Aalto University
School of Science
Degree Programme in Computational and Systems Biology

Debdas Paul

Efficient Parameter Inference for Stochastic Chemical Kinetics

Master's Thesis
Espoo, July 29, 2014

Supervisors:	Professor Erik Aurell, KTH, Sweden
	Professor Harri Lähdesmäki, Aalto University, Finland
Advisor:	Professor Stefan Engblom, Uppsala University, Sweden

Aalto University
School of Science
Degree Programme in Computational and Systems Biology

ABSTRACT OF
MASTER'S THESIS

Author:	Debdas Paul
Title:	Efficient Parameter Inference for Stochastic Chemical Kinetics

Date:	July 29, 2014	**Pages:**	34
Major:	Computational Systems Biology	**Code:**	IL3013
Supervisors:	Professor Erik Aurell Professor Harri Lähdesmäki		
Advisor:	Professor Stefan Engblom, Uppsala University, Sweden		

Parameter inference for stochastic systems is considered as one of the fundamental classical problems in the domain of computational systems biology. The problem becomes challenging and often analytically intractable with the large number of uncertain parameters. In this scenario, Markov Chain Monte Carlo (MCMC) algorithms have been proved to be highly effective. For a stochastic system, the most accurate description of the kinetics is given by the Chemical Master Equation (CME). Unfortunately, analytical solution of CME is often intractable even for considerably small amount of chemically reacting species due to its super exponential state space complexity. As a solution, Stochastic Simulation Algorithm (SSA) using Monte Carlo approach was introduced to simulate the chemical process defined by the CME. SSA is an exact stochastic method to simulate CME but it also suffers from high time complexity due to simulation of every reaction. Therefore computation of likelihood function (based on exact CME) and hence the rejection step (in an acceptance-rejection based MCMC like Metropolis-Hastings) becomes expensive. In this generic work, we introduce different approximations of CME as a pre-conditioning step to the full MCMC in order to make rejection cheaper. The goal is to avoid expensive computation of exact CME as far as possible. We show that, with effective pre-conditioning scheme, one can save a considerable amount of exact CME computations maintaining similar convergence characteristics. Additionally, we investigate three different sampling techniques (dense sampling of the same process, longer time sampling of the same process and i.i.d sampling of different processes) under which convergence of MCMC using exact CME for parameter inference can be analyzed. We find that under i.i.d sampling, better convergence can be achieved than that of other two techniques (atleast for the processes, we have investigated). We verify our theoretical findings for two different fundamental processes: linear birth-death and dimerization. Although, we succeed in saving a considerable amount of CME computations for two simple one-dimesional processes, challenges remain in extending it for higher dimensions which is a non-trivial problem.

Keywords:	stochastic chemical kinetics, systems biology, parameter inference, Markov Chain Monte Carlo, Chemical Master Equation
Language:	English

Acknowledgements

During the culmination of a thesis, a sense emerges that acknowledgements are due, and that is an opportunity to permanently document one's gratitude to people who are special and without whose support this point would be more difficult to reach.

First of all, I would like to express my sincere gratitude to Prof. Stefan Engblom, Uppsala University, Sweden for being my advisor/instructor without whom this work could not be materialized. My sincere thanks goes to Prof. Erik Aurell, Royal Institute of Technology, Sweden and Prof. Harri Lähdesmäki, Aalto University, Finland for being my supervisors.

I would like to thank Michalis, Tiago, Andreas, Kul, George and Bala, Javeria for being wonderful classmates throughout the entire program. Thanks again to Michalis for his critical comments which helps me to improve the text.

Finally, I would like to thank my parents Dr. Kalidas Paul and Mrs. Reba Paul for their constant encouragement and support throughout the entire journey.

Espoo, July 29, 2014

Debdas Paul

Abbreviations and Acronyms

CME	Chemical Master Equation
CLE	Chemical Langevin Equation
RRE	Reaction Rate Equation
SSA	Stochastic Simulation Algorithm
MCMC	Markov Chain Monte Carlo
i.i.d (I.I.D)	Independent Identically Distributed

Contents

Chapter 1

Introduction

Life is an interplay between determinism and randomness. This description of life is dependent on the level of observation and state of the system. When viewed from macroscopic level and the system is in equilibrium, determinism is perceived. In this framework (equilibrium and macroscopic scale), according to the laws of large number, one can expect a negligible fluctuation to a scale of $1/\sqrt{N}$ to a system behavior having N degrees of freedom. But life is also a dissipative system operating (thermodynamically) far from equilibrium [23] and hence can exhibit large fluctuations even in the macroscopic scale [15]. Now imagine, if the number of molecules are significantly less (which is the usual case with most of the fundamental processes inside a living cell), stochasticity takes over. A number of significant publications [2, 3, 16, 21, 22, 28] has already been established the fact. Therefore, stochasticity appears to be inherent in Nature. As an example, we can consider cellular processes which are nothing but set of chemical reactions. The reactions can be modeled using conventional reaction-rate-equations (RRE) as long as the number of interacting molecules is large. The problem occurs when this number is considerably small. We can not apply the classical mass action kinetics to determine the system behavior. Instead, the deterministic reaction rates are replaced by probabilistic reaction rates and the resulting modeling approach becomes Stochastic Chemical Kinetics [7, 17, 27].At this level, the most accurate description of the kinetics is given by the Chemical Master Equation (CME).

Unlike a deterministic system (which requires to solve coupled ODEs), stochastic chemical kinetics relies on generation of random numbers to realize different trajectories corresponding to each simulation. To predict which trajectory is statistically correct (and hence the solution of the stochastic equation) seems to be a difficult task until D.T. Gillespie comes with his

pioneering Monte Carlo technique known as the Stochastic Simulation Algorithm or SSA [8]. The beauty of this algorithm lies in the fact that its random walk realization exactly represents the distribution of the Chemical Master Equation or CME [9] which is a set of coupled first order ODE's describes the time evolution of the probability distribution of system being at a particular state among a discrete set of states. Though an exact description of the system, CME is often hard to solve analytically (most of the times) as well as numerically (sometimes). Moreover, for a bimolecular reaction, it is hard to predict the average behavior of the system [10].

Apart from the modeling approach using CME and its algorithmic treatment as SSA, stochastic chemical kinetics requires an efficient estimation of reaction rates or parameters which is often needed in systems biology involves experiments. Efficient estimation of the reaction rates leads to reconstruction of biochemical networks from experimental data: a fundamental problem in systems biology. In stochastic chemical kinetics paradigm, the continuum assumption of the number of molecules is not valid (as it is small) and hence least square fitting or maximum likelihood are not the best way to estimate the parameters [26]. In this scenario, Monte Carlo based approaches are proved to be efficient. Monte Carlo based methods can be classified into two main categories: Maximum Likelihood methods [20, 26] to obtain maximum likelihood estimates (MLE) of the parameters and Bayesian inference methods to obtain maximum *a posteriori* (MAP) estimates of the parameters. Now one of the techniques for Bayesian inference is Markov Chain Monte Carlo which first creates a random walk or a Markov process having stationary distribution same as the posterior distribution and then Monte Carlo sampling to sample directly from The major assumption underlying the CME is that reactions are occurring in well-mixed environments. Typically, this assumption is valid in sub-micron intracellular compartments since normal diffusion created homogeneity of molecular species over small volumes. The primary reason which has limited the exploitation of the CME approach is the lack of exact solutions. Hence, much of the literature to date have focused on identifying cases where exact solutions of the CME are possible and more generally on obtaining approximate solutions to the moments of the CME using sophisticated mathematical approaches. Here are some key papers relevant to the theme of this work:

Golightly and **Wilkinson** [11] proposes that one should use the Chemical Langevin Equation(CLE) rather than the *Kurtz process* (jump Markov process description) to describe the underlying stochastic process. The strategy is to use maximum likelihood estimation (MLE), where the probability distribution is given by a Gaussian centered at the expectation and spread by the co-variance found using the RRE. This method clearly fails in many

cases, however for simple reaction kinetics this gives a significant speed up. The paper extends into using MCMC for parameter search. Furthermore there is a discussion on unseen states, this is out of the scope of our work and can be omitted. Currently the most accepted MLE method uses CLE approximation.

Now, **Zimmer** and **Sahle** [29] takes a different approach. They construct a new likelihood function which is made up differences between an ODE trajectory, which they call *multiple shooting*, and the observation. This results in a very cheap likelihood to calculate. This method again for simple reaction kinetics behaves well, its resolution increases with sample size. This method of course converges when one increases the number of observations.

Next, **Milner et. al.** [19] extend the paper by Golightly and Wilkinson [11]. The paper uses moment closure to approximate the moments of the RRE. This is currently the best method in the literature. It of course runs into classical problems of approximating via the CLE.

Finally, **Efendiev et. al.** [4] takes on the preconditioned MCMC which demonstrates that using inexpensive coarse-scale computations one can improve the acceptance rate of MCMC up to 10 times.

In this work, we aim at CME and its relevant approximations to use as a precondition for MCMC. This investigation is novel in the sense that CME has never been exploited in this preconditioned fashion for parameter inference. Moreover, this work shows promises to obtain a scale-free preconditoned MCMC with higher acceptance rate with proper choice of preconditioning scheme. It also guides us towards the limitations of CME in general for parameter inference and how we can possibly overcome it by avoiding the exact computations of CME as far as possible.

Chapter 2

Theoretical Concepts and Algorithms

2.1 Stochastic Chemical Kinetics

In cellular systems, we often encounter small number of chemically reacting species. In this mesocopic environment , the interactions (dynamical behavior) between them govern by inherent randomness. Therefore, to model the time evolution of such system, we need a stochastic approach which is precisely the concept behind the **stochastic chemical kinetics**: chemical reactions modeled by stochastic processes. The commonest way for such kind of description is achieved by the **chemical master equation** (CME).

2.1.1 Chemical reactions and master equation

A homogeneous chemical reaction network can described as: $\{D, R\}$, where D is the set of species interacting with each other specified by set of reactions R. The state of the system at time t is $[x, t] = \{0, 1, \dots\}^D$: the number of molecules of each kind. Now, as the system is stochastic, there is some associated probability with which the state transition takes place. We call the **reaction propensities** denoted by $w_r : \mathbb{Z}_+^D \to \mathbb{R}_+$. Now if $\mathbb{N} \in \mathbb{Z}^{D \times R}$ is the **stoichiometric matrix**, we can write:

$$x \xrightarrow{w_r(x)} x - \mathbb{N}_r \tag{2.1}$$

Now, the CME describes the time evolution of the probability of the number of chemical species present at time t. For example, if we have $p(x, t)$ is the probability that x number of molecules is present in the system at time t, we

can write the CME [6, 27] as follows:

$$\frac{\partial p(x,t)}{\partial t} = \sum_{r=1;x+\mathbb{N}_r^- \geq 0}^{R} w_r(x+\mathbb{N}_r)p(x+\mathbb{N}_r,t) - \sum_{r=1;x-\mathbb{N}_r^+ \geq 0}^{R} w_r(x)p(x,t) \quad (2.2)$$

Where, $\mathbb{N}_r = \mathbb{N}_r^+ + \mathbb{N}_r^-$. Let us take the following two examples of a linear birth-death process and dimerization process (which is also going to be used for further analysis later in this thesis)

Example 2.1.1. *A linear birth-death process has the birth and death rate constant or linear. Therefore, we can have the following equation [5, eq : 2.4]:*

$$\phi \xrightarrow{\kappa} x$$
$$x \xrightarrow{\mu x} \phi \quad (2.3)$$

We have $\mathbb{N} = [-1 \; 1]$ $w = [k \; \mu x]^T$. The corresponding master equation becomes [5, eq : 2.5]

$$\frac{\partial p(x,t)}{\partial t} = [x \geq 1]kp(x-1,t) - kp(x,t) + \mu(x+1)p(x+1,t) - \mu x p(x,t) \quad (2.4)$$

Example 2.1.2. *Dimerization is one of the fundamental processes which spans from chemical, biological and physics system. Let us take the following set of reactions from [5, eq : 2.9]:*

$$\phi \xrightarrow{k} x$$
$$x \xrightarrow{\mu x} \phi$$
$$x + x \xrightarrow{\nu x(x-1)} \phi \quad (2.5)$$

The equation can be further simplified (by dropping the mass loss) to:

$$\phi \xrightarrow{k} x$$
$$x + x \xrightarrow{\nu x(x-1)} \phi \quad (2.6)$$

We mention the above simplification as our analysis of dimerization later is based on these equations. Now, we can write the corresponding master equation as:

$$\frac{\partial p(x,t)}{\partial t} = [x \geq 1]k(p(x-1,t) - p(x,t)) + \nu(x+2)(x+1)p(x+2,t) - \nu x(x-1)p(x,t) \quad (2.7)$$

2.2 MCMC and Preconditioned MCMC for Parameter estimation

2.2.1 Monte Carlo Methods

In mathematics, we often encounter complicated functions whose integration over its domain of defintion is not straightforward to compute (high dimensional non-smooth integrands). *Monte-Carlo* methods are the way to obtain the numerical approximation of those kind of integrations by expressing the integral in terms of expectation derive from the *Law of Large Numbers*. For example, consider the following integration:

$$I = \int_\Omega f(\mathbf{x})d\mathbf{x} \tag{2.8}$$

Where, Ω is the domain of the integration for the function $f(\mathbf{x})$. Now, the integration can also be written in the form of expectation of a random variable X with repspect to some probability measure in the following way:

$$E(f(\mathbf{x})) = \int_\Omega f(\mathbf{x})\rho(\mathbf{x})d\mathbf{x} \tag{2.9}$$

If density function $\rho(\mathbf{x}) > 0$ whenever $f(\mathbf{x}) \neq 0$, we can re-write (2.9) as :

$$
\begin{aligned}
E(f(\mathbf{x})) &= \int_\Omega f(\mathbf{x})\rho(\mathbf{x})d\mathbf{x} \\
&= \int_\Omega \frac{f(\mathbf{x})}{q(\mathbf{x})}\rho(\mathbf{x})q(\mathbf{x})d\mathbf{x} \\
&= E\left(\frac{f(\mathbf{X})\rho(\mathbf{X})}{q(\mathbf{X})}\right) = E(g(\mathbf{X}))
\end{aligned}
\tag{2.10}
$$

Now, considering Ω finite, from the *Laws of Large Numbers*, we have for iid r.v $\{\mathbf{X}_i\}_{i=1}^\infty$

$$E(g(\mathbf{X}) = \lim_{N\to\infty} \frac{1}{N}\Sigma_{i=1}^N g(\mathbf{X_i}) \tag{2.11}$$

Therefore, we can first generate $\{\mathbf{X}_i\}_{i=1}^N$ from $\rho(\mathbf{X})$ and then calculate the expectation based on *Laws of Large Numbers* to finally approximate the integration $I_{MC} \sim \frac{1}{N}\Sigma_{i=1}^N g(\mathbf{X_i})$. the *Monte-Carlo* error in estimation is given by:

$$Err = \left| \lim_{N\to\infty} \frac{1}{N}\Sigma_{i=1}^N g(\mathbf{X_i}) - I \right| \tag{2.12}$$

The convergence rate in *Monte-Carlo* approach is of $O(1/\sqrt{N})$ - the reason behind time expensive simulations using *Monte-Carlo* methods. I_{MC} is the unbiased estimator for the integration I and it converges *almost surely* to I (using strong law of large numbers).

Applications of *Monte-Carlo* methods can be well understood through Bayesian inference. Bayesian inference involves expressing the posterior distribution in terms of likelihood and prior. This is acheived through via Bayes' theorem in the following way:

$$f(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\boldsymbol{\theta})f(\mathbf{x}|\boldsymbol{\theta})}{\int f(\boldsymbol{\theta})f(\mathbf{x}|\boldsymbol{\theta})d\boldsymbol{\theta}} \propto f(\boldsymbol{\theta})f(\boldsymbol{x}|\boldsymbol{\theta}) \tag{2.13}$$

Where, $\boldsymbol{\theta}$ is the set of paramters; $f(\boldsymbol{\theta})$ is the prior and $f(\boldsymbol{x}|\boldsymbol{\theta})$ is the likelihood function. $\mathbf{x} = (x_1, x_2, ...x_N)$ conditionally independent and identically distributed (CIIDs). Now once the posterior distribution $f(\boldsymbol{\theta}|\mathbf{x})$ is known, we can derive the mean and variance for each parameters as well as predictive distribution. For example, predictive distribution can be obtained through:

$$f(\mathbf{y}|\mathbf{x}) = \int f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{x})f(\boldsymbol{\theta}|\mathbf{x}) \, d\boldsymbol{\theta} \tag{2.14}$$

Now, using the *Monte-Carlo* trick, we can write:

$$\int f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{x})f(\boldsymbol{\theta}|\mathbf{x}) \, d\boldsymbol{\theta} \sim \frac{1}{N}\Sigma_{i=1}^{N}f(\mathbf{y}|\boldsymbol{\theta}^{(N)}, \mathbf{x}); \boldsymbol{\theta}^{(N)} \sim f(\boldsymbol{\theta}|\mathbf{x}) \tag{2.15}$$

Now summarizing the above discussion, we can describe *Monte-Carlo* methods as to first sample i.i.ds from a target density $\rho(\mathbf{X})$ (posterior for example) defined in a high-dimensional space and finally approximate the density using empirical point-mass function. Now the question arises, what if sampling from $\rho(\mathbf{X})$ is not straightforward (e.g. Cauchy distribution). We need to apply more efficient sample strategies which can evaluate $\rho(\mathbf{X})$ up-to a normalizing constant. This is precisely the key motivation behind our next topic - *Markov Chain Monte Carlo* (MCMC).

2.2.2 Markov Chain Monte Carlo

MCMC explores the state space by constructing an ergodic Markov Chain $\{\eta_i\}_{i=1}^{N}$ whose stationary distribution, say Π is the same as the target distribution $\rho(\mathbf{X})$. After that the chain is simulated until convergence and the next N observed values from the chain approximates a *Monte-Carlo* sample from $\rho(\mathbf{X})$. This immediately raises the question how to construct the chain whose stationary distribution is same as the target distribution. One way

to generate the chain is via the Metropolis-Hastings algorithm. The algorithm was first proposed by Metropolis et al. [18] and later it was extended by Hastings [13]. Unlike Gibb's sampling (another sampling technique) , it avoids any need to sample from complicated distributions. It can be applied to problems where the state is either continuous or discrete, as long as it is possible to compute the ratio of the probabilities, or probability densities, of two states An algorithmic representation is as follows: One of interesting

---

Algorithm 1: The Metropolis-Hastings Algorithm

Given: Observations $\{X(t_i; \boldsymbol{\theta})\}, i = 1 \ldots N$, $\boldsymbol{\theta}$: Unknown parameter vector; A proposal distribution $q(\mathbf{X})$; Target distribution $\rho(\mathbf{X})$
Result: Sample estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$
Initialize: $\boldsymbol{\theta}^1$ at $i = 1$

repeat
    Generate $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}|\boldsymbol{\theta}^i)$
    if $\rho(\boldsymbol{\theta}^i)q(\boldsymbol{\theta}^i|\boldsymbol{\theta}^*) > 0$ **then**
        $\alpha(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*) = \min\left(\frac{\rho(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)}{\rho(\boldsymbol{\theta}^i)q(\boldsymbol{\theta}^i|\boldsymbol{\theta}^*)}, 1\right)$
    else
        $\alpha(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*) = 1$
    end
    $\boldsymbol{\theta}^{i+1} = \begin{cases} \boldsymbol{\theta}^* & \text{with probability } \alpha(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*) \\ \boldsymbol{\theta}^i & \text{with probability } 1 - \alpha(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*) \end{cases}$
    $i = i + 1$
until *Convergence is detected*;

---

facts about MCMC is that it is not Bayesian and hence marginal likelihood does not require but we need to mention the likelihood function. Often the likelihood function is hard to compute and because of multiplication we may encounter very small number difficult to handle during computations. One way to alleviate the problem is to take *log*-likelihood. In this work, we aim to estimate the parameter for stochastic chemical kinetics using MCMC and CME. The problem with MCMC for exact CME is that the rejection step is expensive. Therefore, cheaper likelihood function should be constructed to avoid time expensive rejection step. The idea is to construct a two stage algorithm in which at the first stage we use an approximate prior to make rejection cheaper at the cost of accuracy. In the final step, we compute exact SSA (simulate exact CME) for only few promising samples. This way of *preconditioning* avoids solving CME exactly unless it is deemed necessary. In the next subsection we present a preconditioned version of MCMC [4].

2.2.3 Preconditioned Markov Chain Monte Carlo

Algorithm 2: Preconditioned Markov-Chain-Monte-Carlo

Given: Observations $\{X(t_i; \boldsymbol{\theta}), i = 1 \ldots N$, $\boldsymbol{\theta}$: Unknown parameter vector; A proposal distribution $q(\mathbf{X})$; Target distribution $\rho(\mathbf{X})$

Result: Sample estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$

Initialize: $\boldsymbol{\theta}^1$ at $i = 1$

repeat

    Generate $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}|\boldsymbol{\theta}^i)$

    if $\rho^*(\boldsymbol{\theta}^i)q(\boldsymbol{\theta}^i|\boldsymbol{\theta}^*) > 0$ **then**

        $\alpha^*(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*) = \min \left(\frac{\rho^*(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)}{\rho^*(\boldsymbol{\theta}^i)q(\boldsymbol{\theta}^i|\boldsymbol{\theta}^*)}, 1 \right)$

        Consider $Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i) =$
$\alpha^*(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*)q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i) + \left(1 - \int \alpha^*(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*)q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)d\boldsymbol{\theta}^*\delta_{\boldsymbol{\theta}^i}(\boldsymbol{\theta}^*)\right)$

        if $\rho(\boldsymbol{\theta}^i)Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i) > 0$ **then**

            $\alpha(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*) = \min \left(\frac{\rho(\boldsymbol{\theta}^*)Q(\boldsymbol{\theta}^i|\boldsymbol{\theta}^*)}{\rho(\boldsymbol{\theta}^i)Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)}, 1 \right)$

            $\boldsymbol{\theta}^{i+1} = \begin{cases} \boldsymbol{\theta}^* & \text{with probability } \alpha(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*) \\ \boldsymbol{\theta}^i & \text{with probability } 1 - \alpha(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*) \end{cases}$

            $i = i + 1$

        end

    end

until *Convergence is detected*;

In the above algorithm, we do not need to compute the term $Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)$ using integration. Instead, we can write [4]

$$\frac{Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)}{Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)} = \frac{\rho^*(\boldsymbol{\theta}^i)}{\rho^*(\boldsymbol{\theta}^*)}$$

Therefore,

$$\alpha(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*) = \min \left(\frac{\rho(\boldsymbol{\theta}^*)\rho^*(\boldsymbol{\theta}^i)}{\rho(\boldsymbol{\theta}^i)\rho^*(\boldsymbol{\theta}^*)}, 1 \right)$$

The goal of preconditioning is to increase the acceptance rate *per CME* (in this context).

2.3 Convergence in MCMC with exact CME

Here we consider some simple elementary processes for which we investigate the convergence of MCMC with exact CME.

2.3.1 Methods for convergence

There are three ways to increase data and under which we can consider convergence:

1. More dense sampling of the same process

2. Longer sampling of the same process

3. Sampling of several processes

2.3.2 Analysis of convergence schemes

Analysis of convergence is often a difficult task. Sometimes, we can not prove it analytically. In those cases, we have to find at least one counter example to that particular convergence scheme. Here we follow the same approach. For example, the convergence for pure birth process can not be improved under the scheme 1, hence not appropriate. The detailed explanation is below.

2.3.2.1 Pure birth process

The is because for a homogeneous Poisson process (with mean κt), dense sampling of the same process leads to poor estimate of parameter κ. For example, let us consider the following pure birth process:

$$\phi \xrightarrow{\kappa} X \tag{2.16}$$

Now, this could be represented mathematically by $\{X_t; t\} \sim Po(\kappa t); X_0 = 0$, where $Po()$ denotes the Poisson process, X_t denotes the number of events (birth) at time t. Now, as we increase the data, at infinity the estimate $\hat{\kappa}$ will converge to κ. As ideally $X_t = \kappa t$ (from the definition of $\kappa = $ birth rate) we have the following estimate:

$$\kappa \sim \frac{X_t}{t} \pm \sqrt{Var(X_t/t)} \tag{2.17}$$

Therefore, from the above equation it is evident that the dense sampling of the same process is not the way for the parameter κ to converge as the variance term in the right hand side does not vanish.

2.3.2.2   Linear birth-death process

The linear birth-death process, described in the example 2.1.1 similarly indicates why convergence scheme 2 cannot be expected to work. Let us consider the equations once again:

$$\phi \xrightarrow{\kappa} X$$
$$X \xrightarrow{\nu X} \phi$$

$$(2.18)$$

As linear birth-death process (above) is a $M/M/\infty$ queue, the stationary distribution (at $t \to \infty$) is $\frac{\kappa}{\nu}$. Therefore as we simulate the same process for longer time, the convergence will be towards the ratio $\frac{\kappa}{\nu}$ and therefore, we can not estimate κ and ν separately under the second scheme of convergence. Hence we can essentially only expect convergence $\hat{\theta} \to \theta$ according to the increase of data in the sense of 3, that is, sampling an increasing number of i.i.d. processes. The idea here is to find at least one counter example which does not satisfy a convergence scheme listed above. For the convergence criteria 3, we have to modify our computations for the likelihood function because the simulated path is not continuous due to i.i.d processes. Therefore, the joint *log*-likelihood (as implemented in the algorithm) will be the summation of *log*-likelihood for each process.

2.3.2.3   Dimerization

Dimerization can be approximated with birth-death process (we will see in subsequent sections), therefore expected to be suffered from the problems above. Therefore, we can investigate the convergence only under the scheme 3. We will see in chapter **??**, that it works.

2.4   Exact solution of CME and Preconditioning schemes

In this section, we first describe the exact analytical solution (if exists) for the CME corresponding the processes in discussion and finally, we formulate corresponding preconditioning schemes.
Let us consider the linear-birth death process in example (2.1.1) again. We have the following reactions:

$$\phi \xrightarrow{\kappa} x$$
$$x \xrightarrow{\mu x} \phi$$

$$(2.19)$$

and the corresponding master equation as

$$\frac{\partial p(x,t)}{\partial t} = [x \geq 1]kp(x-1,t) - kp(x,t) + \mu(x+1)p(x+1,t) - \mu x p(x,t) \quad (2.20)$$

Now, if the initial data is not Poisson distributed, we have the following general analytical solution of the above CME due to [25]

$$p(x,t) = \sum_{l=0}^{\min\{x,x(0)\}} \binom{x(0)}{l} e^{-l\mu t}(1 - e^{\mu t})^{x(0)-l} \frac{\lambda(t)^{x-l}e^{\lambda(t)}}{(x-l)!} \quad (2.21)$$

where $\lambda(t) = \frac{k(1-e^{-\mu t})}{\mu}$. Now for dimerization process (below):

$$\phi \xrightarrow{k} x$$
$$x + x \xrightarrow{\nu x(x-1)} \phi \quad (2.22)$$

we have the following CME

$$\frac{\partial p(x,t)}{\partial t} = [x \geq 1]k(p(x-1,t) - p(x,t)) + \nu(x+2)(x+1)p(x+2,t) - \nu x(x-1)p(x,t) \quad (2.23)$$

Equation 2.23 above does not have a simple analytical solution [5], like linear-birth death process. Therefore, we solve it numerically in a one dimensional lattice having N points where N is scaled according to the volume of the system.

2.4.1 Moment based approximations of CME

Here we describe moment based approximations of CMEs used as a preconditioner.

2.4.1.1 Approximation based on expectation

The idea here is to obtain a set of ODE/simple time derivative of the expectation for the corresponding process and after solving, use it as the parameter for a Poisson distribution (approximation). To achieve this, first we need to consider the following lemma due to [5, lemma.$\sim$ 2.1]:

Lemma 2.4.1. *Let p satisfy a proper formulation of the master equation 2.2, Then as long as the both sides make sense, we have:*

$$\sum_{x \geq 0} T(x)\frac{\partial p(x,t)}{\partial t} = \sum_{r=1}^{R} E[(T(X - \mathbb{N}_r) - T(X))w_r(X)], \quad (2.24)$$

where, $T : \mathbb{Z}^{D+} \to \mathbb{R}$ is any suitable test function

Now, let us consider $T(x) = x_i$, then we will have,

$$\sum_{x \geq 0} x \frac{\partial p(x,t)}{\partial t} = \sum_{r=1}^{R} E[(X - \mathbb{N}_r) - X)w_r(X)],$$

$$\frac{dm_i}{dt} = \sum_{r=1}^{R} E[(\mathbb{N}_r)w_r(X)] = \sum_{r=1}^{R} (\mathbb{N}_r)E[w_r(X)] \tag{2.25}$$

Now the left hand side of the equation is nothing but the time derivative of the expectation, let us denote it by m_i. $\mathbb{N} = [-1 \ 1]$ and $w(x) = [k \ \mu x]$, for the linear birth-death process. Putting these values in the final expression of 2.25, and using the following proposition [5, prop.$\sim$ 2.3] based on the assumption of linear propensity function with vanishing higher moments:

Proposition 2.4.2. *Divide the integers* $1...R$ *into two disjoint sets* R_1 *and* R_2 *such that* $\forall r \in R_i : w_r$ *is linear and* $\forall r \in R_i : w_r$ *depends on the dimension of the vanishing higher moments. Then*

$$\frac{dm_i}{dt} = \sum_{r=1}^{R} \mathbb{N}_r^i w_r(m) \tag{2.26}$$

we have,

$$\frac{dm}{dt} = k - \mu m \tag{2.27}$$

Which has the solution:

$$m(t) = \frac{k}{\mu} - \frac{1}{\mu} \exp(-\mu t + C) \tag{2.28}$$

Now, C can be obtained from the initial value $m(0)$. The final dynamic expression for $m(t)$ is:

$$m(t) = \frac{k}{\mu}(1 - \exp(-\mu t)) + m(0)\exp(-\mu t) \tag{2.29}$$

and the correpsonding dynamic solution (Poisson approximation)

$$p(x,t) = \frac{m(t)^x}{x!}\exp(-m(t)) \tag{2.30}$$

For dimerization in 2.22, we have the expression for expection:

$$\frac{dm}{dt} = k - 2\nu m(m-1) \tag{2.31}$$

Therefore, this is equivalent to a birth-death process where μ is replaced by $2\nu(m-1)$. This is called **Explicit linearization**, which we use as one of the preconditioning technique for dimerization.

2.4.1.2 Approximation based on expectation and variance

The approach is the same as above but along with the expectation, we will have one more moment, *variance*. After solving the corresponding ODEs (for both mean and variance), we obtain corresponding Gaussian approximation to the exact CME. To formulate the ODE's, we consider the following proposition [5, prop.$\sim$ 2.5] based on the linearization of the propesnity w_r (using Taylor expnasion)

Proposition 2.4.3. *If all propensities w_r are at most quadratic and if the third central moments may be neglected, then*

$$\frac{dm_i}{dt} = -\sum_{r=1}^{R} n_r^i \left(w_r(m) + \sum_{g,l} \frac{\partial^2 w_r(m)}{\partial x_g \partial x_l} \frac{C_{gl}}{2!} \right)$$

$$\frac{dC_{ij}}{dt} = -\sum_{r=1}^{R} \left(n_r^i \sum_g \frac{\partial w_r(m)}{\partial x_g} \frac{C_{gj}}{1!} + n_r^j \sum_g \frac{\partial w_r(m)}{\partial x_l} \frac{C_{il}}{1!} \right) \tag{2.32}$$

$$+ \sum_{r=1}^{R} n_r^{[i,j]} \left(w_r(m) + \sum_{g,l} \frac{\partial^2 w_r(m)}{\partial x_g \partial x_l} \frac{C_{gl}}{2!} \right)$$

Where, C is the covariance matrix. For our processes, we replace covariance as variance v as we have only one species. Additionally, $n_r[i,j] = n_r^i n_r^j$. Now using the above proposition, we have the ODEs of mean and variance for our linear birth-death process

$$\frac{dm}{dt} = k - \mu m$$
$$\frac{dv}{dt} = k + \mu m - 2\mu v \tag{2.33}$$

Solving the above set of ODEs, we get the following dynamic solution for mean and the variance:

$$m(t) = \frac{k}{\mu}(1 - \exp(-\mu t)) + m(0)\exp(-\mu t)$$
$$v(t) = m(t) - m(0)\exp(-2\mu t) \tag{2.34}$$

Now, we have the approximation as $p(x,t) = \mathcal{N}(m(t), v(t))$, where $\mathcal{N}(\ )$ denotes the Gaussian distribution.

For dimerization, we have the following set of ODEs. Unlike, linear birth-death, it does not have a explicit form of solution and therefore solved nu-

merically.

$$\frac{dm}{dt} = k - 2\nu m(m-1) - 2\nu v$$
$$\frac{dv}{dt} = k + 4\nu m(m-1) + 4\nu v - 4\nu v(2m-1) - 4\nu s$$

(2.35)

s is the third central moment which can be neglected for approximation.

2.4.2 Scaling and Discrete correction for Gaussian

Here we introduce the concept of scaling to set the acceptance rate of full MCMC between $[0.1, 0.4]$. The scaling is introduced during the proposal step for the parameters. In proposal step, we use *log* normal distribution to keep the same support (postive) for proposal and target distribution. Recall the acceptance step of Metropolis-Hastings (MH) algorithm 1, we have

$$\alpha(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*) = \min\left(\frac{\rho(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)}{\rho(\boldsymbol{\theta}^i)q(\boldsymbol{\theta}^i|\boldsymbol{\theta}^*)}, 1\right)$$

Now,as we fomulate the proposal of MH using *log*-normal distribution, we have $\frac{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)}{q(\boldsymbol{\theta}^i|\boldsymbol{\theta}^*)} = \exp(S * \mathcal{N}(0,1))$, where S is the scale factor. It scales the variance $(\sigma^2)$of the Normal random variate by factor of S^2.

Proposition 2.4.4. *If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$*

Proof.

$$
\begin{aligned}
Pr(aX + b < k) \;\; &= Pr(X < (k-b)/a) \\
&= \int_{-\infty}^{(k-b)/a} \frac{1}{\sqrt{2\pi}\sigma} \exp(-(x-\mu)^2/2\sigma^2)dx \\
&= \int_{-\infty}^{k} \frac{1}{\sqrt{2\pi}(a\sigma)} \exp(-(y-(a\mu+b))^2/2a^2\sigma^2)dy \;\; [y = ax+b, dy = adx] \\
&= \text{cdf of } \mathcal{N}(a\mu + b, a^2\sigma^2)
\end{aligned}
$$

$\square$

Now, for our case $\mu = 0, \sigma^2 = 1, a = S, b = 0$. Accordingly, we take the *log*-likelihood of the data instead of simple likelihood. Therefore, the proposal step becomes, $k' = k\exp(S * \mathcal{N}(0,1))$ for k, $\mu' = \mu\exp(S * \mathcal{N}(0,1))$ for μ. For proposal step preserving the ratio $\frac{k}{\mu}$, we have the ratio $\frac{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)}{q(\boldsymbol{\theta}^i|\boldsymbol{\theta}^*)} = \Pi_{j=1}^2 \exp(S*\mathcal{N}(0,1))_j$. All these steps are in accordance with the construction of MH algorithm.

Proof. Let $G(x \to x')$ be the proposal distribution. Let's also assume, $x' = x \exp(S * \mathcal{N}(0,1))$ Then,

$$
\begin{aligned}
G(x \to x') = P(x'|x) \ &= \mathcal{NL}(x', \mu = log(x), \sigma = S) \\
&= \frac{1}{x'\sigma\sqrt{2\pi}} \exp(-\frac{log(x'/x)^2}{2\sigma^2})
\end{aligned}
$$

Similarly we have (due to symmetry of the proposal),

$$
\begin{aligned}
G(x' \to x) = P(x|x') \ &= \mathcal{NL}(x, \mu = log(x'), \sigma) \\
&= \frac{1}{x\sigma\sqrt{2\pi}} \exp(-\frac{log(x/x')^2}{2\sigma^2})
\end{aligned}
$$

Now, $\frac{G(x \to x')}{G(x' \to x)} = x'/x = exp(S * \mathcal{N}(0,1))$. As k and μ are selected independently, we have $\frac{G(k \to k')G(\mu \to \mu')}{G(k' \to k)G(\mu \to \mu')} = (k'\mu')/(k\mu) = \Pi_{j=1}^2 \exp(S * \mathcal{N}(0,1))_j$ □

For dimerization we implement the scaling in the same way as above. Apart from introducing the scaling factor, we also introduce a discrete correction to Gaussian approximation for precondtioning. Experimentally, we find that (see Chapter (3) Results for more details), for fine temporal scale the Gaussian approximation overshoots the target distribution. It seems that as the Gaussian is a continuous distribution, in smaller scale this adjustment is quite important and can not be ignored. For linear birth-death process, the Gaussian is not good in the regime where the exact solution is far from Poissonian. If we consider the exact solution, we find that the term $\exp(-t\mu)$ defines the scale. Hence for about $t > 1/\mu$, Poissonian/Gaussian works similar way. For dimerization, this correction is not well understood. It is experimentally adjusted.

Chapter 3

Results

In this chapter, we analyze the convergence for linear birth-death and dimer-ization processes. Moreover, we also analyze the correctness, convergence and efficiency of corresponding preconditioning schemes.

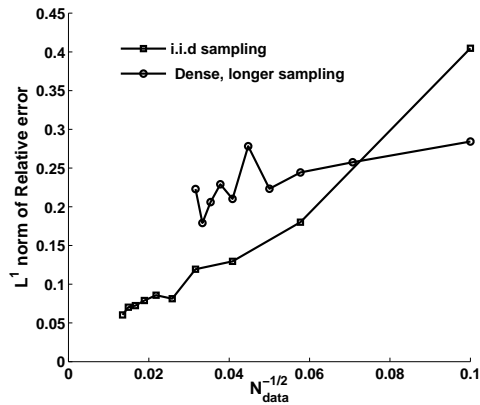## 3.1 Linear birth-death process

3.1.1 I.I.D convergence

Figure 3.1: L^1 norm Relative error in parameter (κ, μ) estimation for a linear birth-death process with exact CME under dense, longer sampling and iid sampling schemes of converegnce

3.1.2   Preconditioned MCMC

3.1.2.1   Correctness of preconditioning schemes
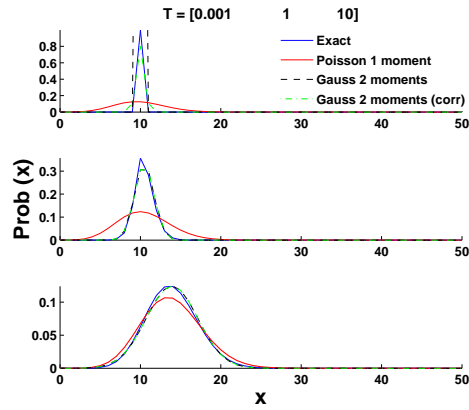
Figure 3.2: Correctness of the different preconditioning schemes (Poisson, Gaussian, Gaussian (with discrete correction) with respect to exact CME

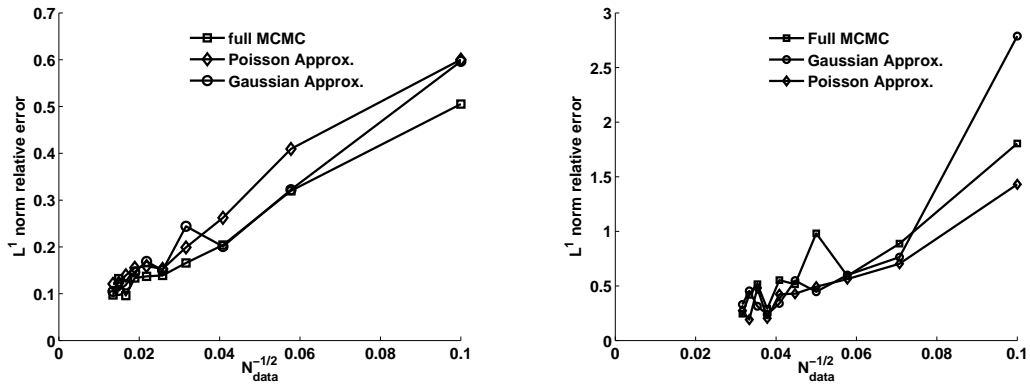#### 3.1.2.2   Convergence of preconditioned MCMC

Figure 3.3: Convergence of the different preconditioning schemes (Poisson, Gaussian (with discrete correction) with respect to exact CME for temporal granularity 1 (left) and 10 (right) respectively
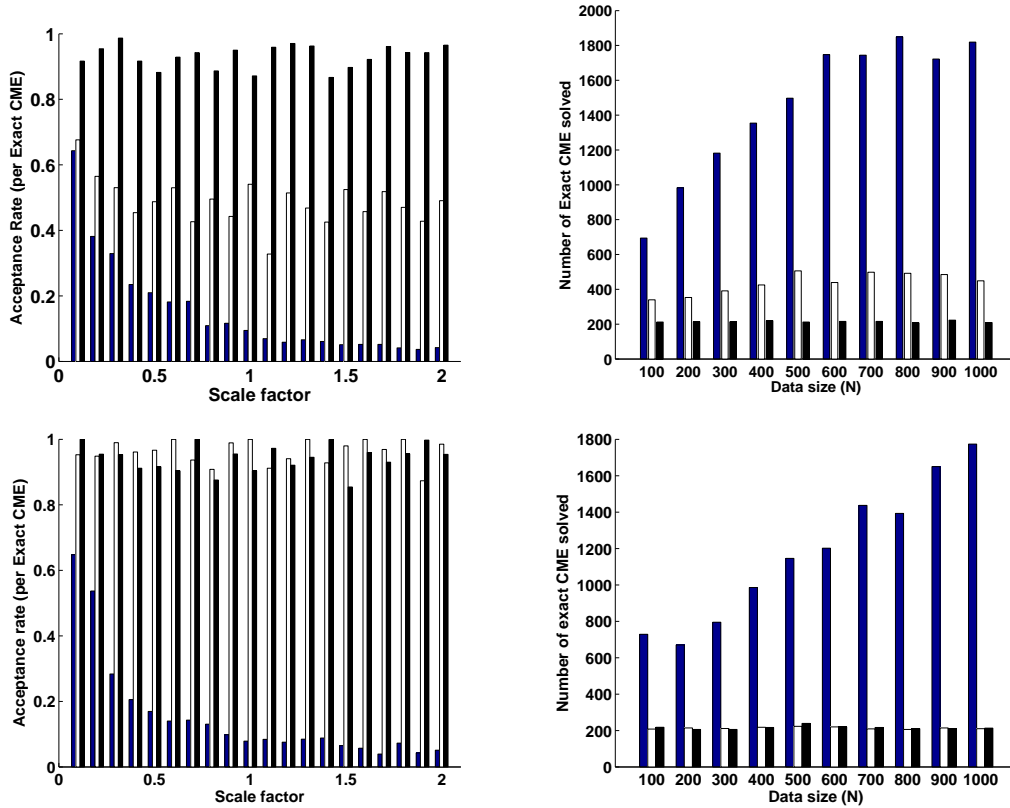
Figure 3.4: Acceptance rate (per exact CME) (left) and the number of exact CME solved for preconditioning and full MCMC (right). Upper left and right figures are for temporal granularity 1 and lower left and right are for granularity 10. Blue, white and black bars indicate the exact CME, the Poisson approximation and the Gaussian approximation for linear birth-death problem respectively

3.1.2.3 Effectiveness of preconditioned MCMC

Remark 1. *Figure 3.1 demonstrates bettter analysis of convergence for linear birth-death process, under iid sampling scheme. Figures 3.2 and 3.3 show granularity wise performace and similar iid convergence chracteristics (with respect to exact CME) for different precondition schemes respectively. Finally, figure 3.4 is quite promising in the sense that, with almost similar error characteristics (iid convergenece), we can save more than 50% of exact CME computations. Moreover, the preconditioning schemes are independent of scaling factor which shows promises to have a scale-free schemes for increasing the acceptance rate. The figures comparing the number of exact CME*

solved are generated by fixing the scale factor to 0.3 because the acceptance rate for exact CME is in [0.1, 0.4] at scale factor 0.3 for both the temporal granularities. Therefore at this scale factor, we can actually compare the efficiency of precondtioning. As the poisson approximation works better (almost similar to that of Gaussian) in coarse scale, the acceptance rate as well as the number of exact CME saved for temoral scale 10 is similar for both the approximations. On fine scale (upper left and right), Gaussian outperforms Poisson approximation. This is in support of the figure 3.2.

3.2 Dimerization process
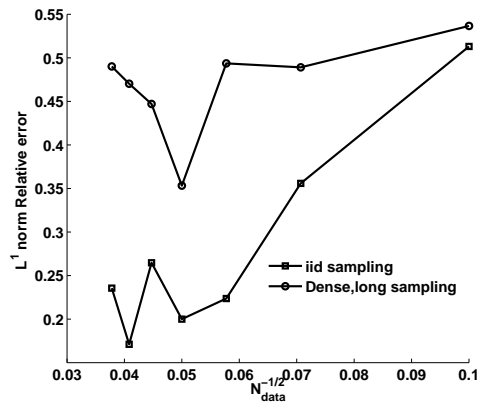
3.2.1 I.I.D convergence

Figure 3.5: L^1 norm Relative error in parameter (κ, ν) estimation for a dimerization process with exact CME under dense, longer sampling and iid sampling schemes of converegnce

3.2.2 Preconditioned MCMC

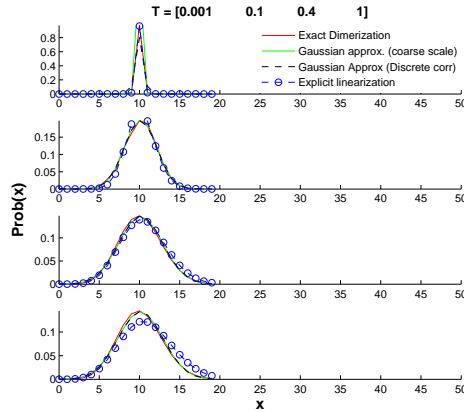### 3.2.2.1 Correctness of preconditioning schemes

Figure 3.6: Correctness of the different preconditioning schemes (Poisson, Gaussian, Gaussian (with discrete correction) with respect to exact CME
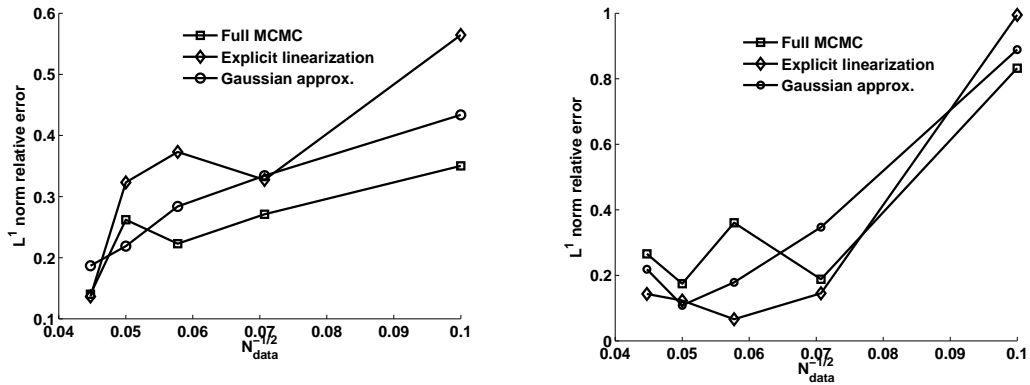
3.2.2.2 Convergence of preconditioned MCMC

Figure 3.7: I.I.D convergence characteristics of different preconditioning schemes (explicit linearization, Gaussian (with discrete correction) with respect to exact CME for granualrity level 0.1 (left) and 0.4 (right) respectively
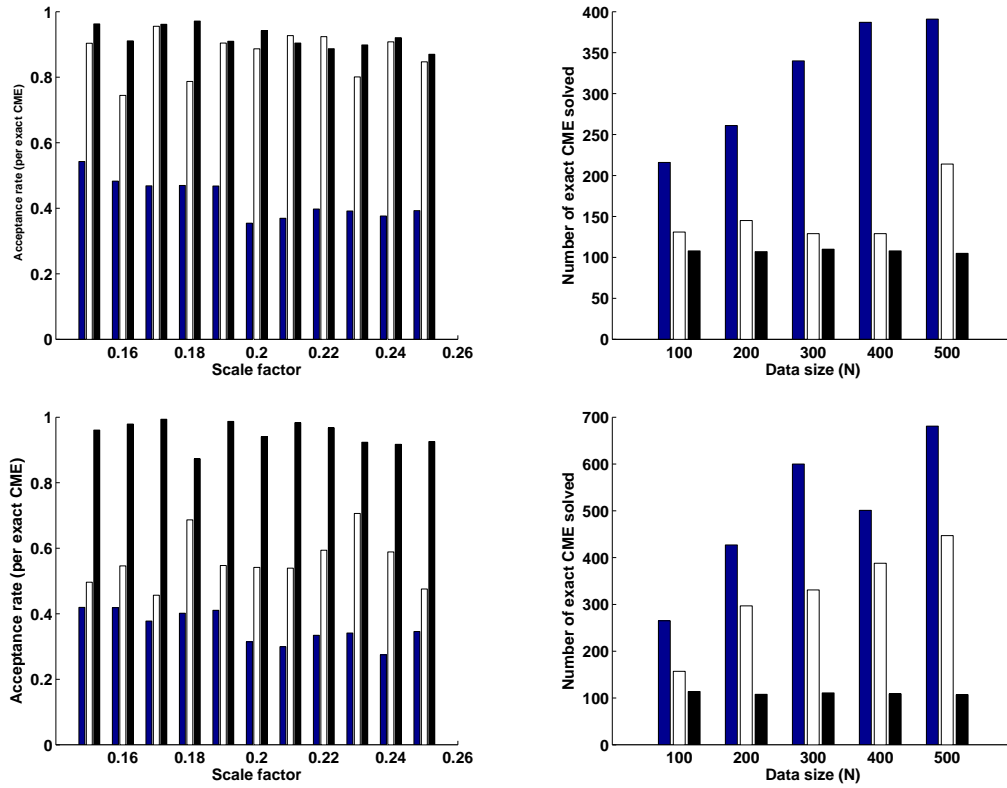
Figure 3.8: Acceptance rate per exact CME (left) and number of exact CME solved in preconditioned and full MCMC (right) at tempral scale 0.1 (upper left and right) and at 0.4 (lower left and right). Blue, white and black bars indicate the exact CME, explicit linearization and the Gaussian approximation (with correction) respectively

3.2.2.3 Effectiveness of preconditioned MCMC

Remark 2. *The experimentation for dimerization yields similar results to that of linear birth-death process. The explicit linearization precondtion scheme performs better in fine scale than corase scale as evident from figure 3.6. Both, explicit linearization and gaussian approximaton follow similar iid convergence characteristics as that of exact CME according to figure 3.7. Finally, figure 3.8 supports 3.7 as we can observe that in finer scale (0.1) explicit linearization and gaussian approximation (with correction) perform similar (although gaussian is slightly better) in terms of saving exact CME computations. In coarser scale (0.4), the performance of explicit liearization based precondtioned MCMC degrades, while that of Gaussian counterpart remains same. We fix the scale factor at 0.2 for both the temporal scale.*

Chapter 4

Discussions and Future directions

The goal of this work is to save time expensive exact CME computations until it is required. The preconditioning steps avoid expensive rejection through various approximations of CME and finally solve the exact CME (using exact SSA) only for the samples which can best approximate the target distribution. The results above shows promises about using preconditioned MCMC and CME for effective parameter inference at least for some simple fundamental processes for which we can construct CME in proper form. As future directions, we can address several issues such as: The above method can be extended to more complicated 2 dimensional problems according to some preliminary results obtained for the following example of bimolecular birth-death reaction [5]:

$$
\begin{aligned}
\phi &\xrightarrow{k_1} x \\
\phi &\xrightarrow{k_2} y \\
x &\xrightarrow{\mu x} \phi \\
y &\xrightarrow{\mu y} \phi \\
x + y &\xrightarrow{\nu xy} \phi
\end{aligned}
\tag{4.1}
$$

This system has 2 species and the correspnding CME will be 2 dimensional. Therefore, we have will have 6 ODEs for moments (2 for means and 4 for covariance. Using 2.32, we can derive the following set of ODEs assuming

$$k_1 = k_2 = k,$$

$$\frac{dm_1}{dt} = k - (C_{12}\nu)/2 - (C_{21}\nu)/2 - m_1\mu - \nu m_1 m_2$$

$$\frac{dm_2}{dt} = k - (C_{12}\nu)/2 - (C_{21}\nu)/2 - m_2\mu - \nu m_1 m_2$$

$$\frac{d\,C_{11}}{dt} = k + (C_{12}\nu)/2 + (C_{21}\nu)/2 + m_1\mu - 2C_{11}(\mu + \nu m_2) - \nu m_1(C_{12} + C_{21}) + \nu m_1 m_2$$

$$\frac{d\,C_{12}}{dt} = C_{12}(\nu/2 - 2\mu - \nu m_1 - \nu m_2) + C_{21}\nu/2 - \nu C_{11}m_2 - \nu C_{22}m_1 + \nu m_1 m_2$$

$$\frac{d\,C_{21}}{dt} = C_{21}(\nu/2 - 2\mu - \nu m_1 - \nu m_2) + C_{12}\nu/2 - \nu C_{11}m_1 - \nu C_{22}m_2 + \nu m_1 m_2$$

$$\frac{d\,C_{22}}{dt} = k + (C_{12}\nu)/2 + (C_{21}\nu)/2 + m_2\mu - 2C_{22}(\mu + \nu m_1) - \nu m_2(C_{12} + C_{21}) + \nu m_1 m_2$$

$$(4.2)$$

After, solving the above system of ODEs, we obtain the following Gaussian approximation of the exact CME. Therefore, figure 4.1 shows promises to
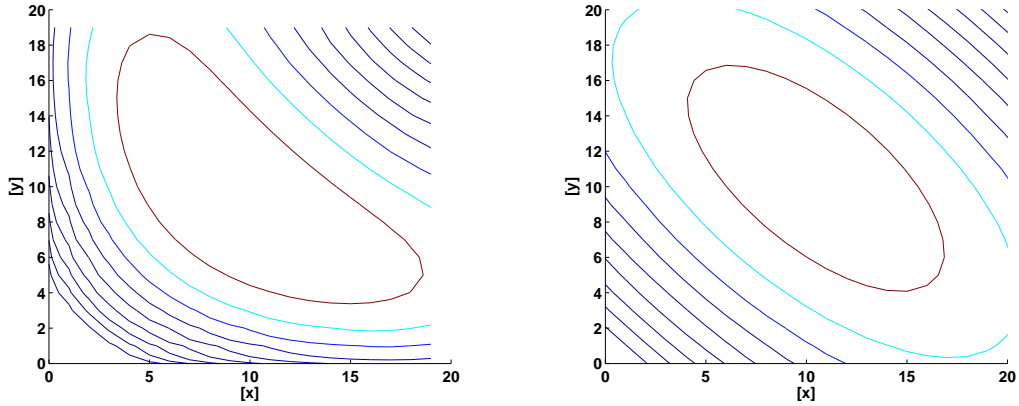

Figure 4.1: Countour plot of probability density of the Gaussian approximation (right) for the exact CME of Bimolecular birth death process (left)

have Gaussian approximation as an effective preconditioner for bimolecular birth-death process.

Generally we use species count based formulation of CME. In [24] the author proposes the formulation of CME based on reaction counts instead of species counts. This formulation is claimed to be effective approximation to species count based CME for certain processes. Our linear birth-death process is one of them. Although this is not a usual approach to CME, it can be interesting

to use this formulation as a preconditioner.

We have used the Next Reaction Method (NRM) by Gillespie to generate a trajectory which uses Realizations of Poisson processes that are consistent for each reaction channel. Now in the paper, [1] , the author propose an efficient modified next reaction method for exact simulation, extended to systems with time dependent propensities as well as system with delays. The idea is to incorporate this algorithm instead of our existing NRM to add more time efficiency especially for systems with delays.

As more complicated exammple of birth-death process we can consider a delayed one. In a very recent paper [12] It is shown mathematically that the difference between delayed birth death process and approximated delayed CLE descriptions converges to 0 with the increase of data. For CME based framework it has not been investigated yet.

Bibliography

[1] ANDERSON, D. F. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *The Journal of chemical physics 127*, 21 (2007), 214107.

[2] ARKIN, A., ROSS, J., AND MCADAMS, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected escherichia coli cells. *Genetics 149*, 4 (1998), 1633–1648.

[3] BLAKE, W. J., KÆRN, M., CANTOR, C. R., AND COLLINS, J. J. Noise in eukaryotic gene expression. *Nature 422*, 6932 (2003), 633–637.

[4] EFENDIEV, Y., HOU, T., AND LUO, W. Preconditioning markov chain monte carlo simulations using coarse-scale models. *SIAM Journal on Scientific Computing 28*, 2 (2006), 776–803.

[5] ENGBLOM, S. Computing the moments of high dimensional solutions of the master equation. *Applied Mathematics and Computation 180*, 2 (2006), 498–515.

[6] GARDINER, C. W., ET AL. *Handbook of stochastic methods*, vol. 3. Springer Berlin, 1985.

[7] GILLESPIE, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics 22*, 4 (1976), 403–434.

[8] GILLESPIE, D. T. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry 81*, 25 (1977), 2340–2361.

[9] GILLESPIE, D. T. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications 188*, 1 (1992), 404–425.

[10] GILLESPIE, D. T. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem. 58* (2007), 35–55.

[11] GOLIGHTLY, A., AND WILKINSON, D. J. Bayesian parameter inference for stochastic biochemical network models using particle markov chain monte carlo. *Interface Focus 1*, 6 (2011), 807–820.

[12] GUPTA, C., LÓPEZ, J. M., AZENCOTT, R., BENNETT, M. R., JOSIĆ, K., AND OTT, W. Modeling delay in genetic networks: From delay birth-death processes to delay stochastic differential equations. *The Journal of Chemical Physics 140*, 20 (2014), 204108.

[13] HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika 57*, 1 (1970), 97–109.

[14] HIGHAM, D. J., AND KHANIN, R. Chemical master versus chemical langevin for first-order reaction networks. *The Open Applied Mathematics Journal 2* (2008), 59–79.

[15] KEIZER, J. *Statistical thermodynamics of nonequilibrium processes.* Springer, 1987.

[16] MCADAMS, H. H., AND ARKIN, A. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences 94*, 3 (1997), 814–819.

[17] MCQUARRIE, D. A. Stochastic approach to chemical kinetics. *Journal of Applied Probability 4*, 3 (1967), 413–478.

[18] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H., AND TELLER, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics 21*, 6 (1953), 1087–1092.

[19] MILNER, P., GILLESPIE, C. S., AND WILKINSON, D. J. Moment closure based parameter inference of stochastic kinetic models. *Statistics and Computing 23*, 2 (2013), 287–295.

[20] POOVATHINGAL, S. K., AND GUNAWAN, R. Global parameter estimation methods for stochastic biochemical systems. *BMC bioinformatics 11*, 1 (2010), 414.

[21] RASER, J. M., AND O'SHEA, E. K. Control of stochasticity in eukaryotic gene expression. *Science 304*, 5678 (2004), 1811–1814.

[22] SAMOILOV, M., PLYASUNOV, S., AND ARKIN, A. P. Stochastic amplification and signaling in enzymatic futile cycles through noise-induced

bistability with oscillations. *Proceedings of the National Academy of Sciences of the United States of America 102*, 7 (2005), 2310–2315.

[23] SCHNEIDER, E. D., AND KAY, J. J. Life as a manifestation of the second law of thermodynamics. *Mathematical and computer modelling 19*, 6 (1994), 25–48.

[24] SUNKARA, V. The chemical master equation with respect to reaction counts. In *Proc. 18th World IMACS/MODSIM Congress* (2009).

[25] SUNKARA, V. *Analysis and Numerics of the Chemical Master Equation.* PhD thesis, Ph. D. thesis, Australian National University, 2013.

[26] TIAN, T., XU, S., GAO, J., AND BURRAGE, K. Simulated maximum likelihood method for estimating kinetic rates in gene expression. *Bioinformatics 23*, 1 (2007), 84–91.

[27] VAN KAMPEN, N. G. *Stochastic processes in physics and chemistry*, vol. 1. Elsevier, 1992.

[28] WEINBERGER, L. S., BURNETT, J. C., TOETTCHER, J. E., ARKIN, A. P., AND SCHAFFER, D. V. Stochastic gene expression in a lentiviral positive-feedback loop: Hiv-1 tat fluctuations drive phenotypic diversity. *Cell 122*, 2 (2005), 169–182.

[29] ZIMMER, C., AND SAHLE, S. Parameter estimation for stochastic models of biochemical reactions. *Journal of Computer Science and Systems Biology 6* (2012), 011–021.