) \\ X_3(t) \end{pmatrix}, \qquad \boldsymbol{W} \; = \; \begin{pmatrix} -c_1 \Delta t \\ c_2 \Delta t \end{pmatrix}. \tag{5.16}$$

The matrix $\boldsymbol{R}$ is of size one, and its element is given by summing up the contributions from reactions R_1 and R_2

$$r_{11} \; = \; X_1(t)X_2(t)c_1\Delta t \left(1 - c_1\Delta t\right) + X_3(t)c_2\Delta t \left(1 - c_2\Delta t\right). \tag{5.17}$$

Utilizing (5.8) and (5.14) and expanding gives us the following expressions for the propagation of the mean and the variance of $X_1(t)$:

$$\mu_1(t + \Delta t) = \mu_1(t) \; - \; E\left(X_1(t)X_2(t)\right) c_1\Delta t \; + \; E\left(X_3(t)\right) c_2\Delta t \tag{5.18}$$

$$\begin{aligned} \sigma_1^2(t + \Delta t) = \sigma_1^2(t) &+ E\left(X_1(t)X_2(t)\right) c_1\Delta t \left(1 - c_1\Delta t\right) \\ &+ \; E\left(X_3(t)\right) c_2\Delta t \left(1 - c_2\Delta t\right) \\ &+ \; \left(c_1\Delta t\right)^2 \left(E\left(X_1^2(t)X_2^2(t)\right) - \left(E\left(X_1(t)X_2(t)\right)\right)^2\right) \\ &+ \; \left(c_2\Delta t\right)^2 \left(E\left(X_3^2(t)\right) - \left(E\left(X_3(t)\right)\right)^2\right) \\ &- \; \left(2c_1\Delta t\right) \left(E\left(X_1^2(t)X_2(t)\right) - E\left(X_1(t)\right) E\left(X_1(t)X_2(t)\right)\right) \\ &+ \; \left(2c_2\Delta t\right) \left(E\left(X_1(t)X_3(t)\right) - E\left(X_1(t)\right) E\left(X_3(t)\right)\right) \\ &- \; \left(2c_1\Delta t c_2\Delta t\right) \left(E\left(X_1(t)X_2(t)X_3(t)\right) - E\left(X_1(t)X_2(t)\right) E\left(X_3(t)\right)\right). \end{aligned} \tag{5.19}$$

In the above expressions $\mu_i(t) = E\left(X_i(t)\right)$, $i = 1, 2, 3$. In addition, we have the constraints $X_1(t) - X_2(t) = \beta_{12}$, $X_1(t) - X_3(t) = \beta_{13}$, on the molecular populations. The evaluation of higher order terms of the form $E\left(X_1^2(t)X_2(t)\right)$ will be discussed in Section 5.3.

5.2.2 Example - Competing Reactions

For the competing reactions

$$R_1 : \; S_1 \; + \; S_2 \quad \xrightarrow{c_1} \quad S_3 \tag{5.20a}$$

$$R_2 : \; S_1 \; + \; S_4 \quad \xrightarrow{c_2} \quad S_5 \tag{5.20b}$$

we track three species S_1, S_2 and S_4. The number of tracked species $N = 3$, the number of reactions $M = 2$ and $\boldsymbol{X}(t) = \begin{pmatrix} X_1(t) & X_2(t) & X_4(t) \end{pmatrix}^T$. The stoichiometries vectors are $\boldsymbol{\nu}_1 = \begin{pmatrix} -1 & -1 & 0 \end{pmatrix}$ and $\boldsymbol{\nu}_2 = \begin{pmatrix} -1 & 0 & -1 \end{pmatrix}$. The matrices $\boldsymbol{H}(t)$ and $\boldsymbol{W}$ are given by

$$\boldsymbol{H}(t) \; = \; \begin{pmatrix} X_1(t)X_2(t) \\ X_1(t)X_4(t) \end{pmatrix}, \qquad \boldsymbol{W} \; = \; \begin{pmatrix} -c_1\Delta t & -c_1\Delta t & 0 \\ -c_2\Delta t & 0 & -c_2\Delta t \end{pmatrix}. \tag{5.21}$$

The matrix

$$\boldsymbol{R}(t) \; = \; \begin{pmatrix} r_{11}(t) & r_{12}(t) & r_{14}(t) \\ r_{21}(t) & r_{22}(t) & r_{24}(t) \\ r_{41}(t) & r_{42}(t) & r_{44}(t) \end{pmatrix}, \tag{5.22}$$

is symmetric and its elements are given by

$$\begin{aligned}
r_{11}(t) &= X_1(t)X_2(t)c_1\Delta t(1 - c_1\Delta t) + X_1(t)X_4(t)c_2\Delta t(1 - c_2\Delta t) \\
r_{12}(t) &= X_1(t)X_2(t)c_1\Delta t(1 - c_1\Delta t) \\
r_{14}(t) &= X_1(t)X_4(t)c_2\Delta t(1 - c_2\Delta t) \\
r_{22}(t) &= X_1(t)X_2(t)c_1\Delta t(1 - c_1\Delta t) \\
r_{24}(t) &= 0 \\
r_{44}(t) &= X_1(t)X_4(t)c_2\Delta t(1 - c_2\Delta t).
\end{aligned}$$

This example will be continued in Section 5.5.1.

5.3 Higher Order Joint Moments

In order to evaluate the expression (5.14), we need to compute the expectations of $\boldsymbol{X}(t)\,\boldsymbol{H}^T(t)$ and $\boldsymbol{H}(t)\,\boldsymbol{H}^T(t)$. These matrices consist of third and higher order terms of the form $X_1(t)X_2(t)X_3(t)$, $X_1^2(t)X_2^2(t)$ and so forth. In general, the expressions for moments of order up to Q consist of moments of order greater than Q. This is the *moment closure* problem and a suitable approximation has to be constructed. We will utilize the multivariate normal distribution approximation for each time step Δt. Gómez-Uribe et al. [32], Goutsias [35] and Lee et al. [47] have implemented moment closure by setting higher order moment terms to zero. It is true that for multivariate normal distribution, the central moments of odd order are zero, and third and higher order cumulants are zero, but the higher order moments contained in the expressions for computations of second moments (covariance matrix in (5.14)) are not necessarily third order central moments or third order cumulants. Singh et al. implemented a *moment closure technique* and obtained expressions for higher order moments that are consistent with lognormal distribution [68], [69]. We will compare the normal and lognormal moment closure expressions for an example presented in [38].

We now derive the moments of multivariate normal distribution, using the moment generating function [36], [74]. The multivariate normal PDF is defined by

$$f_{\boldsymbol{X}(t)}(\boldsymbol{x}(t)) \;=\; \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}(t)|^{1/2}}\,\exp\{-\frac{1}{2}(\boldsymbol{x}(t)-\boldsymbol{\mu}(t))^T\,\boldsymbol{C}^{-1}(t)\,(\boldsymbol{x}(t)-\boldsymbol{\mu}(t))\}. \tag{5.23}$$

Its moment generating function is given by

$$\mathcal{M}_{\boldsymbol{X}(t)}\,(\boldsymbol{\omega}) \;=\; \exp\{\boldsymbol{\mu}^T(t)\,\boldsymbol{\omega}+\frac{1}{2}\boldsymbol{\omega}^T\,\boldsymbol{C}(t)\,\boldsymbol{\omega}\}, \tag{5.24}$$

where $\mathcal{M}_{\boldsymbol{X}(t)}\,(\boldsymbol{\omega})$ is defined by

$$\mathcal{M}_{\boldsymbol{X}(t)}\,(\boldsymbol{\omega}) \;=\; E\left(\exp\{\boldsymbol{\omega}^T\boldsymbol{X}(t)\}\right) \tag{5.25}$$

and $\boldsymbol{\omega}=\begin{pmatrix}\omega_1 & \omega_2 & \ldots & \omega_N\end{pmatrix}$ is a vector of size N. Recall from (4.1) that the joint non-central moments are defined by

$$\gamma^{\boldsymbol{q}}(t) \;=\; E\left(\boldsymbol{X}^{\boldsymbol{q}}(t)\right). \tag{5.26}$$

$E\left(X_i^3(t)\right)$	$\mu_i^3(t) + 3\mu_i(t)\sigma_i^2(t)$
$E\left(X_i^4(t)\right)$	$\mu_i^4(t) + 6\mu_i^2(t)\sigma_i^2(t) + 3\sigma_i^4(t)$
$E\left(X_i^2(t)X_j(t)\right)$	$2\,\mu_i(t)\sigma_{ij}(t) + \left(\mu_i^2(t) + \sigma_i^2(t)\right)\mu_j(t)$
$E\left(X_i(t)X_j(t)X_k(t)\right)$	$\mu_i(t)\mu_j(t)\mu_k(t) + \sigma_{jk}(t)\mu_i(t) + \sigma_{ik}(t)\mu_j(t) + \sigma_{ij}(t)\mu_k(t)$
$E\left(X_i^2(t)X_j^2(t)\right)$	$4\,\mu_i(t)\mu_j(t)\sigma_{ij}(t) + \mu_i^2(t)\mu_j^2(t) + \mu_i^2(t)\sigma_j^2(t)$
	$+\mu_j^2(t)\sigma_i^2(t) + 2\sigma_{ij}^2(t) + \sigma_i^2(t)\sigma_j^2(t)$
$E\left(X_i^2(t)X_j(t)X_k(t)\right)$	$\mu_i^2(t)\mu_j(t)\mu_k(t) + \mu_i^2(t)\sigma_{jk}(t) + 2\mu_i(t)\mu_j(t)\sigma_{ik}(t)$
	$+2\mu_i(t)\mu_k(t)\sigma_{ij}(t) + \sigma_i^2(t)\mu_j(t)\mu_k(t) + \sigma_{jk}(t)\sigma_i^2(t) + 2\sigma_{ij}(t)\sigma_{ik}(t)$

Table 5.1: Higher order expectations expressed in terms of the first and second moments for the multivariate normal distribution.

These moments are computed by utilizing (5.24) according to:

$$\gamma^{\boldsymbol{q}}(t) \;=\; \frac{\partial^{q_1+q_2\ldots q_N}}{\partial\omega_1^{q_1}\partial\omega_2^{q_2}\ldots\partial\omega_N^{q_N}}\mathcal{M}_{\boldsymbol{X}(t)}(\boldsymbol{\omega})\bigg|_{\boldsymbol{\omega}=\boldsymbol{0}}. \tag{5.27}$$

The joint moments required for the computation of expectations in $\boldsymbol{X}(t)\,\boldsymbol{H}^T(t)$ and $\boldsymbol{H}(t)\,\boldsymbol{H}^T(t)$ are listed in Table 5.1.

5.4 Algorithm Implementation

We choose a time step Δt such that the average number of reactions in each reaction channel is less than a specified value. For the three types of reactions

$$R_1 :\; S_1 \;\longrightarrow\; S_2 \tag{5.28a}$$

$$R_2 :\; 2S_1 \;\longrightarrow\; S_2 \tag{5.28b}$$

$$R_3 :\; S_1 \;+\; S_2 \;\longrightarrow\; S_3, \tag{5.28c}$$

choose a time step Δt such that it satisfies

$$c_1 \Delta t \ \leq \epsilon \tag{5.29a}$$

$$x_1(t) c_2 \Delta t \ \leq \epsilon \tag{5.29b}$$

$$K_m(t) c_3 \Delta t \ \leq \epsilon, \qquad K_m(t) = \max(x_1(t), x_2(t)), \tag{5.29c}$$

where ϵ is some specified value. Note that the time step maybe updated once every few recursions. For second order reactions, the time step depends on the molecular populations.

Given the reaction stoichiometries, reaction constants, the initial number of molecules $\boldsymbol{X}(t=0)$ and the stop time T_s, the algorithm is implemented as follows:

Initialization

- Define the matrices $\boldsymbol{X}(t)$ and $\boldsymbol{H}(t)$, for time t.
- Set time $t=0$,

  $\boldsymbol{\mu}(t=0) = \boldsymbol{X}(t=0)$,

  $\boldsymbol{C}(t=0) = \boldsymbol{0}$.

Recursion

Repeat the following steps until stop time $t=T_s$ is reached:

- Update the step size Δt as described above.
- Compute the elements of the matrix $\boldsymbol{W}$ as given by (5.6).
- Compute the elements of the matrix $\boldsymbol{R}(\boldsymbol{t})$ as given by (5.11a) and (5.11b).
- Evaluate the expectations $E_{\boldsymbol{X}(t)}\left(\boldsymbol{H}(t)\right)$ and $E_{\boldsymbol{X}(t)}\left(\boldsymbol{R}(t)\right)$.
- Evaluate the expectations $E_{\boldsymbol{X}(t)}\left(\boldsymbol{X}(t)\,\boldsymbol{H}^T(t)\right)$ and $E_{\boldsymbol{X}(t)}\left(\boldsymbol{H}(t)\,\boldsymbol{H}^T(t)\right)$ from Table 5.1.
- Compute $\boldsymbol{\mu}(t+\Delta t)$ by employing expression (5.8).
- Compute $\boldsymbol{C}(t+\Delta t)$ by employing expression (5.14).
- Advance the time to $t+\Delta t$.

5.5 Simulation Results

As with the first order system, we compare the simulation results from the second order expressions with the SSA, for accuracy and computation time. We track the means and the covariance matrix with time. We also compare plots of normal distribution $\mathcal{N}(\mu_i(t), \sigma_i^2(t))$, for species S_i at time t, using the mean and the variance computed from the RM method, with the distribution generated by the SSA method.

5.5.1 Example - Competing Reactions (continued)

We continue the example from Section 5.2.2. We have

$$\boldsymbol{\mu}(t) = \begin{pmatrix} \mu_1(t) \\ \mu_2(t) \\ \mu_4(t) \end{pmatrix} \quad \boldsymbol{C}(t) = \begin{pmatrix} \sigma_1^2(t) & \sigma_{12}(t) & \sigma_{14}(t) \\ \sigma_{21}(t) & \sigma_2^2(t) & \sigma_{24}(t) \\ \sigma_{41}(t) & \sigma_{42}(t) & \sigma_4^2(t) \end{pmatrix}. \tag{5.30}$$

In order to compute $\boldsymbol{\mu}(t + \Delta t)$ and $\boldsymbol{C}(t + \Delta t)$, we need to evaluate the expectations $\boldsymbol{G}(t) \equiv E_{\boldsymbol{X}(t)}(\boldsymbol{H}(t))$ and $\boldsymbol{B}(t) = E_{\boldsymbol{X}(t)}(\boldsymbol{R}(t))$. In addition, we need the expectations of $\boldsymbol{H}(t)\,\boldsymbol{X}^T(t)$, $\boldsymbol{X}(t)\,\boldsymbol{H}^T(t)$ and $\boldsymbol{H}(t)\,\boldsymbol{H}^T(t)$. The expressions for $\boldsymbol{X}(t)\,\boldsymbol{H}^T(t)$ and $\boldsymbol{H}(t)\,\boldsymbol{H}^T(t)$ are

$$\boldsymbol{X}(t)\boldsymbol{H}^T(t) = \begin{pmatrix} X_1^2(t)X_2(t) & X_1^2(t)X_4(t) \\ X_1(t)X_2^2(t) & X_1(t)X_2(t)X_4(t) \\ X_1(t)X_2(t)X_4(t) & X_1(t)X_4^2(t) \end{pmatrix} \tag{5.31}$$

$$\boldsymbol{H}(t)\boldsymbol{H}^T(t) = \begin{pmatrix} X_1^2(t)X_2^2(t) & X_1^2(t)X_2(t)X_4(t) \\ X_1^2(t)X_2(t)X_4(t) & X_1^2(t)X_4^2(t) \end{pmatrix}. \tag{5.32}$$

The required expectations are evaluated using joint moments of the multivariate normal distribution as described in Section 5.3. Some of these terms are

$$
\begin{aligned}
E\left(X_1(t)X_2(t)\right) &= \sigma_{12}(t) + \mu_1(t)\mu_2(t) && (5.33)\\
E\left(X_1^2(t)X_2(t)\right) &= 2\,\mu_1(t)\sigma_{12}(t) + \left(\mu_1^2(t) + \sigma_1^2(t)\right)\mu_2(t) && (5.34)\\
E\left(X_1^2(t)X_4(t)\right) &= 2\,\mu_1(t)\sigma_{14}(t) + \left(\mu_1^2(t) + \sigma_1^2(t)\right)\mu_4(t) && (5.35)\\
E\left(X_1(t)X_2^2(t)\right) &= 2\,\mu_2(t)\sigma_{21}(t) + \left(\mu_2^2(t) + \sigma_2^2(t)\right)\mu_1(t) && (5.36)\\
E\left(X_1(t)X_2(t)X_4(t)\right) &= \mu_1(t)\mu_2(t)\mu_4(t) + \sigma_{24}(t)\mu_1(t) \\
&\quad + \sigma_{14}(t)\mu_2(t) + \sigma_{12}(t)\mu_4(t). && (5.37)
\end{aligned}
$$

The expressions for $\boldsymbol{\mu}(t)\,\boldsymbol{G}^T(t)$ and $\boldsymbol{G}(t)\,\boldsymbol{G}^T(t)$ are

$$
\boldsymbol{\mu}(t)\boldsymbol{G}^T(t) = \begin{pmatrix} \mu_1(t)E\left(X_1(t)X_2(t)\right) & \mu_1(t)E\left(X_1(t)X_4(t)\right) \\ \mu_2(t)E\left(X_1(t)X_2(t)\right) & \mu_2(t)E\left(X_1(t)X_4(t)\right) \\ \mu_4(t)E\left(X_1(t)X_2(t)\right) & \mu_4(t)E\left(X_1(t)X_4(t)\right) \end{pmatrix} \tag{5.38}
$$

$$
\boldsymbol{G}(t)\boldsymbol{G}^T(t) = \begin{pmatrix} \left(E\left(X_1(t)X_2(t)\right)\right)^2 & E\left(X_1(t)X_2(t)\right)E\left(X_1(t)X_4(t)\right) \\ E\left(X_1(t)X_2(t)\right)E\left(X_1(t)X_4(t)\right) & \left(E\left(X_1(t)X_4(t)\right)\right)^2 \end{pmatrix}. \tag{5.39}
$$

The simulation results from the RM method were compared with the SSA method for the following parameters: $X_1(t=0) = 10,000$, $X_2(t=0) = 4,000$, $X_4(t=0) = 1,000$, $c_1 = 0.0001$ sec^{-1}, $c_2 = 0.0002$ sec^{-1} and $t = 1.0$ sec. In the RM method, $\Delta t = 0.001$ sec. The noisy results from the SSA method in Figure 5.1 are due to insufficient realizations of $1,000$. With $10,000$ realizations in the SSA method (Figure 5.2), the results are less noisy and the RM method is in very good agreement with the SSA method.

The CPU time with the RM method on the Fermat machine is 0.09 sec and with the SSA method it is 1529 sec for $10,000$ realizations. For a value of $\epsilon = 0.001$, we obtain $\Delta t = 0.001$ sec at time $t = 0$. As the CPU time with the RM method is very small, we do not vary the time step during the course of the simulation.

5.5.2 Example - Dimerising Reaction

In the following reaction

$$R_1: \quad 2S_1 \quad \longrightarrow \quad S_2, \tag{5.40}$$

the monomer S_1 dimerises to S_2. We track species S_1, and therefore $N = 1$, the number of reactions $M = 1$ and the state of the system is given by $X_1(t)$. We also have

$$\nu_{11} = -2, \qquad w_{11} = -2c_1\Delta t, \qquad h_1(t) = \frac{1}{2}X_1(t)\left(X_1(t) - 1\right), \tag{5.41}$$

and

$$r_{11}(t) \;=\; 2X_1(t)\left(X_1(t) - 1\right)\; p_{11}\left(1 - p_{11}\right), \tag{5.42}$$

where $p_{11} = 1 - c_1\Delta t$. The expressions for the first and the second moments are

$$\mu_1(t + \Delta t) \;= \mu_1(t) + c_1\Delta t\,\mu_1(t) - c_1\Delta t E_{\boldsymbol{X}(t)}\left(X_1^2(t)\right) \tag{5.43}$$

$$\begin{aligned} E_{\boldsymbol{X}(t+\Delta t)}\left(X_1^2(t + \Delta t)\right) \;&= E\left(X_1^2(t)\right) + 2E_{\boldsymbol{X}(t)}\left(X_1^2(t)\right) c_1\Delta t\left(1 - c_1\Delta t\right) \\ &- \;2\mu_1(t)c_1\Delta t\left(1 - c_1\Delta t\right) - 2c_1\Delta t E\left(X_1^2(t)\left(X_1(t) - 1\right)\right) \\ &+ \;\left(c_1\Delta t\right)^2 E\left(X_1^2(t)\left(X_1(t) - 1\right)^2\right). \end{aligned} \tag{5.44}$$

The simulation of this reaction and its comparison with other methods will be depicted in Chapter 6.

5.5.3 Example - Ten Species

We now illustrate an example with the number of species $N = 10$, the number of reactions $M = 8$. The reactions and the rate constants are listed in Table 5.2 and the initial molecular populations are given in Table 5.3. Results are displayed at time $t = 0.5$ sec in Figure 5.3 and at time $t = 2.0$ sec in Figure 5.4. In the RM method, $\Delta t = 0.01$ sec, and in the SSA method $2,000$ realizations were used. The results from the two methods are in good agreement.

5.5.4 Example - Viral Kinetics

We simulate the Viral Kinetics example described in Section 1.3.1 [73]. The species are $mRNA$, DNA, Pr (protein) and Vr (progeny virus). The reactions and the rate constants are given in Table 5.4. For the reaction constants given in this table, the molecular populations have the steady state values 20, 200 and 10000 for the species $mRNA$, DNA and Pr, respectively. We start simulations with molecular populations of 2, 200 and $10,000$ for $mRNA$, DNA and Pr, respectively. It takes several days for the mRNA population to reach its steady state value. The time step in the RM method is 0.0001 day. The simulations results are displayed in Figures 5.5 and 5.6, at time $t = 1.0$ day, using $80,000$ and $150,000$ realizations, respectively, in the SSA method. The CPU time on the Fermat machine is 0.8 sec using the RM method. Due to very large variance, the plots from SSA simulations are very noisy. The simulation times using the SSA method is about 16 hours with $80,000$ realizations and 31 hours with $150,000$ realizations. Figure 5.7 displays the results using the RM method at time $t = 20.0$ days.

Reaction	Reaction Rate $\left(\text{sec}^{-1}\right)$
$S_1 + S_2 \longrightarrow S_3$	0.001
$S_1 + S_4 \longrightarrow S_5$	0.002
$S_3 \longrightarrow S_1 + S_2$	0.01
$S_5 \longrightarrow S_1 + S_4$	0.01
$S_5 \longrightarrow S_6$	0.0001
$S_6 \longrightarrow S_7 + S_8$	0.0003
$S_7 \longrightarrow S_9$	0.05
$S_8 \longrightarrow S_{10}$	0.05

Table 5.2: The reactions and rate constants for the Ten Species example in Section 5.5.3.

Species	Initial Population
S_1	1000
S_2	400
S_3	0
S_4	200
S_5	0
S_6	1000
S_7	400
S_8	500
S_9	0
S_{10}	0

Table 5.3: Initial molecular populations for the Ten Species example in Section 5.5.3.

Reaction	Reaction Rate $\left(\text{day}^{-1}\right)$
$DNA \longrightarrow mRNA$	1.0
$mRNA \longrightarrow DNA + mRNA$	0.025
$mRNA \longrightarrow Pr + mRNA$	1000
$mRNA \longrightarrow \emptyset$	0.25
$Pr \longrightarrow \emptyset$	1.9985
$DNA + Pr \longrightarrow Vr$	7.5×10^{-6}

Table 5.4: The reactions and rate constants for the Viral Kinetics example in Section 5.5.4.

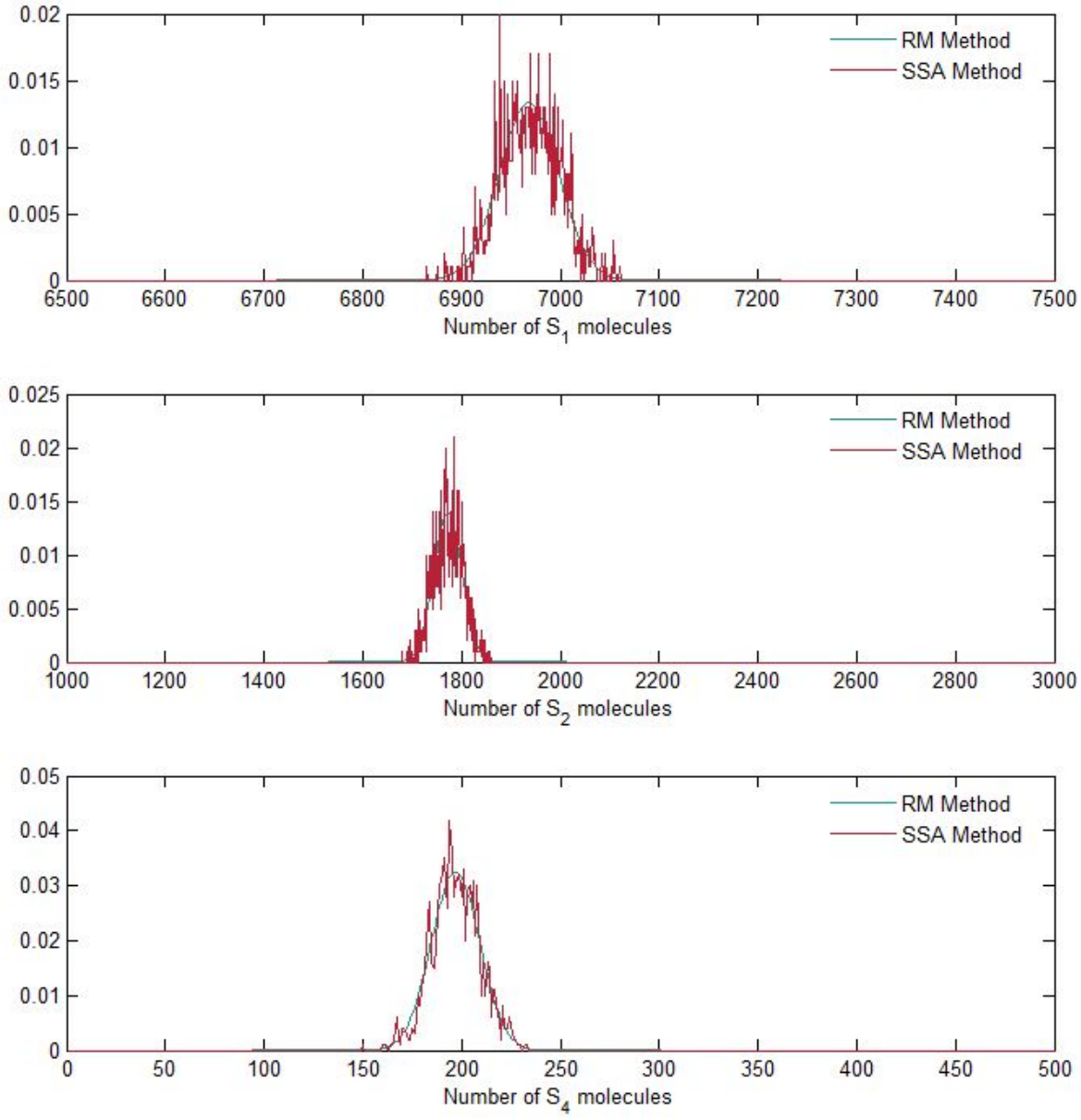

Figure 5.1: Simulation results from the RM (solid line) and the SSA ("noisy" plot) methods for the Competing Reaction system in Section 5.5.1. The $1,000$ realizations used in the SSA method are not adequate.

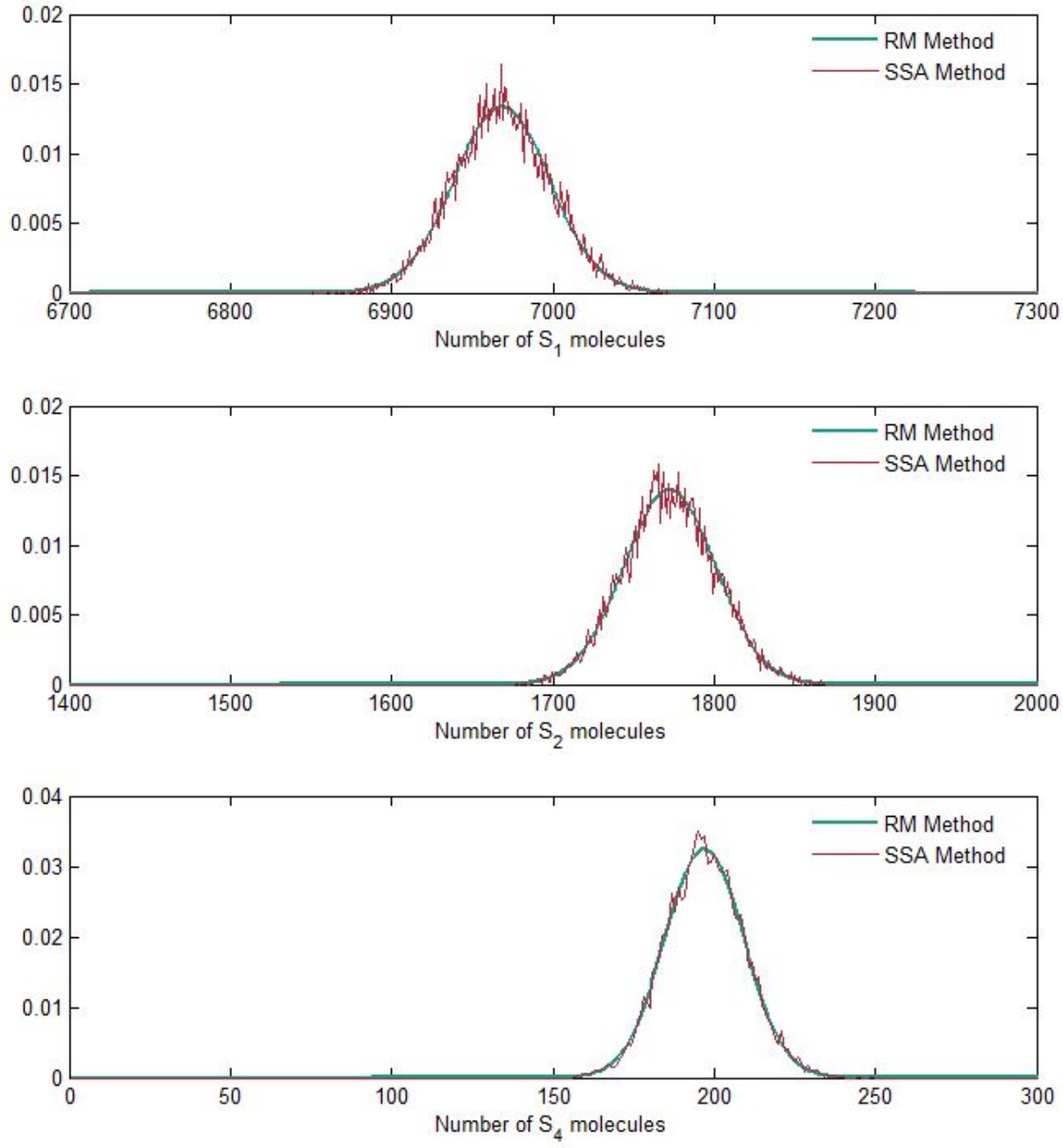

Figure 5.2: Same example as in Figure 5.1, except that the SSA method utilized $10,000$ realizations, and therefore the plots are less "noisy".

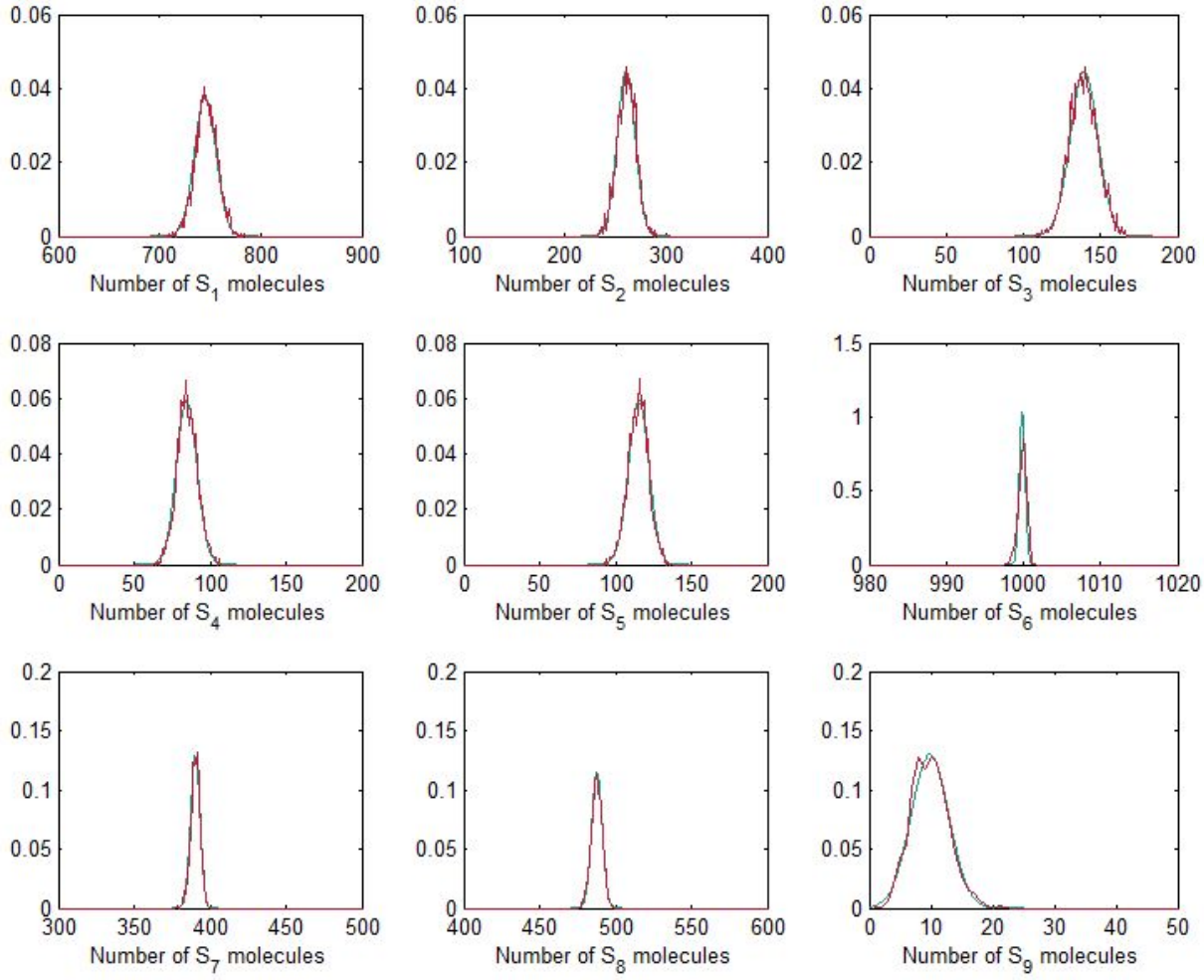

Figure 5.3: Simulation results from the RM (solid line) and the SSA ("noisy" plot) methods, at time $t = 0.5$ sec, for the Ten Species example described in Section 5.5.3.

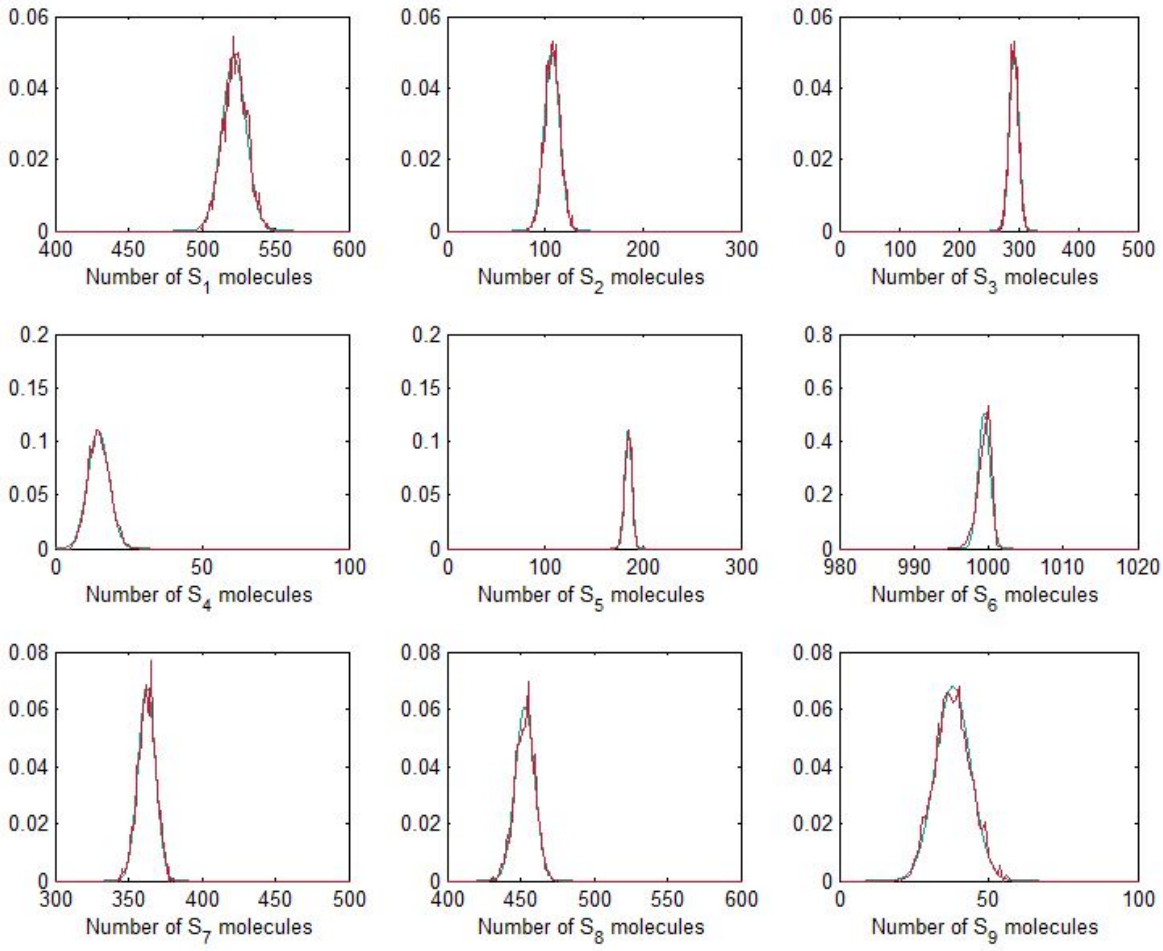

Figure 5.4: Simulation results from the RM (solid line) and the SSA ("noisy" plot) methods, at time $t = 2.0$ sec, for the Ten Species example described in Section 5.5.3.

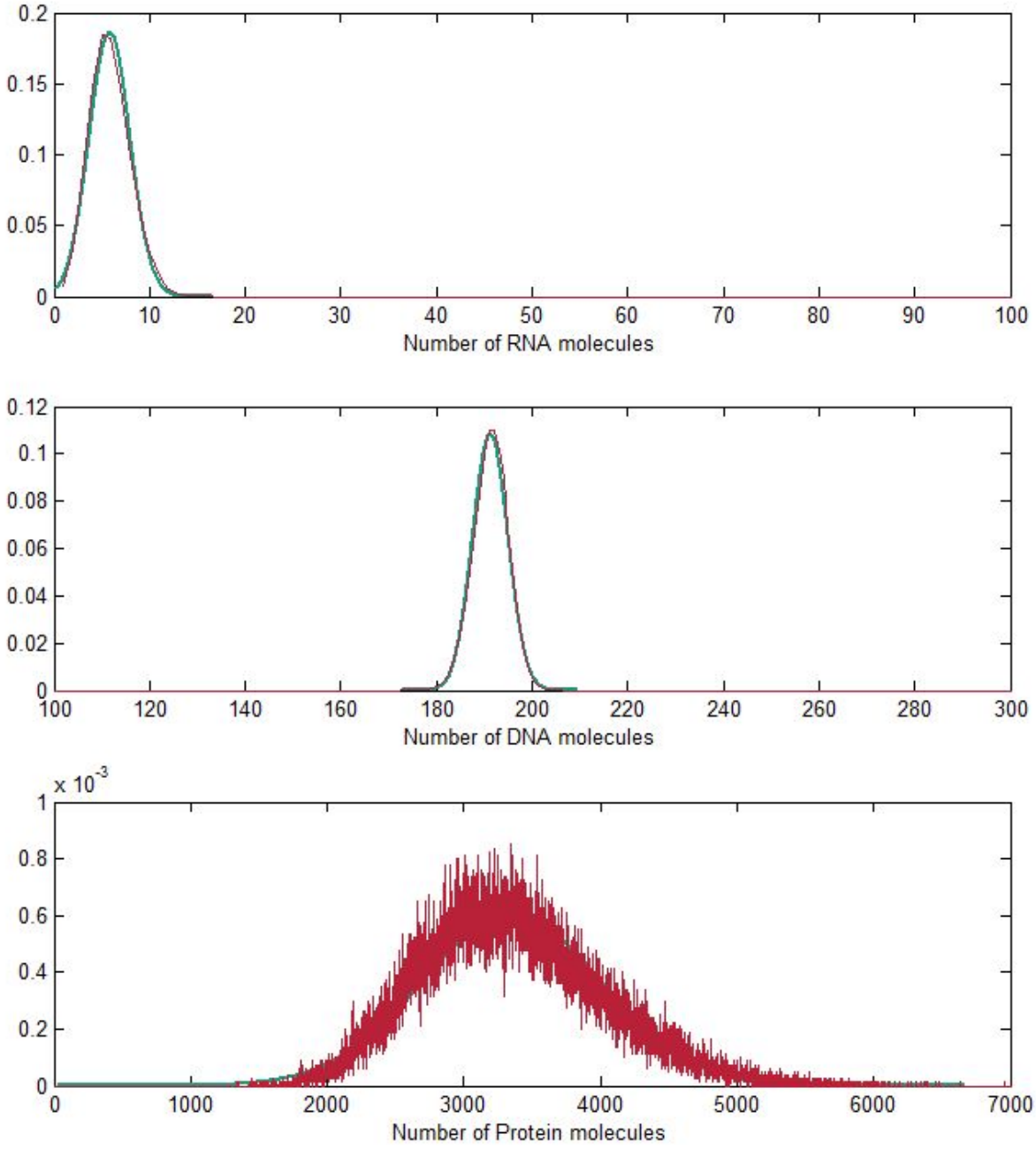

Figure 5.5: Simulation results from the RM (solid line) and the SSA ("noisy" plot) methods, at time $t = 1$ day, for the Viral Kinetics example described in Section 5.5.4. The number of realizations in the SSA method is $80,000$.

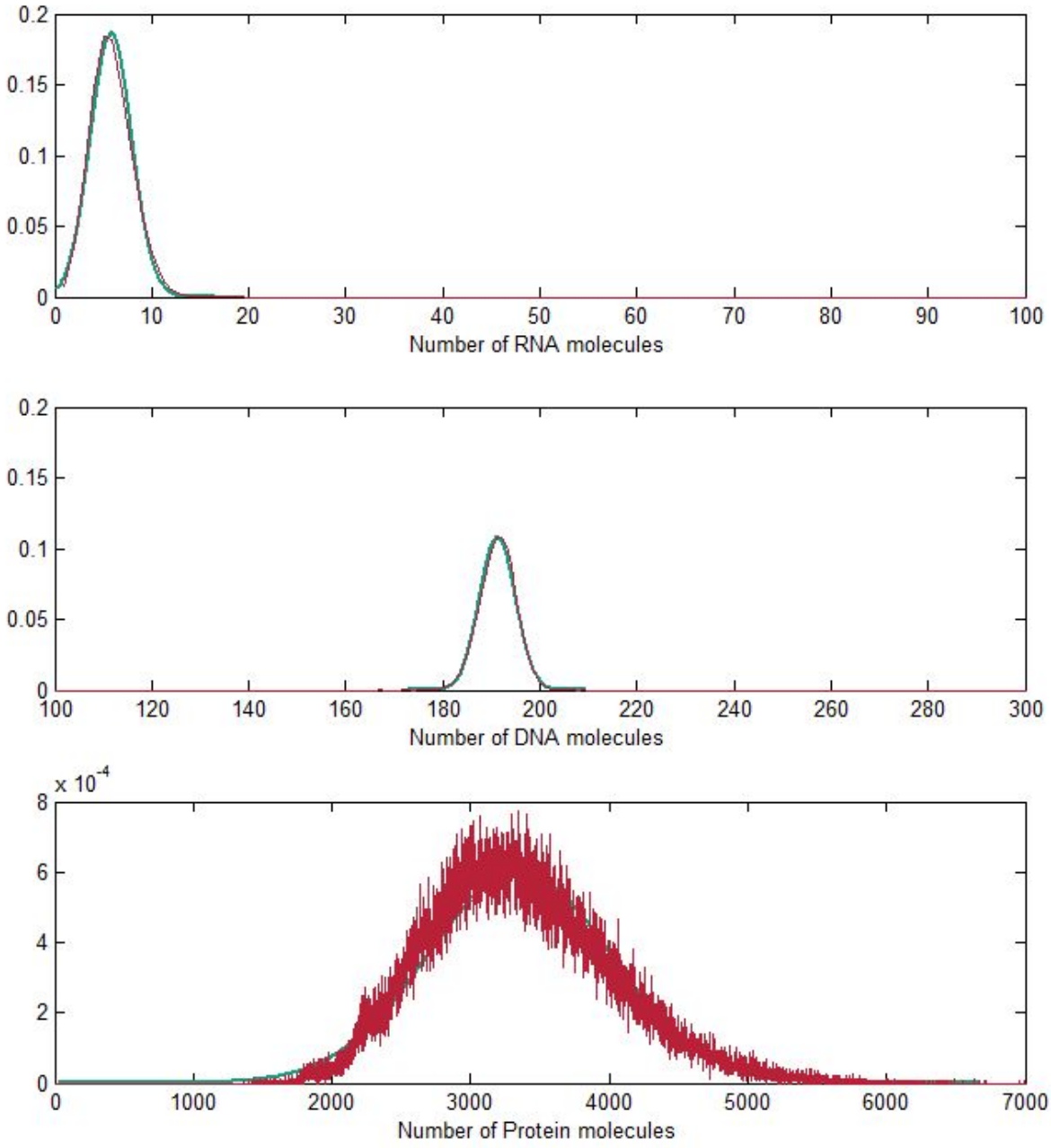

Figure 5.6: Same example as in Figure 5.5, except that the SSA method utilized $150,000$ realizations.

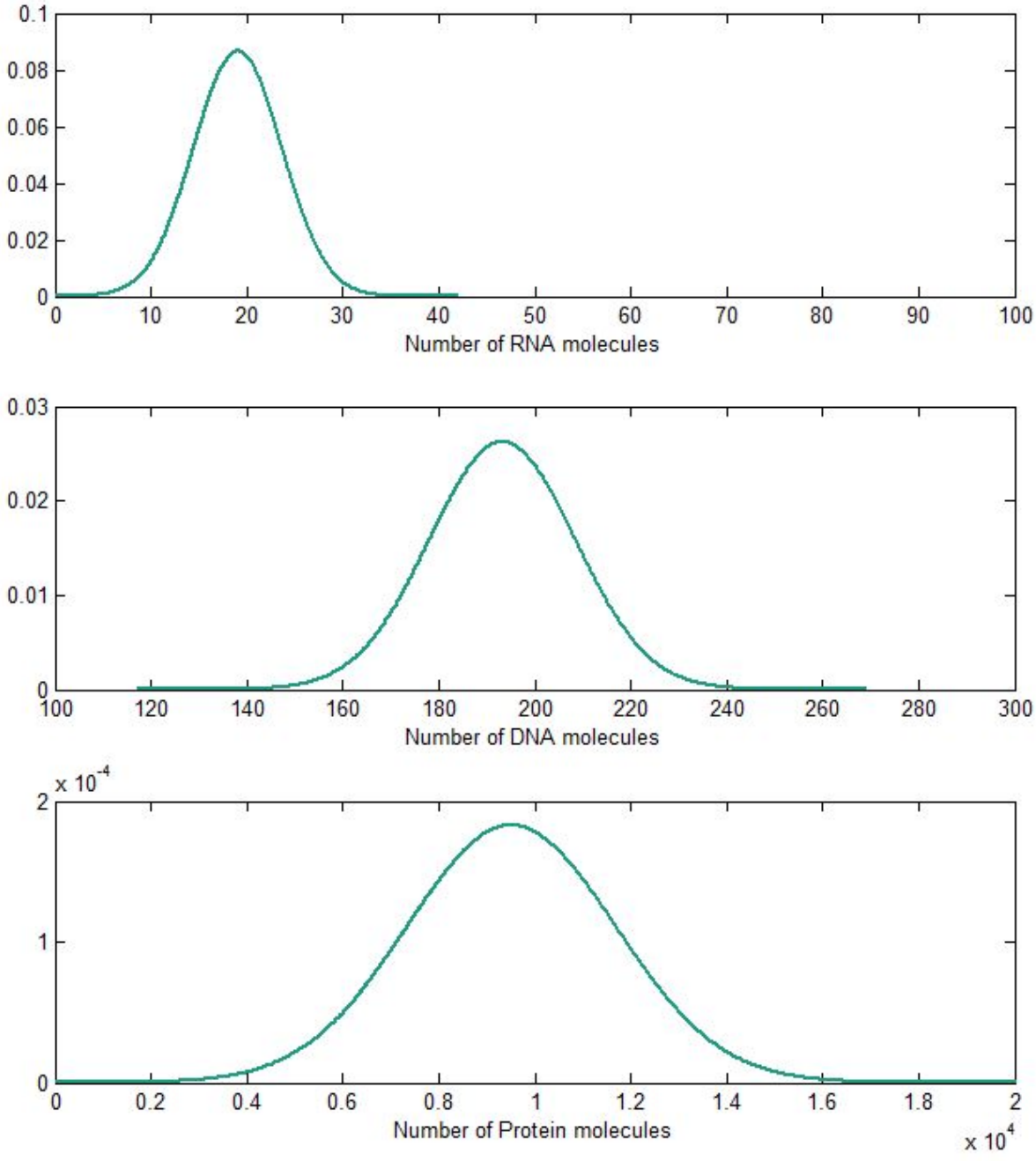

Figure 5.7: Viral Kinetics simulation, using the RM method, at time $t = 20$ days.

Chapter 6

Simulation and Comparison with other methods

6.1 Introduction

As described in Chapter 3, the most popular methods of obtaining the time evolution of the joint probability density in a biochemical system are the Monte Carlo methods, such as the stochastic simulation algorithm and the τ-leap methods.

The moment propagation methods, in which we trace a few lower order moments of $P(\mathtt{x}, t)$, were briefly introduced by Érdi [15] and van Kampen [76]. In 1992, Gillespie utilized the propagator of the Markov process to derive the moment evoluation equations. However, he describes his approximation procedure for moment closure as "lengthy and tedious" (Appendix C, [24]). The procedure consists of solving several sets of differential equations, truncated at different orders, and comparing successive solutions. In their 2007 paper, Gómez-Uribe et al. derived equations to track the means, variances and the covariances, starting from the CME [32]. They term their equations *Mass Fluctuation Kinetics (MFK)*, and they present several examples to illustrate their method. In their 2008 paper, the authors separate the slow and fast reactions to obtain reduced slow and fast CMEs [33]. However, in evaluating the expressions for the second moments, they ignore the third moments and therefore their results are not as accurate as the SSA (Figures 2 and 3, [32]). In 2005, Hespanha et al. [38] and Singh et al. [67] presented a *Stochastic Hybrid Systems (SHS)*

method to compute the first few moments of $P(\mathtt{x}, t)$, and a moment closure technique to evaluate the higher order moments that arise in the expressions for these moments. They illustrate several examples with a few species and reactions. Several other approaches have been presented, such as the Linear Noise Approximation [43] and utilizing the moment generating function [21]. Since the SHS method is better developed than the other methods mentioned above, we will compare the Proposed Method with the SHS.

6.2 Stochastic Hybrid Systems Method

A *hybrid system* is a dynamical system consisting of discrete and continuous states, where the continuous state is described by a differential equation and the discrete states reset the continuous state according to some pre-specified conditions [37], [49]. A *stochastic hybrid system (SHS)* is a hybrid system where both the continuous and the discrete states maybe random [19], [39]. For example, the continuous state maybe described by a stochastic differential equation and the transitions between the continuous states are triggered by a Poisson process. We now present the stochastic hybrid system described by Hespanha et al. and reproduce their results for the evolution of a few lower-order moments. Their work is based on *piecewise deterministic Markov process (PDMP)* introduced by Davis [10]. A PDMP consists of a continuous state described by a deterministic differential equation, and the transition between continuous states described by a stochastic jump process, with state dependent transition probabilities. Davis gives a complete characterization of the extended generator of such processes, which is relevant for computing the necessary expectations.

For the biochemical system defined in Section 4.2 with N molecular species and M reactions, the SHS consists of the following:

- A hybrid state $(\boldsymbol{X}(t), \boldsymbol{\theta}(t))$, where $\boldsymbol{X}(t) : [0, \infty) \to \mathbb{R}^N$ is a stochastic process called the continuous state, $\boldsymbol{\theta}(t) : [0, \infty) \to \boldsymbol{\Theta}$ is a jump process called the discrete state, $\mathbb{R}^N$ denotes the N-dimensional Euclidean space and $\boldsymbol{\Theta}$ denotes a discrete set.

- A differential equation which governs $\boldsymbol{X}(t)$

$$\frac{d\boldsymbol{X}(t)}{dt} = g(\boldsymbol{\theta}, \boldsymbol{X}, t), \qquad g : \boldsymbol{\Theta} \times \mathbb{R}^N \times [0, \infty) \to \mathbb{R}^N. \tag{6.1}$$

- A family of reset maps

$$\boldsymbol{\phi}_m\left(\boldsymbol{\theta}^-, \boldsymbol{X}^-, t\right), \qquad \boldsymbol{\phi}_m : \boldsymbol{\Theta} \times \mathbb{R}^N \times [0,\infty) \to \boldsymbol{\Theta} \times \mathbb{R}^N \tag{6.2}$$

  for the mth reaction defined by the stoichiometry of the species for that reaction. These maps determine how the continuous states are reset.

- A family of transition intensities

$$\lambda_m\left(\boldsymbol{\theta}, \boldsymbol{X}, t\right) = c_m h_m(\boldsymbol{X}, t), \qquad \lambda_m : \boldsymbol{\Theta} \times \mathbb{R}^N \times [0,\infty) \to [0,\infty) \tag{6.3}$$

  for the mth reaction, where c_m is the stochastic rate constant and $h_m(\boldsymbol{X}, t)$ is the number of distinct molecular combinations of reaction R_m as defined in Section 1.4.2.

For the above biochemical system, the continuous state evolves as follows:

$$\frac{d\boldsymbol{X}(t)}{dt} \;=\; 0. \tag{6.4}$$

This is because the molecular populations change only when there is a reaction, and these are set by the reset map defined in (6.2).

In a *polynomial stochastic hybrid system (pSHS)*, the reset maps $\boldsymbol{\phi}_m\left(\boldsymbol{\theta}, \boldsymbol{X}, t\right)$ and the transition intensities $\lambda_m\left(\boldsymbol{\theta}, \boldsymbol{X}, t\right)$ are all polynomial functions of $\boldsymbol{X}(t)$. Let the joint moments about the origin be given by

$$\gamma^{\boldsymbol{q}}(t) \;\equiv\; \gamma^{(q_1, q_2 \ldots q_N)}(t) \;=\; E\left(X_1^{q_1}(t) X_2^{q_2}(t) \ldots X_N^{q_N}(t)\right), \tag{6.5}$$

where $\boldsymbol{q} = (q_1\ q_2\ \ldots\ q_N) \in \mathbb{N}^N$ and $\mathbb{N}$ is the set of positive integers. Recall that the order of the moments Q is defined by $Q = \sum_{n=1}^{N} q_n$. The authors show that the time evolution of $\gamma^{\boldsymbol{q}}(t)$ is described by

$$\frac{d\gamma^{\boldsymbol{q}}(t)}{dt} \;=\; E\left(\sum_{m=1}^{M}\left(\boldsymbol{\phi}_m^{\boldsymbol{q}}\left(\boldsymbol{\theta}, \boldsymbol{X}, t\right) - \boldsymbol{X}^{\boldsymbol{q}}(t)\right)\lambda_m\left(\boldsymbol{\theta}, \boldsymbol{X}, t\right)\right). \tag{6.6}$$

In the above equation, $\boldsymbol{X}^{\boldsymbol{q}}(t)$ denotes the monomial $X_1^{q_1}(t) X_2^{q_2}(t) \ldots X_N^{q_N}(t)$.

6.2.1 Moment Closure

If the biochemical system consists of second order reactions, then the expressions for moments of order up to Q consists of moments of order greater than Q. The authors approximate the higher order moments as nonlinear functions of the lower order moments using a *moment closure technique* [67], [68]. The results for computing these nonlinear functions are presented here. For detailed derivations, we refer you to their papers.

Let $\boldsymbol{\gamma}(t) = \begin{pmatrix}\gamma^{\boldsymbol{q}_1}(t) & \gamma^{\boldsymbol{q}_2}(t) & \dots & \gamma^{\boldsymbol{q}_J}(t)\end{pmatrix}^T$ be a vector which contains all moments of order up to Q and $\bar{\boldsymbol{\gamma}}(t) = \begin{pmatrix}\gamma^{\bar{\boldsymbol{q}}_1}(t) & \gamma^{\bar{\boldsymbol{q}}_2}(t) & \dots & \gamma^{\bar{\boldsymbol{q}}_{J'}}(t)\end{pmatrix}^T$ be a vector which contains moments of order greater than Q, required for the evaluation of the moments in $\boldsymbol{\gamma}(t)$. The sizes of $\boldsymbol{\gamma}(t)$ and $\bar{\boldsymbol{\gamma}}(t)$ are J and J', respectively. The time evolution of $\boldsymbol{\gamma}$, omitting the $t's$, maybe expressed as

$$\frac{d\boldsymbol{\gamma}}{dt} \;=\; \boldsymbol{V}_1\boldsymbol{\gamma} + \boldsymbol{V}_2\bar{\boldsymbol{\gamma}}, \tag{6.7}$$

where the matrices $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are determined from (6.6). The above system is closed by approximating the elements of $\bar{\boldsymbol{\gamma}}$ as nonlinear functions of the elements of $\boldsymbol{\gamma}$. Let $\boldsymbol{\upsilon}$ denote the new approximate system. Then the truncated moment dynamics is given by

$$\frac{d\boldsymbol{\upsilon}}{dt} \;=\; \boldsymbol{V}_1\boldsymbol{\upsilon} + \boldsymbol{V}_2\boldsymbol{\varphi}(\boldsymbol{\upsilon}), \tag{6.8}$$

where $\boldsymbol{\varphi}(\boldsymbol{\upsilon})$ is the vector of *moment closure function*. Its elements, denoted by $\varphi(\boldsymbol{\upsilon})$, are determined by matching the time derivatives of $\boldsymbol{\gamma}$ and $\boldsymbol{\upsilon}$ with deterministic initial conditions. That is

$$\boldsymbol{\gamma}(t_0) \;=\; \boldsymbol{\upsilon}(t_0) \Longrightarrow \left.\frac{d^k\boldsymbol{\gamma}}{dt^k}\right|_{t=t_0} \;=\; \left.\frac{d^k\boldsymbol{\upsilon}}{dt^k}\right|_{t=t_0}, \qquad \forall\; k = 1, 2, \dots K, \tag{6.9}$$

where K is selected to be large enough such that $\boldsymbol{\gamma}$ remains "close" to $\boldsymbol{\upsilon}$. The authors assume that $\varphi(\boldsymbol{\upsilon})$ have the separable form

$$\varphi(\boldsymbol{\upsilon}) \;\equiv\; \varphi(\boldsymbol{\gamma}) \;=\; (\gamma^{\boldsymbol{q}_1})^{\eta_1}(\gamma^{\boldsymbol{q}_2})^{\eta_2}\dots(\gamma^{\boldsymbol{q}_J})^{\eta_J}, \tag{6.10}$$

where $\gamma^{\boldsymbol{q}_j}$, $j = 1 \dots J$, are the elements of $\boldsymbol{\gamma}$ and η_1, η_2, $\dots$ η_J are determined as described below. For any two vectors $\boldsymbol{q} = (q_1, q_2 \dots q_N)$ and $\bar{\boldsymbol{q}} = (\bar{q}_1, \bar{q}_2 \dots \bar{q}_N)$, define

$$\begin{pmatrix}\boldsymbol{q}\\ \bar{\boldsymbol{q}}\end{pmatrix} \;=\; \begin{pmatrix}q_1\\ \bar{q}_1\end{pmatrix}\begin{pmatrix}q_2\\ \bar{q}_2\end{pmatrix}\dots\begin{pmatrix}q_N\\ \bar{q}_N\end{pmatrix}. \tag{6.11}$$

Then for each element $\gamma^{\bar{\boldsymbol{q}}_{j'}}$ in $\bar{\boldsymbol{\gamma}}$, η_1, η_2, $\ldots$ η_J are determined from the equations

$$\begin{pmatrix} \bar{\boldsymbol{q}}_{j'} \\ \boldsymbol{q}_j \end{pmatrix} = \sum_{n=1}^{J} \eta_n \begin{pmatrix} \boldsymbol{q}_n \\ \boldsymbol{q}_j \end{pmatrix} \qquad \forall\, j = 1, 2, \ldots J. \tag{6.12}$$

The authors indicate that their moment closure technique, which gives expressions for higher order moments as functions of lower order moments are consistent with *lognormal* distributions (p. 5004 in [69]).

In their earlier papers (Eq. (14) in [38]), the authors maintain that in order to satisfy the conditions in (6.10) and (6.12), for every moment $\gamma^{\boldsymbol{q}_{j'}}$ in $\bar{\boldsymbol{\gamma}}$, the polynomial $\sum_{i=1}^{\infty} u_{j',i}\, \boldsymbol{x}^{\boldsymbol{q}_i}$ must belong to the linear subspace generated by the polynomials

$$\sum_{i=1}^{\infty} u_{j,i}\, \boldsymbol{x}^{\boldsymbol{q}_{j'} - \boldsymbol{q}_j + \boldsymbol{q}_i}, \qquad j = 1 \ldots J. \tag{6.13}$$

However, in their recent paper, the authors have dropped the condition in (6.13) [69]. We illustrate the method with an example.

6.2.2 Example - Dimerising Reaction

Consider the dimerising reaction

$$R_1 : \quad 2S_1 \quad \longrightarrow \quad S_2. \tag{6.14}$$

The state of the system is $\boldsymbol{X}(t) = X_1(t)$. Omitting the t's, the reset map for the reaction is $\phi_1(\boldsymbol{X}) = X_1 - 2$ and the transition intensity is $\lambda_1(\boldsymbol{X}) = \frac{c_1}{2} X_1(X_1 - 1)$.

Let $\boldsymbol{\gamma} = (\gamma^1 \quad \gamma^2)^T$ and $\bar{\boldsymbol{\gamma}} = (\gamma^3)$. From (6.6), we obtain the expressions for the time evolution of $\boldsymbol{\gamma}$

$$\begin{pmatrix} \frac{d\gamma^1}{dt} \\ \frac{d\gamma^2}{dt} \end{pmatrix} = \begin{pmatrix} c_1 & -c_1 \\ -2c_1 & 4c_1 \end{pmatrix} \begin{pmatrix} \gamma^1 \\ \gamma^2 \end{pmatrix} + \begin{pmatrix} 0 \\ -2c_1 \end{pmatrix} \begin{pmatrix} \gamma^3 \end{pmatrix}. \tag{6.15}$$

In addition, we have, for the time evolution of γ^3

$$\frac{d\gamma^3}{dt} \;=\; 4c_1\gamma^1 - 10c_1\gamma^2 + 9c_1\gamma^3 - 3c_1\gamma^4. \tag{6.16}$$

The expressions in (6.15) and (6.16) do not satisfy the condition in (6.13) as the polynomial

$$p_3(x) \;=\; 4c_1x - 10c_1x^2 + 9c_1x^3 - 3c_1x^4 \tag{6.17}$$

does not belong to the subspace generated by the polynomials

$$\begin{aligned} p_1(x) &= c_1x^3 - c_1x^4 && (6.18)\\ p_2(x) &= -2c_1x^2 + 4c_1x^3 - 2c_1x^4. && (6.19) \end{aligned}$$

The expressions in (6.15) and (6.16) are modified by dropping the lower order moments

$$\begin{pmatrix} \frac{d\gamma^1}{dt} \\ \frac{d\gamma^2}{dt} \end{pmatrix} = \begin{pmatrix} c_1 & -c_1 \\ 0 & 4c_1 \end{pmatrix} \begin{pmatrix} \gamma^1 \\ \gamma^2 \end{pmatrix} + \begin{pmatrix} 0 \\ -2c_1 \end{pmatrix} \begin{pmatrix} \gamma^3 \end{pmatrix}. \tag{6.20}$$

$$\frac{d\gamma^3}{dt} \;=\; 9c_1\gamma^3 - 3c_1\gamma^4. \tag{6.21}$$

The condition in (6.13) is now satisfied.

Further, an expression for γ^3 is obtained using the moment closure technique in (6.10) and (6.12)

$$\gamma^3 \;=\; (\gamma^1)^{\eta_1}(\gamma^2)^{\eta_2}, \tag{6.22}$$

where η_1 and η_2 are obtained by solving the equations

$$\begin{aligned} \eta_1 + 2\eta_2 &= 3 && (6.23)\\ \eta_2 &= 3. && (6.24) \end{aligned}$$

This yields $\eta_1 = -3$ and

$$\gamma^3 \;=\; \left(\frac{\gamma^2}{\gamma^1}\right)^3. \tag{6.25}$$

6.3 Comparison of the RM Method with the SHS Method

We now compare the expressions for the evolution of the first two moments from the RM method with the SHS method. We also compare the simulation results from the two methods and discuss implementation issues.

6.3.1 First Order Reactions

For the first order reactions, the evolution of moments is closed, that is, the computation of moments of order up to Q do not require any higher order moments. We compare the expressions for moments obtained from the RM method with those obtained from the SHS method.

Example - Three Species (continued)

Continuing the example presented in Section 4.4.1, we have the reactions

$$R_1 : \; S_1 \xrightarrow{c_1} S_2 \tag{6.26a}$$

$$R_2 : \; S_2 \xrightarrow{c_2} S_3. \tag{6.26b}$$

The state of the system is $\boldsymbol{X}(t) = \begin{pmatrix} X_1(t) & X_2(t) & X_3(t) \end{pmatrix}^T$. Omitting the t's, the reset maps for the two reactions are

$$\boldsymbol{\phi}_1(\boldsymbol{X}) = \begin{pmatrix} X_1 - 1 \\ X_2 + 1 \\ X_3 \end{pmatrix}, \qquad \boldsymbol{\phi}_2(\boldsymbol{X}) = \begin{pmatrix} X_1 \\ X_2 - 1 \\ X_3 + 1 \end{pmatrix}. \tag{6.27}$$

The transition intensities are

$$\lambda_1(\boldsymbol{X}) = c_1 X_1, \qquad \lambda_2(\boldsymbol{X}) = c_2 X_2. \tag{6.28}$$

From (6.6), we have the following expression for evaluating moments

$$\begin{aligned}\frac{d\gamma^{\boldsymbol{q}}}{dt} &= E\Big(c_1 X_1\left((X_1-1)^{q_1}(X_2+1)^{q_2}X_3^{q_3} - X_1^{q_1}X_2^{q_2}X_3^{q_3}\right) + \\ &\quad c_2 X_2\left(X_1^{q_1}(X_2-1)^{q_2}(X_3+1)^{q_3} - X_1^{q_1}X_2^{q_2}X_3^{q_3}\right)\Big). \end{aligned} \tag{6.29}$$

We track a total of nine first and second order moments, and therefore $J = 9$. The required expressions from the SHS method are

$$\frac{d\gamma^{(1,0,0)}}{dt} = -c_1\gamma^{(1,0,0)} \tag{6.30}$$

$$\frac{d\gamma^{(0,1,0)}}{dt} = c_1\gamma^{(1,0,0)} - c_2\gamma^{(0,1,0)} \tag{6.31}$$

$$\frac{d\gamma^{(0,0,1)}}{dt} = c_2\gamma^{(0,1,0)} \tag{6.32}$$

$$\frac{d\gamma^{(2,0,0)}}{dt} = c_1\gamma^{(1,0,0)} - 2c_1\gamma^{(2,0,0)} \tag{6.33}$$

$$\frac{d\gamma^{(0,2,0)}}{dt} = c_1\gamma^{(1,0,0)} + c_2\gamma^{(0,1,0)} + 2c_1\gamma^{(1,1,0)} - 2c_2\gamma^{(0,2,0)} \tag{6.34}$$

$$\frac{d\gamma^{(0,0,2)}}{dt} = c_2\gamma^{(0,1,0)} + 2c_2\gamma^{(0,1,1)} \tag{6.35}$$

$$\frac{d\gamma^{(1,1,0)}}{dt} = -c_1\gamma^{(1,0,0)} + c_1\gamma^{(2,0,0)} - (c_1+c_2)\gamma^{(1,1,0)} \tag{6.36}$$

$$\frac{d\gamma^{(1,0,1)}}{dt} = -c_1\gamma^{(1,0,1)} + c_2\gamma^{(1,1,0)} \tag{6.37}$$

$$\frac{d\gamma^{(0,1,1)}}{dt} = c_1\gamma^{(1,0,1)} + c_2\gamma^{(0,2,0)} - c_2\gamma^{(0,1,1)} - c_2\gamma^{(0,1,0)} \tag{6.38}$$

Applying (4.17) and (4.39), the RM method yields the following expressions for first moments

$$\mu_1(t+\Delta t) = e^{-c_1\Delta t}\mu_1(t) \tag{6.39}$$

$$\mu_2(t+\Delta t) = \left(1-e^{-c_1\Delta t}\right)\mu_1(t) + e^{-c_2\Delta t}\mu_2(t) \tag{6.40}$$

$$\mu_3(t+\Delta t) = \left(1-e^{-c_2\Delta t}\right)\mu_2(t) + \mu_3(t). \tag{6.41}$$

And from (4.27), the RM method yields the following expressions for the second moments

$$E\left(X_1^2(t+\Delta t)\right) \;=\; E\left(X_1(t)\right)e^{-c_1\Delta t}\left(1-e^{-c_1\Delta t}\right)+E\left(X_1^2(t)\right)\left(e^{-c_1\Delta t}\right)^2 \tag{6.42}$$

$$\begin{aligned} E\left(X_1(t+\Delta t)X_2(t+\Delta t)\right) \;&=\; -E\left(X_1(t)\right)e^{-c_1\Delta t}\left(1-e^{-c_1\Delta t}\right) \\ &\quad + E\left(X_1^2(t)\right)e^{-c_1\Delta t}\left(1-e^{-c_1\Delta t}\right) \\ &\quad + E\left(X_1(t)X_2(t)\right)e^{-c_1\Delta t}e^{-c_2\Delta t}. \end{aligned} \tag{6.43}$$

We rewrite (6.33) from the SHS method, replacing the differentials with finite differences

$$E\left(X_1^2(t+\Delta t)\right) \;=\; E\left(X_1^2(t)\right)+c_1\Delta t E\left(X_1(t)\right)-2c_1\Delta t E\left(X_1^2(t)\right). \tag{6.44}$$

We also rewrite (6.42) from the RM method, using the Taylor series expansion of $e^{-c_1\Delta t}$

$$\begin{aligned} E\left(X_1^2(t+\Delta t)\right) \;&\approxeq\; E\left(X_1(t)\right)\left(c_1\Delta t-\frac{3}{2}(c_1\Delta t)^2+\ldots\right)\;+ \\ &\qquad E\left(X_1^2(t)\left(1-2c_1\Delta t+2(c_1\Delta t)^2\right)+\ldots\right). \end{aligned} \tag{6.45}$$

The above expression contains additional second and higher order terms in $c_1\Delta t$, compared with (6.44), which entails shorter time steps and consequently longer computation times with the SHS method.

We compare the simulation results obtained by the SHS method with the RM and the SSA methods, for the same parameters as in Section 4.4.1. The differential equations in the SHS method were solved using the Euler method. We recapitulate that the initial number of molecules are $\boldsymbol{X}(t=0)=\begin{pmatrix}10000 & 500 & 0\end{pmatrix}^T$ and the stochastic rate constants are $c_1=0.1$ sec^{-1} and $c_2=1.0$ sec^{-1}. The species are tracked from $t=0$ sec to $t=10$ sec, with $\Delta t=0.005$ sec for the RM method, $\Delta t=0.00001$ sec for the SHS method and the number of realizations being $5,000$ for the SSA method. The average number of time steps with the SSA method is $12,721$. From Figure 6.1, we see that the estimated moments obtained by the RM and SHS methods are in good agreement. In Figure 6.2, the time step in the SHS method was changed to $\Delta t=0.0005$, whereas the parameters of the other methods remained unchanged. With the larger time step, the SHS method does not perform as well.

6.3.2 Second Order Reactions

For the second order reactions, the computation of moments of order up to Q require moments of order greater than Q. Therefore, in addition to determining the moment evolution equations from (6.6), we also require expressions for the higher order moments, which are evaluated from (6.10) and (6.12). We compare the dimerising reaction example.

6.3.3 Example - Dimerising Reaction (continued)

In the SHS method, we will use expression (6.15) instead of (6.20), as the authors dropped condition (6.13) in a recent paper [69]. The expressions for the first and the second moments are

$$\frac{d\gamma^1}{dt} = c_1\gamma^1 - c_1\gamma^2 \tag{6.46}$$

$$\frac{d\gamma^2}{dt} = -2c_1\gamma^1 + 4c_1\gamma^2 - 2c_1\gamma^3. \tag{6.47}$$

Rewriting the above expressions, replacing the differentials with finite differences, yields

$$\mu_1(t+\Delta t) = \mu_1(t) + c_1\Delta t\mu_1(t) - c_1\Delta t E\left(X_1^2(t)\right). \tag{6.48}$$

$$\begin{aligned} E\left(X_1^2(t+\Delta t)\right) &= E\left(X_1^2(t)\right) - 2c_1\Delta t\mu_1(t) \\ &\quad + 4c_1\Delta t E\left(X_1^2(t)\right) - 2c_1\Delta t E\left(X_1^3(t)\right), \end{aligned} \tag{6.49}$$

where the expression for $E\left(X_1^3(t)\right)$ is given by (6.25).

In the RM method, we rewrite expression (5.43) for the mean and expand expression (5.44) for the second moment

$$\mu_1(t+\Delta t) = \mu_1(t) + c_1\Delta t\,\mu_1(t) - c_1\Delta t E\left(X_1^2(t)\right) \tag{6.50}$$

$$
\begin{aligned}
E\left(X_1^2(t+\Delta t)\right) &= E\left(X_1^2(t)\right) - 2c_1\Delta t\mu_1(t) \\
&\quad + 4c_1\Delta t E\left(X_1^2(t)\right) - 2c_1\Delta t E\left(X_1^3(t)\right) \\
&\quad + 2\left(c_1\Delta t\right)^2\mu_1(t) - \left(c_1\Delta t\right)^2 E\left(X_1^2(t)\right) \\
&\quad - \left(c_1\Delta t\right)^2 E\left(X_1^3(t)\right) + \left(c_1\Delta t\right)^2 E\left(X_1^4(t)\right), \qquad (6.51)
\end{aligned}
$$

where $E\left(X_1^3(t)\right)$ and $E\left(X_1^4(t)\right)$ are evaluated from

$$
\begin{aligned}
E\left(X_1^3(t)\right) &= \mu_1^3(t) + 3\mu_1(t)\sigma_1^2(t) \qquad (6.52) \\
E\left(X_1^4(t)\right) &= \mu_1^4(t) + 6\mu_1^2(t)\sigma_1^2(t) + 3\sigma_1^4(t). \qquad (6.53)
\end{aligned}
$$

Once again, the expression for $E\left(X_1^2(t+\Delta t)\right)$ from the RM method contains additional terms, compared with the SHS method.

Figure 6.3 compares the simulation results, from the two methods, with the SSA. The parameters are $X_1(t=0) = 1,500$, $X_2(t=0) = 0$, $c_1 = 0.002$ sec^{-1} and the simulation is performed up to time $t=1$ sec. The number of realizations in the SSA method is $3,000$. In the RM method $\Delta t = 0.002$ sec. The results from the SHS method compare well with the SSA with $\Delta t = 0.0001$ sec (top plot), but not with the larger $\Delta t = 0.002$ sec (bottom plot). This is due to the additional higher order terms in the RM method, and different moment closure expressions in the two methods. The RM method utilizes the normal distribution for the moment closure approximation, whereas the moment closure expressions in the SHS method correspond to lognormal distribution.

6.3.4 Implementation Issues

We now discuss the implementation issues with the SHS method, for examples consisting of a large number of molecules and species, such as the first order complex example presented in 4.4.2 and the second order example with ten species presented in 5.5.3.

In order to execute the complex first order example given in Section 4.4.2, with the number of species $N = 100$ and the number of reactions $M = 200$, using the SHS method, the expression (6.6) for the evolution of moments will consist of 200 terms. Many of these terms would be zero. Nonetheless, we would require a total of N equations for the first moments and $N(N+1)/2$ equations for the second moments. This gives a total of $5,150$

equations similar to those in (6.30) through (6.38). This is much more arduous to implement than the RM method. For a given biochemical system with first order reactions, the RM method employs expressions (4.6a) and (4.6b) for computing $\boldsymbol{P}$ and (4.25a) and (4.25b) for computing $\boldsymbol{B}(t)$. The mean and the covariance at each time step are obtained using the recursive expressions in (4.17) and (4.29).

In order to execute the second order example given in Section 5.5.3, with the number of species $N = 10$ and the number of reactions $M = 8$, using the SHS method, we have to derive a total of 55 equations, for the evolution of the first and the second moments. In addition we have to solve the system of equations in (6.12), to determine the expressions for higher order moments. These are many more additional steps compared to the RM method. For a given system of second order reactions, the RM method is implemented as described in Section 5.4. That is, given the reaction network and rate constants, define matrices $\boldsymbol{W}$, $\boldsymbol{X}(t)$ and $\boldsymbol{H}(t)$, compute the higher order expectations utilizing Table 5.1, and compute $\boldsymbol{\mu}(t + \Delta t)$ and $\boldsymbol{C}(t + \Delta t)$, recursively, by employing expressions (5.8) and (5.14), without deriving or solving any differential equations.

To recapitulate, the RM method produces the same accuracy as the SHS method, in fewer time steps, due to additional higher order terms and a different moment closure technique. The moment closure techniques from various methods require further investigation. For a given biochemical system, with known reactions and rate constants, the RM method can be implemented by utilizing the given recursive expressions, whereas the SHS method requires additional steps or additional software to derive the differential equations, determine the expressions for higher order moments and solve the resulting equations.

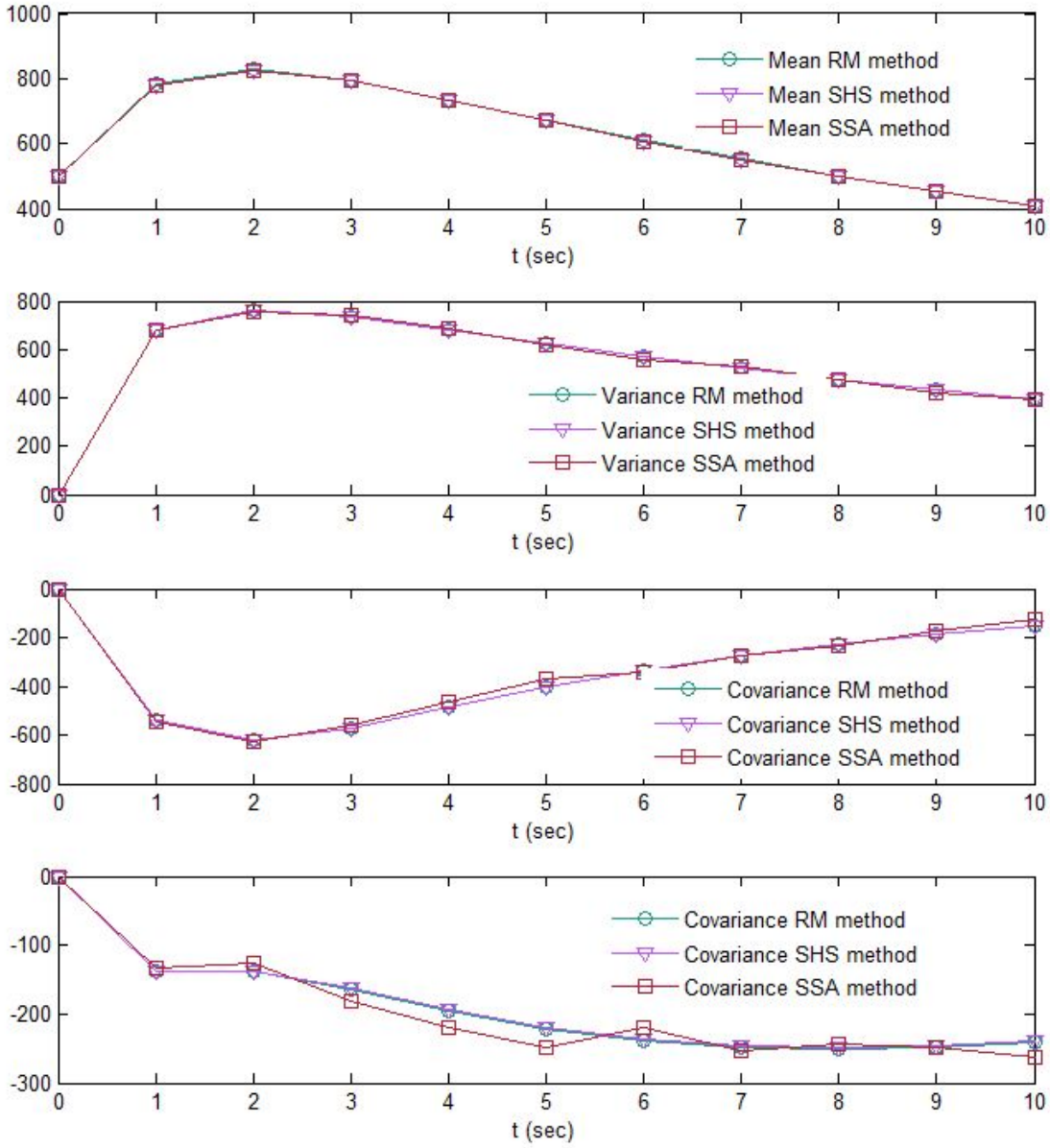

Figure 6.1: Comparison of the results from the RM, SHS and the SSA methods for the Three Species example given in Section 6.3.1. The curves represent the computed means $\mu_2(t)$ (top plot), the variances $\sigma_2^2(t)$ (second plot), and the covariances $\sigma_{21}(t)$ (third plot) and $\sigma_{23}(t)$ (bottom plot). Δt is 0.005 sec with the RM method and 0.00001 sec with the SHS method.

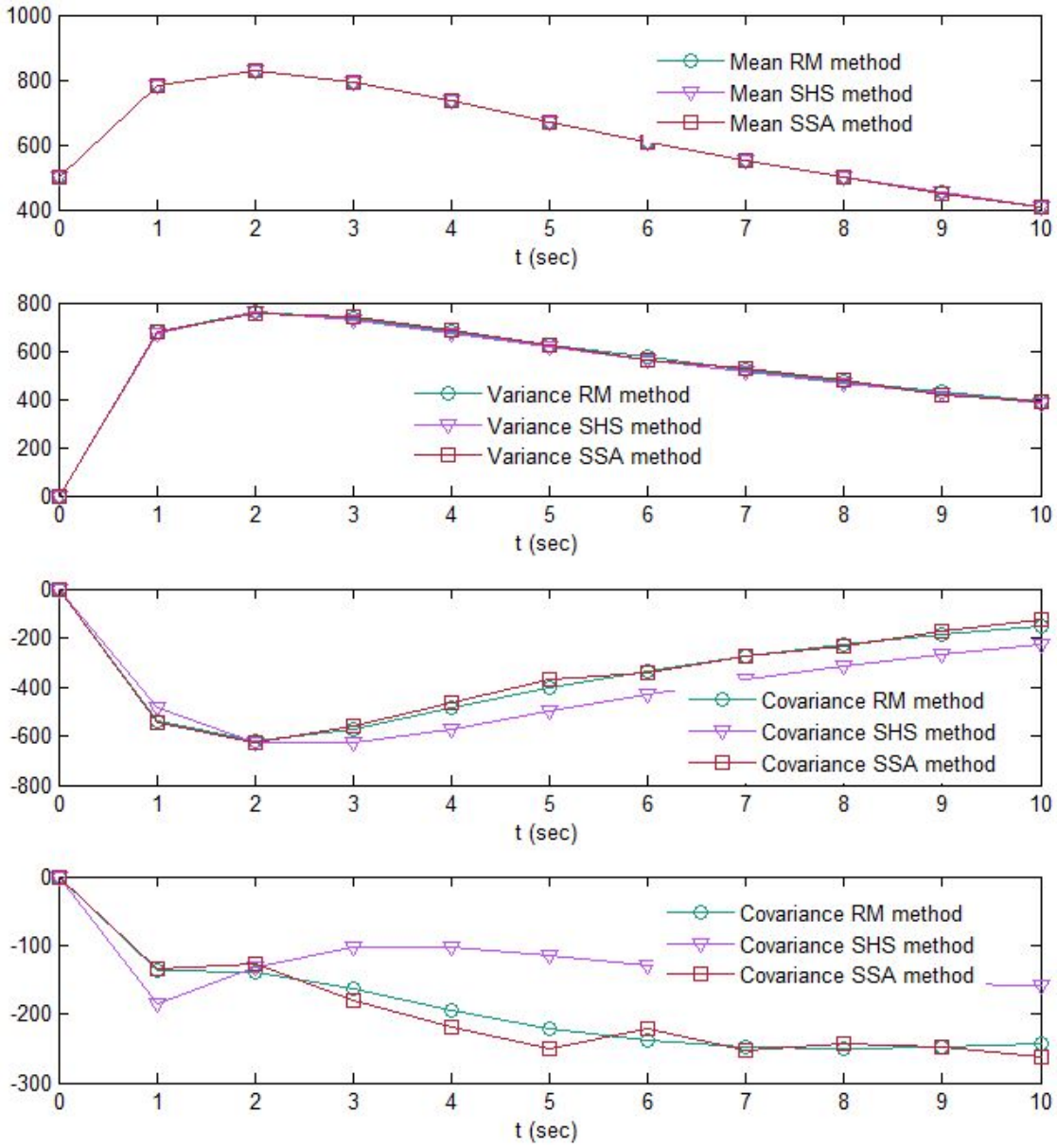

Figure 6.2: Same example and parameters as in Figure 6.1, except that in the SHS method $\Delta t = 0.0005$ sec. With a larger Δt, the SHS method does not agree well with the RM and the SSA methods.

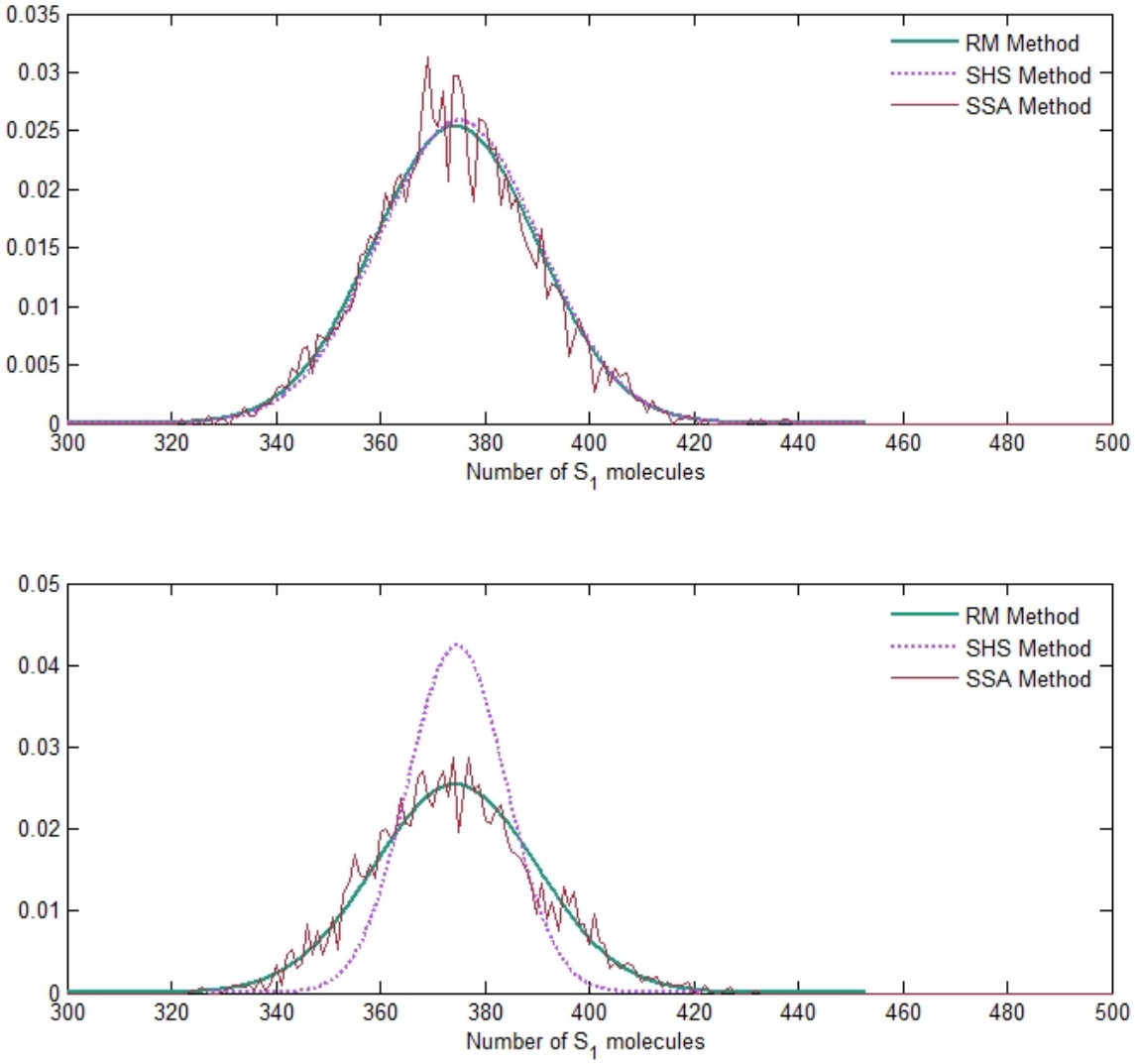

Figure 6.3: Simulation results from the RM (solid line), the SHS (dotted line) and the SSA ("noisy" plot) methods for the Dimerising Reaction example described in Section 6.3.3. In the top plot, Δt is 0.002 sec in the RM method and 0.0001 sec in the SHS method. When Δt is changed to 0.002 sec in the SHS method (bottom plot), it doesn't agree very well with the SSA.

Chapter 7

Conclusion and Future Work

In the forward problem of biochemical reaction systems, we have knowledge of all the reactions and the associated rate constants. We want to determine the joint probability distribution of the populations of all the molecular species at any time instant. The present state-of-the-art approaches for stochastic simulation of such systems are based on Monte Carlo methods, such as the Stochastic Simulation Algorithm (SSA) and its accelerated versions. The Monte Carlo methods provide approximations of the complete distribution, but they are computationally infeasible for complex biochemical systems consisting of a large number of species and reactions.

An alternate approach is to determine the evolution of a few lower order moments with time. In this dissertation, we presented a new method, called the Recursive Moment (RM) method, for propagating the first two moments of the joint probability distribution of the molecular populations in complex biochemical systems. The proposed method is applicable to very large systems. Compared with the SSA, the RM method yields significant savings in computation time. In contrast to other moment propagation methods, the recursive expressions in our method can be implemented by specifying rate constants and reaction stoichiometries, without having to derive or solve any differential equations.

An important direction for continued work on the RM method includes evaluation of its numerical accuracy. We would like to be in a position to predict the accuracy of the proposed method. Therefore, it is important to study the accuracy as a function of various factors, including the value of the time step, the values of the reaction constants, the types of reactions, and the population sizes of the species. Ideally, we would like to be able

to determine upper bounds of performance given a particular set of reactions, molecular populations, and choices of time step.

There are cases where the distributions of the molecular populations are not unimodal. More specifically, there may be scenarios where the PMFs may be multimodal, for example, the bimodal distribution in the Schlögl model [66]. This implies that tracking only the first two moments will not be sufficient. Thus, it is important to expand the work presented in the dissertation so that one can track higher order moments. The approach used for deriving the proposed method is amenable to generalize the method. The goal, however, is to obtain a procedure that is not cumbersome for execution. Clearly, once we have higher order moments, we will be able to deduce much more about the joint distribution of the species.

An important extension of the proposed work is to apply the method in practice. To that end, simulations of complex metabolic and signaling pathways, such as the cell signaling pathway in cancer and stem cells as described in [13], are particularly attractive. Since such pathways are models developed by biologists, it would be interesting to test them against experimental data. Therefore, developing ways of testing proposed models using the simulation tools would add considerable value to the analysis.

Finally, in the problem studied here, it was assumed that we have knowledge of the reaction network and all the rate constants. However, these quantities are usually unknown, but what is available instead are the time series data of biochemical pathways. Further, it is difficult to obtain data for all the species sampled at high enough frequencies, and one may only have observations of a subset of the species. In future work, one would like to examine statistical techniques for inferring parameters from data in challenging scenarios. The class of Bayesian methods offer approaches for tackling many difficult problems [6], [18]. With partially observed data and for systems with a large number of species (high dimension), computational techniques such as the Markov Chain Monte Carlo (MCMC) can be used to solve such problems [64].

Bibliography

[1] B. Alberts, A. Johnson, J. Lewis, and M. Raff, *Molecular Biology of the Cell*, Garland Science, fourth edition, 2002.

[2] A. Arkin, J. Ross, and H. H. McAdams, "Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *escherichia coli* cells," *Genetics*, vol. 149, pp. 1633–1648, 1998.

[3] A. F. Bartholomay, "Stochastic models for chemical reactions: I. theory of the unimolecular reaction process," *Bulletin of Mathematical Biophysics*, vol. 20, no. 3, pp. 175–190, 1958.

[4] D. Bleeker and G. Csordas, *Basic partial differential equations*, International Press, 1996.

[5] J. M. Bower and H. Bolouri, Eds., *Computational Modeling of Genetic and Biochemical Networks*, MIT Press, 2004.

[6] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Wiley-Interscience, 1992.

[7] Y. Cao, H. Lee, and L. Petzold, "Efficient formulation of the stochastic simulation algorithm for chemically reacting systems," *The Journal of Chemical Physics*, vol. 121, no. 9, pp. 4059–4067, 2004.

[8] G. Casella and R. L. Berger, *Statistical Inference*, Duxbury Press, second edition, 2001.

[9] A. Chatterjee, D. G. Vlachos, and M. A. Katsoulakis, "Binomial distribution based τ-leap accelerated stochastic simulation," *The Journal of Chemical Physics*, vol. 122, no. 2, pp. 024112 (7), 2005.

[10] M. H. A. Davis, "Piecewise-deterministic markov processes: A general class of non-diffusion stochastic models," *Journal of the Royal Statistical Society. Series B*, vol. 46, no. 3, pp. 353–388, 1984.

[11] H. de Jong, "Modeling and simulation of genetic regulatory systems: A literature review," *Journal of Computational biology*, vol. 9, no. 1, pp. 67–103, 2002.

[12] J. L. Doob, "Markoff chains–denumerable case," *Transactions of the American Mathematical Society*, vol. 58, no. 3, pp. 455–473, 1945.

[13] O. Dreseen and A. H. Brivanlou, "Signaling pathways in cancer and embryonic stem cells," *Stem Cell Review*, vol. 3, pp. 7–17, 2007.

[14] J. Elf, P. Lötstedt, and P. Sjöberg, "Problems of high dimension in molecular biology," *Proceedings of the* 19^{th} *GAMM-Seminar, Leipzig*, pp. 21–30, 2003.

[15] P. Érdi and J. Tóth, *Mathematical models of chemical reactions: Theory and applications of deterministic and stochastic models*, Manchester University Press, 1989.

[16] W. Feller, *An introduction to probability theory and its applications, Vol. 2*, Wiley, 1970.

[17] C. Gadgil, "Stochastic modeling of biological reactions," *Journal of the Indian Institute of Science*, vol. 88, no. 1, pp. 45–55, 2008.

[18] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC, second edition, 2004.

[19] M. K. Ghosh, A. Arapostathis, and S. I. Marcus, "Ergodic control of switching diffusions," *SIAM Journal on Control and Optimization*, vol. 35, no. 6, pp. 1952–1988, 1997.

[20] M. A. Gibson and J. Bruck, "Efficient exact stochastic simulation of chemical systems with many species and many channels," *The Journal of Physical Chemistry*, vol. 104, no. 9, pp. 1876 (14), 2000.

[21] C. S. Gillespie, "Moment-closure approximations for mass-action models," *IET Systems Biology*, vol. 3, pp. 52–58, 2009.

[22] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, pp. 403–434, 1976.

[23] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.

[24] D. T. Gillespie, *Markov Processes: An Introduction for Physical Scientists*, Academic Press, 1992.

[25] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Physica A: Statistical Mechanics and its Applications*, vol. 188, pp. 404–425, 1992.

[26] D. T. Gillespie, "The chemical langevin equation," *The Journal of Chemical Physics*, vol. 113, no. 1, pp. 297–306, 2000.

[27] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *The Journal of Chemical Physics*, vol. 115, no. 4, pp. 1716–1733, 2001.

[28] D. T. Gillespie, "The chemical langevin and fokker-planck equations for the reversible isomerization reaction," *The Journal of Physical Chemistry*, vol. 106, no. 4, pp. 5063–5071, 2002.

[29] D. T. Gillespie, "Improved leap-size selection for accelerated stochastic simulation," *The Journal of Chemical Physics*, vol. 119, no. 16, pp. 8229–8234, 2003.

[30] D. T. Gillespie, "Stochastic chemical kinetics," in *Handbook of Materials Modeling*, S. Yip, Ed., pp. 1735–1752. Springer, 2005.

[31] D. T. Gillespie, "Stochastic simulation of chemical kinetics," *Annual Review of Physical Chemistry*, vol. 58, pp. 35–55, 2007.

[32] C. A. Gómez-Uribe and G. C. Verghese, "Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations," *The Journal of Chemical Physics*, vol. 126, no. 2, pp. 024109 (12), 2007.

[33] C. A. Gómez-Uribe and G. C. Verghese, "Enhanced identification and exploitation of time scales for model reduction in stochastic chemical kinetics," *The Journal of Chemical Physics*, vol. 129, no. 24, pp. 244112 (16), 2008.

[34] P. J. E. Goss and J. Peccoud, "Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets," *Proceedings of the NAtional Academy of Sciences*, vol. 95, no. 12, pp. 6750–6755, 1998.

[35] J. Goutsias, "Classical versus stochastic kinetics modeling of biochemical reaction systems," *Biophysical Journal*, vol. 92, pp. 2350–2365, 2007.

[36] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, Oxford University Press, third edition, 2001.

[37] R. L. Grossman, A. Nerode, A. P. Ravn, and H. Rischel, Eds., *Hybrid Systems (Lecture Notes in Computer Science)*, Springer-Verlag, 1993.

[38] J. P. Hespanha and A. Singh, "Stochastic models for chemically reacting systems using polynomial stochastic hybrid systems," *International Journal of Robust and Nonlinear Control*, vol. 15, pp. 669–689, 2005.

[39] J. Hu, J. Lygeros, and S. Sastry, "Towards a theory of stochastic hybrid systems," in *Hybrid Systems: Computation and Control*, N. Lynch and B. H. Krogh, Eds., pp. 160–173. Springer-Verlag, 2000.

[40] K. Ishida, "Stochastic model for bimolecular reaction," *The Journal of Chemical Physics*, vol. 41, no. 8, pp. 2472–2478, 1964.

[41] T. Jahnke and W. Huisinga, "Solving the chemical master equation for monomolecular reaction systems analytically," *Journal of Mathematical Biology*, vol. 54, pp. 1–26, 2007.

[42] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, "Stochasticity in gene expression: From theories to phenotypes," *Nature Reviews Genetics*, pp. 451–464, 2005.

[43] M. Khammash and H. E. Samad, "Stochastic modeling and analysis of genetic networks," *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, pp. 2320–2325, 2005.

[44] A. M. Kierzek, J. Zaim, and P. Zielenkiewicz, "The journal of biological chemistry," *The Effect of Transcription and Translation Initiation Frequencies on the Stochastic Fluctuations in Prokaryotic Gene Expression*, vol. 276, no. 11, pp. 8165–8172, 2001.

[45] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Springer-Verlag, 1992.

[46] I. J. Laurenzi, "An analytical solution of the stochastic master equation for reversible biomolecular reaction kinetics," *The Journal of Chemical Physics*, vol. 113, pp. 3315–3322, 2000.

[47] C. H. Lee, K. H. Kim, and P. Kim, "A moment closure method for stochastic reaction networks," *The Journal of Chemical Physics*, vol. 130, no. 13, pp. 134107, 2009.

[48] A. J. Lotka, *Elements of Physical Biology*, Williams and Wilkins, 1925.

[49] J. Lygeros, "Lecture notes on hybrid systems," available at `http://robotics.eecs.berkeley.edu/ sastry/ee291e/lygeros.pdf`, 2004.

[50] V. Mane, M. F. Bugallo, and P. M. Djurić, "Stochastic modeling of second order reactions using a moment propagation method," in *IEEE International Workshop on Genomic Signal Processing and Statistics*, May 2009, pp. 1–4.

[51] V. Mane, M. F. Bugallo, and P. M. Djurić, "A stochastic approach to studying biochemical reactions without monte carlo simulations," in *IEEE/SP 15th Workshop on Statistical Signal Processing*, August 2009, pp. 549–552.

[52] H. H. McAdams and A. Arkin, "Stochastic mechanisms in gene expression," *Proceedings of the National Academy of Sciences*, vol. 94, no. 3, pp. 814–819, 1997.

[53] H. H. McAdams and A. Arkin, "It's a noisy business!," *Trends in Genetics*, vol. 15, no. 2, pp. 65–69, 1999.

[54] D. A. McQuarrie, "Kinetics of small systems," *The Journal of Chemical Physics*, vol. 38, no. 2, pp. 433–436, 1963.

[55] D. A. McQuarrie, "Stochastic approach to chemical kinetics," *Journal of Applied Probability*, vol. 4, no. 3, pp. 413–478, 1967.

[56] A. Novick and M. Weiner, "Enzyme induction as an all-or-none phenomenon," *Proceedings of the National Academy of Sciences*, vol. 43, pp. 553–566, 1957.

[57] B. Øksendal, *Stochastic Differential Equations*, Springer, sixth edition, 2002.

[58] M. F. Pettigrew and H. Resat, "Multinomial tau-leaping method for stochastic kinetic simulations," *The Journal of Chemical Physics*, vol. 126, no. 8, pp. 084101 (15), 2007.

[59] S. Ramsey, D. Orrell, and H. Bolouri, "Dizzy: stochastic simulation of large-scale genetic regulatory networks," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 415–436, 2005.

[60] C. V. Rao, D. M. Wolf, and A. P. Arkin, "Control, exploitation and tolerance of intracellular noise," *Nature*, vol. 420, pp. 231–237, 2002.

[61] V. N. Reddy, M. N. Liebman, and M. L. Mavrovouniotis, "Qualitative analysis of biochemical reaction systems," *Computers in Biology and Medicine*, vol. 26, pp. 9–24, 1996.

[62] A. Rényi, "Treatement of chemical reactions by means of the theory of stochastic processes," *Magyar Tud. Akad. Alkalm. Mat. Int. Közl*, vol. 2, pp. 93–101, 1954.

[63] A. Ribeiro, R. Zhu, and S. A. Kauffman, "A general modeling strategy for gene regulatory networks with stochastic dynamics," *Journal of computational Biology*, vol. 13, no. 9, pp. 1630–1639, 2006.

[64] C. P. Roberts and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, second edition, 2004.

[65] M. R. Roussel and R. Zhu, "Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression," *Physical Biology*, vol. 3, pp. 274–284, 2006.

[66] F. Schlögl, "On thermodynamics near a steady state," *Zeitschirft für Physik A*, vol. 248, no. 5, pp. 446–458, 1971.

[67] A. Singh and J. P. Hespanha, "Models for multi-specie chemical reactions using polynomial stochastic hybrid systems," in *Proceedings of the 44th IEEE Conference on Decision and Control and the European Control Conference*, 2005, pp. 2969–2974.

[68] A. Singh and J. P. Hespanha, "A derivative matching approach to moment closure for the stochastic logistic model," *Bulletin of Mathematical Biology*, vol. 69, no. 6, pp. 1909–1925, 2007.

[69] A. Singh and J. P. Hespanha, "Stochastic hybrid systems for studying biochemical processes," *Philosophical Transactions of the Royal Society: A*, vol. 368, no. 1930, pp. 4995–5011, 2010.

[70] P. Sjöberg, P. Lötstedt, and J. Elf, "Fokker-planck approximation of the master equation in molecular biology," *Computing and Visualization in Science*, vol. 12, no. 1, pp. 37–50, 2009.

[71] G. B. Skinner, *Intorduction to Chemical Kinetics*, Academic Press, 1974.

[72] A. Slepoy, A. P. Thompson, and S. J. Plimpton, "A constant-time kinetic monte carlo algorithm for simulation of large biochemical reaction networks," *The Journal of Chemical Physics*, vol. 128, no. 20, pp. 205101 (8), 2008.

[73] R. Srivastava, M. S. Peterson, and W. E. Bentley, "Stochastic kinetic analysis of the escherichia coli stress circuit using σ^{32}-targeted antisense," *Biotechnology and Bioengineering*, vol. 75, pp. 120–129, 2001.

[74] A. Stuart and K. Ord, *Kendall's Advanced Theory of Statistics, Volume I: Distribution Theory*, Wiley, sixth edition, 2009.

[75] T. Tian and K. Burrage, "Binomial leap methods for simulating stochastic chemical kinetics," *The Journal of Chemical Physics*, vol. 121, no. 21, pp. 10356 (9), 2004.

[76] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, third edition, 2007.

[77] V. Volterra, "Variations and fluctuations of the number of individuals in animal species living together," in *Animal Ecology*, pp. 409–448. McGraw-Hill, 1931.

[78] D. J. Wilkinson, *Stochastic Modeling for System Biology*, Chapman & Hall/CRC, 2006.

[79] X. Zhang, K. De Cock, M. F. Bugallo, and P. M. Djurić, "A general method for the computation of probabilities in systems of first order chemical reactions," *The Journal of Chemical Physics*, vol. 122, no. 10, pp. 104101 (8), 2005.

Appendix A

Acronyms

CME	chemical master equation
FPE	Fokker-Planck equation
ODE	ordinary differential equation
PDE	partial differential equation
PMF	probability mass function (discrete random variable)
PDF	probability density function (continuous random variable)
MGF	moment generating function
MFK	mass fluctuation kinetics
RM	recursive method
RRE	reaction rate equation
SDE	stochastic differential equation
SHS	stochastic hybrid system
SSA	stochastic simulation algorithm

Appendix B

Notation

$a_m(\boldsymbol{x}, t)$	reaction propensity in R_m.
α_i	population of species S_i at time $t = 0$.
$\boldsymbol{\alpha}$	state of the system at time $t = 0$.
$b(\cdot)$	SDE drift coefficient.
c_m	stochastic rate constant of reaction R_m.
C_i	sum of the stochastic rate constants in which species S_i is a source.
$\boldsymbol{C}(t)$	covariance of $\boldsymbol{X}(t)$
$\delta(t)$	Dirac delta function.
Δt	step size of the recursive method.
$\mathcal{G}(s_i, t)$	probability generating function of $P(x_i, t)$.
$\mathcal{G}(\boldsymbol{s}, t)$	probability generating function of $P(\mathtt{x}, t)$.
$\gamma^{\boldsymbol{q}}(t)$	elements of $\boldsymbol{\gamma}(t)$
$\gamma'^{\,\boldsymbol{q}}(t)$	elements of $\boldsymbol{\gamma}'(t)$
$\boldsymbol{\gamma}(t)$	vector of non-central moments
$\boldsymbol{\gamma}'(t)$	vector of central moments
$\Gamma(\cdot)$	Gamma function.
$h_m(\boldsymbol{x}, t)$	number of distinct molecular combinations in R_m.
k_m	deterministic rate constant of reaction R_m.

M	number of reactions.
$\mathcal{M}_{\boldsymbol{X}(t)}(\boldsymbol{\omega})$	moment generating function of $\boldsymbol{X}(t)$.
$\boldsymbol{\mu}(t)$,	mean of $\boldsymbol{X}(t)$
n_A	Avogadro number
N	number of species or the number of tracked species.
$\mathbb{N}$	set of positive integers.
$\mathcal{N}(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2.
ν_{mi}	stoichiometry of species S_i in reaction R_m.
$\boldsymbol{\nu}_m$	stoichiometry vector in reaction R_m.
$P(\mathtt{x}, t)$, $P(\mathtt{x}(t))$	pdf or pmf of of $\boldsymbol{X}(t)$.
$P_{l\|l-1}(\cdot)$	conditional pmf of a random vector at time t_l given its pmf at time t_{l-1}.
Q	order of the moments in $\gamma^{\boldsymbol{q}}(t)$.
R_m	mth reaction, $m = 1 \dots M$.
$\mathbb{R}$	set of real numbers.
S_i, S_j	ith and jth species, $i = 1 \dots N$, $j = 1 \dots N$.
t or t_l	time index
$u(\cdot)$	SDE diffusion coefficient.
V	volume of biochemical reaction system
$\mathcal{W}(\cdot)$	Wiener process.
$\boldsymbol{x}(t)$	a realization of $\boldsymbol{X}(t)$.
$X_i(t)$	population of species S_i at time t.
$\boldsymbol{X}(t)$	state of the system at time t, with elements $X_i(t)$.