APPROVAL SHEET

Title of Dissertation: Parametric Sensitivity Analysis of Stochastic Reaction Networks

Name of Candidate: Ting Wang
Doctor of Philosophy, 2015

Dissertation and Abstract Approved: MRathinam
Dr. Muruhan Rathinam
Professor
Department of Mathematics and Statistics

Date Approved: 10/2/15

NOTE: *The Approval Sheet with the original signature must accompany the thesis or dissertation. No terminal punctuation is to be used.

ABSTRACT

Title of dissertation: Parametric Sensitivity Analysis of Stochastic Reaction Networks

Ting Wang, Doctor of Philosophy, 2015

Dissertation directed by: Dr. Muruhan Rathinam
Department of Mathematics and Statistics

Reaction networks are systems consisting of several species interacting with each other through a set of predefined reaction channels. Models of real world reaction systems often contain several parameters which play a significant role in determining the system's dynamics. Therefore, parametric sensitivity analysis is an essential tool for the modeling and parameter estimation process. Due to the complex and random nature of the reaction systems, among all approaches for sensitivity analysis, Monte Carlo simulation is the most suitable for the parametric sensitivity analysis because its complexity does not grow dramatically as the problem dimension grows. Most Monte Carlo methods for sensitivity analysis can be classified into three categories, the pathwise derivative (PD), the finite difference (FD) and the Girsanov transformation (GT). Comparisons of these methods for specific examples have been done by many researchers, which showed that when applicable, the PD method and FD method tend to outperform the GT method. However, to the best of our knowledge, no existing literature studies these observations from a theoretical point of view. In this thesis, we provide a theoretical justification for these obser-

vations in terms of system size asymptotic analysis. We also examine our result by testing several numerical examples. Other than the analysis for the efficiency of these Monte Carlo estimators, we also provide some sufficient conditions which guarantee the validity of the GT method. Finally, for an ergodic system, there exists a steady state distribution and hence it is reasonable for us to consider the steady state sensitivity estimation problem. We establish an asymptotic correlation result and use this result to justify the ensemble-averaged correlation function method introduced in the literature.

Parametric Sensitivity Analysis of Stochastic Reaction Networks

by

Ting Wang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore County in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Dr. Muruhan Rathinam, Chair/Advisor
Dr. Weining Kang
Dr. Anindya Roy
Dr. Rouben Rostamian
Dr. Avanti Athreya

Acknowledgments

First of all, I would like to express my sincere gratitude to my advisor Dr. Muruhan Rathinam for his continuous support and encouragement of my study and research in the past five years. I have been amazingly fortunate to have an advisor who has not only the immerse knowledge to give me big pictures of my research but also the patience to sit down to walk me through every line of my code. This dissertation would not have been possible without his guidance and training. To me, he is not my advisor, but my family.

Besides my advisor, I would like to thank other members of my dissertation committee, Dr. Rostamian, Dr. Roy, Dr. Kang and Dr. Athreya for their valuable time and insightful feedback and input throughout this dissertation process.

I am also grateful to all the professors who have taught me in my graduate studies. My sincere thanks goes to Dr. Rostamian who taught me advanced C programming and Dr. Kang who taught me lots of tricks in probability theory.

Finally, but most importantly, none of this would have been possible without the unconditional love and support from my parents and fiancee. I would like to express my heart-felt gratitude to my parents who sacrificed a lot to support me throughout writing this dissertation, my job searching and my life in general.

Table of Contents

Chapter 1: Introduction

1.1 Stochastic Reaction Networks

Reactions networks are an important modeling tool in areas such as systems biology, epidemiology and chemical engineering. Traditionally, these have been modeled as a deterministic system whose variables are assumed to evolve continuously. That is, the time evolution of the number of molecules X is governed by the following ordinary differential equations

$$\frac{dX(t)}{dt} = \sum_{j=1}^{m} \nu_j \bar{a}_j(X(t)), \tag{1.1}$$

where $\bar{a}_j$ are known as the deterministic reaction rate function and ν_j are the stoichiometirc vectors (we will introduce it later). The above equation is known as the **reaction rate equation** (RRE). This deterministic model works reasonably well when the molecular copy number of all reactants in the network is large. However, if the system is so small that one can not ignore the discrete nature of the interactions, the RRE fails to describe the true dynamics of the system. To circumvent this issue, a continuous time Markov chain (CTMC) is often used to model a system with small populations.

Let us formally describe stochastic reaction networks. Consider a chemical re-

action system consisting of m reaction channels and n chemical species $\{S_1, \cdots, S_n\}$. The n-dimensional vector $X(t)$ with the state space being a subset of $\mathbb{Z}_+^n$ characterizes the state of the system where each entry $X_i(t)$ represents the number of molecules of the species S_i at time t. We assume $X(t)$ is a CTMC with state space $\mathbb{Z}_+^n$.

The firing of a reaction channel $j \in \{1, \cdots, m\}$ at time t causes the state to be incremented by the *stoichiometric vector* ν_j such that $X(t) = X(t-) + \nu_j$. Thus we assume that X is *càdlàg*, i.e., paths of X are right continuous with left-hand limits. The process X is assumed to be a continuous time Markov chain (CTMC) with **infinitesimal generator** $\mathcal{A}$ given by

$$(\mathcal{A}f)(x) = \sum_{j=1}^{m} a_j(x)(f(x+\nu_j) - f(x)) \tag{1.2}$$

for any bounded $f : \mathbb{R}^n \to \mathbb{R}$.

Associated with each reaction channel is an **intensity function** also known as **propensity function** in the chemical kinetics literature $a_j(x), j = 1, \cdots, m$, which is such that for any small time increment Δt we have

$$\begin{aligned} &\mathbb{P}(X(t+\Delta t) = x + \nu_j | X(t) = x) = a_j(x)\Delta t + o(\Delta t) \\ &\mathbb{P}(X(t+\Delta t) = x | X(t) = x) = 1 - \sum_{j=1}^{m} a_j(x)\Delta t + o(\Delta t). \end{aligned} \tag{1.3}$$

For $j = 1, \cdots, m$ we denote by $R_j(t)$ the number of firings of the j-th reaction channel during $(0, t]$. Thus $X(t) = X(0) + \nu R(t)$ for $t \geq 0$, where ν is the **stoichiometric matrix** whose j column is ν_j and $R(t) = (R_1(t), \cdots, R_m(t))^T$. We note that $R(0) = 0$ and $R_j(t) - R_j(t-)$ is either 0 or 1. Following the terminology of [8] we note that $R_j(t)$ are counting processes which admit the $\mathcal{F}_t$-predictable **inten-**

sity process $a_j(X(t-))$ where $\mathcal{F}_t$ is the filtration generated by X and R. In the description of these processes $X(t)$ and $R(t)$, there are parameters $c = (c_1, \ldots, c_p)$ involved. Therefore, it is more accurate to write $X(t, c)$ and $R(t, c)$. We assume that these parameters enter only via the intensity functions and not via the stoichiometric vectors ν_j. Thus $a_j = a_j(x, c)$ for $j = 1, \ldots, m$ while the ν_j for $j = 1, \ldots, m$ do not depend on c.

To familiarize the readers with all the above definitions involving the stochastic reaction networks, we give the following example

$$S_1 + S_2 \xrightarrow{c_1} S_3, \qquad S_3 \xrightarrow{c_2} S_1 + S_2.$$

The state (or population) and reaction count of this system are characterized by $X(t)$ and $R(t)$, respectively. The stoichiometric vectors for the system are $\nu_1 = [-1, -1, 1]$ for the forward reaction and $\nu_2 = [1, 1, -1]$ for the backward reaction. For the forward reaction, we have x_1 ways to choose one S_1 and x_2 ways to choose S_2, hence its intensity function is $a_1(x, c) = c_1 x_1 x_2$. Similarly, the intensity function for the second reaction is $a_2(x) = c_2 x_3$.

1.2 Random Time Change Representation

Recall that the CTMC $X(t)$ can be written as

$$X(t, c) = X(0, c) + \sum_{j=1}^{m} \nu_j R_j(t, c)$$

where the species count $R_j(t,c)$ is a counting process. It turns out that the $R_j(t,c)$ can be characterized in terms of independent unit rate Poisson processes Y_j as

$$R_j(t,c) = Y_j\left(\int_0^t a_j(X(s,c),c)ds\right).$$

We define

$$S_j(t,c) = \int_0^t a_j(X(s,c),c)ds \tag{1.4}$$

and refer to $S_j(t,c)$ as the **internal times**, which are in fact dimensionless quantities. Hence $X(t,c)$ can be written as

$$X(t,c) = X(0,c) + \sum_{j=1}^{m} \nu_j Y_j(S_j(t,c)) \tag{1.5}$$

where Y_j are independent unit rate Poisson processes. The above representation is known as the **random time change representation** (RTC). Also, for any bounded function $f : \mathbb{Z}_+^n \to \mathbb{R}$. Let $X(t,c)$ be a solution of (1.5), we have

$$f(X(t,c)) - f(X(0,c)) - \int_0^t \mathcal{A}f(X(s,c))ds \tag{1.6}$$

is a zero mean martingale, which is known as the **Dynkin's martingale** (see [10]). We refer the interested reader to [10] for a rigorous derivation of this representation.

Since $X(t,c)$ is a stepwise function in t, we have the following representation for the internal times

$$S_j(t,c) = S_j(T_{i-1},c) + a_j(X(T_{i-1},c),c)(t - T_i) \tag{1.7}$$

for $j = 1, \cdots, m$ and $T_{i-1} \le t < T_i$, where T_i is the time at which a reaction of any type fires. This form is crucial in the derivation of the RTC simulation algorithm described in Chapter 1 and the RPD sensitivity algorithm in Chapter 3.

1.3 Chemical Master Equation

The distribution of the state vector X is governed by a set of differential equations. Denoting $p(x,t) = \mathbb{P}(X(t) = x)$, one can show that

$$\frac{dp(x,t)}{dt} = \sum_{j=1}^{m} (p(x - \nu_j, t)a_j(x - \nu_j) - p(x,t)a_j(x)) \quad x \in \mathbb{Z}_+^n. \tag{1.8}$$

This equation is known as the **chemical master equation** (CME) or Kolmogorov's forward equation in most probabilistic literature [14]. One may attempt to solve the CME to obtain the distribution of state vector X at a particular time t. However, for most reaction networks, the corresponding CME is an infinite dimensional system and hence difficult or impossible to solve analytically. Some numerical methods such as the finite state projection (FSP) [21] algorithm truncates the state space and solve the truncated system to approximate the solution of the CME. It is even capable of telling us how closely the truncated space approximation matches the true solution. However, it is still computationally prohibitive for FSP to handle a network with large species count. Therefore, a more realistic approach is using Monte Carlo methods to simulate the exact sample paths and use them to approximate the distribution. We will talk about these simulation methods for CTMC in the next section.

1.4 Stochastic Simulation Algorithms

To simulate the time evolution of the stochastic reaction system, one needs to keep track of the holding time (or inter-arrival time) and the index of the cor-

responding reaction J_i $(i = 1, 2, \cdots)$. Let ΔT_x denote the holding time of the CTMC X after entering the state x. One can easily show that ΔT_x has exponential distribution with rate $a_0(x)$, where $a_0(x) = \sum_{j=1}^{m} a_j(x)$.

Hence, we can generate ΔT_x easily from a uniformly distributed random number r_1 in $(0, 1]$ using

$$\Delta T_x = \frac{1}{a_0(x)} \ln\left(\frac{1}{r_1}\right). \tag{1.9}$$

The probability that it is the j^*th reaction channel fires at this moment is simply

$$\mathbb{P}(J_i = j^*) = \frac{a_{j^*}(x)}{a_0(x)}$$

and one can generate J_i by

$$\sum_{j=1}^{j^*-1} a_j(x) < r_2 a_0(x) \leq \sum_{j=1}^{j^*} a_j(x) \tag{1.10}$$

Given the distribution of the holding time and the index of the corresponding reaction, one can easily simulate the CTMC $X(t)$ using the following **stochastic simulation algorithm** (SSA) [12]. The above algorithm is also known as the Gillespie's direct SSA. Figure 1.1 is a plot of one exact sample path for the reaction network

$$S_1 \xrightarrow{c_1} S_2, \qquad S_2 \xrightarrow{c_2} S_1$$

using the direct SSA.

There are other versions of SSA such as Gibson and Bruck's next reaction method in [11]. In this section, we present another SSA, i.e., the random time change (RTC) simulation algorithm from [26], which is also known as the modified next reaction method in [3]. We will use this version of SSA when we describe the

Algorithm 1 Stochastic simulation algorithm

Require: N_{tr}, terminal time T, initial state x_0

for $k = 1 : N_{\text{tr}}$ **do**

Initialized the time $t = 0$ and $x = x_0$

while $t < T$ **do**

Evaluate the intensity $a_j(x)$ and their sum $a_0(x)$

Generate random number r_1 and r_2 from uniform distribution $U[0, 1]$

Generate the holding time ΔT_x from (1.9) and the index of reaction j^* from (1.10)

Update $t \leftarrow t + \Delta T_x$ and $x \leftarrow x + \nu_{j^*}$

end while

end for

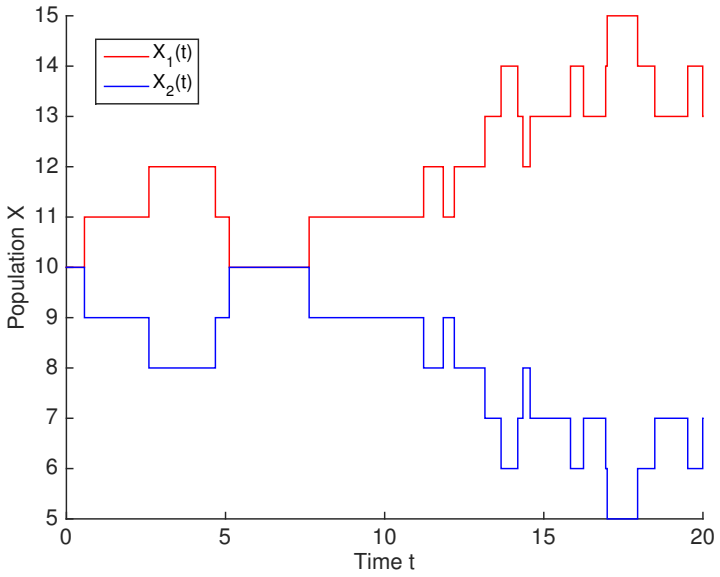


Figure 1.1: *Realization of one path up to* $t = 20$ *using SSA. The initial population* $X(0) = [10, 10]$*, parameters are set to be* $c_1 = 0.05$ *and* $c_2 = 0.05$*.*

common reaction path method [26] in Chapter 3. Recall that the RTC representation uses m independent Poisson processes to characterize the CTMC X. Let us denote the internal jump times of the Poisson process Y_j by I_i^j, where $j = 1, \cdots, m$ and $i = 1, 2, \cdots$. Note that $\{I_i^j\}$ is an increasing sequence, i.e.,

$$I_1^j < I_2^j < I_3^j \cdots$$

for each j. Also recall the internal time S_j we define in 1.2, we define $I_+^j(t, c)$ to be the internal time of the next firing of reaction channel j at time t, that is,

$$I_+^j(t) = \min\{I_l^i | I_l^i > S_j(t), l = 1, 2, \cdots\}. \tag{1.11}$$

Given T_i is the physical time of the ith firing of any reaction channel, one can use the collection $I_+^j(T_i)$ and J_{i+1} to determine the next firing time and its type, and $J_{i+1} \in \{1, 2, \cdots, m\}$ is the index of the reaction channel that fires at time T_{i+1}. Using these notations defined above and assuming that we know $T_1, \cdots, T_i$ and $J_1, \cdots, J_i$ for some i, (1.4) implies

$$T_{i+1} = T_i + \min\left\{ \frac{I_+^j(T_i) - S_j(T_i)}{a_j(X(T_i))} \middle| j = 1, \cdots, m \right\}. \tag{1.12}$$

We comment that the minimum is unique with probability 1 because $I_+^j(T_i)$ has continuous joint distribution given a constant T_i. To implement the RTC simulation algorithm, we need to generate m streams of unit exponential random numbers E_i^j such that $I_{i+1}^j - I_i^j = E_i^j$. Also we use k_j to denote the index of the jth stream of exponential random numbers. The RTC simulation algorithm is outlined in algorithm 2.

Algorithm 2 RTC simulation algorithm

Require: N_{tr}, terminal time T, initial state x_0

for $k = 1 : N_{\mathrm{tr}}$ **do**

Generate m independent exponential random numbers E_1^j for $j = 1, \cdots, m$

Initialize $i = 0$, $T_i = 0$, $t = 0$, $x = x_0$, $k_j = 1$, $S_j = 0$ and $I_+^j = E_1^j$

while $t < T$ **do**

Evaluate the intensity $a_j(x)$

Compute T_{i+1} from (1.12) and let j^* be the index of the minimum

Update $S_j \leftarrow S_j + a_j(x)(T_{i+1} - T_i)$ for $j = 1, \cdots, m$

Update $k_{j^*} \leftarrow k_{j^*} + 1$

Generate a unit exponential random number $E_{k_{j^*}}^{j^*}$ and update $I_+^{j^*} \leftarrow I_+^{j^*} + E_{k_{j^*}}^{j^*}$

Update $t = T_{i+1}$, $x \leftarrow x + \nu_{j^*}$, $i \leftarrow i + 1$

end while

end for

1.5 Contribution of this Thesis

In this thesis, we are concerned with the determination of parametric sensitivities of the form

$$\frac{\partial}{\partial c_k}\mathbb{E}(f(X(t,c))),$$

for $k = 1, \cdots, p$, where $t > 0$ and $f : \mathbb{R}^n \to \mathbb{R}$. We will provide some preliminaries (or tools) in Chapter 2 for our analysis in this thesis.

Generally, there is no analytical way to solve the sensitivity problem (unless the intensities are affine and f is a polynomial). Therefore, Monte Carlo simulation based methods are often used to estimate the sensitivity. These methods can be roughly divided into two categories - finite perturbation and infinitesimal perturbation methods. Within the finite perturbation category, there are several methods in current use based on different ways for coupling. Similarly, in the infinitesimal perturbation category, one can deal with the sensitivity problem using Girsanov transformation (Girsanov transformation method) or pathwise derivative (regularized pathwise derivative method). We will describe these methods in Chapter 3.

The major contribution of this thesis are in Chapter 4. It has been widely observed by researchers that when applicable, the finite difference and pathwise derivative methods tend to have lower variance and hence more efficient than the Girsanov transformation method in most cases. However, to the best of our knowledge, there is no existing literature that explains the larger variance of the Girsanov transformation method. In Chapter 4, we provide theoretical explanation for the observation in terms of the system size N, which is roughly the total molecular copy

number in the system. The approach we will take for this analysis is to construct a family of processes in terms of the system size N. We will eventually show that the estimator variance for these estimators are of different orders of N as $N \to \infty$.

The next contribution of this thesis is the rigorous derivation of the Girsanov transformation method. The development of the Girsanov transformation method is described in [22] is based on certain assumptions which need to be verified. In Chapter 5, we study the validity of these assumptions. Theorem 5.2.1 and Theorem 5.3.1 together provide some sufficient conditions for the validity of the Girsanov transformation method for sensitivity.

Finally, in Chapter 6, we extend the sensitivity estimation problem at a finite time horizon to that at an infinite time horizon, i.e., the sensitivity at the steady state. One often uses the time average to approximate the steady state mean ($t \to \infty$). Problems of this approach is that one has to take a large relaxation time to guarantee the accuracy. Moreover, the estimator variance normally grows dramatically in terms of time t. Therefore, variance reduction techniques are important for estimation problems at the steady state. We present a result concerning the asymptotic behavior of $\mathbb{E}[f(X(t))Z(t_0)]$ as $t \to \infty$ and apply this result to rigorously justify a variance reduction technique for steady state sensitivity problem.

Chapter 2: Preliminaries

In this chapter, we mainly describe some mathematical tools for our analysis in later chapters. Since the population process $X(t)$ is a jump process, much of the analysis in this thesis uses the Skorohod space $D[0,T]$ and we briefly describe the Skorohod metric in Section 2.1 for readers who are not familiar with this topic. In Section 2.2, we review the law of stochastic mass action and introduce the system size N into the intensity functions, which leads to the so-called density dependent processes. The functional law of large numbers and the functional central limit theorem will be discussed in Section 2.3 to relate the stochastic model and the deterministic model. These two theorems are crucial tools for our analysis in Chapter 5. Finally we cite the martingale functional central limit theorem in Section 2.4 which will also be used in Chapter 4.

2.1 Skorohod Space $D^n[0,T]$

Most stochastic processes arising in applications do not have continuous paths, which makes the continuous function space $C[0,T]$ (or $C[0,\infty)$) not sufficient to model these processes. For example, the population process $X(t)$ we consider in this thesis is a right continuous process with left limits, that is, $X(t) : [0,T] \to \mathbb{R}^n$

satisfies

$$\lim_{s \to t^+} X(s) = X(t) \qquad \text{and} \qquad \lim_{s \to t^-} X(s) = X(t-) \quad \text{exists}$$

Such a process is called a càdlàg process and we denote the space of càdlàg functions by $D^n[0,T]$.

To define the topology on the Skorohod space, we need a metric on $D^n[0,T]$. Let λ be a strictly increasing, continuous mapping from $[0,T]$ onto $[0,T]$. Note that this implies $\lambda(0) = 0$ and $\lambda(T) = T$. We denote Λ to be the class of such functions. For any $x, y \in D^n[0,T]$, we define the distance between x and y as

$$d(x,y) = \inf_{\lambda \in \Lambda} \{\|\lambda - I\| \vee \|x - y\lambda\|\},$$

where I is the identity map on $[0,T]$ and $\|\cdot\|$ is the supremum norm. Here the mapping λ serves as a time deformation function. The distance d can be shown to be a metric and is known as the Skorohod metric [7]. Now with the Skorohod metric, we can define the Borel σ-field of $D^n[0,T]$ by $\mathcal{D}$. We denote a sequence of σ-fields generated by the process $X(t)$ by $\mathcal{D}_t, t \leq T$, i.e., the natural filtration generated by X.

2.2 Law of Stochastic Mass Action

We describe the stochastic form of the law of mass action that commonly arises in stochastic chemical kinetics. If we divide the stoichiometric vector ν_j into two parts, such that $\nu_j = \nu'_j - \nu''_j$, where

ν'_j : the vector number of molecules of each species that are created in the jth reaction,

ν_j'' : the vector number of molecules of each species that are consumed in the jth reaction,

then the intensity of jth reaction is

$$a_j^N(x,c) = \frac{c_j}{N^{|\nu_j''|-1}} \prod_i^n \binom{x_i}{\nu_{ij}''} \tag{2.1}$$

where $|\nu_j''| = \sum_{i=1}^n \nu_{ij}''$ and N is the volume of the system times Avogadro's number (see [13] for derivation from first principles). We make the following remarks:

- here $\prod_i^n \binom{x_i}{\nu_{ij}''}$ is the number of ways to choose ν_{ij}'' reactant molecules out of the x_i number of total molecules from reactant species i for $i = 1, \cdots, n$.
- $\frac{1}{N^{|\nu_{ij}''|-1}}$ accounts for the effect of system volume on the probability.
- c_j is a constant of proportionality, we call it the deterministic parameter, which is not the same as the stochastic parameter we defined in the introduction (it is $c_j^N = \frac{c_j}{N^{|\nu_j''|-1}}$ indeed). Since they only differ by a scaling factor $N^{|\nu_j''|-1}$, we abuse notation by using the same c_j for them. The sensitivity problem we defined before is with respect to the stochastic parameter c_j^N (though we also use c_j there). For our analysis in Chapter 4, we actually take derivative with respect to the deterministic parameter c_j.

We give an example to illustrate the relationship between the intensity a_j^N and the deterministic reaction rate $\bar{a}_j$. Let us consider the reaction

$$2S \to 3S$$

with parameter c. Suppose the population of S is Nx, where x is the concentration. Then by the law of stochastic mass action, the intensity for this reaction is

$$a^N(Nx, c) = \frac{c}{2N} Nx(Nx - 1).$$

In this case, the deterministic parameter is c and the stochastic parameter $c^N = c/N$. To obtain the deterministic reaction rate, we normalize the above form of intensity by the volume N and take the limit as $N \to \infty$

$$\lim_N \frac{a^N(Nx, c)}{N} = \frac{c}{2} x^2.$$

We note that the right hand side is exactly the deterministic reaction rate $\bar{a}(x, c)$. More explicitly, the above two quantities differ by $cx/2N$, which is of order N^{-1}. In general, the intensity function a_j^N depend on N and x in a specific manner as follows.

$$\frac{a_j^N(Nx, c)}{N} = \bar{a}_j(x, c) + \mathcal{O}\left(\frac{1}{N}\right)$$

where $\bar{a}_j$ is the deterministic reaction rate as is in (1.1). The coefficient for the term $\mathcal{O}(N^{-1})$ depends on the concentration x and parameter c. Moreover, this coefficient is continuous in x. We can summarize the relationship between the stochastic intensity and deterministic reaction rate as follows. Suppose K is a compact set for x, for each $c > 0$, there exists a constant B_K such that

$$\left| \frac{a_j^N(Nx, c)}{N} - \bar{a}_j(x, c) \right| \leq \frac{B_K(c)}{N}. \tag{2.2}$$

We refer to this form of intensity as **density dependence** and thus call the family of processes X^N indexed by $N \geq 1$ corresponding to the family of intensity functions

a_j^N **density dependent processes** (see [10]). This density dependence leads to a deterministic limiting behavior in the large system size ($N \to \infty$). As a result it is instructive to study the family of processes X^N. Following (1.5) one can represent all these processes on the same sample space via the stochastic equation

$$X^N(t,c) = X^N(0,c) + \sum_{j=1}^{m} Y_j \left(\int_0^t a_j^N(X^N(s,c),c)\, ds \right) \nu_j, \quad N \geq 1, \tag{2.3}$$

where Y_j are independent unit rate Poisson processes. We also define the corresponding family of vector reaction count processes $R^N(t,c)$ whose jth component $R_j^N(t,c)$ counts the number of reaction events of type j that occurred during $(0,t]$. Thus

$$R_j^N(t,c) = Y_j \left(\int_0^t a_j^N(X^N(s,c),c)\, ds \right), \quad N \geq 1, \;\; j = 1, \ldots, m.$$

We also define the centered processes $M^N(t,c) = (M_1^N(t,c), \ldots, M_m^N(t,c))$ by

$$M_j^N(t,c) = R_j^N(t,c) - \int_0^t a_j^N(X^N(s,c),c) ds, \quad N \geq 1, \;\; j = 1, \ldots, m. \tag{2.4}$$

It can be shown that for each N, $M_j^N(t,c)$ is a zero mean local martingale and it becomes a martingale when $\int_0^t \mathbb{E}[a_j^N(X^N(s,c),c)] ds < \infty$ (see [8]).

2.3 Functional Law of Large Numbers and Functional Central Limit Theorem

In this section, we study the limiting behavior of $X^N(t,c)$ when N approaches infinity (see [10] for details). Our analysis for estimator efficiency will rely on the following two theorems. For brevity, we will omit the parameter c when we write

various quantities since the following analysis is for a fixed c. Let us define the density

$$X_N(t) = \frac{X^N(t)}{N}.$$

We note that X_N can be interpreted as the *concentration* of molecules at time t for system size N. First we state the law of large numbers (or fluid limit theorem) for density dependent processes.

Theorem 2.3.1. *(Law of large numbers [10]) Suppose for each compact $K \subset \mathbb{R}^n$, $F(x) = \sum_{j=1}^m \nu_j \bar{a}_j(x)$ is Lipschitz on K, that is, for each $x, y \in K$, there exists some constant M_K such that*

$$|F(x) - F(y)| \leq M_K |x - y|. \tag{2.5}$$

Suppose $X_N(0) = x_0 \in \mathbb{R}^n_+$ for all N. Let X be the solution for the equation

$$X(t) = x_0 + \int_0^t F(X(s))ds, \tag{2.6}$$

where we assume the solution exists on $[0, t]$. Then

$$\lim_N \sup_{s \leq t} |X_N(s) - X(s)| = 0 \quad a.s..$$

The proof of this theorem given in [10] is rather brief. We present a more detailed proof here. First we state the following two lemmas. The proof of the first lemma can be found in many real analysis textbooks.

Lemma 2.3.1. *Let $f_n : [0, T] \to \mathbb{R}$ be a sequence of increasing functions (increasing in t) which converges pointwise to $f : [0, T] \to \mathbb{R}$ which is continuous. Then f_n converges to f uniformly on $[0, T]$.*

Lemma 2.3.2. *(LLN for Poisson process) If Y is a unit rate Poisson process, then for each $t > 0$,*

$$\lim_{N\to\infty} \sup_{t\le T} \left| \frac{Y(Nt)}{N} - t \right| = 0 \quad a.s.$$

Proof. For each t, the independent increment property of Poisson process implies the law of large number,

$$\frac{Y(Nt)}{N} = \sum_{n=1}^{N} \frac{Y(nt) - Y((n-1)t)}{N} \to t \quad a.s..$$

Now by Lemma 2.3.1, the monotone property of $Y(Nt)$ implies the uniformity. □

Now we are ready to prove the law of large numbers. For convenience, we define the centered Poisson process $\tilde{Y}(t) = Y(t) - t$.

Proof. (of Theorem 2.3.1) Denote the state space of the Markov process X_N by $E_N = \{N^{-1}k : k \in \mathbb{Z}^n\}$. Choose $\eta > 0$ and for each N, define

$$\Gamma_N = E_N \cap \{y \in \mathbb{R}^n : \inf_{s\le t} |y - X(s)| < \eta\} \neq \emptyset.$$

Let K be the closure of $\{y \in \mathbb{R}^n : \inf_{s\le t} |y - X(s)| < \eta\}$. Note that $\Gamma_N \subset K$ for each N. Define the first exit time for the set Γ_N,

$$\tau_N = \inf\{t \ge 0 : X_N(t) \in \Gamma_N^c\}.$$

Now let us consider the truncated Markov processes $X_N(t \wedge \tau_N)$, where

$$\begin{aligned} X_N(t\wedge\tau_N) =& x_0 + \sum_{j=1}^{m} \nu_j N^{-1} \tilde{Y}_j\Big(\int_0^{t\wedge\tau_N} a_j^N(X^N(s))ds\Big) \\ &+ \sum_{j=1}^{m} \nu_j N^{-1} \int_0^{t\wedge\tau_N} a_j^N(X^N(s))ds. \end{aligned} \tag{2.7}$$

Define $A_j = \sup_{x\in K} \bar{a}_j(x)$ and let

$$\epsilon_N(t) = \sup_{s\le t} \left| X_N(s\wedge\tau_N) - x_0 - \sum_{j=1}^m \nu_j \int_0^{s\wedge\tau_N} \bar{a}_j(X_N(u))du \right|.$$

From (2.7),

$$\begin{aligned}
\epsilon_N(t) \le & \sup_{s\le t} \left| \sum_{j=1}^m \nu_j N^{-1} \tilde{Y}_j\Big(\int_0^{s\wedge\tau_N} a_j^N(X^N(u))du \Big) \right| \\
& + \sup_{s\le t} \left| \sum_{j=1}^m \nu_j N^{-1} \int_0^{s\wedge\tau_N} a_j^N(X^N(u))du - \sum_{j=1}^m \nu_j \int_0^{s\wedge\tau_N} \bar{a}_j(X_N(u))du \right| \\
\le & \sum_{j=1}^m |\nu_j| N^{-1} \sup_{s\le t} \left| \tilde{Y}_j\Big(\int_0^{s\wedge\tau_N} a_j^N(X^N(u))du \Big) \right| \\
& + \sum_{j=1}^m |\nu_j| \int_0^{t\wedge\tau_N} \left| N^{-1} a_j^N(X^N(u)) - \bar{a}_j(X_N(u)) \right| du \\
\le & \sum_{j=1}^m |\nu_j| N^{-1} \sup_{s\le t} \left| \tilde{Y}_j(NA_j s + B_K s) \right| + \frac{B_K}{N} t \sum_{j=1}^m |\nu_j|
\end{aligned} \tag{2.8}$$

where the last inequality is by (2.2). Taking $N\to\infty$, by Lemma 2.3.2,

$$\lim_{N\to\infty} \sup_{s\le t} \left| N^{-1} \tilde{Y}_j(NA_j s + B_K s) \right| = 0 \quad a.s., \tag{2.9}$$

which shows that $\lim_N \epsilon_N(t) \to 0$ for each fixed t.

Now, for each $s \le t$,

$$\left| X_N(s\wedge\tau_N) - x_0 - \sum_{j=1}^m \nu_j \int_0^{s\wedge\tau_N} \bar{a}_j(X_N(u))du \right| < \epsilon_N(t).$$

This and the Lipschitz condition (2.5) imply

$$|X_N(s\wedge\tau_N) - X(s\wedge\tau_N)| \le \epsilon_N(t) + \int_0^{s\wedge\tau_N} M_K |X_N(u) - X(u)| du.$$

By Gronwall's lemma,

$$\begin{aligned}
& \sup_{s\le t} |X_N(s\wedge\tau_N) - X(s\wedge\tau_N)| \\
& \qquad \le \epsilon_N(t) e^{M_K t} \\
& \qquad < \eta,
\end{aligned} \tag{2.10}$$

for sufficiently large N and hence sufficiently small $\epsilon_N(t)$.

Now for these sufficiently large N,

$$|X_N(s \wedge \tau_N)| - X(s \wedge \tau_N)| < \eta, \quad s \leq t,$$

we have $X_N(s \wedge \tau_N) \in \Gamma_N$ and $\tau_N > s$. Therefore, (2.10) becomes

$$\sup_{s \leq t} |X_N(s) - X(s)| < \eta,$$

hence the fluid limit theorem. □

Recall that $F(x) = \sum_{j=1}^m \nu_j \bar{a}_j(x)$ and X is the solution to

$$X(t) = x_0 + \int_0^t F(X(s))ds.$$

Define $V_N(t) = \sqrt{N}(X_N(t) - X(t))$,

$$\begin{aligned} V_N(t) =& \sum_{j=1}^m \nu_j N^{-1/2} \tilde{Y}_j(\int_0^t a_j^N(X^N(s))ds) \\ &+ \sum_{j=1}^m \nu_j N^{-1/2} \int_0^t a_j^N(X^N(s)) - N\bar{a}_j(X(s))ds \\ =& \sum_{j=1}^m \nu_j N^{-1/2} \tilde{Y}_j(\int_0^t a_j^N(X^N(s))ds) + \sqrt{N} \int_0^t F(X_N(s)) - F(X(s))ds \\ &+ \sum_{j=1}^m \nu_j \sqrt{N} \int_0^t a_j^N(X^N(s))/N - \bar{a}_j(X_N(s))ds. \end{aligned}$$

The above equation suggests a weak limit for V_N,

$$V(t) = \sum_{j=1}^m \nu_j W_j(\int_0^t \bar{a}_j(X(s))ds) + \int_0^t \partial F(X(s))V(s)ds \tag{2.11}$$

where W_j are independent standard Brownian motions and ∂F is the Jocobian of F. The rigorous proof can be found in [10]. We state this functional central limit theorem as follows.

Theorem 2.3.2. *Suppose that the $\bar{a}_j$ and ∂F are continuous. Then $V_N \Rightarrow V$ in $D^n[0,T]$, where V is the solution to* (2.11).

2.4 Martingale Functional Central Limit Theorem

We state an important result which will be used in Chapter 5 (See [32]).

Theorem 2.4.1. *Let M^N be a sequence of $\mathbb{R}^m$-valued local martingales. Suppose the expectation of the maximum jump in M^N is asymptotically negligible, that is, for each $t > 0$*

$$\lim_N \mathbb{E}\{\sup_{s \le t} |M^N(s) - M^N(s-)|\} = 0,$$

and, for each pair $1 \le i, j \le m$ and each $t > 0$,

$$[M_i^N, M_j^N](t) \to d_{i,j}(t),$$

where the convergence is in probability and $D(t) = \{d_{i,j}(t)\}$ is deterministic and continuous matrix-valued function. Then $M^N \Rightarrow M$ in $D^m[0,T]$, where M is Gaussian with independent increments having mean vector and covariance matrix

$$\mathbb{E}M(t) = 0 \quad \textit{and} \quad \mathbb{E}[M(t)M(t)^T] = D(t).$$

Chapter 3: Parametric Sensitivity Methodologies

Monte Carlo simulation methods for sensitivity analysis can be roughly divided into two categories : the finite perturbation and the infinitesimal perturbation. In Section 3.2, we will describe various finite difference methods based on difference coupling schemes. For the methods from the infinitesimal perturbation category, we will review the Girsanov transformation method in Section 3.3 and the pathwise derivative method in Section 3.4.

3.1 Parametric Sensitivity

Recall that we denote the Markov process by $\{X(t,c) : t \geq 0\}$ and its associated intensity functions by $a_j(x,c)$ for $j = 1, \cdots, m$. In general, a stochastic reaction network has several parameters. We shall focus on one scalar parameter c. Given a function $f : \mathbb{Z}^n \to \mathbb{R}$ and a final time $T \geq 0$, the **parametric sensitivity** is defined as

$$s_j(c) = \frac{\partial}{\partial c_j}\mathbb{E}[f(X(T,c))].$$

This partial derivative evaluates how the expected value of output $f(X(T,c))$ changes with respect to the perturbation of parameter c. To simplify notations, without loss of generality, we assume that the derivative is always taken with respect to c_1. In

the rest of this paper, we abuse of notation by writing the parametric sensitivity as

$$s = \frac{\partial}{\partial c}\mathbb{E}[f(X(T,c))] \tag{3.1}$$

Readers should keep in mind that s is the parametric sensitivity with c_1. Since we can use SSA to simulate the path of $X(t,c)$ easily, finite difference scheme (FD) can be used to approximate the sensitivity s as

$$s = \frac{1}{h}\mathbb{E}[f(X(T,c+h)) - f(X(T,c))] \tag{3.2}$$

and its associated crude Monte Carlo estimator is

$$S_{\mathrm{FD}} = \frac{1}{h\mathrm{Ntr}}\sum_{i=1}^{\mathrm{Ntr}}(f(X^{(i)}(T,c+h)) - f(X^{(i)}(T,c))) \tag{3.3}$$

where $X^{(i)}(T,c), i = 1, \cdots, \mathrm{Ntr}$ as well as $X^{(i)}(T,c+h), i = 1, \cdots, \mathrm{Ntr}$ are an i.i.d. sample of $X(T,c)$ and $X(T,c+h)$ respectively. N_{tr} is the sample size. Finite difference is easy to implement, however, it has several drawbacks. The most significant one is that finite difference schemes always introduce a bias into the sensitivity and this bias cannot be eliminated even by taking infinitely many samples. There is a bias-variance trade-off issue for finite difference methods. One can take small perturbation h to reduce bias, however, the associate estimator variance will blow up as $h \to 0$ which requires us to generate much larger number of samples to maintain the estimated sensitivity within a desired confidence interval.

Instead of approximating the sensitivity via finite perturbation, another category of methods for sensitivity analysis estimate the sensitivity directly and we refer it as the infinitesimal perturbation scheme. Within the infinitesimal perturbation

category, there are two different points of view to model the system's dependence on parameters:

- Process X is fixed at c_0 while the underlying probability measure P is varied with c (i.e.,$P = P(c), X = \{X(t, c_0, \omega); t \leq T\}$);
- Probability measure P is fixed while process X is varied with c (i.e.,$P = P, X = \{X(t, c, \omega); t \leq T\}$).

The above two points of view are equivalent but lead to two completely different infinitesimal perturbation methods. With the first point of view, we obtain the method known as the Girsanov transformation (GT) or likelihood ratio method [22]. In the GT approach, the derivative operator in (3.1) can be "absorbed" into the final formula

$$s = \mathbb{E}_{c_0}[f(X(T))Z(T)] \tag{3.4}$$

and its crude Monte Carlo estimator is

$$S_{\mathrm{GT}} = \frac{1}{\mathrm{Ntr}} \sum_{i=1}^{\mathrm{Ntr}} f(X^{(i)}(T)Z^{(i)}(T)) \tag{3.5}$$

where $\{Z(t) : t \leq T\}$ is the weight process and the subscript c_0 is used to emphasize that the underlying probability measure is $P(c_0)$. Unlike the FD method, GT has the advantage of being an unbiased estimator and achieving the sensitivity with respect to several parameters in a single run. However, it has been observed by several researchers that the GT estimator has larger variance than the other estimators [22, 29, 31]. Therefore, we suggest a modified version of GT by centering $f(X(t))$ at $\mathbb{E}f(X(t))$ and we refer to it as the **centered Girsanov transformation** algorithm.

A method very similar to this was proposed by Warren and Allen in [30]. The Monte Carlo estimator for CGT is

$$S_{\mathrm{CGT}} = \frac{1}{N_{\mathrm{tr}}} \sum_{i=1}^{N_{\mathrm{tr}}} (f(X^{(i)}(T)) - \overline{f(X(T))}) Z^{(i)}(T), \tag{3.6}$$

where $\overline{f(X(T))}$ is the sample mean $N^{-1} \sum_{i=1}^{N_{\mathrm{tr}}} f(X^{(i)}(T))$. We will explore the GT method and CGT method in greater detail later in this chapter.

Under the second point of view, another approach in the infinitesimal perturbation category estimates the sensitivity using pathwise differentiation ([29]). However, a direct pathwise differentiation is not applicable for sensitivity analysis in the context of stochastic reaction networks. Therefore, the original sensitivity problem was modified to the following **regularized pathwise derivative (RPD)**

$$s = \frac{\partial}{\partial c} \mathbb{E} \left[\int_{T-w}^{T+w} \frac{1}{2w} f(X(t,c))\, dt \right],$$

where w is the half regularization window size. It has been observed that RPD is an efficient method. However, this method also produces a bias and its efficiency depends crucially on the selection of parameter w. On the other hand, RPD is not applicable to certain type of problems. For example, in the reaction system

$$S_1 \to \emptyset, \quad 2S_1 \to \emptyset$$

with initial population $X_1(0) = 2$, if the first reaction fires first, it will prevent the occurrence of the second reaction. In this case, RPD is not applicable for sensitivity estimation any more. Unfortunately, this type of situation is common for the stochastic reaction networks, making RPD only applicable to a small set of

problems. A recent hybrid pathwise sensitivity method proposed by David Anderson [1] extends the applicability of RPD by using multilevel Monte Carlo approach.

3.2 Finite Difference

3.2.1 Independent Random Number Method

The simplest way to estimate the sensitivity s is to simulate the processes $X(t, c+h)$ and $X(t, c)$ independently, which is often referred as **independent random number (IRN)** method in stochastic simulation literature. We briefly describe the implementation of IRN as follows.

Algorithm 3 Independent random number algorithm

Require: c_0, h, N_{tr}, T

for $k = 1 : N_{\text{tr}}$ **do**

Initialized the $U[0, 1]$ random number generator with random seed

Set $c = c_0$ and run SSA to compute $X(T, c_0)$

Re-initialize the $U[0, 1]$ random number generator with a different seed

Set $c = c_0 + h$ and run SSA to compute $X(T, c_0 + h)$

Compute the sensitivity $(f(X(T, c_0+h)) - f(X(T, c_0)))/h$ for the ith trajectory

end for

IRN is easy to implement, however, it often leads to large variance of the estimator of S_{FD}. This fact can be seen from the following analysis. Let us denote $S_1 = f(X(T, c + h)), \quad S_2 = f(X(T, c))$. Since $S_{FD} = (S_1 - S_2)/h$, the variance of

S_{FD} can be written as follows

$$\frac{1}{h^2}(\mathrm{Var}(S_1) + \mathrm{Var}(S_2) - 2\mathrm{Cov}(S_1, S_2)). \tag{3.7}$$

Observing that in the IRN method, the covariance part vanishes since S_1 and S_2 are independent. One can couple processes $f(X(t,c))$ and $f(X(t,c+h))$ in some intelligent ways to increase the covariance and hence reduce the estimator variance. Different coupling schemes between $f(X(t,c))$ and $f(X(t,c+h))$ lead to different FD estimators.

3.2.2 Common Random Number Method

The idea behind **common random number (CRN)** method is simple. By using the same sequence of random numbers, one introduces positive covariance between $f(X(t,c))$ and $f(X(t,c+h))$ provided that h is sufficiently small.

Algorithm 4 Common random number algorithm

Require: $c_0, h, N_{\mathrm{tr}}, T$

for $k = 1 : N_{\mathrm{tr}}$ **do**

Initialized the $U[0,1]$ random number generator with seed ω;

Set $c = c_0$ and run SSA to compute $X(T, c_0)$;

Re-seed the $U[0,1]$ random number generator with the same seed ω;

Set $c = c_0 + h$ and use the same random seed run SSA to compute $X(T, c_0 + h)$;

Compute the sensitivity $(f(X(T, c_0+h)) - f(X(T, c_0)))/h$ for the ith trajectory.

end for

3.2.3 Common Reaction Path Method

The CRN can be implemented with any variants of SSA like Gillespie's direct method, Gibson and Bruck's next reaction method or Anderson's modified next reaction method. In this section, we describe a special form of CRN, i.e., the implementation of CRN in the context of RTC simulation algorithm proposed by Rathinam et al ([26]), which is known as the **common reaction path (CRP)** method. Recall that the RTC simulation algorithm simulates trajectories by keeping track of the collection $(I_+^j(T_i), J_{i+1})$ for $i = 1, 2, \cdots$, where T_i is the random times at which the ith reaction event of any type occurs, J_{i+1} is the index of the reaction channel which fires immediately after T_i and $I_+^j(T_i)$ is the internal firing time after T_i for the jth reaction channel . The basic idea for CRN method is to use the same underlying random numbers to reduce the variance. While in CRP, one utilizes the structure of random time change representation and use the same driving Poisson processes, i.e., the same paths of the Poisson processes are used to simulate the perturbed and nominal trajectories. The coupling can be seen from the equation (3.8).

$$\begin{aligned} X(t, c+h) &= X(0, c+h) + \sum_{j=1}^{m} \nu_j Y_j \left(\int_0^t a_j(X(s, c+h), c+h) ds \right) \\ X(t, c) &= X(0, c) + \sum_{j=1}^{m} \nu_j Y_j \left(\int_0^t a_j(X(s, c), c) ds \right) \end{aligned} \tag{3.8}$$

where Y_j are independent unit-rate processes. Heuristically, one can expect that $X(t, c+h)$ and $X(t, c)$ will be highly correlated since they are driven by the same Poisson processes. We outline the CRP algorithm before we discuss more details.

Algorithm 5 Common reaction path algorithm

Require: c_0, h, N_{tr}, T

Need m parallel streams of independent random number to generate $E_1^j, E_2^j, \cdots$, $j = 1, \cdots, m$

for $k = 1 : N_{\text{tr}}$ **do**

Set $c = c_0$ and run RTC simulation algorithm to compute $X(T, c_0)$

Reseed the streams to generate the same $E_1^j, E_2^j, \cdots$ for $j = 1, \cdots, m$

Set $c = c_0 + h$ and run RTC simulation algorithm to compute $X(T, c_0 + h)$ using $E_1^j, E_2^j, \cdots$ for $j = 1, \cdots, m$

Compute the sensitivity $(f(X(T, c_0 + h)) - f(X(T, c_0)))/h$ for the ith trajectory

end for

Recall that in the RTC simulation algorithm, we assume the existence of m streams of unit exponential random numbers E_i^j for $i = 1, 2, \cdots$ and $j = 1, \cdots, m$. These random numbers represent the inter-arrival times of the unit rate Poisson processes Y_j and relate the internal firing times I_i^j by $I_{i+1}^j - I_i^j = E_i^j$. We use these random numbers E_i^j to implement the RTC simulation algorithm. Therefore, the implementation of CRP is slightly harder than that of CRN. Instead of using a single stream of unit rate exponentials, CRP requires m parallel streams of exponentially distributed random numbers for implementation. The reason is that by using M parallel streams of random numbers, one can ensure that each of the reaction paths will be identical between the perturbed process the nominal process, that is, they will share the same internal firing times I_i^j. Therefore, the two processes will be tightly coupled, see [26] for a detailed discussion on this point. On the other hand,

CRP tends to be more efficient than CRN with other simulation algorithms because the jump times for CRP algorithm are continuous with respect to the parameters c while this is not true for general CRN. We will not explore this property in details here, readers who are interested in the proof are referred to the appendix B in [26].

3.2.4 Coupled Finite Difference Method

The **coupled finite difference (CFD)** methods proposed by Anderson [2] also takes advantages of the coupling property between the perturbed and nominal processes. We use the same example from [2] to motivate the CFD method. Let Y_1 and Y_2 be independent unit-rate Poisson processes and construct another two processes

$$Z_1(t) = Y_1(13t) + Y_2(0.1t)$$

$$Z_2(t) = Y_1(13t)$$

By the additivity property of Poisson processes, $Z_1(t)$ and $Z_2(t)$ are Poisson processes with rate 13 and 13.1. Note that both Z_1 and Z_2 have the same source of jumps from $Y_1(13t)$ but Z_1 has an extra source of jumps from $Y_2(0.1t)$. Since $Y_1(13t)$ has much larger rate than $Y_2(0.1t)$, we expect Z_1 and Z_2 will jump together for most of the time and hence be tightly coupled. Motivated by this simple example, CFD

uses the following coupling scheme

$$\begin{aligned} X(t,c+h) =& X(0,c+h) + \sum_{j=1}^{m} Y_{j,1}\left(\int_0^t b_j(s)ds\right) \\ &+ \sum_{j=1}^{m} Y_{j,2}\left(\int_0^t a_j(X(s,c+h),c+h) - b_j(s)ds\right) \\ X(t,c) =& X(0,c) + \sum_{j=1}^{m} Y_{j,1}\left(\int_0^t b_j(s)ds\right) \\ &+ \sum_{j=1}^{m} Y_{j,3}\left(\int_0^t a_j(X(s,c),c) - b_j(s)ds\right) \end{aligned} \tag{3.9}$$

where $b_j(s) = a_j(X(s,c+h),c+h) \wedge a_j(X(s,c),c)$ and $Y_{j,1}, Y_{j,2}, Y_{j,3}$ are independent unit-rate Poisson processes. The processes $Y_{j,1}\left(\int_0^t b_j(s)ds\right)$ have the same contribution to $X(t,c+h)$ and $X(t,c)$, but each of which have some extra source of jump from $Y_{j,2}$ and $Y_{j,3}$ respectively. On the other hand, the marginal processes of $(X(t,c+h), X(t,c))$ generated by (3.9) have the same distribution as the respective processes generated from (3.8), this fact guarantees the correctness of the CFD algorithm.

The implementation of CFD has the advantage that it does not require m parallel streams of exponentially distributed random numbers as CRP. More importantly, numerical examples shows that CFD is more efficient than CRP in most cases and it is more analytically tractable. In fact, under some mild assumptions, one can derive an explicit upper bound for the estimator variance of CFD (see [2]). Interested readers can refer to [2] for more details.

3.3 Girsanov Transformation Method

In this section, we briefly review the GT method in a heuristic manner. The validity of using Girsanov's change of measure theory for sensitivity analysis will be discussed in Chapter 5. Let us consider the general probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the space of càdlàg functions $(D[0,T], \mathcal{D}, P(c))$ for each c, where the $P(c)$ is the law induced by the process $X(t,c)$, i.e.,

$$P(c, A) = \mathbb{P}(\omega | X(\ldots, c, \omega) \in A)$$

for all $A \in \mathcal{D}$. Suppose we are interested in the sensitivity at $c = c_0$, i.e.,

$$s = \frac{\partial}{\partial c}\bigg|_{c=c_0} \mathbb{E}[f(X(T,c))].$$

We can rewrite the above formula as

$$\frac{\partial}{\partial c}\bigg|_{c=c_0} \int_\Omega f(X(T,c))\mathbb{P}(d\omega) = \frac{\partial}{\partial c}\bigg|_{c=c_0} \int_{D[0,T]} f(x(T))P(c, dx)$$

If $P(c)$ is absolutely continuous with respect to $P(c_0)$, then the Radon-Nikodym derivative (or likelihood ratio) is

$$l(t, x, c) = \frac{dP(c)}{dP(c_0)}\bigg|_{\mathcal{D}_t}$$

and therefore

$$\frac{\partial}{\partial c}\bigg|_{c=c_0} \int_{D[0,T]} f(x(T))P(c, dx) = \frac{\partial}{\partial c}\bigg|_{c=c_0} \int_{D[0,T]} f(x(T))l(T, x, c)P(c_0, dx).$$

Now suppose that we can interchange the derivative with integral, then

$$s = \int_{D[0,T]} f(x(T))z(T, x, c_0)P(c_0, dx)$$

where $z(t,x,c_0) = \left.\dfrac{\partial}{\partial c}\right|_{c=c_0} l(t,x,c)$. Now, we switch back to the space $(\Omega, \mathcal{F}, \mathbb{P})$ and write the sensitivity as

$$s = \int_\Omega f(X(T,c_0))Z(T,\omega,c_0)\mathbb{P}(d\omega),$$

where $Z(t,\omega,c_0) = z(t,X(\omega,c0),c_0)$. Since c_0 is fixed, we denote $Z(t,\omega,c_0) = Z(t)$ for simplicity. Therefore, the final formula for sensitivity is

$$s = \mathbb{E}[f(X(T,c_0))Z(T)]. \tag{3.10}$$

The weight process can be explicitly written as [26, 29]

$$Z(t) = \sum_j \int_{(0,t]} \frac{\dfrac{\partial}{\partial c}a_j(X(s-),c)}{a_j(X(s-),c)} dR_j(s) - \sum_j \int_0^t \frac{\partial}{\partial c}a_j(X(s-),c)ds.$$

In general, $Z(t)$ is a zero-mean local martingale. We summarize the implementation of the GT in Algorithm 6.

From this algorithm, we can see that the implementation of GT is no harder than that of SSA. Also, unlike the FD method, GT can estimate the sensitivity with respect to several parameters simultaneously and it is an unbiased method. However, for most numerical examples we have tested, GT tends to be inefficient in the sense that it often leads to large estimator variance. Therefore, modifications to GT which reduce the variance would be desirable. For this purpose, we suggest a modified version of GT by centering the GT estimator at $\mathbb{E}f(X(T))$, we refer to it as the CGT method (a very similar method is proposed by Allen and Warren in [30]). This modification is reasonable because for each c, $Z(t)$ is a zero mean local martingale, therefore

$$\mathbb{E}[f(X(t))Z(t)] = \mathbb{E}[(f(X(t)) - \mathbb{E}f(X(t)))Z(t)].$$

Algorithm 6 Girsanov transformation sensitivity algorithm

Require: c_0, N_{tr}, T

for $i = 1 : N_{\text{tr}}$ **do**

Initialize $t = 0$, $X(t) = x_0$ and $Z_j = 0$ for $j = 1, \cdots, M$

while $t < T$ **do**

Calculate $a_j(X(t), c_0)$ and $\dfrac{\partial a_j}{\partial c}$ for $j = 1, \cdots, M$

Set $a_0 = \sum_{j=1}^{M} a_j$

Generate r_1 and r_2 from $U[0, 1]$

Set the inter-arrival time $\tau = -\log(r_1)/a_0$

Find the reaction type $j^* \in [1, \ldots, M]$ which satisfies

$$\sum_{j=1}^{j^*-1} a_j < r_2 a_0 \leq \sum_{j=1}^{j^*} a_j$$

Set $\Delta Z_{j^*} = \dfrac{\partial a_{j^*}}{\partial c}\dfrac{1}{a_{j^*}} - \tau\dfrac{\partial a_{j^*}}{\partial c}$ and $\Delta Z_j = -\tau\dfrac{\partial a_j}{\partial c}$ for $j \neq j^*$

Update $Z \leftarrow Z + \sum_{j=1}^{M} \Delta Z_j$, $t \leftarrow t + \tau$ and $X(t) \leftarrow X(t) + \nu_{j^*}$

end while

Compute the sensitivity $f(X)Z$ for the ith trajectory

end for

For most numerical examples we tested, the CGT method is more efficient than the GT method. However, it is not necessarily the case when we choose small initial population x_0. We provide a theoretical justification for this observation in Chapter 4, which is also the main contribution of this paper.

Finally, we would like to point out that the above derivation for GT method is not rigorous. The first place which needs some justifications is the change of measure step. Given the two probability measures $P(c)$ and $P(c_0)$, it is not clear if they are equivalent to each other, not to mention finding the formula for the Radon-Nikodym derivative $L(t, c)$ and then obtaining a formula for the weight process $Z(t, c)$. In fact, this relates to the Novikov type condition problem in stochastic analysis, which is crucial for the validity of changing measure. On the other hand, we have not provided any conditions for differentiating inside the integral. We will give a set of conditions in Chapter 5 that validate these two steps.

3.4 Regularized Pathwise Derivative Method

In this section, we will review the regularized pathwise derivative (RPD) method from [29]. To derive RPD method, we work in the abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which carries the m independent Poisson processes Y_j. Recall that in Chapter 2 we denote the internal time of reaction channel j

$$S_j(t, c) = \int_0^t a_j(X(s, c), c)ds$$

and we express the RTC representation as

$$X(t,c) = X(0,c) + \sum_{j=1}^{M} \nu_j Y_j(S_j(t,c)),$$

where $Y_j, j = 1, \cdots, M$ are independent unit-rate Poisson random variable associated with each reaction channel. Also recall that we denote the random jump times of the Poisson process Y_j by I_i^j where $j = 1, \cdots, M, i = 1, 2, \cdots$. We can use the collection (T_i, J_i) to keep track of the time evolution of the reaction system (see Section 1.4 for definition). We remind the reader that T_i, J_i and S_j all depend on the parameter c, but we suppress it for notation convenience. T_i follows the following recursive formula

$$T_{i+1} = T_i + \min\left\{ \frac{I_+^j - S_j(T_i)}{a_j(X(T_i))} \middle| j = 1, \cdots, m \right\}$$

where $I_+^j(t)$ is the internal time for the next potential firing of reaction channel j at physical time t, i.e.,

$$I_+^j(t) = \min\{I_l^i | I_l^i > S_j(t), l = 1, 2, \cdots\}$$

for $j = 1, \cdots, m$.

Under the assumption that $a_j(x,c)$ are smooth functions of c, one can show that T_i is piecewise differentiable with respect to c. Moreover, $X(T_i)$ and $I_+^j(T_i)$ are locally constant in c. Hence we can differentiate (1.12) and obtain the recursive formula:

$$\frac{\partial T_{i+1}}{\partial c} = \frac{\partial T_i}{\partial c} - \frac{I_+^{j^*}(T_i) - S_{j^*}(T_i)}{(a_{j^*}(X(T_i,c),c))^2} \frac{\partial a_{j^*}(X(T_i,c),c)}{\partial c} - \frac{\dfrac{dS_{j*}(T_i)}{dc}}{a_{j^*}(X(T_i,c),c)} \tag{3.11}$$

where j^* is the index of minimum in (1.12). The derivative $dS_{j^*}(T_i)/dc$ can be found by differentiating the recursive formula (1.7),

$$\frac{dS_j(T_{i+1})}{dc} = \frac{dS_j(T_i)}{dc} + \frac{\partial a_j(X(T_i,c),c)}{\partial c}(T_{i+1}-T_i) + a_j(X(T_i,c),c)\left(\frac{\partial T_{i+1}}{\partial c} - \frac{\partial T_i}{\partial c}\right). \tag{3.12}$$

With the initial conditions

$$\frac{\partial T_0}{\partial c} = 0, \frac{dS_j(T_0)}{dc} = 0, j = 1, \cdots, m. \tag{3.13}$$

we can solve for $\partial T_i/\partial c, i = 0, 1, \cdots$ recursively.

Next we shall derive a formula for the sensitivity estimator in term of $\partial T_i/\partial c$. The pathwise derivative approach requires us to compute

$$\mathbb{E}\left[\frac{\partial f(X(T,c))}{\partial c}\right].$$

Unfortunately, in the context of stochastic reaction networks, it is with probability one that $X(T)$ is locally a constant in c and hence the pathwise derivative would be zero in a neighborhood of this c. To resolve this issue, Rathinam et al [29] modify this problem by introducing a regularization window $2w$

$$\frac{\partial}{\partial c}\mathbb{E}\left[\int_{T-w}^{T+w} \frac{1}{2w} f(X(t))dt\right].$$

With this modification, we introduce some bias into the estimator but we expect that this bias is negligible when w is small enough.

Following [29], we define

$$\Delta_i(c) = f(X(T_i(c),c)) - f(X(T_{i-1}(c)),c).$$

To derive the final sensitivity estimator, we rewrite the process $f(X(t,c))$ as follows

$$f(X(t,c)) = f(x_0) + \sum_{i=1}^{\infty} \Delta_i(c) I_{[T_i,\infty)}(t).$$

Observe that for each t, there are only finitely many terms in the above formula are nonzero. Therefore,

$$\frac{\partial}{\partial c}\Big|_{c=c_0} \int_{T-w}^{T+w} f(X(t,c))dt = \sum_{i=1}^{\infty} \frac{\partial}{\partial c}\Big|_{c=c_0} \Delta_i(c) \int_{T-w}^{T+w} I_{[T_i,\infty)}(t)dt.$$

Note that $\partial\Delta_i(c)/\partial c = 0$ since $X(T_{i-1})$ and $X(T_i)$ are locally a constant in c. Hence

$$\frac{\partial}{\partial c}\Big|_{c=c_0} \int_{T-w}^{T+w} f(X(t,c))dt = \sum_{i=1}^{\infty} \Delta_i(c_0) \frac{\partial}{\partial c}\Big|_{c=c_0} \int_{T-w}^{T+w} I_{[T_i,\infty)}(t)dt.$$

Finally, one can show that

$$\frac{\partial}{\partial c}\Big|_{c=c_0} \int_{T-w}^{T+w} I_{[T_i,\infty)}(t)dt = -\frac{\partial T_i(c_0)}{\partial c}$$

for $T_i(c_0) \in (T-w, T+w)$ and zero if $T_i(c_0) < T-w$ or $T_i(c_0) > T+w$. Let us define i_l and i_u to be the lowest and highest indices such that $T_i(c_0) \in (T-w, T+w)$, then the regularized pathwise sensitivity is given by

$$\frac{1}{2w} \frac{\partial}{\partial c}\Big|_{c=c_0} \int_{T-w}^{T+w} f(X(t,c))dt = -\frac{1}{2w} \sum_{i=i_l}^{i_u} \Delta_i(c_0) \frac{\partial T_i(c_0)}{\partial c}. \tag{3.14}$$

We skip the implementation of RPD algorithm here. Readers who are interested in it can refer to [29] for details.

Chapter 4: Estimator Efficiency Analysis

This chapter contains the main results we established regarding the efficiency of various Monte Carlo based sensitivity estimators. It has been observed that when applicable, the PD or FD methods tend to be more efficient than GT methods in most cases [22, 26, 29]. However, to our knowledge, no existing literature provides theoretical justifications for these observations. Moreover, the CGT method we mentioned in Chapter 3 outperforms the GT method in many examples we tested. We will provide theoretical explanations for these observations in terms of the system size N. First, we motivate our analysis in Section 4.1. In Section 4.2, we make some assumptions to facilitate our analysis. We establish some important results regarding the weight process $Z^N(t)$ in Section 4.3 and these results will be used to show the efficiency of various estimators in terms of the system size N. Since our analysis is for fixed parameter c, we will omit the c for convenience in this chapter except when it appears explicitly outside $X(t), X^N(t), R^N(t)$ or $R(t)$.

4.1 Motivating Example

To motivate our analysis in this chapter, we consider the following pure death model

$$S \xrightarrow{c} \emptyset$$

with intensity function $a(x, c) = cx$. We want to understand the sensitivity of the expected number of molecules S with respect to the parameter c. The population process $X(t)$ and the weight process $Z(t)$ associated with the GT method in this example are

$$\begin{aligned} X(t) &= x_0 - \int_{(0,t]} dR(s) \\ Z(t) &= \int_{(0,t]} \frac{1}{c} dR(s) - \int_0^t X(s) ds. \end{aligned} \tag{4.1}$$

One can use the Ito formula for processes driven by finite variation processes (see [28]) to write down the equation for $X(t)Z(t)$,

$$X(t)Z(t) = \int_{(0,t]} \left(\frac{X(s-)}{c} - Z(s-) - \frac{1}{c} \right) dR(s) - \int_0^t X(s)^2 \, ds.$$

Taking expected value of both sides leads to (after simplification)

$$\begin{aligned} \mathbb{E}(X(t)Z(t)) &= \int_0^t \mathbb{E}\left(\left(\frac{X(s-)}{c} - Z(s-) - \frac{1}{c} \right) cX(s) \right) ds - \int_0^t \mathbb{E}(X(s)^2) ds \\ &= -c \int_0^t \mathbb{E}(X(s)Z(s)) \, ds - \int_0^t \mathbb{E}X(s) \, ds. \end{aligned} \tag{4.2}$$

Also note that

$$\mathbb{E}X(t) = x_0 - \int_0^t \mathbb{E}X(s) \, ds. \tag{4.3}$$

The last equation and (4.2) form a system of linear ODEs. For the initial condition, we assume that the initial population is deterministic, i.e., $X(0) = x_0$, then $\mathbb{E}X(0) =$

x_0 and $\mathbb{E}[X(0)Z(0)] = 0$. One can easily solve this ODE system to obtain

$$\mathbb{E}(X(t)Z(t)) = \frac{x_0}{1-c}(e^{-t} - e^{-ct}).$$

Similarly, to find an analytical formula for $\mathbb{E}(X(t)^2 Z(t)^2)$, one can derive a linear system of ODEs involving variables of the form $\mathbb{E}(X(t)^\alpha Z(t)^\beta)$, where α and β are integers satisfying $\alpha \leq 2, \beta \leq 2$. The variances of GT and CGT estimators can be shown to be

$$\begin{aligned} Var(X(t)Z(t)) = \frac{1}{c^2}(&e^{-2ct}x_0^3 - 4e^{-2ct}x_0^2 + 3e^{-2ct}x_0 + 3e^{-2ct}x_0^2t^2c^2 \\ &- 2e^{-3ct}x_0 + 3e^{-3ct}x_0^2 + e^{-ct}x_0^2 - e^{-ct}x_0 \\ &+ e^{-ct}x_0t^2c^2 - 4e^{-2ct}t^2c^2x_0 - e^{-3ct}x_0^3), \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} Var((X(t) - \mathbb{E}[X(t)])Z(t)) = \frac{1}{c^2}(&-2e^{-2ct}x_0^2 + 3e^{-2ct}x_0 + e^{-2ct}x_0^2t^2c^2 \\ &- 2e^{-3ct}x_0 + e^{-3ct}x_0^2 + e^{-ct}x_0^2 \\ &- e^{-ct}x_0 + e^{-ct}x_0t^2c^2 - 4e^{-2ct}x_0t^2c^2). \end{aligned} \tag{4.5}$$

Depending on the values of c, t and x_0 the variance of the GT method may be more or less than that of CGT. However it is instructive to focus on the dependence on the initial state x_0. The variance of GT estimator contains x_0^3 while that of CGT estimator only has x_0^2 involved in the formula. If x_0 is modestly large (say 100), a significant amount of variance reduction can be expected using CGT. On the other hand, if we consider the variance of the FD estimator, we can bound the variance as

$$\frac{1}{h^2}\mathrm{Var}(X(t, c+h) - X(t, c)) \leq \frac{2}{h^2}\left(\mathrm{Var}(X(t, c+h)) + \mathrm{Var}(X(t, c))\right).$$

To derive a formula for the right hand side, we need the first and second moment of $X(t,c)$ and $X(t,c+h)$. Without loss of generality, we can only derive formula for the moments of $X(t,c)$. $\mathbb{E}X(t,c)$ can be derived by taking expectation for $X(t)$ in (4.1) and it turns out to be

$$\mathbb{E}X(t,c) = x_0 e^{-ct}.$$

Now for the second moment, we apply the Ito formula to obtain

$$X(t,c)^2 = x_0^2 + \int_0^t (1 - 2X(s-))dR(s).$$

After taking expectation and plugging in $\mathbb{E}X(t,c)$, we have

$$\frac{d}{dt}\mathbb{E}X(t,c)^2 = -2c\mathbb{E}X(t,c)^2 + cx_0 e^{-ct}$$

with initial condition $X(0)^2 = x_0^2$. Solving this differential equation, we have

$$\mathbb{E}X(t,c)^2 = (x_0 e^{ct} + x_0^2 - x_0)e^{-2ct}$$

and hence

$$\mathrm{Var}(X(t,c)) = x_0(e^{-ct} - e^{-2ct}).$$

Since $\mathrm{Var}(X(t,c))$ depends on x_0 linearly, we expect the variance of the FD estimator to be even smaller than that of the CGT.

These observations motivate the analysis in the rest of this paper. We note that x_0 can be thought of as a measure of *system size*. We will show that when the system size grows the variance of the FD methods grow the smallest in terms of system size, while the variance of the CGT method grows modestly and the variance of the GT method grows the fastest for more general chemical reaction systems.

4.2 Running Assumptions

To facilitate the analysis in this chapter, we shall make some explicit assumptions which shall hold throughout the rest of this thesis.

Assumption 1. *We assume the following form of parameter dependence on the intensity function. For each $j = 1, \cdots, m$ and N,*

$$a_j^N(x, c) = c_j b_j^N(x), \tag{4.6}$$

where $b_j^N : \mathbb{R}^n \to \mathbb{R}$ are such that b_j^N restricted to $\mathbb{Z}_+^N$ are nonnegative. This also implies that there are precisely m parameters, one for each reaction j.

For the analysis in this thesis we need not to assume the stochastic mass action form, but merely the density dependence we described in Chapter 2. We restated it as our Assumption 2 here.

Assumption 2. *We suppose that for each $j = 1, \cdots, m$, there exists a function (deterministic reaction rate) $\bar{a}_j(x)$ such that for each compact $K \subset \mathbb{R}_+^n$, the collection of functions $a_j^N(Nx) - N\bar{a}_j(x)$ is uniformly bounded for $x \in K$ and $N \geq 1$.*

We note that this implies that for each compact set $K \subset \mathbb{R}_+^n$ there exists $B_K > 0$ such that

$$\left| \frac{a_j^N(Nx)}{N} - a_j(x) \right| \leq \frac{B_K}{N}, \quad N \geq 1,\ x \in K,\ j = 1, \ldots, m. \tag{4.7}$$

Here the constant B_K depends implicitly on c.

In order to satisfy the conditions stated in Theorem 2.3.1 we shall assume the following.

Assumption 3. *For each $j = 1, \ldots, m$, the functions $\bar{a}_j : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable. This automatically implies the Lipschitz condition in Theorem 2.3.1.*

The following assumption is used to facilitate the analysis in this thesis. Several, but not all examples in applications satisfy this assumption.

Assumption 4. *We take $X_N(0) = x_0$ (deterministic) for all N. We assume that the sequence of concentration processes X_N is uniformly bounded, that is, there exists a constant Γ such that for all $t \geq 0$,*

$$|X_N(t)| \leq \Gamma \quad a.s. \tag{4.8}$$

for all $N \geq 1$.

We note that if there exists a strictly positive vector $\alpha \in \mathbb{R}^m_+$ so that $\alpha^T \nu_j \leq 0$ for each j then this assumption is satisfied. We note that a form of converse of this statement is also true [27].

Now we turn our attention to the sensitivity. Given $f : \mathbb{R}^n \to \mathbb{R}$, we are interested in computing the sensitivity

$$\frac{\partial}{\partial c}\mathbb{E}(f(X^N(t,c))),$$

where $c \in (0, \infty)$ is a parameter. In view of Assumption 1, without loss of generality, we shall take $c = c_1$. Then we note that the GT sensitivity estimator is $f(X^N(t))Z^N(t)$ and the CGT estimator is $[f(X^N(t)) - \mathbb{E}(f(X^N(t))]Z^N(t)$, where we note that $Z^N(t) = M_1^N(t)/c_1$ in this case.

As we are concerned with families of processes indexed by N, it makes sense

to consider a corresponding family of functions $f^N : \mathbb{R}^n \to \mathbb{R}$ instead of one function f and make reasonable assumptions on f^N and f.

To motivate the assumption we make on f^N and f we note that we shall be concerned with $f^N(X^N(t)) = f^N(NX_N(t))$ which we wish to compare with $f(X(t))$. When $f^N(x) = x_i$, one of the components of x, we have

$$f^N(NX_N(t))/N = X_{Ni}(t) \to X_i(t) = f(X(t)),$$

with $f(x) = x_i$. Alternatively, if $f^N(x) = x_i^\alpha$ for some $\alpha > 0$ we have

$$f^N(NX_N(t))/N^\alpha = (X_{Ni}(t))^\alpha \to (X_i(t))^\alpha = f(X(t)),$$

with $f(x) = x_i^\alpha$. If however $f^N(x) = x_i^2 + x_i$ then we have

$$f^N(NX_N(t))/N^2 = (X_{Ni}(t))^2 + X_{Ni}(t)/N \to (X_i(t))^2 = f(X(t)),$$

where $f(x) = x_i^2$. In this case we note that $f^N(Nx)/N^2 - f(x) = x_i/N$ which tends to 0 as $1/N$, uniformly for x in a compact set. Motivated by this we impose the following assumption.

Assumption 5. *We assume that there exist a function f and a constant $\alpha > 0$ such that for each compact set $K \subset \mathbb{R}^n_+$,*

$$\left|f^N(Nx)/N^\alpha - f(x)\right| \le \frac{L_K}{\sqrt{N}}, \quad x \in K, \;\; N \ge 1 \tag{4.9}$$

for some constant $L_K > 0$.

We remark that the $O(1/\sqrt{N})$ behavior is adequate for our proofs.

4.3 Strong and Weak Limit for the Weight Process Z^N

In this section we derive results concerning the $N \to \infty$ limit for the various relevant processes. Throughout the rest of the paper $X(t)$ will denote the solution to the equation

$$X(t) = x_0 + \sum_{j=1}^{m} \nu_j \int_0^t a_j(X(s))ds, \tag{4.10}$$

where $x_0 \in \mathbb{R}^n_+$ is fixed (given the solution exists).

Lemma 4.3.1. *We have for each $j = 1, \cdots, m$, there exists $A_j > 0$ such that for all $t > 0$*

$$\frac{a_j^N(NX_N(t))}{N} \leq A_j \quad a.s..$$

Proof. By (4.8) in Assumption 4, the processes X_N are contained in a compact set of $\mathbb{R}^n$, say K, therefore for each j we have the estimation

$$\sup_{t\geq 0} \frac{a_j^N(NX_N(t))}{N} \leq \sup_{x\in K} \frac{a_j^N(Nx)}{N}$$

Since $N^{-1}a_j^N(Nx)$ converges uniformly to $a_j(x)$ for x in K by (4.7), it is apparent that $\sup_{x\in K} N^{-1}a_j^N(Nx)$ is bounded by continuity of a_j. Hence $\sup_{t\geq 0} N^{-1}a_j^N(NX_N(t))$ is bounded by a constant A_j. □

Lemma 4.3.2. *For $j = 1, \cdots, m$, and $t > 0$, we have*

$$\sup_{s\leq t} \left| \frac{a_j^N(NX_N(s))}{N} - a_j(X(s)) \right| \to 0, \quad a.s.$$

Proof. Write

$$\begin{aligned} &\left| \frac{a_j^N(NX_N(s))}{N} - a_j(X(s)) \right| \\ \leq &\left| \frac{a_j^N(NX_N(s))}{N} - a_j(X_N(s)) \right| + \left| a_j(X_N(s)) - a_j(X(s)) \right|. \end{aligned}$$

The first part on the right hand side converges to zero uniformly for s in $[0,t]$ because of Assumptions 2 and 4. To see the second part on the right hand side converges uniformly to 0 on $[0,t]$, note that by Assumption 3 and Assumption 4, a_j is Lipschitz continuous on the compact set K (which contains X_N and X), hence the result follows by Theorem 2.3.1. □

We define the sequence of scaled processes $R_N(t)$ by $R_N(t) = R^N(t)/N$.

Lemma 4.3.3. *For each $j = 1, 2, \cdots, m$ and $t > 0$*

$$\sup_{s\le t}\left|R_{Nj}(s) - \int_0^s a_j(X(u))du\right| \to 0 \quad a.s.$$

Proof. Recall that $R_j^N(t) = Y_j\left(\int_0^t a_j^N(NX_N(s))ds\right)$. For each $j = 1, \cdots, m$,

$$\begin{aligned}
&\sup_{s\le t}\left|\frac{1}{N}Y_j\left(\int_0^s a_j^N(NX_N(u))du\right) - \int_0^s a_j(X(u))du\right| \\
\le&\sup_{s\le t}\left|\frac{1}{N}Y_j\left(\int_0^s a_j^N(NX_N(u))du\right) - \frac{1}{N}\int_0^s a_j^N(NX_N(u))du\right| \\
&+ \int_0^t \left|\frac{1}{N}a_j^N(NX_N(u)) - a_j(X(u))\right| du.
\end{aligned}$$

The second term on the right hand side converges to zero by Lemma 4.3.2. Setting $\tilde{Y}(t) = Y(t) - t$, the first term on the right can be written and then bounded as

$$\sup_{s\le t}\left|\frac{1}{N}\tilde{Y}_j\left(\int_0^s a_j^N(NX_N(u))du\right)\right| \le \sup_{s\le t}\left|\frac{1}{N}\tilde{Y}_j\left(NA_j s\right)\right| \quad \text{a.s.}$$

where the last term converges to zero by the law of large numbers for Poisson process. □

Lemma 4.3.4. *For a given $t > 0$, suppose that f is continuous at $X(t)$. Then*

$$\lim_{N\to\infty} |f^N(NX_N(t))/N^\alpha - f(X(t))| = 0, \quad a.s. \tag{4.11}$$

Proof. Write

$$\left|f^N(NX_N(t))/N^\alpha - f(X(t))\right| \leq \left|f^N(NX_N(t))/N^\alpha - f(X_N(t))\right| + \left|f(X_N(t)) - f(X(t))\right|.$$

The first term converges to zero almost surely by Assumption 4 and (4.9) in Assumption 5. The second term converges to zero by the continuity assumption on f since $X_N(t)$ converges to $X(t)$ almost surely. □

Recall the definition of M^N,

$$M^N(t) = R^N(t) - \int_0^t a^N(NX_N(s))ds.$$

Note that in general $M^N(t)$ is an m-dimensional local martingale for each N, but by Lemma 4.3.1 it follows that $\mathbb{E}[R_j^N(t)] \leq NA_jt$ for all $t > 0$ which makes $M^N(t)$ a martingale. We define the scaled processes $M_N = N^{-1}M^N$ and $Z_N = N^{-1}Z^N$. We note that $Z^N(t) = M_1^N(t)/c_1$ and $Z_N(t) = M_{N1}(t)/c_1$.

Let us denote by $D^m[0,\infty)$ the space of càdlàg functions mapping from $[0,\infty)$ to $\mathbb{R}^m$ endowed with the Skorohod topology. We provide a lemma on the weak convergence of M_N.

Theorem 4.3.1. *Let $D(t) = (d_{ij}(t))$ be a $m \times m$ matrix of functions, where*

$$d_{ij}(t) = \begin{cases} \int_0^t \bar{a}_j(X(s))ds & i = j \\ 0 & i \neq j \end{cases} \tag{4.12}$$

Then $\sqrt{N}M_N \Rightarrow \bar{M}$ on $D^m[0,\infty)$, where $\bar{M}(t)$ is an m-dimensional Gaussian process with independent increments having mean vector and covariance matrix

$$\mathbb{E}[\bar{M}(t)] = (0,\cdots,0), \quad \mathbb{E}[\bar{M}(t)\bar{M}(t)^T] = D(t). \tag{4.13}$$

Furthermore, the scaled Girsanov sensitivity (or weight) process $\sqrt{N}Z_N \Rightarrow U$ *on* $D[0,\infty)$*, where*

$$U(t) = \frac{1}{c_1}\bar{M}_1(t). \tag{4.14}$$

Also since U *has continuous sample paths, for each* $t > 0$ *we have*

$$\sqrt{N}Z_N(t) \Rightarrow U(t).$$

Proof. The proof relies on Theorem 2.4.1 the martingale FCLT. Note that each jump of $\sqrt{N}M_N$ has size $1/\sqrt{N}$, therefore,

$$\lim_{N\to\infty} \mathbb{E}\left[\sup_{s\le t}\left|\sqrt{N}M_N(s) - \sqrt{N}M_N(s-)\right|\right] = 0.$$

Also, for each pair (i,j) with $i,j = 1,\cdots,m$, and each t, since the jump size for M_{Nj} is always N^{-1} and there are no simultaneous jumps, we have the following quadratic covariation

$$\left[\sqrt{N}M_{Ni}, \sqrt{N}M_{Nj}\right](t) = \begin{cases} R_{Nj} & i = j \\ 0 & i \neq j \end{cases} \tag{4.15}$$

By Lemma 4.3.3, $R_{Nj}(t)$ converges almost surely to $d_{jj}(t) = \int_0^t a_j(X(s))ds$. Then, for each pair (i,j),

$$\left[\sqrt{N}M_{Ni}, \sqrt{N}M_{Nj}\right](t) \to d_{ij}(t)$$

almost surely and hence in probability. Thus, the weak convergence of M_N follows from the martingale FCLT. □

Lemma 4.3.5. *For each* $p \ge 1$*, there exists a constant* $\beta(p)$ *such that for all* $t > 0$

$$\limsup_N \mathbb{E}\left(\sup_{s\le t}\left|\sqrt{N}M_N(s)\right|\right)^p \le \beta(p)t^{p/2}. \tag{4.16}$$

Proof. Observe that the quadratic variation of $\sqrt{N}M_N$ is

$$\left[\sqrt{N}M_N, \sqrt{N}M_N\right](t) = N^{-1}\sum_{j=1}^{m} Y_j\left(\int_0^t a_j^N(NX_N(s))ds\right).$$

By the Burkholder-Davis-Gundy inequality (see [23]), there exists a constant $C(p)$ such that

$$\begin{aligned}\mathbb{E}\left(\sup_{s\le t}\left|\sqrt{N}M_N(s)\right|\right)^p \le& C(p)\mathbb{E}\left(\frac{1}{N}\sum_{j=1}^{m} Y_j\left(\int_0^t a_j^N(NX_N(s))ds\right)\right)^{p/2}\\ \le& C(p)\mathbb{E}\left(\frac{1}{N}\sum_{j=1}^{m} Y_j\left(NA_jt\right)\right)^{p/2}\\ \le& C(p)N^{-p/2}\left(\mathbb{E}\left(\sum_{j=1}^{m} Y_j(NA_jt)\right)^p\right)^{1/2},\end{aligned}$$

where we have used Lemma 4.3.1.

Hence,

$$\limsup_N \mathbb{E}\left(\sup_{s\le t}\left|\sqrt{N}M_N(s)\right|\right)^p \le \limsup_N C(p)N^{-p/2}\left(\mathbb{E}\left(\sum_{j=1}^{m} Y_j(NA_jt)\right)^p\right)^{1/2}.$$

First we observe that for $j = 1, \cdots, m$, the pth moment of the Poisson random variable $Y_j(NA_jt)$ is a polynomial of degree p in NA_jt. Also, noting that Y_j are independent, we obtain that the right hand side is bounded by a term $\beta(p)t^{p/2}$, where $\beta(p)$ is a constant.

□

Since $Z^N(t) = {c_1}^{-1}M_1^N(t)$, we immediately have the following property regarding process Z_N.

Lemma 4.3.6. *For each $p \ge 1$, there exists a constant $\gamma(p)$ such that for all $t > 0$,*

$$\limsup_N \mathbb{E}\left(\sup_{s\le t}\sqrt{N}\left|Z_N(s)\right|\right)^p \le \gamma(p)t^{p/2}. \tag{4.17}$$

Define the process $V_N(t) = \sqrt{N}(X_N(t) - X(t))$. Let us consider the moment of this process on a compact time interval.

Lemma 4.3.7. *For each $p \geq 1$, there exist constants $\bar{\beta}(p), K(p)$ such that for all $t > 0$*

$$\limsup_N \sup_{s \leq t} \mathbb{E}\left(|V_N(s)|^p\right) \leq \bar{\beta}(p) t^{p/2} e^{K(p)t^p}.$$

Proof. Recall that

$$X_N(s) = x_0 + \nu R_N(s)$$

and

$$X(s) = x_0 + \int_0^s \nu a(X(u))du,$$

where ν is the n by m dimensional stoichiometric matrix. One can write V_N as follows,

$$\begin{aligned} V_N(s) =& \sqrt{N}\nu R_N(s) - \sqrt{N}\int_0^s \nu a(X(u))du \\ =& \sqrt{N}\nu\left(R_N(s) - \int_0^s \frac{a^N(NX_N(u))}{N}du\right) \\ &+ \sqrt{N}\nu\left(\int_0^s \frac{a^N(NX_N(u))}{N} - a(X(u))du\right). \end{aligned}$$

Note that we denote $M_N(s) = R_N(s) - \int_0^s N^{-1}a^N(NX_N(u))du$, hence

$$|V_N(s)| \leq \|\nu\|\left|\sqrt{N}M_N(s)\right| + \|\nu\|\int_0^s \sqrt{N}\left|\frac{a^N(NX_N(u))}{N} - a(X(u))\right|du.$$

We note that

$$\begin{aligned} \sqrt{N}\left|\frac{a^N(NX_N(u))}{N} - a(X(u))\right| \leq& \sqrt{N}\left|\frac{a^N(NX_N(u))}{N} - a(X_N(u))\right| \\ &+ \sqrt{N}\left|a(X_N(u)) - a(X(u))\right|. \end{aligned}$$

Since X_N lies in a compact set K according to Assumption 4, we have for all $u > 0$,

$$\left| \frac{a^N(NX_N(u))}{N} - a(X_N(u)) \right| \leq \frac{B_K}{N}$$

where we have used Assumption 2 and abuse of notation using B_K from (4.7).

On the other hand, for each $j = 1, \cdots, m$, by Assumption 3, a_j is continuously differentiable and hence it is Lipschitz continuous on the compact set K. Hence, there exists a Lipschitz constant C_j such that for all $u > 0$,

$$|a_j(X_N(u)) - a_j(X(u))| \leq C_j \left| X_N(u) - X(u) \right|.$$

It follows that there exists a constant C such that

$$|a(X_N(u)) - a(X(u))| \leq C \left| X_N(u) - X(u) \right|.$$

Therefore,

$$\begin{aligned} |V_N(s)| \leq & \|\nu\| \left(\left| \sqrt{N} M_N(s) \right| + N^{-1/2} B_K s + C \int_0^s \sqrt{N} \left| X_N(u) - X(u) \right| du \right) \\ = & \|\nu\| \left(\left| \sqrt{N} M_N(s) \right| + N^{-1/2} B_K s + C \int_0^s \left| V_N(u) \right| du \right), \end{aligned}$$

Applying the inequality $(a+b+c)^p \leq 3^p(a^p+b^p+c^p)$ and the Holder's inequality, we obtain

$$|V_N(s)|^p \leq (3\|\nu\|)^p \left(\left| \sqrt{N} M_N(s) \right|^p + N^{-p/2} (B_K s)^p + C^p s^{p-1} \int_0^s |V_N(u)|^p du \right).$$

Taking expected value of both sides, for $s \in [0, t]$

$$\begin{aligned} \mathbb{E}|V_N(s)|^p \leq & (3\|\nu\|)^p \left(\mathbb{E} \left| \sqrt{N} M_N(s) \right|^p + N^{-p/2} (B_K t)^p \right) \\ & + (3\|\nu\|)^p C^p s^{p-1} \left(\int_0^s \mathbb{E}|V_N(u)|^p du \right). \end{aligned}$$

To estimate the first term of the right hand side, recall that in the proof of Lemma 4.3.5,

$$\mathbb{E}\left(\sup_{s\leq t}\left|\sqrt{N}M_N(s)\right|\right)^p \leq C(p)N^{-p/2}\left(\mathbb{E}\left(\sum_{j=1}^{m}Y_j(NA_jt)\right)^p\right)^{1/2}.$$

For convenience, let us denote

$$\Phi_N(t) = C(p)N^{-p/2}\left(\mathbb{E}\left(\sum_{j=1}^{m}Y_j(NA_jt)\right)^p\right)^{1/2}.$$

Therefore,

$$\mathbb{E}|V_N(s)|^p \leq (3\|\nu\|)^p\left(\Phi_N(t) + N^{-p/2}(B_Kt)^p + C^p s^{p-1}\left(\int_0^s \mathbb{E}|V_N(u)|^p du\right)\right).$$

We note that $\mathbb{E}|V_N(s)|^p$ is continuous in s and applying the Gronwall inequality, we obtain for $s \leq t$,

$$\mathbb{E}|V_N(s)|^p \leq (3\|\nu\|)^p\left(\Phi_N(t) + N^{-p/2}(B_Kt)^p\right)e^{(3\|\nu\|)^pC^ps^p}.$$

Taking supremum over $s \in [0,t]$ and then taking $\limsup_N$, the result follows from same considerations as in the proof of Lemma 4.3.5. □

4.4 Estimator Efficiency

Now we are in a good position to use the results we established in the last section to analyze the efficiency of various estimator. A measure of accuracy of a Monte Carlo estimator S is the ratio of its standard deviation to the absolute value of its expected value, i.e.,

$$\frac{\sqrt{\mathrm{Var}(S)}}{|\mathbb{E}(S)|}.$$

We refer to this quantity as the **relative standard error (RSE)**.

In this section, we study the system size dependence of the sensitivity

$$\frac{\partial}{\partial c}\mathbb{E}(f^N(X^N(t))),$$

and the variances and RSEs of the GT, CGT and FD estimators. In the context of stochastic mass action form of intensities given by (2.1), we note that $c = c_j$ is the deterministic parameter while $c'_N = c_j/N^{|\nu''_j|-1}$ is the stochastic parameter. In practice, one would compute sensitivity with respect to the stochastic parameter c'_N. The difference between the sensitivity with respect to the stochastic parameter and with respect to the deterministic parameter is merely a scaling factor $N^{|\nu''_j|-1}$. Therefore, the RSE is unchanged regardless of whether one considers the sensitivity with respect to the stochastic parameter or the deterministic parameter. From an analytical point of view, it is convenient to study the sensitivity with respect to the deterministic parameter.

Recall that the sensitivity estimator of Girsanov transformation method is

$$f^N(X^N(t,c))Z^N(t,c)$$

where $f^N : \mathbb{R}^n \to \mathbb{R}$.

Theorem 4.4.1. *(***Scaling of the sensitivity***) In addition to our running assumptions, we assume that f in* (4.9) *is continuously differentiable. Then for each $t \geq 0$*

$$\sup_{s\leq t}\mathbb{E}(f^N(X^N(s))Z^N(s)) = \mathcal{O}(N^\alpha).$$

That is, the sensitivity is asymptotically $\mathcal{O}(N^\alpha)$ uniformly on $[0,t]$.

Proof. It is sufficient to show that $\sup_{s\leq t}\mathbb{E}(f^N(X^N(s))Z^N(s))/N^\alpha$ is bounded in N. Instead of working with $\mathbb{E}(f^N(X^N(s))Z^N(s))/N^\alpha$, we use

$$\mathbb{E}\left(\frac{f^N(X^N(s))}{N^\alpha}Z^N(s) - f(X(s))Z^N(s)\right)$$

because they are equal but the latter is easier to work with.

Note that f is continuously differentiable hence Lipschitz on the compact set K corresponding to Assumption 4. Denote by C_K the Lipschitz constant for f. Using the assumptions on f^N and f and writing X^N in terms of V_N as before as

$$X^N(s) = NX(s) + \sqrt{N}V_N(s),$$

leads to

$$\begin{aligned}
&\left|\frac{f^N(NX(s)+\sqrt{N}V_N(s))}{N^\alpha} - f(X(s))\right||Z^N(s)| \\
\leq&\left|\frac{f^N(NX(s)+\sqrt{N}V_N(s))}{N^\alpha} - f\left(X(s)+\frac{V_N(s)}{\sqrt{N}})\right)\right||Z^N(s)| \\
&+\left|f\left(X(s)+\frac{V_N(s)}{\sqrt{N}}\right) - f(X(s))\right||Z^N(s)| \\
\leq&\frac{L_K}{\sqrt{N}}|Z^N(s)| + C_K|V_N(s)|\frac{|Z^N(s)|}{\sqrt{N}} \\
\leq&L_K\sqrt{N}|Z_N(s)| + \frac{1}{2}C_K\left(|V_N(s)|^2 + N|Z_N(s)|^2\right).
\end{aligned}$$

The result follows from Lemmas 4.3.6 and 4.3.7. □

Remark: We believe that under the N^α scaling, the sensitivity of the stochastic process should limit to the sensitivity of the fluid limit ODE as $N\to\infty$. The above result is weaker than that.

Theorem 4.4.2. *(***Second moment of GT estimator***) In addition to our running assumptions, we assume that f in* (4.9) *is bounded on every compact set and for a*

given $t > 0$, f is continuous at $X(t)$. Then we have as $N \to \infty$,

$$N^{-1-2\alpha}\mathbb{E}\left\{(f^N(X^N(t)))^2(Z^N(t))^2\right\} \to (f(X(t)))^2\frac{1}{c_1}\int_0^t a_1(X(s))ds. \tag{4.18}$$

Furthermore, for each $t > 0$,

$$\sup_{s\le t}\mathbb{E}\left((f^N(X^N(s)))Z^N(s)\right)^2 = \mathcal{O}(N^{2\alpha+1}).$$

Proof. Lemma 4.3.6 implies the uniformly integrability of $N^{-1}(Z^N(t))^2$. By Assumption 4 and (4.9) we have that $(f^N(X^N(t)))^2/N^{2\alpha}$ is a uniformly bounded sequence. Thus $N^{-1-2\alpha}(f^N(X^N(t)))^2(Z^N(t))^2$ is uniformly integrable.

By Lemma 4.3.4 we have that $N^{-2\alpha}(f^N(X^N(t)))^2$ converges to $(f(X(t)))^2$ almost surely. We also have that $N^{-1}Z^N(t)$ converge weakly to $U(t)$. Thus by Theorem A.1.1 in appendix and the continuous mapping theorem we have that

$$N^{-1-2\alpha}(f^N(X^N(t)))^2(Z^N(t))^2 \Rightarrow (f(X(t)))^2U^2(t).$$

By Theorem 3.5 from [7], we note that if a uniformly integrable sequence converges weakly then it converges in the mean, hence the result (4.18) follows.

Also, recall that $(f^N(X^N(t)))^2/N^{2\alpha}$ is uniformly bounded, hence

$$N^{-2\alpha-1}\sup_{s\le t}\mathbb{E}\left((f^N(X^N(s)))Z^N(s)\right)^2 \le \tilde{C}\mathbb{E}(\sup_{s\le t}\sqrt{N}|Z_N(s)|)^2.$$

Taking $\limsup_N$ and applying Lemma 4.3.6 yields the seconds result. □

Note that the above theorem does not assume f is continuously differentiable. However, to state the result regarding the estimator variance for GT method, we still need to assume continuous differentiability on f so that we can use Theorem 4.4.1.

Corollary 4.4.1. *(***Variance of GT estimator***) Suppose f in (4.9) is continuous differentiable, then for given $t > 0$, the estimator variance of GT method is asymptotically $\mathcal{O}(N^{2\alpha+1})$ uniformly on $[0, t]$.*

Next, we will explore the variance of the centered Girsanov transformation approach.

Theorem 4.4.3. *(***Second Moment of CGT estimator***) In addition to our running assumptions, we assume that f in (4.9) is continuously differentiable. Then for each $t > 0$,*

$$\sup_{s\leq t} \mathbb{E}\left((f^N(X^N(s)) - \mathbb{E}f^N(X^N(s)))Z^N(s)\right)^2 = \mathcal{O}(N^{2\alpha}).$$

Proof. Write

$$\begin{aligned}
&\mathbb{E}\left(\left|\frac{f^N(X^N(s))}{N^\alpha} - \mathbb{E}\left(\frac{f^N(X^N(s))}{N^\alpha}\right)\right|^2 (Z^N(s))^2\right)\\
\leq& 2\mathbb{E}\left(\left|\frac{f^N(X^N(s))}{N^\alpha} - f(X(s))\right|^2 (Z^N(s))^2\right)\\
&+ 2\mathbb{E}\left(\left|f(X(s)) - \mathbb{E}\left(\frac{f^N(X^N(s))}{N^\alpha}\right)\right|^2 (Z^N(s))^2\right)\\
\leq& 2\mathbb{E}\left(\left|\frac{f^N(X^N(s))}{N^\alpha} - f(X(s))\right|^2 (Z^N(s))^2\right)\\
&+ 2\mathbb{E}\left(\left|\frac{f^N(X^N(s))}{N^\alpha} - f(X(s))\right|^2\right)\mathbb{E}(Z^N(s))^2,
\end{aligned}$$

where the last inequality is true due to the fact that $f(X(t))$ is deterministic. Using similar argument as in the proof of Theorem 4.4.1, the first term on the right-hand side can be bounded by

$$4L_K^2\mathbb{E}\left(|\sqrt{N}Z_N(s)|\right)^2 + 4C_K^2\mathbb{E}\left(|V_N(s)|\sqrt{N}|Z_N(s)|\right)^2.$$

Similarly, the second term on the right-hand side can be bounded by

$$4L_K^2\mathbb{E}\left(\sqrt{N}|Z_N(s)|\right)^2 + 4C_K^2\mathbb{E}|V_N(s)|^2\mathbb{E}\left(\sqrt{N}|Z_N(s)|\right)^2.$$

Both of the above terms are bounded in N uniformly on $[0,t]$ by Lemma 4.3.6 and 4.3.7. □

Combining this result with Theorem 4.4.1, the following corollary is immediate.

Corollary 4.4.2. *(***Variance of CGT estimator***) For any given $t > 0$, the estimator variance of CGT method is asymptotically $\mathcal{O}(N^{2\alpha})$ uniformly on $[0,t]$.*

Theorem 4.4.4. *(***Variance of FD estimator***) Suppose f^N satisfies* (4.9) *and f is continuously differentiable. Then for each $t > 0$ and $h > 0$,*

$$\sup_{s\le t} Var\left(X^N(s,c+h) - X^N(s,c)\right) = \mathcal{O}(N^{2\alpha-1}).$$

That is, the estimator variance of FD method is asymptotically $\mathcal{O}(N^{2\alpha-1})$.

Proof. Note that

$$\mathrm{Var}\left(X^N(s,c+h) - X^N(s,c)\right) \le 2\mathrm{Var}\left(X^N(s,c+h)\right) + 2\mathrm{Var}\left(X^N(s,c)\right),$$

hence it is sufficient to show that $\mathrm{Var}\left(X^N(t,c)\right) = \mathcal{O}(N^{2\alpha-1})$. We write

$$\frac{1}{N^{2\alpha-1}}\mathrm{Var}\left(X^N(s,c)\right) = N\mathbb{E}\left(\left|\frac{f^N(X^N(s,c))}{N^\alpha} - \mathbb{E}\left(\frac{f^N(X^N(t,c))}{N^\alpha}\right)\right|^2\right).$$

One can estimate the right hand side by using the same argument as is in Theorem 4.4.3 to get an upper bound $8L_K^2 + 8C_K^2\mathbb{E}\left(|V_N(s)|\right)^2$, which is bounded in N uniformly on $[0,t]$ by Lemma 4.3.7. □

Remark: Based on Theorem 4.4.1, Corollary 4.4.1, Corollary 4.4.2 and Theorem 4.4.4, we may expect the RSEs of the GT, CGT and FD methods to scale as $\mathcal{O}(N^{1/2})$, $\mathcal{O}(1)$ and $\mathcal{O}(N^{-1/2})$, respectively. Since in Theorem 4.4.1, we do not have an exact limit for the sensitivity itself, this conclusion is not rigorously proven. Nevertheless, our numerical results in the next section support this expectation.

4.5 Numerical Examples

We illustrate the dependence of RSE of various sensitivity estimators on the system size N via numerical examples. When comparing the GT or CGT methods with FD or RPD methods, we must bear in mind that while GT and CGT do not have method parameters, the FD method has a perturbation parameter h and the RPD method has a window size parameter w, making the comparison not straightforward. A proper practical comparison involves choosing parameters h and w to obtain an acceptable bias. We do not pursue such a detailed comparison here as we are focused solely on the dependence on system size N.

We note that in the very large system size limit, the stochastic system behaves nearly deterministically and hence none of these stochastic sensitivity methods are needed; traditional ODE sensitivity methods would do. However, when the system size N is modestly large, say $N = 100$, the system may not be approximated by the ODE and our asymptotic analysis will be relevant in this regime. Our numerical results below show this.

4.5.1 Reversible Isomerization Network

The reversible isomerization model consists of two species S_1 and S_2 and involves the following two reactions:

$$S_1 \xrightarrow{c_1} S_2, \qquad S_2 \xrightarrow{c_2} S_1. \tag{4.19}$$

In the model with system size N, the intensity functions for processes R_1^N and R_2^N are

$$a_1^N(X^N(t), c) = c_1 X_1^N(t),$$

$$a_2^N(X^N(t), c) = c_2 X_2^N(t),$$

respectively. The stoichiometric vectors are $\nu_1 = [-1, 1]^T$ and $\nu_2 = [1, -1]^T$.

In this example, the expectation of the population of species at a fixed time t can be computed analytically:

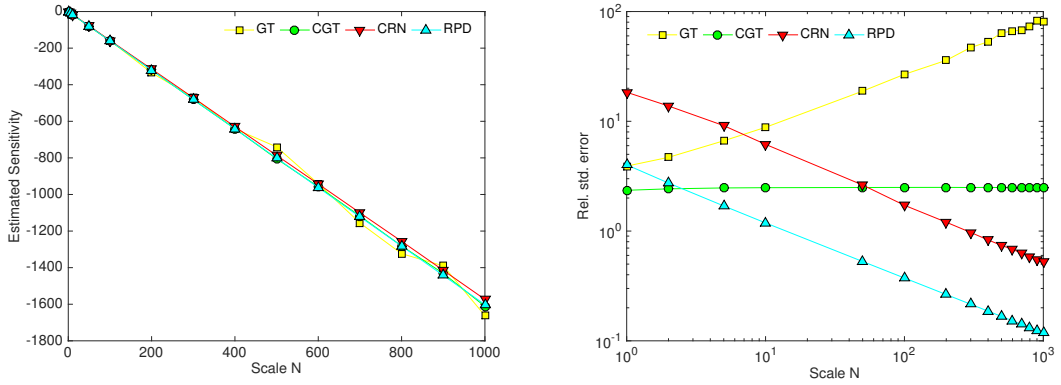
$$E[X_1^N(t)] = X_1^N(0) + \frac{1 - e^{-(c_1+c_2)t}}{c_1 + c_2}(c_2 X_2^N(0) - c_1 X_1^N(0)), \tag{4.20}$$

$$E[X_2^N(t)] = X_2^N(0) + \frac{1 - e^{-(c_1+c_2)t}}{c_1 + c_2}(c_2 X_2^N(0) - c_1 X_1^N(0)), \tag{4.21}$$

where $X_1^N(0)$ and $X_2^N(0)$ are assumed to be deterministic. One can compute the exact sensitivities by differentiating (4.20) and (4.21) with respect to parameters. In the numerical tests considered here, we choose parameters $c_1 = 0.3$ and $c_2 = 0.2$ and the initial population $X_1^N(0) = N$ and $X_2^N(0) = N$, where N is the system size parameter. We set the terminal time $T = 10$ and compute the sensitivity for various system size parameter $N = 1, 2, 5, 10, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000$. We use four different methods here, namely GT, CGT, CRN and RPD. We note

that by CRN we mean the common random number (one-sided) finite difference method in conjunction with Gillespie's SSA ([26]). The perturbation parameter for CRN method is $h = 0.01$ for parameter $c_1 = 0.3$ and the window size parameter $w = 1.0$ for terminal time $T = 10$. The number of trajectories for simulation is $\text{Ntr} = 10^6$ for each system size N.

The first output function we consider here is $f^N(x) = x_1$ for all N, that is, we compute the sensitivity of $\mathbb{E}(X_1^N(T))$ with respect to parameter c_1. Obviously, conditions in Assumption 5 are satisfied with $\alpha = 1$ and $f(x) = x_1$. We examine the growth of sensitivity of $\mathbb{E}(X_1^N(T))$ with respect to c_1 in terms of N using 10^6 independent trajectories. The computed sensitivity is shown in Fig 4.1(a) and Fig 4.1(b) shows the loglog plot of RSE of all four methods.

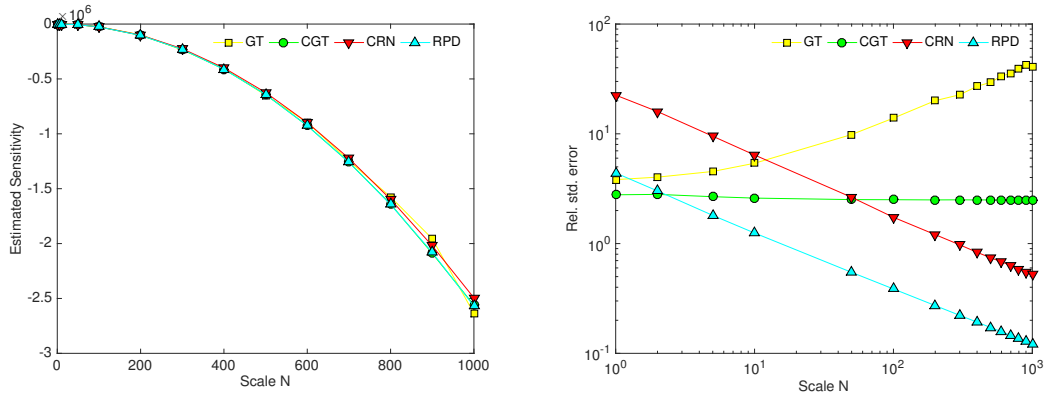


(a) Sensitivity

(b) RSE

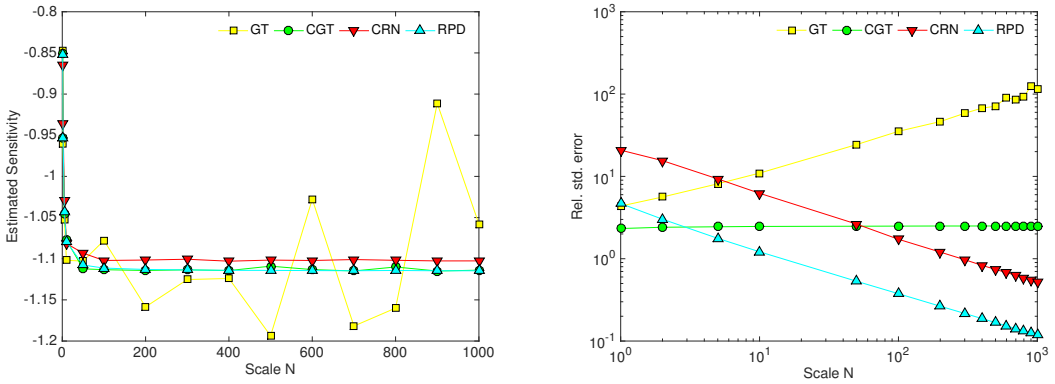
Figure 4.1: *Sensitivity of* $\mathbb{E}(X_1^N(T))$ *with respect to* c_1 *at final time* $T = 10$ *for reversible isomerization model.*

The second output function we use for testing is $f^N(x) = x_1^2$ for all N. By (4.9), $f(x) = x_1^2$ and $\alpha = 2$ in Assumption 5. See Figure 4.2 for sensitivity and



(a) Sensitivity (b) RSE

Figure 4.2: *Sensitivity of* $\mathbb{E}(X_1^N(T))^2$ *with respect to* c_1 *at final time* $T = 10$ *for reversible isomerization model.*



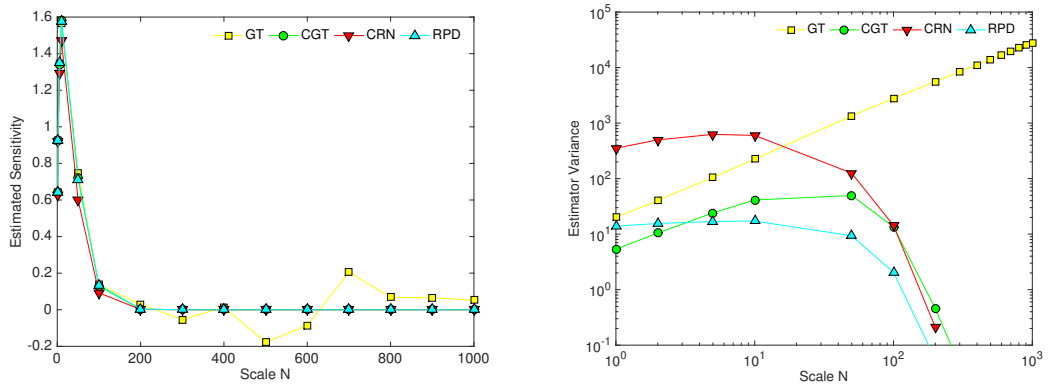
(a) Sensitivity (b) RSE

Figure 4.3: *Sensitivity of* $\mathbb{E}(\sin(X_1^N(T)/N))$ *with respect to* c_1 *at final time* $T = 10$ *for reversible isomerization model.*

RSE. The third output function is $f^N(x) = \sin(x_1/N)$ and so $f(x) = \sin x_1$. It can be seen that for this case, $\alpha = 0$ in Assumption 5. Plot for the numerical result in shown in Figure 4.3.

The last output function we consider here is the indicator function $f^N(x) =$

$1_{\{x_1 \le x_2\}}$, which does not satisfy the conditions in our theorems since $f = 1_{\{x_1 \le x_2\}}$ is not continuously differentiable. However, numerical tests still show similar behavior as indicated by our theorems. Note that the sensitivity approaches zero as N increases to ∞ and hence RSE is not well defined for large N. Instead, we plot the estimator variance against N in Figure 4.4(b).



(a) Sensitivity (b) Variance

Figure 4.4: *Sensitivity of* $\mathbb{P}(X_1^N(T) \le X_2^N(T))$ *with respect to* c_1 *at final time* $T = 10$ *for reversible isomerization model.*

Finally, the Table 4.1 summarizes the rate of growth (as a power of N) of the numerically estimated RSE for the different estimators considered above. The numerical results are in agreement with the theory.

4.5.2 Decaying-Dimerizing Network

As a second numerical example, let us consider the decaying-dimerizing model described in (4.22).

$$S_1 \xrightarrow{c_1} \emptyset, \qquad 2S_1 \xrightarrow{c_2} S_2, \qquad S_2 \xrightarrow{c_3} 2S_1, \qquad S_2 \xrightarrow{c_4} S_3, \tag{4.22}$$

Table 4.1: *Observed slopes (via regression for large N) for the loglog plots of RSE for reversible isomerization model, that is, R_1, R_2 and R_3 are the observed asymptotic order of the estimator RSE (as a power of N) for $\mathbb{E}(X_1^N(T))$, $\mathbb{E}(X_1^N(T))^2$ and $\mathbb{E}(\sin(X_1^N(T)/N))$, respectively. R is the theoretical slope.*

	R	R_1	R_2	R_3
GT	0.5	0.4992	0.4895	0.5724
CGT	0	-0.0004	-0.0008	0.0009
CRN	-0.5	-0.5156	-0.5160	-0.5162
RPD	-0.5	-0.5005	-0.5000	-0.5000

The stoichiometric vectors are $\nu_1 = [-1, 0, 0]^T$, $\nu_2 = [-2, 1, 0]^T$, $\nu_3 = [2, -1, 0]^T$ and $\nu_4 = [0, -1, 1]^T$. We set the initial population to be $X_1^N(0) = 10N, X_2^N(0) = 0, X_3^N(0) = 0$. Using the stochastic mass action form (2.1), the intensity for processes R_1^N, R_2^N, R_3^N and R_4^N are

$$a_1^N(X^N(t), c) = c_1 X_1^N(t),$$

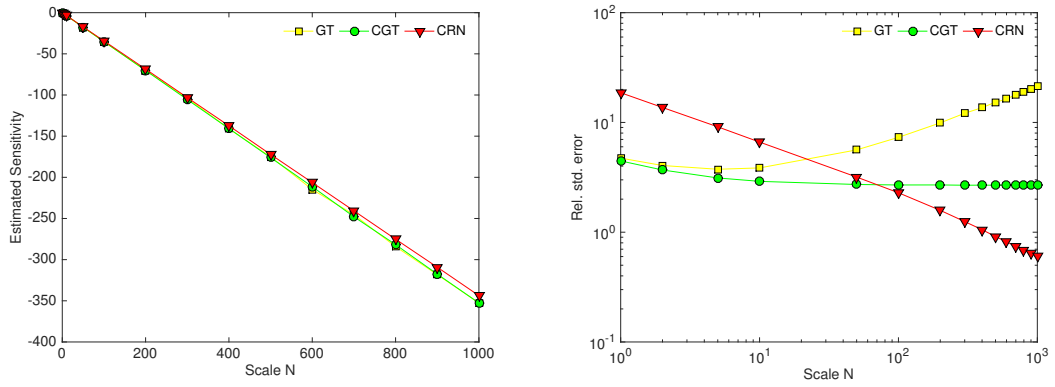
$$a_2^N(X^N(t), c) = \frac{c_2}{2N} X_1^N(t)(X_1^N(t) - 1),$$

$$a_3^N(X^N(t), c) = c_3 X_2^N(t),$$

$$a_4^N(X^N(t), c) = c_4 X_2^N(t).$$

We set the parameters as follows, $c_1 = 1.0$, $c_2 = 0.002$, $c_3 = 0.5$ and $c_4 = 0.04$. Note that the intensity for the second reaction is not linear, hence an analytical formula for the sensitivity is not attainable. We examine the sensitivity and RSE

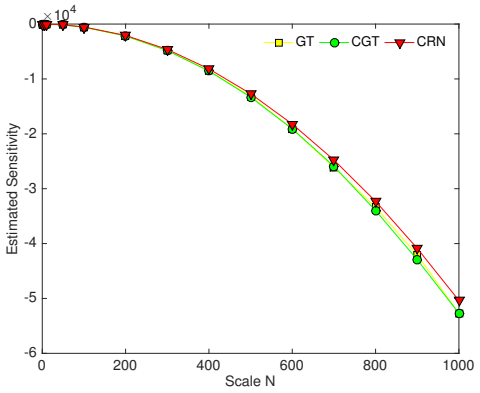
for $\mathbb{E}[f^N(X_1^N)]$ with respect to c_1. For the CRN method, we use one-sided finite difference scheme and perturb the parameter c_1 by $h = 0.01$. Note that RPD is not applicable for this example since the firing of the first reaction will prevent the second reaction to happen when the population of S_1 is 1 (see [29]), therefore we only examine the efficiency of GT, CGT and CRN here. For each system size N, the number of trajectories we use for simulation is Ntr $= 10^6$. Plots of the sensitivity and RSE are shown in Figure 4.5, 4.6 and 4.7 for $\mathbb{E}(X_1^N(T))$, $\mathbb{E}(X_1^N(T))^2$ and $\mathbb{E}(\sin(X_1^N(T)/N))$, respectively. The rate of growth (as a power of N) of the numerically estimated RSE are summarized in Table 4.2.



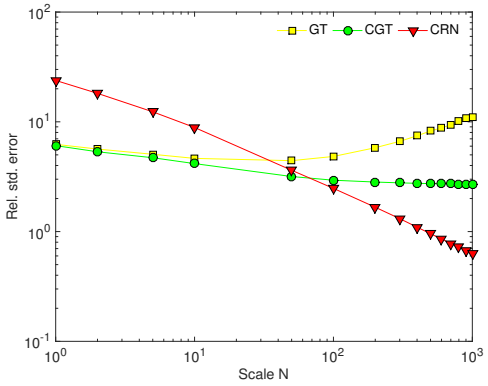
(a) Sensitivity

(b) RSE

Figure 4.5: *Sensitivity of* $\mathbb{E}[X_1^N(T)]$ *with respect to* c_1 *at final time* $T = 5$ *for decaying-dimerizing model.*

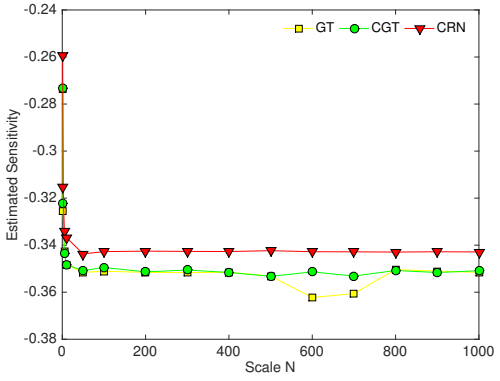


(a) Sensitivity

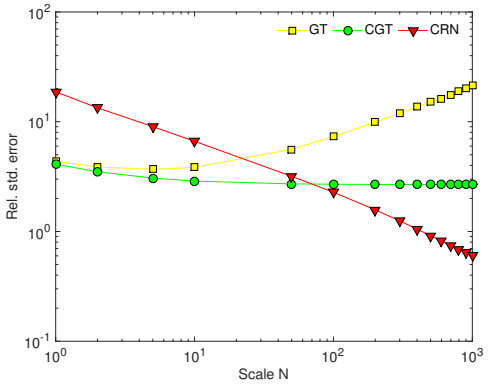


(b) RSE

Figure 4.6: *Sensitivity of* $\mathbb{E}(X_1^N(T))^2$ *with respect to* c_1 *at final time* $T = 5$ *for decaying-dimerizing model.*



(a) Sensitivity



(b) RSE

Figure 4.7: *Sensitivity of* $\mathbb{E}(\sin(X_1^N(T)/N))$ *with respect to* c_1 *at final time* $T = 5$ *for decaying-dimerizing model.*

Table 4.2: *Observed slopes (via regression) for the loglog plots for RSE for decaying-dimerizing model, that is, R_1, R_2 and R_3 are the observed asymptotic order of the estimator RSE (as a power of N) for $\mathbb{E}(X_1^N(T))$, $\mathbb{E}(X_1^N(T))^2$ and $\mathbb{E}(\sin(X_1^N(T)/N))$, respectively. R is the theoretical slope.*

	R	R_1	R_2	R_2
GT	0.5	0.4689	0.4100	0.4737
CGT	0	-0.0040	-0.0257	-0.0008
CRN	-0.5	-0.6022	-0.6068	-0.6009

Chapter 5: Validity of Girsanov Transformation Method

In this chapter, we provide sufficient conditions for the validity of applying GT method for sensitivity analysis. We reviewed the GT method in Chapter 3, where we assumed that it was valid to apply Girsanov's change of measure theory and take the derivative inside the integral. However, these two steps need to be justified carefully in order to guarantee that the GT method is valid. In Section 5.1, we present some important results regarding the change of measure from [8]. In Section 5.2, we provide a Novikov type condition under which the Radon-Nikodym derivative becomes a martingale instead of a local martingale. Finally, we give a sufficient condition for differentiating inside the integral. Throughout this chapter, we use an aabstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the process $X(t, c)$ maps from this space to the Skorohod space $(D^n[0, T], \mathcal{D}, P)$.

5.1 Change of Measure

The first step we shall verify for GT method is changing the measure from $\mathbb{P}$ to some other measure $\mathbb{P}(c)$, where $\mathbb{P}(c)$ is suppose to be the measure associated with parameter c after we apply Girsanov transformation. Let us construct $\mathbb{P}(c)$ from $\mathbb{P}$.

Suppose we are interested in the sensitivity at $c = c_0$, i.e.,

$$\left.\frac{\partial}{\partial c}\right|_{c=c_0} \mathbb{E}f(X(t,c)).$$

For this fixed $c_0 > 0$, we call it the **reference parameter** and denote a neighborhood of it by I_{c_0}, where

$$I_{c_0} = (c_0 - \epsilon, c_0 + \epsilon)$$

and $c_0 > \epsilon$ (so that the lower bound for I_{c_0} is positive). Recall that by definition $R(t, c_0)$ has the $(\mathbb{P}, \mathcal{F}_t)$ predictable intensity $a_j(X(t-, c_0), c_0)$. Our goal in this section is to use the **change of intensity** theorem from [8] (VI. $T3$ theorem) to construct a measure $\mathbb{P}(c)$ for any $c \in I_{c_0}$ and conclude that $R(t, c_0)$ has the $(\mathbb{P}(c), \mathcal{F}_t)$ predictable intensity $a_j(X(t-, c_0), c)$.

For this purpose, first we define an auxiliary (multidimensional) process μ as follows. Given $c \in I_{c_0}$, suppose for all $x \in \mathbb{Z}_+^n$,

$$a_j(x, c_0) = 0 \text{ if and only if } a_j(x, c) = 0 \tag{5.1}$$

for all $j = 1, 2, \cdots, m$. Note that this is true when the intensities are of the form $a_j(x, c) = c_j b_j(x)$ which holds in the stochastic mass action form. Now for each $j = 1, \cdots, m$, based on the above assumptions, the following process is well-defined (c_0 is fixed), for $c \in I_{c_0}$

$$\mu_j(t, c) = \frac{a_j(X(t-, c_0), c)}{a_j(X(t-, c_0), c_0)}. \tag{5.2}$$

In the case that $a_j(X(t-, c_0), c_0) = 0$, by assumption we have $a_j(X(t-, c_0), c) = 0$ as well, so we can simply define $\mu_j(t, c)$ to be any positive constant. Thus for each $c \in I_{c_0}$, $\mu_j(t, c)$ is $\mathcal{F}_t$-predictable by left continuity and for each $t > 0$, we have $0 < \mu_j(t, c) < \infty$ almost surely.

Next, we define the stochastic process $L(t,c)$ as follows

$$L(t,c) = \prod_{j=1}^{m} \left(\left(\prod_{n=1}^{R_j(t,c_0)} \mu_j(T_j^n, c) \right) \exp\left(\int_0^t (1 - \mu_j(s,c_0)) a_j(X(s,c_0),c_0)\, ds \right) \right),$$

where T_j^n is the nth jump time of reaction j and μ_j are nonnegative, $\mathcal{F}_t$ predictable processes as we defined previously. By convention, we take the product $\prod_{n=1}^{R_j(t,c_0)}$ to be 1 if $R_j(t,c_0) = 0$. It can be shown that this L can be written as the solution of the following SDE,

$$L(t,c) = 1 + \sum_{j=1}^{m} \int_{(0,t]} L(s-,c)(\mu_j(s,c) - 1) dM_j(s,c_0),$$

where $M(t,c_0) = R(t,c_0) - \int_0^t a(X(s,c_0),c_0) ds$ as we defined before. For all $t \geq 0$ and all $j = 1, \cdots, m$, if

$$\int_0^t a_j(X(s,c_0),c_0)\, ds < \infty, \quad \mathbb{P}\text{-a.s.}, \tag{5.3}$$

then M is a local martingale. Hence it is reasonable for us to define

$$Y(t,c) = \sum_{j=1}^{m} \int_{(0,t]} (\mu_j(s,c) - 1)\, dM_j(s,c_0), \tag{5.4}$$

where Y is also a local martingale because the integrand above is left continuous ([23]). We note that L is known as the stochastic exponential of Y (see [24]),

$$L(t,c) = 1 + \int_{(0,t]} L(s-,c)\, dY(s,c).$$

$L(t,c)$ is a local martingale with respect to $(\mathbb{P}, \mathcal{F}_t)$ by the following theorem.

Theorem 5.1.1. *([8], VI. T2 Theorem) For the counting process*

$$R(t,c_0) = (R_1(t,c_0), \cdots, R_m(t,c_0))$$

adapted to $\mathcal{F}_t$*, let* $a_j(X(t-,c_0),c_0)$ *be the predictable* $(\mathbb{P},\mathcal{F}_t)$ *predictable intensity of* $R_j(t,c_0)$*,* $1 \le j \le m$*. For any* $c \in I_{c_0}$*, let* $\mu_j(t,c)$*,* $1 \le j \le m$ *be nonnegative,* $\mathcal{F}_t$ *predictable processes such that for all* $t > 0$*,* $\mathbb{P}$ *almost surely we have*

$$\int_0^t \mu_j(s,c)a_j(X(s-,c_0),c_0)ds < \infty.$$

Then $L(t,c)$ *is a* $(\mathbb{P},\mathcal{F}_t)$*-nonnegative local martingale.*

Next theorem shows that if $L(t,c)$ is a martingale, then the desired probability measure $\mathbb{P}(c)$ is well defined and $L(t,c)$ is the Radon-Nikodym derivative between $\mathbb{P}(c)$ and $\mathbb{P}$.

Theorem 5.1.2. *([8], VI. T*3 *Theorem) Under the same conditions as in Theorem 5.1.1, suppose moreover that* $\mathbb{E}L(T) = 1$*. Define the probability measure* $\mathbb{P}(c)$ *such that*

$$\mathbb{P}(c,A) = \int_A L(T,c)\mathbb{P}(d\omega)$$

for $A \in \mathcal{F}_T$*. Then for each* $1 \le j \le m$*,* $R_j(t,c_0)$ *has the* $(\mathbb{P}(c),\mathcal{F}_t)$*-intensity* $a_j(X(t-,c_0),c) = \mu_j(t,c)a_j(X(t-,c_0),c_0)$*.*

We shall use the above facts to show that

$$\mathbb{E}^{\mathbb{P}}[f(X(T,c))] = \mathbb{E}^{\mathbb{P}(c)}[f(X(T,c_0)], \tag{5.5}$$

where $\mathbb{E}^{\mathbb{P}}$ represents the expectation is taken with respect to the measure $\mathbb{P}$. It is equivalent to show that the probability law of $X(c)$ under the measure $\mathbb{P}$ is same as the probability law of $X(c_0)$ under the measure $\mathbb{P}(c)$. In fact, this can be seen from the Kolmogorov's forward equation (1.8) which governs the distribution of the

process X. We recall the equation here for reader's convenience

$$\frac{dp(x,t)}{dt} = \sum_{j=1}^{m}(p(x-\nu_j,t)a_j(x-\nu_j) - p(x,t)a_j(x)) \quad x \in \mathbb{Z}_+^n,$$

where $p(x,t) = \mathbb{P}(X(t) = x)$. Since $R(t,c_0)$ has $(\mathbb{P}(c),\mathcal{F}_t)$ predictable intensity $a_j(X(t-,c_0),c)$ by Theorem 5.1.2 and $R(t,c)$ has $(\mathbb{P},\mathcal{F}_t)$ predictable intensity $a_j(X(t-,c),c)$ by definition, they have the same Kolmogorov's forward equation and thus we obtain

$$\mathbb{P}(X(t,c) = x) = \mathbb{P}(c)(X(t,c_0) = x),$$

which implies (5.5). Therefore, to justify the change of intensity step for GT method, it is adequate to provide a sufficient condition for $L(t,c)$ to be a martingale on $[0,T]$. This leads to the so-called Novikov type condition problem. We will explore this problem in the next section.

5.2 Novikov Type Condition

In this section, we provide a Novikov type condition for the validity of applying GT method for sensitivity estimation. The result is based on the theorem from [24] (restated as A.2.1 in appendix) which discusses the Novikov type condition for martingales with jumps. Recall that we defined the process Y in terms of processes M and μ_j as

$$Y(t,c) = \sum_{j=1}^{m}\int_{(0,t]}(\mu_j(s,c)-1)\,dM_j(s,c_0)$$

and its stochastic exponential $L(t,c)$ satisfies

$$L(t,c) = 1 + \sum_{j=1}^{m}\int_{(0,t]} L(s-,c)(\mu_j(s,c)-1)dM_j(s,c_0).$$

Next theorem shows that under suitable conditions on the intensity functions, $L(t,c)$ turns out to be a zero mean martingale and hence we can define a new probability measure in terms of $L(t,c)$.

Theorem 5.2.1. *Let $a_j(X(t-,c_0),c_0)$ be the $(\mathbb{P},\mathcal{F}_t)$-intensity of $R_j(t,c_0)$. Suppose the following hold:*

1. *For all $c \in I_{c_0}$, and for all $t \geq 0$ and all $j = 1,\cdots,m$*

$$\int_0^t a_j(X(s,c_0),c)\,ds < \infty \quad \mathbb{P}\text{-a.s.}$$

2. *For all $c \in I_{c_0}$, all $x \in \mathbb{Z}_+^n$ and $j = 1,\cdots,m$*

$$a_j(x,c_0) = 0 \text{ if and only if } a_j(x,c) = 0.$$

3. *For all $c \in I_{c_0}$,*

$$\mathbb{E}\left[\exp\left(\sum_{j=1}^m \int_0^T (\mu_j(s,c)-1)^2 a_j(X(s,c_0),c_0)\,ds\right)\right] < \infty.$$

Then for each $j = 1,\cdots,m$, $a_j(X(t-,c_0),c)$ is the $(\mathbb{P}(c),\mathcal{F}_t)$-intensity of $R_j(t,c_0)$ over $[0,T]$, where

$$\frac{d\mathbb{P}(c)}{d\mathbb{P}} = L(T,c).$$

Proof. First note that under the first condition, $M(t,c_0)$ is a local martingale (under $\mathbb{P}$) for any $c > 0$ hence $Y(t,c)$ is also a local martingale (under $\mathbb{P}$) since the integrand μ_j are left continuous (see for example [23]). Secondly, the jump of $Y(t,c)$ is locally bounded since we can stop the processes μ_j by its left continuity. Therefore, by Lemma A.2.1 in the appendix, $Y(t,c)$ is a locally square integrable martingale.

Next, we need to decompose $Y(t,c)$ into its continuous part Y^c and purely discontinuous part Y^d and then check the condition (A.1) from Theorem A.2.1. Generally, it is hard to obtain these two parts for an arbitrary local martingale. However, since $Y(t,c)$ is a process of finite variation, $Y(t,c)$ itself is a purely discontinuous martingale (see [18, 23]), that is, we have $Y^d = Y, Y^c = 0$ and

$$\langle Y^d, Y^d \rangle = \langle Y, Y \rangle,$$

where $\langle Y, Y \rangle$ is the predictable quadratic variation process of Y (the compensator of $[Y, Y]$). By Theorem A.2.1, it remains to show that

$$\mathbb{E}\left[\exp\left(\langle Y, Y \rangle(T)\right)\right] < \infty,$$

then the process $L(t,c)$ is a $(\mathbb{P}, \mathcal{F}_t)$ martingale. We note that

$$\begin{aligned}
[Y,Y](t) &= \sum_{j=1}^{m} \int_{[0,t]} (\mu_j(s,c) - 1)^2 \, dR_j(s, c_0) \\
&= \sum_{j=1}^{m} \int_{[0,t]} (\mu_j(s,c) - 1)^2 \, dM_j(s, c_0) \\
&\quad + \sum_{j=1}^{m} \int_0^t (\mu_j(s,c) - 1)^2 a_j(X(s,c_0), c_0) \, ds.
\end{aligned}$$

Here we used the property $[R_j, R_k] = R_j$ if $j = k$ and $[R_j, R_k] = 0$ if $j \neq k$. Observe that the process defined by the first term on the right hand side is a local martingale and the process defined by second term is increasing and predictable. Hence, by Doob-Meyer decompostion (see [23]), the second term on the right hand side is the compensator of $[Y, Y]$, that is,

$$\langle Y, Y \rangle(t) = \sum_{j=1}^{m} \int_0^t (\mu_j(s,c) - 1)^2 a_j(X(s,c_0), c_0) \, ds.$$

Therefore, under the third condition, we conclude that $L(t,c)$ is a martingale for any $c \in I_{c_0}$.

□

Using the above change of measure criterion, the sensitivity is

$$\left.\frac{\partial}{\partial c}\right|_{c=c_0} \mathbb{E}f(X(T,c)) = \left.\frac{\partial}{\partial c}\right|_{c=c_0} \int_\Omega f(X(T,c_0))L(T,c)\,dP(c),$$

hence we validate the change-of-measure step for GT method. We will give conditions under which one can take derivative inside the integral in the next section.

5.3 A Sufficient Condition for GT

In this section, we provide a sufficient condition for differentiating inside the integral. We restrict the intensity function to be of the form

$$a_j(x,c) = c_j b_j(x)$$

for $j = 1, 2, \cdots, m$. Since c_0 is fixed, we will simply write $X(t)$ for $X(t,c_0)$ and similarly $R(t)$ for $R(t,c_0)$ and so on. Without loss of generality, we assume that we are interested in the sensitivity of $f(X(T))$ with respect to parameter c_1. Using the explicit form of $L(T,c)$ from [8],

$$L(T,c) = \prod_{j=1}^{m} \left(\left(\prod_{n=1}^{R_j(T)} \mu_j(T_j^n, c) \right) \exp\left(\int_0^T (1-\mu_j(s,c)) a_j(X(s), c_0)\,ds \right) \right),$$

where T_j^n is the nth jump time of reaction j and we take the product $\prod_{n=1}^{R_j(T)}$ to be 1 if $R_j(T) = 0$. Under the previous assumption on intensity, the above formula is

simplified to

$$L(T,c) = \left(\frac{c_1}{c_{01}}\right)^{R_1(T)} \exp\left(\int_0^T (c_{01}-c_0)(b_1(X(s))-b_1(X(s)))\,ds\right). \tag{5.6}$$

Note that

$$\frac{\partial}{\partial c_1}\ln L(T,c) = \frac{1}{c_1}R_1(T) - \int_0^T b_1(X(s))\,ds$$

hence,

$$\frac{\partial}{\partial c_1}L(T,c) = L(T,c)\left(\frac{1}{c_1}R_1(T) - \int_0^T b_1(X(s),c)\,ds\right).$$

Recall that $c_1 \in I_{c_{01}} = (c_{01}-\epsilon, c_{01}+\epsilon)$. If $c_1 \le c_{01}$,

$$L(T,c) \le \exp\left(\int_0^T (c_{01}-c_1)(b_1(X(s))-b_1(X(s)))\,ds\right),$$

and if $c_1 \ge c_{01}$,

$$L(T,c) \le \left(\frac{c_1}{c_{01}}\right)^{R_1(T)}.$$

Hence

$$\begin{aligned} L(T,c) &\le \exp\left(\int_0^T (c_{01}-c_0)(b_1(X(s))-b_1(X(s)))\,ds\right) + \left(\frac{c_1}{c_{01}}\right)^{R_1(T)} \\ &\le \exp\left(\epsilon\int_0^T b_1(X(s))\,ds\right) + \left(1+\frac{\epsilon}{c_{01}}\right)^{R_1(T)}. \end{aligned} \tag{5.7}$$

In the context of Theorem A.3.1, we take G to be

$$G(c,\omega) = f(X(T))L(T,c),$$

keep in mind $c = c_1$ a scalar. Then a Lipschitz constant $K(\omega)$ for G on the interval $I_{c_{01}}$ is

$$f(X(T))\left(\frac{R_1(T)}{c_{01}-\epsilon} + \int_0^T b_1(X(s))\,ds\right)\left(\exp\left(\epsilon\int_0^T b_1(X(s))\,ds\right) + \left(1+\frac{\epsilon}{c_{01}}\right)^{R_1(T)}\right).$$

We need the above Lipschitz constant to be integrable, so we need integrability on the following four terms,

$$f(X(T))R_1(T)\exp\left(\epsilon\int_0^T b_1(X(s))\,ds\right),$$
$$f(X(T))R_1(T)\left(1+\frac{\epsilon}{c_{01}}\right)^{R_1(T)},$$
$$f(X(T))\left(\int_0^T b_1(X(s))\,ds\right)\exp\left(\epsilon\int_0^T b_1(X(s))\,ds\right),$$
$$f(X(T))\left(\int_0^T b_1(X(s))\,ds\right)\left(1+\frac{\epsilon}{c_{01}}\right)^{R_1(T)}.$$

Using the inequality $3abc \le a^3+b^3+c^3$, it is suffice to impose that

$$\mathbb{E}(|f(X(T))|)^3 < \infty, \tag{5.8}$$

$$\mathbb{E}\left(\exp\left(3\epsilon\int_0^T b_1(X(s))\,ds\right)\right) < \infty, \tag{5.9}$$

$$\mathbb{E}\left(1+\frac{\epsilon}{c_{01}}\right)^{3R_1(T)} < \infty, \tag{5.10}$$

then we have the Lipschitz constant $K(\omega)$ is integrable. Also note that under the above integrability conditions, condition 4 of Theorem A.3.1 from appendix is immediate. This discussion leads to the following theorem.

Theorem 5.3.1. *Assume that the intensity is of the form $a_j(x,c) = c_j b_j(x)$ for $j = 1,\cdots,m$ and* (5.8) (5.9) (5.10) *hold true. Then*

$$\frac{\partial}{\partial c}\bigg|_{c=c_{01}}\int_\Omega f(X(T))L(T,c)\,dP(c_0) = \int_\Omega f(X(T))Z(T)\,dP(c_0),$$

where

$$Z(T) = \sum_{j=1}^m\int_0^T \frac{\frac{\partial a_j}{\partial c}(X(s-,c_0),c_0)}{a_j(X(s-c_0),c_0)}dR_j(s) - \sum_{j=1}^m\int_0^T \frac{\partial a_j}{\partial c}(X(s-,c_0),c_0)ds. \tag{5.11}$$

Chapter 6: Sensitivity at the Steady State

In this chapter, we briefly discuss the sensitivity estimation problem at the steady state. It is a well known result that if the CTMC $X(t,c)$ is irreducible, aperiodic and positive recurrent, then it has a unique steady state (or stationary) distribution π_c (see [20]). That is, there exists a unique measure on the positive integer lattice $\mathbb{Z}_+^n$ such that

$$\sum_{x\in\mathbb{Z}_+^n} \pi_c(x) = 1$$

and

$$\pi_c(x) = \lim_{t\to\infty} \mathbb{P}(X(t,c) = x).$$

Given that a unique steady state of a CTMC $X(t,c)$ exists, the steady state parametric sensitivity can be formulated as follows,

$$\left.\frac{\partial}{\partial c}\right|_{c=c_0} \pi_c(f) = \left.\frac{\partial}{\partial c}\right|_{c=c_0} \sum_{x\in\mathbb{Z}_+^n} f(x)\pi_c(x). \tag{6.1}$$

6.1 Drift-Diffusivity Condition

In this section, we cite the drift-diffusivity (DD) condition and some results regarding the long time behavior of the distribution of $X(t,c)$ from [5]. First we make some special requirement for the state space S: S is the smallest non-empty

subset of $\mathbb{Z}_+^n$ satisfying that if $x \in S$ and $a_j(x) > 0$ for some $j = 1, \cdots, m$, then $x + \nu_k \in S$. That is, S contains all the points to which the CTMC $X(t)$ can reach.

DD condition: For a strictly positive vector $v \in \mathbb{R}^m$, there exist positive constants c_1, c_2, c_3, c_4 and c_5 such that for all $x \in S$,

$$\sum_{j=1}^{m} a_j(x)\langle v, \nu_j\rangle \leq c_1 - c_2\langle v, x\rangle$$

$$\sum_{j=1}^{m} a_j(x)\langle v, \nu_j\rangle^2 \leq c_3 + c_4\langle v, x\rangle + c_5\langle v, x\rangle^2.$$

For this vector $v \in \mathbb{R}^m$, define the v-norm

$$\|x\|_v = \sum_{j=1}^{m} v_j|x_j|,$$

which will be used throughout this chapter. We also define $r_{max} = 1 + \frac{2c_2}{c_5}$ if $c_5 > 0$ or ∞ if $c_5 = 0$. Under the DD condition, one can show that the rth moment of $X(t, c)$ is uniformly bounded (in t) and the upper bound is function of initial value x_0 (Theorem 2 in [5]). If we further assume that $r < r_{\max} - 1$, then the rth moment of $X(t, c)$ converges to the rth moments of the stationary distribution π_c (Theorem 5 in [5]). These results are crucial for establishing the long time behavior of stochastic reaction networks. For our purpose, we cite the ergodicity result from [5] here.

Theorem 6.1.1. *Suppose the DD condition holds and there exists some positive integer $r < r_{max}$ such that*

$$f(x) \leq C(1 + \|x\|_v^r)$$

for all $x \in \mathbb{Z}_+^n$. Then $\sum_{x\in\mathbb{Z}_+^n} |f(x)|\pi_c(x) < \infty$. Moreover, for any $x_0 \in \mathbb{Z}_+^n$, if $X(0, c) = x_0$, then

$$\lim_{t\to\infty} \mathbb{E}f(X(t, c)) = \sum_{x\in\mathbb{Z}_+^n} f(x)\pi_c(x).$$

6.2 Asymptotic Correlation

We give the pure death and birth example to motivate our analysis in this section. The pure death and birth system consists of two reaction channels

$$S \xrightarrow{c_1} \emptyset \qquad \emptyset \xrightarrow{c_2} S$$

with intensities $a_1(x) = c_1 x$ and $a_2(x) = c_2$. For any fixed $t_0 > 0$, we want to see how the quantity $\mathbb{E}[X(t)Z(t_0)]$ grows as $t \to \infty$, where Z is the weight process for the sensitivity of $\mathbb{E}[X(t)]$ with respect to c_1. Similar to the argument we applied for the example in Chapter 3, first we write the processes X and Z as

$$\begin{aligned} X(t) &= X(0) - \int_0^t dR_1(s) + \int_0^t dR_2(s) \\ Z(t) &= \int_0^t \frac{1}{c_1} dR_1(s) - \int_0^t X(s) ds. \end{aligned} \tag{6.2}$$

We define $Z'(t)$ as follows

$$Z'(t) = \int_0^t \frac{1}{c_1} \mathbb{I}_{[0,t_0]} dR_1(s) - \int_0^t X(s) \mathbb{I}_{[0,t_0]} ds$$

Since eventually we will take $t \to \infty$, we assume $t \geq t_0$ and hence by Ito's formula

$$\begin{aligned} X(t)Z'(t) = &\int_0^t (X(s-) - 1)(Z'(s-) + \frac{1}{c_1}\mathbb{I}_{[0,t_0]}) - X(s-)Z'(s-) dR_1(s) \\ &+ \int_0^t (X(s-) + 1)Z'(s-) - X(s-)Z'(s-) dR_2(s) \\ &- \int_0^t X^2(s-)\mathbb{I}_{[0,t_0]} ds. \end{aligned} \tag{6.3}$$

Taking expectation of both sides leads to

$$\begin{aligned} \mathbb{E}[X(t)Z'(t)] = &- \int_0^t c_1 \mathbb{E}[X(s)Z'(s)] ds - \int_0^{t_0} \mathbb{E}[X(s)] ds + \int_0^{t_0} \mathbb{E}[X(s)^2] ds \\ &+ c_2 \int_0^t \mathbb{E}[Z'(s)] ds - \int_0^{t_0} \mathbb{E}[X(s)^2] ds \end{aligned} \tag{6.4}$$

which can be simplified to

$$\mathbb{E}[X(t)Z'(t)] = -\int_0^t c_1\mathbb{E}[X(s)Z'(s)]ds - \int_0^{t_0}\mathbb{E}[X(s)]ds. \tag{6.5}$$

The associated ode is

$$\frac{d}{dt}\mathbb{E}[X(t)Z'(t)] = -c_1\mathbb{E}[X(t)Z'(t)] \tag{6.6}$$

with initial condition at $t = t_0$ (note $Z(t_0) = Z'(t_0)$)

$$\mathbb{E}[X(t_0)Z(t_0)] = -\int_0^{t_0} c_1\mathbb{E}[X(s)Z(s)]ds - \int_0^{t_0}\mathbb{E}[X(s)]ds. \tag{6.7}$$

Hence the solution of (6.5) is

$$\mathbb{E}[X(t)Z'(t)] = \mathbb{E}[X(t_0)Z(t_0)]e^{-c_1(t-t_0)}. \tag{6.8}$$

Taking $t \to \infty$, we have

$$\mathbb{E}[X(t)Z(t_0)] \to 0.$$

Motivated by the above example, we prove the following result regarding the asymptotic correlation between $X(t)$ and $Z(t_0)$ for a fixed t_0 as $t \to \infty$.

Theorem 6.2.1. *Assume the DD condition and suppose $f : S \to \mathbb{R}$ is a function such that for some positive integer $r < r_{max} - 1$, there exists a constant $C > 0$ satisfying $|f(x)| \le C(1 + \|x\|_v^r)$ for all $x \in S$. For any $t_0 > 0$, the process $f(X(t))$ is asymptotically uncorrelated with the $Z(t_0)$, that is,*

$$\lim_{t\to\infty}\mathbb{E}[f(X(t))Z(t_0)] = 0.$$

Proof. First we condition the left hand side at $\mathcal{F}_{t_0}$. By the property of conditional expectation, we have

$$\mathbb{E}[f(X(t))Z(t_0)] = \mathbb{E}[Z(t_0)\mathbb{E}[f(X(t))|\mathcal{F}_{t_0}]].$$

Since X is a Markov process, we can further rewrite the above equation as

$$\mathbb{E}[f(X(t))Z(t_0)] = \mathbb{E}[Z(t_0)\mathbb{E}[f(X(t))|X(t_0)]].$$

Now note that $Z(t)$ always have zero mean, hence we can subtract $\mathbb{E}Z(t_0)\pi(f)$ from both sides of the above equation and write

$$\mathbb{E}[f(X(t))Z(t_0)] = \mathbb{E}[Z(t_0)(\mathbb{E}[f(X(t))|X(t_0)] - \pi(f))],$$

where $\pi(f) = \sum_{y\in S} f(y)\pi(y)$. Applying the Cauchy-Schwarz inequality, we have

$$\mathbb{E}[f(X(t))Z(t_0)] \leq \sqrt{\mathbb{E}[Z(t_0)^2]}\sqrt{\mathbb{E}[K(t-t_0, X(t_0))^2]}, \tag{6.9}$$

where $K(s,x) = \mathbb{E}[f(X(s+t_0))|X(t_0) = x] - \pi(f)$. By the proposition $S2$ in [5], for a function f which satisfies the above assumption, we have

$$\lim_{t\to\infty} \mathbb{E}[f(X(t))|X(0) = x_0] = \pi(f) < \infty$$

for any $x_0 \in S$. For any $x \in S$, conditioning on $X(t_0) = x$,

$$\lim_{t\to\infty} K(t-t_0, x) = \lim_{t\to\infty} \mathbb{E}[f(X(t))|X(t_0) = x] - \pi(f) = 0.$$

That is, $\lim_{t\to\infty} K(t-t_0, X(t_0)) = 0$ almost surely. On the other hand,

$$\mathbb{E}[f(X(t))|X(t_0) = x] \leq C\mathbb{E}[\|X(t)\|_v^r|X(t_0) = x] + C.$$

By theorem $S2$ from [5], $\mathbb{E}[\|X(t)\|_v^r|X(t_0) = x]$ is bounded uniformly (in t) by a constant $C_r(x)$. Hence, for any $t \geq 0$,

$$K(t-t_0, X(t_0)) \leq C + CC_r(X(t_0)) + \pi(f). \tag{6.10}$$

It can be shown that $C_r(X(t_0))$ is a polynomial of $\|X(t_0)\|_v$ and hence integrable by theorem $S2$ from [5]. Applying dominate convergence theorem to $K(t-t_0, X(t_0))$ leads to the result. □

We will show in next section that this result can be applied to reduce variance for sensitivity analysis.

6.3 Ensemble-Averaged Correlation Function Method

In this section, we use Theorem 6.1.1 to extend the GT method for steady state sensitivity estimation, which turns out to be the ensemble-averaged correlation function method in [30]. Recall that by 6.1.1, under suitable conditions, one can use the sensitivity at a finite time horizon to approximate the steady state sensitivity, that is,

$$\lim_{t\to\infty} \mathbb{E}(f(X(t,c))) = \lim_{t\to\infty} \sum_{x\in S} f(x) p^c_{x_0}(t,x) = \sum_{x\in S} f(x)\pi_c(x)$$

where $p^c_{x_0}(x,t) = \mathbb{P}(X(t,c) = x | X(0) = x_0)$. For a bounded system (i.e., $X(t,c)$ is uniformly bounded for any fixed c), we can show that

$$\left.\frac{\partial}{\partial c}\right|_{c=c_0} \lim_{t\to\infty} \mathbb{E}(f(X(c,t))) = \lim_{t\to\infty} \left.\frac{\partial}{\partial c}\right|_{c=c_0} \mathbb{E}(f(X(c,t))).$$

Therefore,

$$\left.\frac{\partial}{\partial c}\right|_{c=c_0} \pi_c(f) = \lim_{t\to\infty} \mathbb{E}(f(X(t,c_0))Z(t,c_0)). \tag{6.11}$$

For t large, we expect that $\mathbb{E}(f(X(t,c_0))Z(t,c_0))$ approximates the steady state sensitivity. CGT estimator can be used to reduce variance,

$$\left.\frac{\partial}{\partial c}\right|_{c=c_0} \pi_c(f) = \lim_{t\to\infty} \mathbb{E}((f(X(t,c_0)) - \mathbb{E}f(X(t,c_0)))Z(t,c_0)). \tag{6.12}$$

To achieve desired accuracy, one has to either simulate longer trajectory (t large) or run a large amount of trajectories (N_{tr}). However, the variance of CGT estimator grows dramatically in terms of the final time t, which makes this approximation not efficient. To reduce the variance further, note that

$$\begin{aligned}\left.\frac{\partial}{\partial c}\right|_{c=c_0} \pi_c(f) &= \lim_{t\to\infty} \mathbb{E}_{c_0}[(f(X(t)) - \mathbb{E}f(X(t)))Z(t)] \\ &= \lim_{t\to\infty} \mathbb{E}_{c_0}[(f(X(t)) - \mathbb{E}f(X(t)))(Z(t) - Z(t_0))] \\ &\quad + \lim_{t\to\infty} \mathbb{E}_{c_0}[f(X(t))Z(t_0)]\end{aligned}$$

and we can drop the last term because of Theorem 6.2.1. Therefore, for large t,

$$\left.\frac{\partial}{\partial c}\right|_{c=c_0} \pi_c(f) \approx \mathbb{E}_{c_0}[(f(X(t)) - \mathbb{E}f(X(t)))(Z(t) - Z(t_0))].$$

To see why this estimator is better than the original one, note that $Z(t)$ is a martingale and hence

$$\begin{aligned}\mathbb{E}(Z(t) - Z(t_0))^2 &= \mathbb{E}Z(t)^2 + \mathbb{E}Z(t_0)^2 - 2\mathbb{E}[Z(t)Z(t_0)] \\ &= \mathbb{E}Z(t)^2 - \mathbb{E}Z(t_0)^2 \\ &\leq \mathbb{E}Z(t)^2\end{aligned}$$

The last equality can be shown by conditioning.

Let us revisit the pure death birth system

$$S \xrightarrow{c_1} \emptyset \qquad \emptyset \xrightarrow{c_2} S$$

and apply the ensemble-averaged correlation function method to estimate the steady state sensitivity. It can be shown that the steady state sensitivity of X with respect to the parameters c_1 and c_2 are

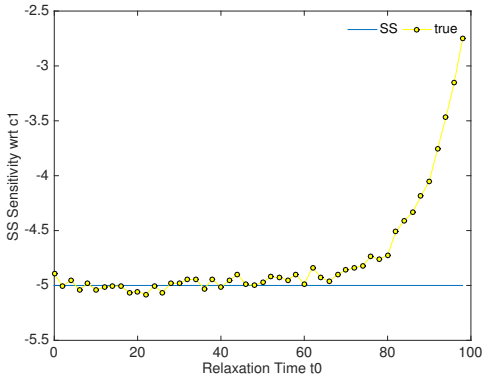
$$s_1 = -\frac{c_2}{c_1^2}$$

and

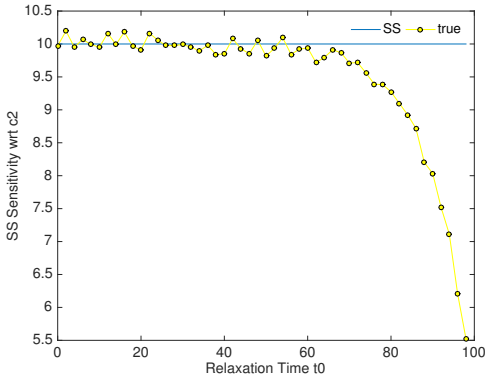
$$s_2 = \frac{1}{c_1}$$

respectively. We set the initial population to be $X(0) = 0$ and parameters $c_1 = 0.1$, $c_2 = 0.05$. We fix the terminal time $T = 100$ and use the sensitivity at T to approximate the steady state sensitivity. The relaxation time t_0 for this network changes from 0 to 98.

The following Figure 6.1 shows the approximated steady state sensitivity with respect to c_1 and c_2. Figure 6.2 compares the variance of the GT, the CGT and the ensemble-averaged correlation function method. One can easily observe that the variance of the ensemble-averaged correlation function estimator decays linearly as the relaxation time t_0 increase. However, if we take large t_0 to reduce the variance, we actually increase the bias. This facts can be seen from (6.8), which tells us that the bias is $\mathbb{E}[X(t_0)Z(t_0)]e^{-c_1(t-t_0)}$. Therefore, the ensemble-averaged correlation function suffers a bias-variance trade-off issue. We will not study the optimal choice of the relaxation time here, we encourage the interested reader to explore this aspect.

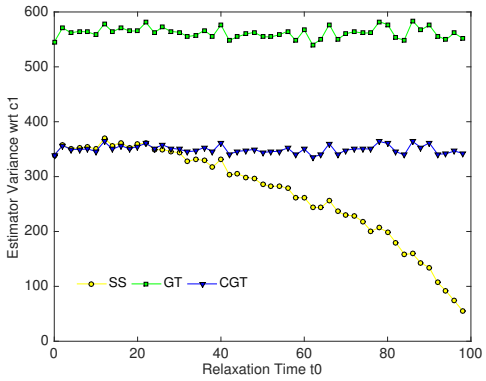


(a) Sensitivity wrt c_1

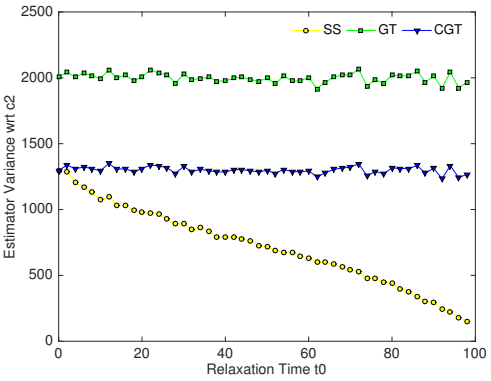


(b) Sensitivity wrt c_2

Figure 6.1: *Steady state sensitivity of the pure death and birth network. The horizontal axis represents the relaxation time t_0 and vertical axis represents the estimated steady state sensitivity. The blue line shows the true sensitivity when $c_1 = 0.1$, $c_2 = 0.05$.*



(a) Estimator Variance for Sensitivity wrt c_1



(b) Estimator Variance for Sensitivity wrt c_2

Figure 6.2: *Variance of the ensemble-averaged correlation function estimator. The horizontal axis represents the relaxation time t_0 and vertical axis represents the variance of the estimator.*

Chapter A: Appendix

A.1 Weak Convergence for joint Distribution

We give a theorem regarding the weak convergence of joint distribution.

Lemma A.1.1. *Suppose that α_n and β_n are random variables taking values on a metric space $(\mathcal{S}, \rho)$. If $\alpha_n \Rightarrow \alpha$ and $\rho(\alpha_n, \beta_n) \Rightarrow 0$, then $\beta_n \Rightarrow \alpha$.*

Proof. See Theorem 3.1 in [7]. □

Theorem A.1.1. *Let X_n and Y_n be $\mathbb{R}^m$ valued and $\mathbb{R}^k$ valued sequences of random variables on the same sample space. Suppose X_n converges to X in probability where X is deterministic and $Y_n \Rightarrow Y$. Then $(X_n, Y_n) \Rightarrow (X, Y)$ in $\mathbb{R}^{m+k}$.*

Proof. Let $x \in \mathbb{R}^m$ be such that $X = x$ almost surely. First we show that $(X, Y_n) \Rightarrow (X, Y)$. If $f : \mathbb{R}^{m+k} \to \mathbb{R}$ is bounded and continuous then so is $g : \mathbb{R}^k \to \mathbb{R}$ defined by $g(y) = f(x, y)$. Since $Y_n \Rightarrow Y$ we have that

$$\mathbb{E}(g(Y_n)) = \mathbb{E}(f(X, Y_n)) \to \mathbb{E}(g(Y)) = \mathbb{E}(f(X, Y)).$$

Now $\|(X_n, Y_n) - (X, Y_n)\| = \|X_n - X\|$ and since $X_n \to X$ in probability, $\|X_n - X\| \to 0$ in probability (implies convergence in distribution). Thus by Lemma A.1.1 we have that $(X_n, Y_n) \Rightarrow (X, Y)$. □

A.2 Novikov Type Criterion

Lemma A.2.1. *Let Y be a local martingale such that $Y(0)$ is square integrable and such that the process $\Delta Y(t) = Y(t) - Y(t-)$ is locally bounded, then Y is a locally square integrable martingale. That is, there exists a localizing sequence τ_n satisfying $\lim \tau_n = \infty$ such that for each n, $Y(t \wedge \tau_n)$ is a martingale and*

$$\sup_{t \geq 0} \mathbb{E}(Y^2(t \wedge \tau_n)) < \infty.$$

Proof. Without loss of generality, we assume Y is a martingale. For each $t > 0$, write $Y(t) = Y(t-) + \Delta Y(t)$. Since ΔY is locally bounded, there is a localizing sequence S_n such that $\Delta Y(t \wedge S_n)$ is uniformly bounded. Similarly, since $Y(t-)$ is left continuous, we can use the sequence of stopping time

$$T_n = \inf\{t > 0 : X(t-) \geq n\}$$

to localize it so that it is uniformly bounded. Now for each n, take $\tau_n = T_n \wedge S_n$, it is clear that $Y(t \wedge \tau_n)$ is a martingale and $\sup_{t \geq 0} \mathbb{E}(Y^2(t \wedge \tau_n)) < \infty$. □

Now we state a Novikov type criterion for cadlag martingales from [24]. Also see [23] and [18] for background.

Theorem A.2.1. *(Theorem 9 in [24]) Let Y be a locally square integrable martingale such that its jump $\Delta Y > 1$. If*

$$\mathbb{E}\left(e^{\frac{1}{2}\langle Y^c, Y^c\rangle(T) + \langle Y^d, Y^d\rangle(T)}\right) < \infty, \tag{A.1}$$

where Y^c and Y^d are the continuous and the purely discontinuous parts of Y, $\langle \cdot, \cdot \rangle$ is the predictable quadratic variation process. Then the stochastic exponential of Y, denoted by $L(t)$ defined by

$$L(t) = 1 + \int_{[0,t]} L(s-)\, dY(s)$$

is a martingale on $[0, T]$, where T can be infinity.

A.3 Differentiating Inside an Integral

Theorem A.3.1. *(See [4]) Suppose $G(c, \omega)$ is a random variable for each c in some interval of the real line. Let c_0 be a specific value of c. Suppose the following hold:*

1. *For a set of ω with probability one, $G(c, \omega)$ is differentiable with respect to c at $c = c_0$.*

2. *There exists an interval (c_l, c_u) containing c_0 (independent of ω) on which $G(c, \omega)$ is Lipschitz (in c) for a set of omega with probability one, with constant K which may depend on ω. That is, for any c_1, c_2 in the interval (c_l, c_u), the following holds:*

$$|G(c_1, \omega) - G(c_2, \omega)| \leq K(\omega)|c_1 - c_2|.$$

3. *$\mathbb{E}(K)$ is finite.*

4. *$\mathbb{E}(|G(c, \omega)|)$ is finite for all c in (c_l, c_u).*

Then the following holds:

$$\left.\frac{d}{dc}\right|_{c=c_0} \mathbb{E}(G(c)) = \mathbb{E}\left(\left.\frac{d}{dc}\right|_{c=c_0} G(c)\right).$$

Bibliography

[1] D. F. Anderson and E.S. Wolf, *Hybrid Pathwise Sensitivity Methods for Discrete Stochastic Models of Chemical Reaction Systems*, J. Chem. Phys., 142, 034103 (2015)

[2] D. F. Anderson, *An Efficient Finite Difference Method for Parameter Sensitivities of Continuous Time Markov Chains*, SIAM J. Numer. Anal., 50(5), (2012), pp 2237–2258.

[3] D. F. Anderson, *A Modified Next Reaction Method for Simulating Chemical Systems with Time Dependent Propensities and Delays*, J. Chem. Phys., 127, 214107 (2007).

[4] S. Asmussen and P.W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Springer, New York, 2007.

[5] Corentin Briat, Ankit Gupta and Mustafa Khammash *A Scalable Computational Framework for Establishing Long-Term Behavior of Stochastic Reaction Networks*, PLoS Comput Biol 10(6): e1003669. doi:10.1371/journal.pcbi.1003669.

[6] K. Ball, T. G. Kurtz, L. Popovic and G. Rempala, *AsymptoticAnalysis of Multiscale Approximations to Reaction Networks*, Ann. Appl. Probab., 16(4), (2006), pp 1925–1961.

[7] P. Billingsley, *Convergence of Probability Measures*, Second ed, John Wiley & Sons, New York, 1999.

[8] P. Bremaud, *Point processes and Queues: Martingale Dynamics*, Springer Verlag, New York, 1981.

[9] T. Cao, D.T. Gillespie and L.R. Petzold, *The Slow-scale Stochastic Simulation Algorithm*, J. Chem. Phys., 122, 014116 (2005).

[10] S.N. Ethier and T.G. Kurtz, *Markov processes: Characterization and convergence*, Second ed, John Wiley & Sons, New York, 2005.

[11] M.A. Gibson and Jehoshua Bruck, *Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels*, J. Phys. Chem. A, 2000, 104(9), pp 1876–1889.

[12] D.T. Gillespie, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comput. Phys., 22(4) (1976), pp 403–434.

[13] D.T. Gillespie, *Exact Stochastic Simulation of Coupled Chemical Reactions*, J. Phys. Chem., 81(25), (1977), pp 2340–2361.

[14] D.T. Gillespie, *Approximate Accelerated Stochastic Simulation of Chemically Reacting Systems*, J. Chem. Phys., 115, 1716 (2001).

[15] P. Glasserman, *Gradient Estimation Via Perturbation Analysis*, Springer, New York, 1990.

[16] A. Gupta and M. Khammash, *Unbiased Estimation of Parameter Sensitivities for Stochastic Chemical Reaction Networks*, SIAM J. Sci. Comput., 35(6), A2598A2620.

[17] A. Gupta and M. Khammash, *An Efficient and Unbiased Method for Sensitivity Analysis of Stochastic Reaction Networks*, Journal of The Royal Society Interface 11.101 (2014): 20140979.

[18] J. Jacod and A.N. Shiryaev, *Limit Theorems for Stochastic Processes*, Second ed, Springer, 2002.

[19] H. Kang, T. G. Kurtz, *Separation of Time-scale and Model Reduction for Stochastic Reaction Networks*, Ann. Appl. Probab., 23(2), (2013), pp 529–583.

[20] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, Second ed, Academic Press, 1975.

[21] Brian Munsky and Mustafa Khammash, *The Finite State Projection Algorithm for the Solution of the Chemical Master Equation*, J. Chem. Phys., 123, 044104(2006).

[22] S.Plyasunov and A.P. Arkin, *Efficient stochastic sensitivity analysis of discrete event systems*, J. Comput. Phys., 221(2) (2007), pp 724-738.

[23] P. Protter, *Stochastic integration and differential equations*, Second ed, Springer Verlag, New York, 2005.

[24] P. Protter and K. Shimbo, *No Arbitrage and General Semimartingales*, IMS Collections, Markov Processes and related Topics, Vol.4(2008), pp 267-283.

[25] Y. Pantazis and M. A. Katsoulakis, *A Relative Entropy Rate Method for Path Space Sensitivity Analysis of Stationary Complex Stochastic Dynamics*, J. Chem. Phys., 138, 054115 (2103).

[26] M.Rathinam, P.W. Sheppard, and M.Khammash, *Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks*, J. Chem. Phys., 132, 034103(2010).

[27] M. Rathinam, *Moment growth bounds on continuous time Markov processes on non-negative integer lattices*, Quarterly of Applied Mathematics (2015).

[28] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes and martingales: Volume 2, Itô Calculus*, Second ed, Cambridge University Press, 2000.

[29] P.W. Sheppard, M. Rathinam and M. Khammash, *A Pathwise Derivative Approach to the Computation of Parameter Sensitivities in Discrete Stochastic Chemical Systems*, J. Chem. Phys., 136, 034115 (2012).

[30] P.B. Warren and R.J. Allen, *Steady-state Parameter Sensitivity in Stochastic Modeling Via Trajectory Reweighting*, J. Chem. Phys., 136, 104106 (2012).

[31] T. Wang and M. Rathinam, *Efficiency of the Girsanov Transformation Approach for Parametric Sensitivity Analysis of Stochastic Reaction Networks*, arXiv:1412.1005, under revision.

[32] W. Whitt, *Proofs of the martingale FCLT*, Prob. Surveys, 4 (2007), pp 268–302.