Electronic Theses and Dissertations   Student Works

8-2015

Comparison of Two Parameter Estimation Techniques for Stochastic Models

Thomas C. Robacker
East Tennessee State University

Recommended Citation

Comparison of Two Parameter Estimation Techniques for Stochastic Models

————————

A thesis

presented to

the faculty of the Department of Mathematics and Statistics

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Mathematical Sciences

————————

by

Thomas Robacker

August 2015

————————

Michele Joyner, Ph.D.

Jeff Knisley, Ph.D.

Ariel Cintron-Arias, Ph.D.

Keywords: Parameter estimation, stochastic models, continuous-time Markov

chains, MCR method, ordinary least squares (OLS)

ABSTRACT

Comparison of Two Parameter Estimation Techniques for Stochastic Models

by

Thomas Robacker

Parameter estimation techniques have been successfully and extensively applied to deterministic models based on ordinary differential equations but are in early development for stochastic models. In this thesis, we first investigate using parameter estimation techniques for a deterministic model to approximate parameters in a corresponding stochastic model. The basis behind this approach lies in the Kurtz limit theorem which implies that for large populations, the realizations of the stochastic model converge to the deterministic model. We show for two example models that this approach often fails to estimate parameters well when the population size is small. We then develop a new method, the MCR method, which is unique to stochastic models and provides significantly better estimates and smaller confidence intervals for parameter values. Initial analysis of the new MCR method indicates that this method might be a viable method for parameter estimation for continuous-time Markov chain models.

DEDICATION

This work is dedicated to my family and my lovely girlfriend Caryn Brehm.

ACKNOWLEDGMENTS

I would first like to acknowledge my thesis advisor, Dr. Michele Joyner. She has been an outstanding mentor and person to collaborate with. Dr. Joyner introduced me to stochastic modeling and inverse problems which has become a life long research interest of mine. I would also like to acknowledge the entire Department of Mathematics and Statistics and thank my thesis committee. Additionally, thanks go to Dr. Chris Wallace in the Department of Computing for ensuring the Knightrider cluster works smoothly for numeric computation and Kimberly Brockman in the School of Graduate Studies for all her help in everything graduate school related.

TABLE OF CONTENTS

LIST OF TABLES

1 INTRODUCTION

Many natural phenomena are genuinely stochastic. For example, the spread of a disease or epidemic and the competition process between two species are stochastic in nature. In modeling these phenomena, often times a deterministic approach is taken. This is a good approximation when the system modeled involves a large population of individuals or objects. However, for situations in which there may be small population sizes, the deterministic model may be insufficient and possibly misleading. In such instances, the corresponding stochastic model may be a better approach.

Parameter estimation refers to the process of using data sampled from a process to estimate the parameters of a mathematical model of that process. This process is also known as the inverse problem. An inverse problem is a framework used to convert observed measurements into information about some system or model. An inverse problem is a transformation from data to model parameters via the interaction of the system we are studying. That is, it relates the model of the phenomena to actual observed data. This is contrary to the forward problem which is the transformation of the model and its parameters to data we observe. It is the inverse problem in the form of parameter estimation for stochastic models that hold our interest in this thesis.

The implementation of parameter estimation to stochastic models is in early development [6]. We investigate parameter estimation for such models using well-established methods for deterministic systems. In addition, to find a better method for handling small populations, we present a new method of parameter estimation unique to stochastic models called the MCR method.

2   EXAMPLE MODELS

In this chapter we introduce stochastic models formally and present the two example models we investigate for parameter estimation. We develop the deterministic along with the stochastic models and discuss the difference between the two types of models.

2.1   Stochastic Models

There are several types of stochastic models. The stochastic models we consider in this thesis are continuous-time Markov chain (CTMC) models. We first define what a stochastic process is and then what it means for a stochastic process to be a CTMC. Definition 2.1 is from L. Allen, 2011.

Definition 2.1 *A stochastic process is a collection of random variables $\{X_t(s) : t \in T, s \in S\}$, where T is some index set and S is the common sample space of the random variables. For each fixed t, $X_t(s)$ denotes a single random variable defined on S. For each fixed $s \in S$, $X_t(s)$ corresponds to a function defined on T that is called a sample path or a stochastic realization of the process.*

Definition 2.2 *The stochastic process $\{X(t) : t \in [0, \infty]\}$, is called a continuous-time Markov chain (CTMC) if it satisfies the following condition:*

$$\mathrm{Prob}\{X_{t+s} = j | X_s = i, X_u = x_u, 0 \leq u < s\} = \mathrm{Prob}\{X_{t+s} = j | X_s = i\}$$

for all $s, t \geq 0$, $i, j, x_u \in S$ and $0 \leq u < s$.

The latter condition is known as the Markov property. The transition at time $t+s$ to state j depends only on the value of the state at time t and does not depend on any other history of the process. This is also referred to as the memoryless property.

2.2   The SIS Model

When modeling the spread of a disease with a very long infectious period or a disease in a very large population, dynamic changes in the population itself cannot be ignored. In a large community the susceptible population might be augmented fast enough for the epidemic to be maintained for a long time without introducing new infectious individuals into the community. Such a disease is called *endemic* [3].

An alternative way of achieving endemicity is to retain the assumption of a closed population (N constant), but to suppose that the infected individuals lose their immunity after some time. This model, called the SIS epidemic model, will be the topic of this section. The SIS model has been applied to diseases such as influenza or the common cold as well as some sexually transmitted diseases [3].

2.2.1   The Deterministic SIS Model

This model is referred to as an SIS epidemic model because susceptible individuals (S) become infected (I) but do not develop immunity after they recover. They can immediately become infected again, $S \to I \to S$. Individuals that become infected are also infectious. That is, they can transmit the infection to others. We assume we have a closed homogeneously mixing population consisting of N individuals. The population remains constant for all time since the number of births equals the number
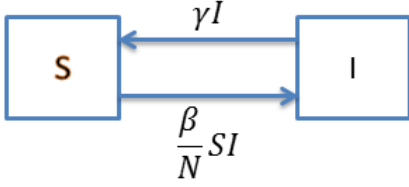
Figure 1: Compartmental diagram of the SIS epidemic model.

of deaths, $S + I = N$. The model has a compartmental diagram shown in Figure 1 which illustrates the transitions between the two states, S and I.

The differential equations, i.e. the deterministic model, for the SIS epidemic model are clearly

$$\frac{dS}{dt} = \gamma I - \frac{\beta}{N} SI \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta}{N} SI - \gamma I \tag{2}$$

with $S(0) + I(0) = N$. The parameter β is the transmission rate, the number of contacts per time that result in an infection of a susceptible individual. The parameter $1/\gamma$ is the average length of the infectious period.

Since individuals are either susceptible or infectious, it is sufficient to keep track of the number of individuals in the infectious state. This is clear since $S = N - I$. The explicit solution to this system is

$$I(t) = \frac{\left(1 - \frac{\gamma}{\beta}\right) I_0 e^{(\beta-\gamma)t}}{1 - \frac{\gamma}{\beta} + I_0 \left(e^{(\beta-\gamma)t} - 1\right)}.$$

A quantity of particular interest is the basic reproduction number $\mathcal{R}_0$. It determines the dynamics of the system. If the whole population is susceptible and one infected and infectious individual is introduced into the population, then $\mathcal{R}_0$ represents the

average number of successful contacts (β) during the period of infectivity ($1/\gamma$) that will result in a new infectious individual [1]. The basic reproduction number is given by $\mathcal{R}_0 = \frac{\beta}{\gamma}$. If $\mathcal{R}_0 \leq 1$ then $I(t) \to 0$ as $t \to \infty$. On the other hand, if $R_0 > 1$ then $I(t) \to 1 - \frac{\gamma}{\beta} > 0$ as $t \to \infty$.

2.2.2 The Stochastic SIS Model

In the stochastic SIS epidemic model, transitions no longer occur with certainty. Instead, the model deals with the probability of a transition during a small interval of time Δt. Let $I(t)$ denote the random variable for the number of infected individuals at time t. The state space for $I(t)$ is $\{0, 1, 2, \ldots, N\}$. The transition probabilities are

$$\text{Prob}\left\{\Delta I(t) = j | I(t) = i\right\} = \begin{cases} \frac{\beta}{N}i(N-i)\Delta t + o(\Delta t), & j = 1 \\ \gamma i \Delta t + o(\Delta t), & j = -1 \\ 1 - [\gamma i + \beta i(N-i)]\Delta t + o(\Delta t), & j = 0 \\ o(\Delta t), & j \neq 0, 1, -1 \end{cases}$$
$$(3)$$

where $\Delta I(t) = I(t + \Delta t) - I(t)$ and $i \in \{0, 1, \ldots, N\}$. Here, $o(\Delta t)$ means that $\lim_{\Delta t \to 0} o(\Delta t)/\Delta t = 0$ or $o(\Delta t)$ approches zero faster than Δt. This is also what is meant by a probability being neglible. For instance, in the SIS model the probability that there is a transition other than $j = 0, 1, -1$ is negligble in time Δt.

In Figure 2 we plot ten stochastic realizations of the SIS model with $N = 1250$ and $I_0 = 0.04N$ with parameters $\beta = 0.125$ and $\gamma = 0.1$. These are the parameters we will use throughout the thesis. The dashed curves show the corresponding deterministic solution. Notice that $\mathcal{R}_0 = \frac{\beta}{\gamma} = 1.25$ so that $\lim_{n \to \infty} I_n = N(1 - 1/\mathcal{R}_0) = 250$ for the deterministic solution as depicted in Figure 2.

14
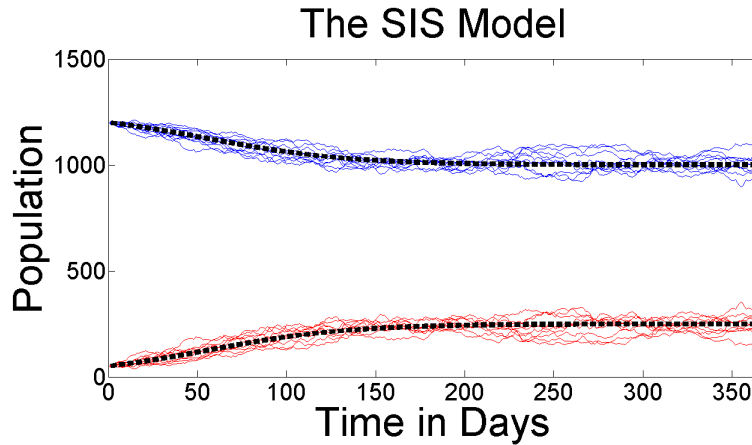
Figure 2: Ten stochastic realizations of the SIS model with $N = 1250$ and $I = 0.04N$ with parameters $\beta = 0.125$, $\gamma = 0.1$. The upper curves are the susceptible population and the red curves are the infected individuals. The black curves are the deterministic solutions.

2.3   The Lotka-Volterra Predator-Prey Model

The second model we will consider in this thesis is the Lotka-Volterra Predator-Prey model. The simplest model of predator and prey interaction includes only natural growth or decay and the predator-prey interaction. The deterministic model can be developed from first principles as in [5] and many other elementary texts on differential equations. We will summarize the deterministic model as in [5] and [1].

2.3.1   The Deterministic Lotka-Volterra Predator-Prey Model

Let $x(t)$ and $y(t)$ denote the population sizes for the prey and predator at time t, respectively. The deterministic Lotka-Volterra predator-prey model is the system of

ODEs

$$\begin{aligned}
\frac{dx}{dt} &= x\left(a_{10} - \frac{a_{12}}{N}y\right) \\
\frac{dy}{dt} &= y\left(\frac{a_{21}}{N}x - a_{20}\right)
\end{aligned}$$
(4)

where the parameters $a_{ij} > 0$, $x(0) > 0$, $y(0) > 0$. The parameter a_{10} represents the combination of the natural birth and death rate of the prey. The parameter a_{12} represents a death rate in the prey due to interaction with predators, and a_{21} represents a birth rate for the predator due to the same interaction with the prey. Finally, the parameter a_{20} represents the combination of the natural birth and death rate of the predator.

2.3.2 The Stochastic Predator-Prey Model

Now we develop the stochastic model for the predator-prey process as in [1]. Let $X(t)$ and $Y(t)$ denote random variables for the population sizes of the prey and predator at time t, respectively. The transition probabilities are

$$\begin{aligned}
&\text{Prob}\left\{\Delta X(t) = i, \Delta Y(t) = j | X(t) = x, Y(t) = y\right\} \\
&= \begin{cases}
a_{10}x\Delta t + o(\Delta t), & (i,j) = (1,0) \\
\frac{a_{12}}{N}xy\Delta t + o(\Delta t), & (i,j) = (0,1) \\
\frac{a_{21}}{N}xy\Delta t + o(\Delta t), & (i,j) = (-1,0) \\
a_{20}y\Delta t + o(\Delta t), & (i,j) = (0,-1) \\
1 - x[a_{10} + a_{21}y]\Delta t & \\
\quad -y[a_{20} + a_{21}x]\Delta t + o(\Delta t), & (i,j) = (0,0) \\
o(\Delta t), & \text{otherwise.}
\end{cases}
\end{aligned}$$
(5)

where $\Delta X(t) = X(t + \Delta t) - X(t)$ and $\Delta Y(t) = Y(t + \Delta t) - Y(t)$.

In Figure 3 we plot ten stochastic realizations of the Predator-Prey model and the deterministic system with $N = 60$, $X(0) = 0.75N$ with parameters $a_{10} = 0.50$, $a_{12} =$

16

0.05, $a_{12} = 0.01$, $a_{20} = 0.20$. These are the parameters we will use throughout the thesis.
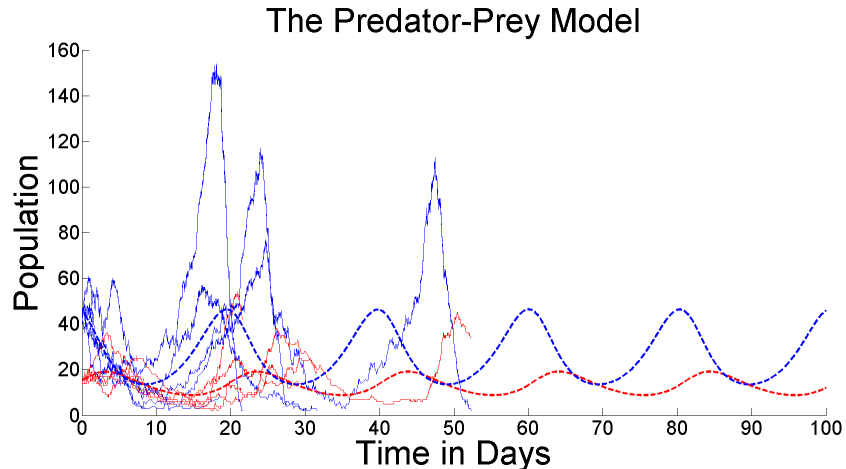
Figure 3: Ten realizations of the Predator-Prey model and its deterministic solution where $N = 60$, $X(0) = 0.75N$ with parameters $a_{10} = 0.50$, $a_{12} = 0.05$, $a_{12} = 0.01$, $a_{20} = 0.20$. The blue curves represent the prey population and the red curves represent the predator population.

2.4 The Gillespie Algorithm

The solutions to the deterministic systems above occur with absolute certainty. The stochastic models, however, always have unique outcomes for any given realization. There are several methods one can use to generate a realization of a stochastic model. In this thesis we implement the standard Gillespie algorithm which is also known as the stochastic simulation algorithm (SSA) [2]. This is the standard algorithm used to simulate CTMC models [4].

The Gillespie algorithm can be summarized as follows:

17

Step 1: Set the initial condition(s) for each state at $t = 0$.

Step 2: For the given state $\boldsymbol{x}$ of the system, calculate the sum of all transition rates, $\lambda_{\boldsymbol{x}} = \sum_{i=1}^{m} \lambda_i(\boldsymbol{x})$ where $i = 1, 2, \ldots, m$ and m represents the total number of transitions in the given model.

Step 3: Draw Δt from an exponential distribution with parameter $\lambda_{\boldsymbol{x}}$.

Step 4: Generate a random number r from a uniform distribution on $(0, 1)$ and choose the transition as follows: If $0 < r \leq \lambda_1(\boldsymbol{x})/\lambda_{\boldsymbol{x}}$, choose transition 1; if $\lambda_1(\boldsymbol{x})/\lambda_{\boldsymbol{x}}$

$< r \leq (\lambda_1(\boldsymbol{x}) + \lambda_2(\boldsymbol{x}))/\lambda_{\boldsymbol{x}}$ choose transition 2, and so on.

Step 5: Let transition η be the transition chosen in **Step 4**. Update the time by setting $t = t + \Delta t$ and update the system state based on the transition η.

Step 6: Iterate **Step 2** through **Step 5** until $t \geq t_{stop}$.

In the next section, we use the Gillespie algorithm to generate several realizations of the example models.

2.5   Data Sets for Parameter Estimation

Here we present several stochastic realizations, or data sets, we will initially use to compare our parameter estimation techniques. Throughout the thesis, as explained earlier, we will use the parameters $\beta = 0.125$ and $\gamma = 0.1$ for the SIS model. We consider populations of size 125, 1250, and 12500. For $N = 125$ and $N = 1250$, we will look at three different synthetic data sets and one data set for $N = 12500$ for

illustrative purposes. The proportion of initial infective to susceptible individuals will remain the same for each population size at $I_0 = 0.04N$. Figure 4 plots the three data sets for the SIS model with size $N = 125$. Figure 5 plots the three data sets for the SIS model with size $N = 1250$. Figure 6 plots the data set for the SIS model with size $N = 12500$.

Throughout the thesis, we will use the parameters $a_{10} = 0.50$, $a_{12} = 0.05$, $a_{21} = 0.01$, $a_{20} = 0.20$ for the Predator-Prey model. We consider populations of size 60 and 600. For $N = 60$ we will examine three different synthetic data sets and two data sets for $N = 600$. The proportion of initial predators will remain constant for each population size at $Y_0 = 0.25N$. Figure 7 plots the three data sets for the Predator-Prey model with size $N = 60$. Figure 8 plots the two data sets for the Predator-Prey model with size $N = 600$.

In order to ensure we are able to replicate our results in Matlab and keep track of our data, we set the seed for the pseudo-random number generator. The seed number simply allows one to repeat arrays of random numbers. This gives a convenient labeling for the synthetic data sets with seed numbers.
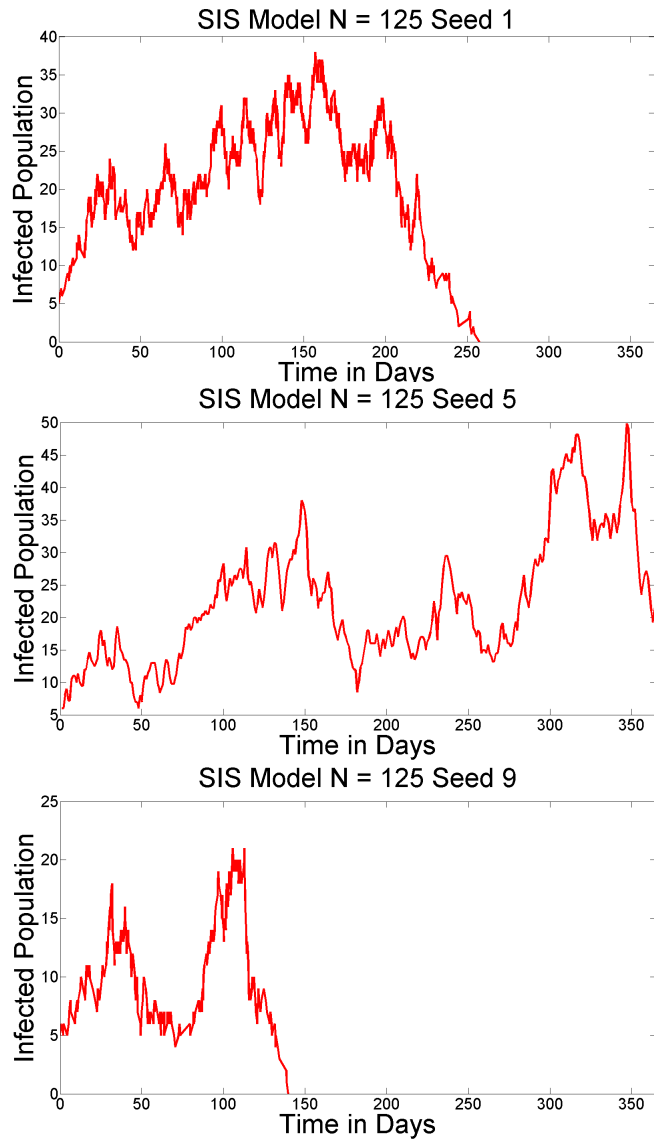
Figure 4: Three data sets for the SIS model labeled Seed 1, Seed 5, and Seed 9 for population size $N = 125$.
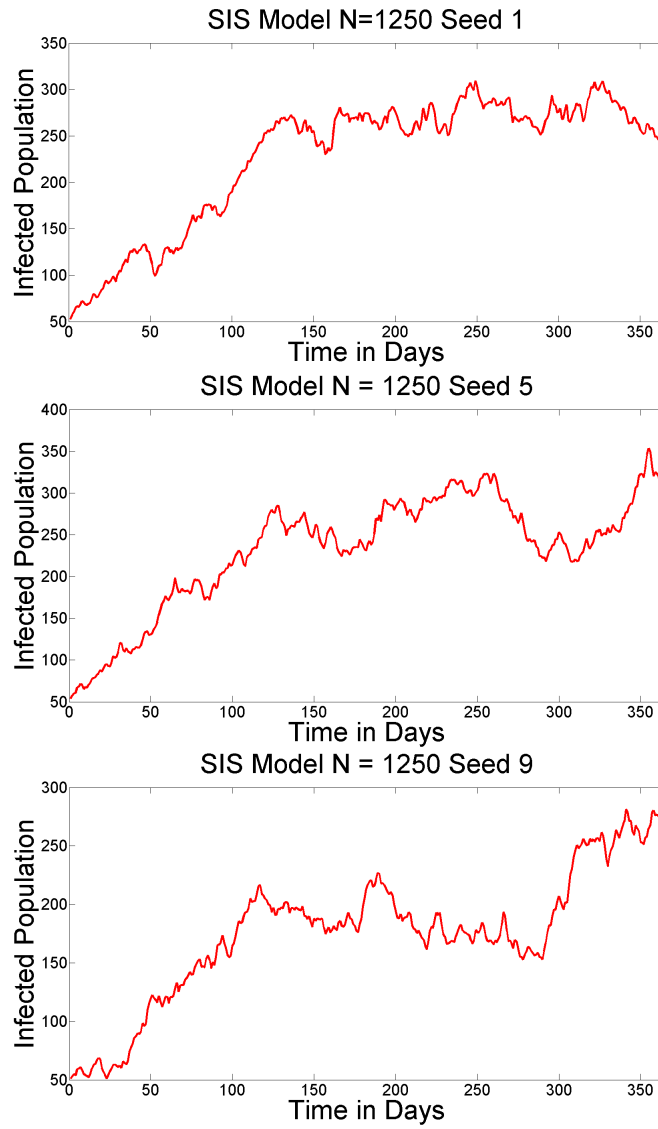
Figure 5: Three data sets for the SIS model labeled Seed 1, Seed 5, and Seed 9 for population size $N = 1250$.
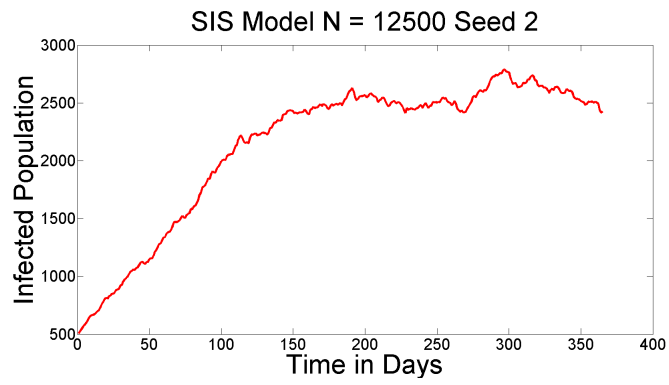
Figure 6: A data set for the SIS model labeled Seed 1 for population size $N = 12500$.
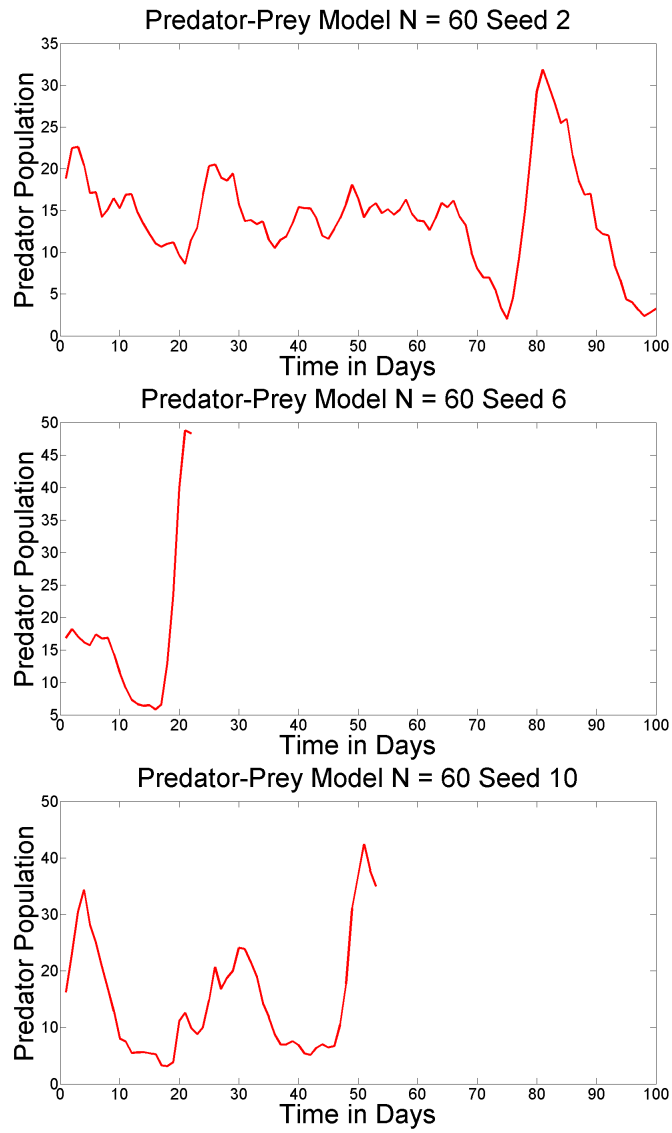
Figure 7: Three data sets for the Predator-Prey model labeled Seed 2, Seed 6, and Seed 10 for population size $N = 60$.
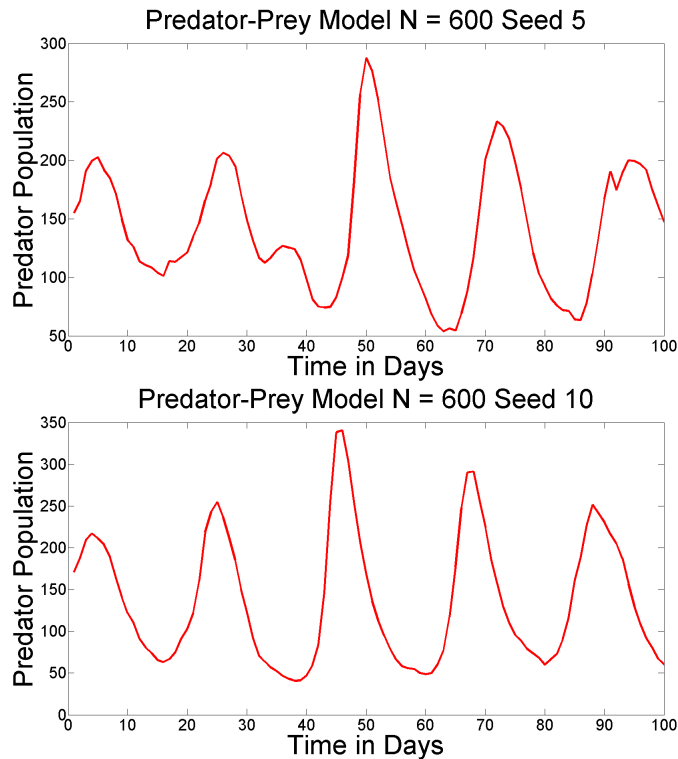
Figure 8: Two data sets for the Predator-Prey model labeled Seed 5 and Seed 10 for population size $N = 600$.

There are several observations to make regarding the characteristics of the data sets in Figures 4 through Figure 8. First, all the data sets are generated from the exact same set of parameters for their respective model equations, either Eq. (3) for the SIS model or Eq. (5) for the Predator-Prey model. The large variability in each population size is evident. Consider the small SIS model with $N = 125$. Each realization is significantly different from the other. Additionally, in a stochastic realization the population has a non-zero probability of going to zero or vanishing. In the case of the small SIS model with $N = 125$ this happens for both data sets Seed 1 and Seed 9. This is not observed in the deterministic solution which persists

indefinitely. This is an advantage of the stochastic model over the deterministic model in modeling realistic scenarios. In the next section we proceed to the goal of estimating parameters.

3 ESTIMATING PARAMETERS IN STOCHASTIC MODELS USING ORDINARY DIFFERENTIAL EQUATIONS

In this chapter we present the parameter estimation problem for using techniques common to deterministic systems. In particular, we develop the ordinary least squares (OLS) method. We present confidence intervals for the parameters of interest and the average computational time required for the method. We will see that there is a significant drawback of using deterministic methods when the population size is small. First, we present a powerful result from [4] which justifies using ODEs for parameter estimation of stochastic models in a special case.

Theorem 3.1 *(Kurtz Limit Theorem) Let $\boldsymbol{X}(t)$ be a continuous-time Markov chain. Suppose that $\lim_{M \to \infty} \boldsymbol{X}^M(0) = \boldsymbol{x}_0$ and for any compact set $\Omega \in \mathbb{R}^n$ there exists a positive constant η_Ω such that*

$$|\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{g}(\hat{\boldsymbol{x}})| \leq \eta_\Omega |\boldsymbol{x} - \hat{\boldsymbol{x}}|,$$

for $\boldsymbol{x}, \hat{\boldsymbol{x}} \in \Omega$. Then we have

$$\lim_{M \to \infty} \sup_{t \leq t_f} |\boldsymbol{X}^M(t) - \boldsymbol{x}(t)| = 0 \tag{6}$$

almost surely for all $t_f > 0$, where $\boldsymbol{x}$ denotes the unique solution to the ODE

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{g}(\boldsymbol{x}), \quad \boldsymbol{x}(0) = \boldsymbol{x}_0.$$

The parameter M can be interpreted as the total number of individuals in the population, even if it is dynamic [4] and the function $\boldsymbol{g}$ represents the right hand side of

the ODE model; for instance, $\boldsymbol{g}$ may represent the right hand side of the SIS model in equation (2). The Kurtz Limit Theorem justifies the use of ODEs for modeling stochastic effects when the population size is sufficiently large. Specifically, equation (6) implies that as the population M tends to infinity the difference between the CTMC model and the corresponding deterministic solution approaches zero. This gives justification for approximating stochastic models as deterministic systems when M is large.

3.1   ODE Estimation Techniques

Our interest is in estimating the parameters for any particular stochastic realization, or data set, from our two example models, the SIS epidemic and Lotka-Volterra Predator-Prey models. We consider the parameter estimation problem, and proceed as in [4], in the context of a parameterized dynamical system

$$\frac{d\boldsymbol{x}(t)}{dt} \;=\; \boldsymbol{g}(t, \boldsymbol{x}(t), \boldsymbol{\theta}), \tag{7}$$

$$\boldsymbol{x}(t_0) \;=\; \boldsymbol{x}_0, \tag{8}$$

where $\boldsymbol{g}$ is a function giving us the right hand side of the deterministic ODEs, $\boldsymbol{x}$ is the state vector, and $\boldsymbol{\theta}$ the vector of parameters. For example, in the SIS model, the state vector is $[S\ I]^T$ and the parameter vector is $[\beta\ \gamma]^T$.

One statistical model for the observation process is of the form

$$\boldsymbol{X}_j = \boldsymbol{f}(t_j; \boldsymbol{\theta}_0) + \boldsymbol{\mathcal{E}}_j, \quad j = 1, \ldots, n, \tag{9}$$

where $\boldsymbol{\mathcal{E}}_j$ is assumed to be normally distributed with unknown variance. This is the familiar ordinary least squares formulation. In words, the observation process is the

assumption that our observed data, the stochastic realization in our case, is the model output, $\boldsymbol{f}(t_j; \boldsymbol{\theta}_0)$, plus some measurement error, $\boldsymbol{\mathcal{E}}_j$. For the statistical model given by Eq. (9), we define the vector of optimal parameter values as

$$\boldsymbol{\theta}_{OLS} = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \tag{10}$$

where

$$J(\boldsymbol{\theta}) = \sum_{j=1}^{N} [\boldsymbol{X_j} - \boldsymbol{f}(t_j; \boldsymbol{\theta})]^2 \tag{11}$$

denotes the cost function. Then $\boldsymbol{\theta}_{OLS}$ can be viewed as minimizing the distance between the data and model. We note that $\boldsymbol{\theta}_{OLS}$ is a random vector since $\boldsymbol{\mathcal{E}}_j$ is a random variable. Hence if $\{x_j\}_{j=1}^{N}$ are realizations of the random variables $\{X_j\}_{j=1}^{N}$, then solving

$$\hat{\boldsymbol{\theta}}_{OLS} = \arg\min_{\boldsymbol{\theta}} \sum_{j=1}^{N} [\boldsymbol{x_j} - \boldsymbol{f}(t_j; \boldsymbol{\theta})]^2 \tag{12}$$

provides a realization for $\boldsymbol{\theta}_{OLS}$. Throughout the thesis we will often drop the subscript OLS for the estimates when the context is clear and simply use $\hat{\boldsymbol{\theta}}$. Ordinary least squares is a commonly used method for parameter estimation in deterministic systems. The method is but one statistical observation model which can be generalized as in [4].

Now we summarize the algorithm for parameter estimation using ordinary differential equations. We estimate parameters for the stochastic models given by Eq. (3) and Eq. (5) by assuming the stochastic model can be estimated using the deterministic models given by Eq. (2) and Eq. (4) for the SIS and Predator-Prey models respectively. This is justified by Kurtz Limit Theorem. That is, in Eq. (12), we assume

$\boldsymbol{f}$ is the output from a deterministic model even though the data, $\boldsymbol{x}_j$, is stochastic in nature. We use the synthetic data described in Section 2.5. For each data set, with a given seed and population size, we generate an initial guess for the parameter estimation, $\boldsymbol{\theta}_0$, normally distributed about the true parameter value $\boldsymbol{\theta}_t$. For test purposes, we consider a variance of 0.01 about $\boldsymbol{\theta}_t$. We then use the built-in Matlab function `fminsearch` to minimize the cost function given by Eq. (11) to estimate the optimal parameter values $\hat{\boldsymbol{\theta}}$. The function `fminsearch` uses the Nelder-Mead simplex algorithm. We note that there are other minimization algorithms which can be employed, but we do not explore those algorithms in this thesis. In the next section we show the results for each of the example models.

3.2   Parameter Estimation for Several Data Sets

This section implements the algorithm for parameter estimation using ordinary differential equations for the data sets in Figures 4 through Figure 8. Tables 1 through Table 3 give the results of the parameter estimation for the very large, large, and small SIS models with $N = 12500$, $N = 1250$, and $N = 125$, respectively.

Table 1: Estimated parameter values for the very large $N = 12500$ SIS model.

	β			γ		
Data Set	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 2	0.125	0.125	.01 %	0.1	0.0991	0.86 %

Table 2: Estimated parameter values for the large $N = 1250$ SIS model.

Data Set	β			γ		
	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 1	0.125	0.1150	8.02 %	0.1	0.0894	10.59 %
Seed 5	0.125	0.1376	10.07 %	0.1	0.1074	7.42 %
Seed 9	0.125	0.1523	21.82 %	0.1	0.1269	26.91 %

Table 3: Estimated parameter values for the small $N = 125$ SIS model.

Data Set	β			γ		
	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 1	0.125	2.2483	1698.64 %	0.1	1.9894	1889.70 %
Seed 5	0.125	0.1422	13.76 %	0.1	0.1127	12.70 %
Seed 9	0.125	2.8421e-16	100.00 %	0.1	0.0038	96.20 %

From these results we can see that the OLS method worked very well for the very large population with $N = 12500$, mediocore for the large model with $N = 1250$, and unacceptably poor for two cases regarding the small model with $N = 125$. The results can be understood intuitively by re-examining Figures 4 through 6. The worst estimates result from the realizations whose infected populations vanish early. That is, the small data sets Seed 1 and Seed 9 whereas the realizations that persist through 365 days have better estimates. As we would expect from Kurtz Limit Theorem, the very large population with $N = 12500$ provides an accurate estimate.

Table 4 and Table 5 give the results of the parameter estimation for the small and large Predator-Prey models with $N = 600$ and $N = 60$, respectively.

Table 4: Estimated parameter values for the large $N = 600$ Predator-Prey model.

	a_{10}			a_{12}		
Data Set	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 5	0.50	0.4252	14.96%	0.05	0.0414	17.18 %
Seed 10	0.50	0.6167	23.34%	0.05	0.0629	25.81 %

	a_{21}			a_{20}		
Data Set	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 5	0.01	0.0112	11.53%	0.20	0.1954	2.31 %
Seed 10	0.01	0.0164	63.76%	0.20	.1789	10.66 %

Table 5: Estimated parameter values for the small $N = 60$ Predator-Prey model.

	a_{10}			a_{12}		
Data Set	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 2	0.50	0.6190	23.8%	0.05	0.0584	16.80 %
Seed 6	0.50	2.0583	311.66%	0.05	0.1687	237.40 %
Seed 10	0.50	.6819	36.38%	0.05	0.0494	1.20 %

	a_{21}			a_{20}		
Data Set	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 2	0.01	0.0081	19.00%	0.20	0.1676	16.20 %
Seed 6	0.01	0.027	170.00%	0.20	0.1229	38.55 %
Seed 10	0.01	.0144	44.00%	0.20	0.1510	24.50 %

The results for the Predator-Prey model are similar to the SIS model. The larger population estimates tend to be more accurate estimates of the original parameters. However, in some cases the small population has nearly as small an error as the large population. Again, this can be intuitively understood by re-examining Figure 5 and Figure 4. In the next section we test the parameter estimation more rigorously and construct confidence intervals for parameter values.

3.3 Confidence Intervals for Deterministic Estimation

This section tests the algorithm for parameter estimation using ordinary differential equations rigorously by constructing 95% confidence intervals for parameter estimates for 1000 sample data sets. The following algorithm summarizes the construction of the confidence intervals:

Step 1: Initialize the seed value to $k = 1$.

Step 2: Generate sample data for the stochastic model using seed value k and exact parameter values.

Step 3: Generate a normally distributed initial guess, $\boldsymbol{\theta}_0$, roughly approximating the true parameter value, $\boldsymbol{\theta}_t$.

Step 4: Solve for the best estimate $\hat{\boldsymbol{\theta}}_k$ using the cost function in Eq. (11) where $\boldsymbol{x}_j$ is the data generated in **Step 2**.

Step 5: Update the seed value $k = k + 1$ and repeat **Steps 2-4** until $k = M$ (We use M=1000 in this thesis).

Step 6: Construct the confidence intervals using the mean and standard deviation of the values $\boldsymbol{\theta}_k, \ k = 1, \ldots, M$ as defined below.

In constructing the confidence intervals the Central Limit Theorem implies that they may be calculated by

$$\hat{\boldsymbol{\theta}} \pm z^* \frac{\boldsymbol{s}}{\sqrt{N}} \tag{13}$$

where $\hat{\boldsymbol{\theta}}$ is the mean of the $\hat{\boldsymbol{\theta}}_k$, z^* is the critical value, $\boldsymbol{s}$ is the vector sample standard deviation, and N is the size of the data. This holds for large samples; in particular, for $N = 365$ in the case of the SIS model we can employ this formulation. For the Predator-Prey model we have at most $N = 100$ data points and must use the Student's t-distribution to construct confidence intervals as

$$\hat{\boldsymbol{\theta}} \pm t \frac{\boldsymbol{s}}{\sqrt{N}} \tag{14}$$

where t is the critical value from the t-distribution with degrees of freedom $N - 1$.

We now present the confidence intervals for each example model. Table 6 through Table 8 shows the confidence intervals for the SIS model with populations $N = 12500$, $N = 1250$ and $N = 125$, respectively. For each interval we calculate the maximum relative error, which is the interval endpoint which yields the maximum relative error for the parameter. The average time taken to estimate the parameters for a single realization are 22.29 seconds, 13.36 seconds, and 7.70 seconds for $N = 12500$, $N = 1250$ and $N = 125$, respectively.

Table 6: Confidence intervals for the $N = 12500$ SIS model.

Parameter	True Value	Confidence Interval	Max Rel. Error
β	0.125	(0.1241, 0.1272)	1.90 %
γ	0.1	(0.0993, 0.1019)	1.76 %

Table 7: Confidence intervals for the $N = 1250$ SIS model.

Parameter	True Value	Confidence Interval	Max Rel. Error
β	0.125	(0.1255, 0.1368)	9.44 %
γ	0.1	(0.1008, 0.1101)	10.10 %

Table 8: Confidence intervals for the $N = 125$ SIS model.

Parameter	True Value	Confidence Interval	Max Rel. Error
β	0.125	(0.3875, 1.0049)	703.92 %
γ	0.1	(0.3782, 0.9834)	883.40 %

As with the specific data sets we examined we see that the estimation in general is robust for a very large population $N = 12500$. We are 95% confident that the estimated parameters β and γ will be at most 1.90% and 1.76%, respectively, away from the true parameters in relative error. The large population also faired well. The small population in general, however, produces unacceptable estimates 95% of the time. Clearly, the deterministic method for parameter estimation failed for the small population and succeeded for the large populations which we expect by Kurtz Limit Theorem.

We present the confidence intervals for the Predator-Prey model with $N = 600$ and $N = 60$ in Table 9 and Table 10. The average time for parameter estimation for populations $N = 600$ and $N = 60$ are 99.17 seconds and 91.75 seconds, respectively.

Table 9: Confidence intervals for the $N = 600$ Predator-Prey model.

Parameter	True Value	Confidence Interval	Max Rel. Error
a_{10}	0.50	(0.5010, 0.5359)	7.18 %
a_{12}	0.05	(0.0506, 0.0546)	9.20 %
a_{21}	0.01	(0.0106, 0.0106)	6.00 %
a_{20}	0.20	(0.2003, 0.2163)	8.15 %

Table 10: Confidence intervals for the $N = 60$ Predator-Prey model.

Parameter	True Value	Confidence Interval	Max Rel. Error
a_{10}	0.50	(0.4642, 0.5817)	16.34 %
a_{12}	0.05	(0.0348, 0.0438)	30.40 %
a_{21}	0.01	(1988.7811, 12155.4461)	12.15e7 %
a_{20}	0.20	(1491.9326, 9116.9226)	45.585e5 %

Clearly, there is a major difference in the estimation for the two Predator-Prey models. As with the SIS model, the Predator-Prey model also has unacceptable estimates for the small population but strong estimates for the larger population.

We would like to be able to estimate parameters for small populations. We now know that this is virtually impossible given our confidence intervals: we are 95% confident that the parameter estimates will be unacceptable. In the next chapter we introduce a new method unique to stochastic models that allows us to estimate parameters for both the large and, of more interest, the small populations that we have in this chapter.

4 PARAMETER ESTIMATION USING THE MCR METHOD

This chapter develops a new method of estimating parameters unique to stochastic models. It provides significantly better estimates and smaller confidence intervals for parameter values. Although Kurtz Limit Theorem allows us to approximate a stochastic model with a corresponding deterministic model when the population is sufficiently large, there is a practical motivation for desiring a method for small populations. For instance, the SIS model may be a suitable model for the spread of a disease in an intensive care unit within a hospital setting. In such a scenario the population of individuals involved is likely small, and we would be interested in modeling the epidemic.

4.1   The MCR Method

The MCR method is an acronym for Minimum-Cost-Realization. The name comes from the algorithm it utilizes. One of the main differences between the MCR method and the deterministic method can be seen in the cost function $J(\boldsymbol{\theta})$. For the deterministic methods above we used

$$J(\boldsymbol{\theta}) = \sum_{j=1}^{N} \left[\boldsymbol{x_j} - \boldsymbol{f}(t_j; \boldsymbol{\theta}) \right]^2 .$$

where $\boldsymbol{f}(t_j; \boldsymbol{\theta})$ represents the deterministic model output with parameters $\boldsymbol{\theta}$. The cost function for the MCR method is defined as

$$J^{MCR}(\boldsymbol{\theta}) = \min_{n \in 1,2,\ldots,n_s} J_n(\boldsymbol{\theta}). \tag{15}$$

where

$$J_n(\boldsymbol{\theta}) = \sum_{j=1}^{n} [\boldsymbol{x_j} - \boldsymbol{h}(t_j; \boldsymbol{\theta})]^2 \tag{16}$$

with $n = 1, 2, \ldots, n_s$ where $\boldsymbol{h}(t_j; \boldsymbol{\theta})$ represents the stochastic model output in lieu of the deterministic model and n_s denotes the number of realizations chosen for comparison. This is analagous to the cost function $J(\boldsymbol{\theta})$ we used for parameter estimation with ODEs. The motivation behind this method lies in the fact that realizations of a stochastic model can be significantly different. Therefore, for a given parameter estimate, $\boldsymbol{\theta}$, we try to determine the realization that "best fits" the data in a least squares sense. Using this "best fit" for a given parameter estimate, we seek the optimal parameter values. Eq. (15) can be thought of as choosing which one of the n_s realizations is a best fit to the data set in terms of $J_n(\boldsymbol{\theta})$. We henceforth drop the superscript MCR when the context is clear. We now summarize the algorithm for implementing the MCR method:

Step 1: Generate a normally distributed guess, $\boldsymbol{\theta}_0$ about $\boldsymbol{\theta}_t$.

Step 2: Use `fminsearch` to estimate

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}_0).$$

where $J(\boldsymbol{\theta})$ is given by Eq. (15). In order to calculate $J(\boldsymbol{\theta}_0)$, the following steps are implemented:

Step 2.1: Generate n_s data sets from the parameters $\boldsymbol{\theta}_0$. Bin the n_s data sets to match the size of $\boldsymbol{x}$.

37

Step 2.2: Calculate $J_n(\boldsymbol{\theta}_0)$ for $n = 1, 2, \ldots, n_s$:

$$J_n(\boldsymbol{\theta}_0) = \sum_{j=1}^{N} [\boldsymbol{x_j} - \boldsymbol{h}(t_j; \boldsymbol{\theta}_0)]^2$$

Step 2.3: Set cost function as

$$J(\boldsymbol{\theta}_0) = \min_{n \in 1, 2, \ldots, n_s} J_n(\boldsymbol{\theta}_0).$$

To illustrate the MCR method we implement the above algorithm with $n_s = 10$ for data set Seed 1 in Figure 3. The algorithm generates $n_s = 10$ stochastic realizations using parameters $\boldsymbol{\theta}_0$; these are the black and red curves in Figure 9 where we omitted a few curves for clarity. The red curve represents the realization that is the best fit to the data. Examining the figure gives some intuition to the algorithm: the best fit realization here appears to be very similar to the data. The estimated parameter values are $\beta = 0.1212$ and $\gamma = 0.1064$ with relative error 3.2% and 6.4%, respectively. The algorithm took 7.97 seconds. This is a disparity over the deterministic method in the previous section from a small sample which gave an estimate of $\beta = 2.2483$ and $\gamma = 1.9894$ with relative error 1698.64% and 1889.70%, respectively.
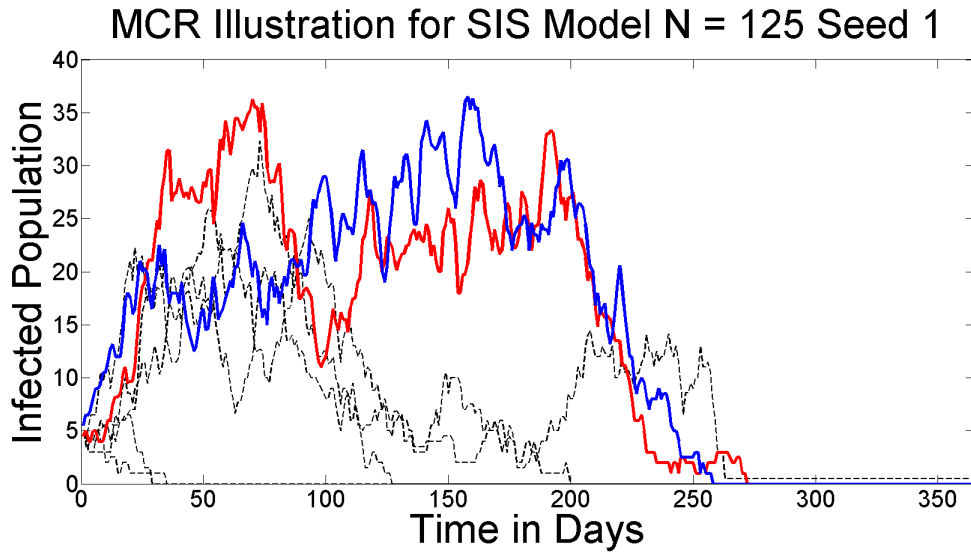
Figure 9: Illustration of the MCR Method. The blue curve represents data set Seed 1. The red curve is the realization that best fits Seed 1. The black curves are several other realizations that were not the best fit.

Table 11 and Table 12 give the results of the parameter estimation for the large and small SIS models with $N = 1250$ and $N = 125$ with $n_s = 10$, respectively. The large Seeds 1, 5, and 9 took 266.32 seconds, 335.88 seconds, and 245.02 seconds, respectively. The small Seeds 1, 5, and 9 took 7.96 seconds, 20.47 seconds, and 5.41 seconds, respectively.

Table 11: MCR Estimated parameter values for the small $N = 125$ SIS model.

Data Set	β			γ		
	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 1	0.125	0.1212	3.04%	0.1	0.1064	6.40 %
Seed 5	0.125	0.1362	8.96 %	0.1	0.1061	6.10 %
Seed 9	0.125	0.1034	17.28 %	0.1	0.1033	3.3 %

39

Table 12: MCR Estimated parameter values for the large $N = 1250$ SIS model.

Data Set	β			γ		
	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 1	0.125	0.1153	7.76%	0.1	0.0893	10.70 %
Seed 5	0.125	0.1315	5.20 %	0.1	0.1028	2.80 %
Seed 9	0.125	0.1023	18.16 %	0.1	0.0815	18.50 %

Notice that for the small SIS population data set, Seed 1 in Table 11, the estimate for β has a relative error of only about 3% compared to over 1600% using the deterministic method as seen in Table 3. There is a similar improvement in the estimate of γ with the slightly more than 6% relative error using the MCR method compared with more than 1800% relative error using the deterministic method. Similar results were found for the other small population data sets.

It was still difficult to accurately estimate β for Seed 9 data, but the MCR method still shows a remarkable improvement in accuracy when compared to the deterministic method. Interestingly, the MCR method only showed a slight improvement over the deterministic method for the large population, $N = 1250$, with a total average increase in accuracy of 4% across the 5 estimates in which there was an improvement (the estimate for γ was not improved for Seed 1).

Tables 13 and Table 14 give the results of the parameter estimation for the large and small Predator-Prey models for $N = 600$ and $N = 60$ with $n_s = 10$, respectively. The large Seeds 2 and 10 took 245.02 seconds and 808.69 seconds, respectively. The small Seeds 2, 6, and 10 took 15.98 seconds, 21.69 seconds, and 17.88 seconds, respectively.

Table 13: MCR Estimated parameter values for the $N = 600$ Predator-Prey model.

Data Set	a_{10}			a_{12}		
	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 5	0.50	0.4862	2.64%	0.05	0.0533	6.60 %
Seed 10	0.50	0.4863	2.76 %	0.05	0.0533	6.60 %
	a_{21}			a_{20}		
Data Set	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 5	0.01	0.0106	6.00%	0.20	0.1614	19.3 %
Seed 10	0.01	0.0107	7.00 %	0.20	0.1586	20.70 %

Table 14: MCR Estimated parameter values for the $N = 60$ Predator-Prey model.

Data Set	a_{10}			a_{12}		
	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 2	0.50	0.4640	7.2 %	0.05	0.0520	4.00 %
Seed 6	0.50	0.4780	4.4 %	0.05	0.0536	7.20 %
Seed 10	0.50	0.4669	6.62 %	0.05	0.0492	1.60 %
	a_{21}			a_{20}		
Data Set	Actual	Estimate	Rel. Error	Actual	Estimate	Rel. Error
Seed 2	0.01	0.0101	1.00 %	0.20	0.1787	10.65 %
Seed 6	0.01	0.0100	0.20 %	0.20	0.1837	8.15 %
Seed 10	0.01	0.0106	6.00 %	0.20	0.1416	29.2 %

Notice that for the small Predator-Prey population data set, Seed 6 in Table 14, the estimate for a_{12} is only about 4% compared to over 300% for the deterministic method as seen in Table 5. Simliar results hold for every parameter in each data set besides a_{20} for Seed 10. Similar claims hold for the large $N = 600$ Predator-Prey population upon inspection of Tables 4 and Table 13. There is, however, a small discrepancy in the parameter a_{20} but the improved estimates with the MCR method for the other parameters is outstanding. As with the SIS model the MCR method still shows a remarkable improvement in accuracy when compared to the deterministic method. In order to ensure the MCR method works more generally we

construct confidence intervals for parameter estimates in the next section.

4.2   Confidence Intervals for The MCR Method

In this section we construct confidence intervals for the small SIS and Predator-Prey models as was done for the deterministic method in Chapter 3. This will allow us to compare the two methods' effectiveness in parameter estimation for these two example models. In constructing confidence intervals we use 1000 implementations of the MCR method using $n_s = 10$. We implement the same algorithm as in Section 3.3 where $J(\boldsymbol{\theta})$ is given by Eq. (15) instead of Eq. (11).

Table 15 shows the confidence intervals for a population of $N = 125$. The average time to compute a single estimate was 15.60 seconds. We notice that for the small SIS model, there is a 95% confidence of the exact parameters having less than 11% relative error. This is compared to the results in Table 8 using the deterministic approach in which the maximum relative error in the parameter values is more than 700%. Confidence intervals for the large population SIS and the Predator-Prey example model will be presented in a future publication.

The MCR method relies on a cost function which compares n_s realizations of the stochastic model to the given data set. Therefore, we also address the effect of n_s on the estimated values. If one were to estimate the parameters one time, the result may be different than another time since the 10 realizations are randomly chosen and therefore will be different from one estimation to the next. Thus one question we address is how much variation is expected across runs. In Table 11, we see that for the data set labeled Seed 1, there is approximately a 3% and 6% relative error in β and

γ, respectively, when we estimate parameters once. Table 16 repeats this estimation 1000 times and determines a confidence interval for the parameter values for Seed 1 as well as Seeds 5 and 9. The maximum relative error (with 95% confidence) is still less than 10% for β and approximately 6% for γ with Seed 1. Seed 9 is the hardest to estimate producing a potential error as large as 13% (which is still significantly smaller than the deterministic approach). This is probably due to the fact that Seed 9 data has the shortest duration of all three data sets (the one in which the infected population vanishes before 150 days in Figure 4). The average time taken to estimate the parameters for a set of $n_s = 10$ are 31.31 seconds for Seed 1, 47.75 seconds for Seed 5, and 18.77 seconds for Seed 9.

Table 15: MCR Confidence intervals for the $N = 125$ SIS model with $n_s = 10$.

Parameter	True Value	Confidence Interval	Max Rel. Error
β	0.125	(0.1124, 0.1166)	10.08 %
γ	0.1	(0.1010, 0.1073)	7.30%

Table 16: MCR Confidence intervals for the $N = 125$ SIS data sets with $n_s = 10$.

Data Set	Parameter	True Value	Confidence Interval	Max Rel. Error
Seed 1	β	0.125	(0.1128, 0.1170)	9.76 %
	γ	0.1	(0.0941, 0.0975)	5.90 %
Seed 5	β	0.125	(0.1224, 0.1248)	2.08 %
	γ	0.1	(0.0945, 0.0963)	5.5 %
Seed 9	β	0.125	(0.1084, 0.1122)	13.28 %
	γ	0.1	(0.0997, 0.1035)	3.5 %

Table 17 gives the confidence intervals for the Predator-Prey model with $N = 600$. Comparing this with Table 10 we see an astonishing improvement in estimating the parameters a_{21} and a_{20} with the MCR method. The MCR method estimated each

parameter with maximum relative error no greater than 4%; an improvement across the board compared to the deterministic method.

Table 17: MCR Confidence intervals for the $N = 60$ Predator-Prey model.

Parameter	True Value	Confidence Interval	Max Rel. Error
a_{10}	0.50	(0.4899, 0.5014)	2.02 %
a_{12}	0.05	(0.0507, 0.0519)	3.80 %
a_{21}	0.01	(0.0098, 0.0100)	2.00 %
a_{20}	0.20	(0.2046, 0.1981)	2.30 %

The other question that arises is the choice of n_s. In the calculations already presented using $n_s = 10$ gave significantly improved accuracy while also increasing computational time compared to the deterministic approach. We used Seed 1 to test $n_s = 100$ which had an average time for estimation of 317.47 seconds. Interestingly, there does not appear to be a significant advantage in using $n_s = 100$ for this data set, Seed 1. However, the maximum relative error with $n_s = 100$ appears to be halfed compared to the confidence intervals with $n_s = 10$. Since there is already a small error in the estimates using $n_s = 10$, the reduction in error when using $n_s = 100$ may not be worth the increase in computational time required when using $n_s = 100$. For data sets which have a shorter time span, such as Seed 9, this increase in computational time may be worth an increase in accuracy if the accuracy results in a maximum relative error less than 10%. This is a topic worth investigating in the future.

Table 18: MCR Confidence intervals for the $N = 125$ SIS model with $n_s = 100$.

Data Set	Parameter	True Value	Confidence Interval	Max Rel. Error
Seed 1	β	0.125	(0.1178, 0.1203)	5.76 %
	γ	0.1	(0.0976, 0.0997)	2.4 %

It is clear that the MCR method proves useful in estimating parameters for the

small and large SIS model as well as the small Predator-Prey model. Note that although several of the intervals do not contain the actual parameter value, their proximity to the actual parameter is superb and each interval is small. The confidence intervals presented for the MCR method give us assurance that the method is robust for these two example models and it clearly shows that it is preferable over the deterministic method for parameter estimation.

5 CONCLUSIONS

We implemented a well-established method for estimating parameters of a deterministic system to purely stochastic data. We determined that the method failed to estimate parameters when the population size of our two example models, the SIS epidemic and Lotka-Volterra Predator-Prey models, was sufficiently small. In order to achieve successful estimates for small populations we developed a new method unique to stochastic models: the MCR method. We showed that the MCR method is significantly more effective in estimating parameters for both small and large populations in our example models. This initial analysis of the MCR method shows that it may be a viable method for parameter estimation for continuous-time Markov chain models.

In the future, it will be necessary to further test the capabilities of the MCR method. This includes investigating criteria such as different initial parameter estimates, the number of realizations to implement in finding a best fit to the data, and if there is a significant difference in increasing these realizations at the expense of computation time. Additionally, the MCR method may prove viable for many other models and applications.

BIBLIOGRAPHY

[1] L. Allen. *An Introduction to Stochastic Processes with Applications to Biology, 2nd Edition,* Taylor and Francis Group, LLC, 2011.

[2] D.T. Gillespie. *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J Comp Phys. 22 (1976), 403434.

[3] H. Andersson, T. Britton. *Stochastic Epidemic Models and Their Statistical Analysis*, Springer-Verlag New York, 2000.

[4] H.T. Banks, et al. *Modeling and Inverse Problems in the Presence of Uncertainty*, Taylor & Francis Group, 2014.

[5] Robert L. Borrelli, Courtney S. Coleman. *Differential Equations: A Modeling Perspective*. John Wiley & Sons, Inc. 2004.

[6] Zimmer C, Sahle S. *Parameter Estimation for Stochastic Models of Biochemical Reactions*, J Comput Sci Syst Biol 6: 011-021, doi: 10.4172/jcsb.1000095, 2012.

VITA

THOMAS CHRISTOPHEL ROBACKER

Education:	M.S. Mathematical Sciences,
	East Tennessee State University,
	Johnson City, Tennessee, 2015
	B.S. Physics and Mathematics,
	University of Tennessee,
	Knoxville, Tennessee, 2013
Professional Experience:	Upward Bound Instructor,
	East Tennessee State University,
	Johnson City, Tennessee, Summer 2015
	Teaching Associate,
	East Tennessee State University,
	Johnson City, Tennessee, Summer 2013–2015
Awards:	Faculty Award: Outstanding Graduate Student,
	East Tennessee State University,
	Johnson City, Tennessee, 2014–2015
	UT Physics Fellowship, Summer 2011, 2012
	University of Tennessee
	Knoxville, Tennessee