# Abstract

GOYAL, LOVELY. Statistical Inference for Non-linear Mixed Effects Models Involving Ordinary Differential Equations. (Under the direction of Dr. Sujit K. Ghosh.)

In the context of nonlinear mixed effect modeling, "within subject mechanisms" are often represented by a system of nonlinear ordinary differential equations (ODE), whose parameters characterize the different characteristics of the underlying population. These models are useful because they offer a flexible framework where parameters for both individuals and population can be estimated by combining information across all subjects. Estimating parameters for these models becomes challenging in the absence of any analytical solution for the system of ODEs involved in the modeling.

In this dissertation we proposed two estimation approaches (i) Bayesian Euler's Approximation Method (BEAM) and (ii) Splines Euler's Approximation Method (SEAM). While we proposed SEAM only for the fixed effect models, BEAM is described for fixed as well as mixed effects models. Both of these approaches involve the likelihood approximation based on the naive Euler's numerical approximation method, thereby providing an analytic closed form approximation for the mean function. SEAM combines the Euler's approximation with Spline interpolation to obtain the parameter estimates for each subject separately. On the other hand, BEAM combines the likelihood approximation with the existing Bayesian hierarchical modeling framework to obtain the parameter estimates.

For illustration purposes, we presented the real data analyses and simulation studies for both fixed and mixed effects models and compared the results with estimates from the NLS method (fixed effects model) and from the NLME method (mixed effects model). For both type of models, proposed methodologies provide competitive results in terms of estimation accuracy and efficiency. The Bayesian Euler's approximation method was also used to estimate parameters involved in an HIV model, for which an analytical closed form mean function is not available.

# Statistical Inference For Non-linear Mixed Effects Models Involving Ordinary Differential Equations


by

Lovely Goyal


a dissertation submitted to the graduate faculty of

north carolina state university

in partial fulfillment of the

requirements for the degree of

doctor of philosophy

statistics

raleigh

March 16, 2006

approved by:

| | |
|---|---|
| Dr. Sujit K. Ghosh (Chair) | Dr. Marie Davidian |
| Dr. Subhashis Ghosal | Dr. Kevin Gross |

# Dedication

To my Mummy, Babuji, Mamiji, Mamaji and Kartik.

# Biography

Lovely Goyal was born on November 14, 1978 in India. She spent first twenty two years of her life in Lucknow, a northern city of India. She finished her Bachelor of Science (Mathematics, Statistics) and Master of Science (Statistics) from University of Lucknow. She joined the Department of Statistics at North Carolina State University as a graduate student in August 2001 and since August 2003 she has been working on her dissertation, under the supervision of Dr. Sujit K. Ghosh. During her graduate study, she worked as a graduate industrial trainee in BD Technologies and GlaxoSmithKline. After receiving her doctoral degree, she will join Amgen Inc. as a senior biostatistician.

# Acknowledgements

I had a wonderful time at North Carolina State University and it has been my privilege to be called as a NC State student. Today when I am just about to finish my graduate studies, there are some people whom I want to thank for making my stay at NC State so wonderful and for helping me to achieve my dream.

First of all, I would like to express my deepest thanks and sincere appreciation to my adviser Dr. Sujit K. Ghosh, who provided me with his excellent guidance throughout my dissertation. I am very lucky to have an adviser who has always been so motivating, understanding and showed incredible patience whenever I needed it most. Without his help and guidance, I could not have accomplished it.

I want to thank Dr. Marie Davidian for serving in my committee and for being such a great teacher. Her "Preparation for Research" class has been one of the best courses I have taken so far and I enjoyed that class a lot even when it was a morning 8:10 a.m. class. I would also like to thank Dr. Subhashis Ghosal and Dr. Kevin Gross for serving in my committee and for their helpful suggestions during the oral prelim exam. I am grateful that they took the time to serve in my committee.

I have been fortunate to have an opportunity to learn from the wonderful faculty at NC State and in particular, I would like to thank Dr. Sastry Pantula, Dr. William Swallow and Dr. Bibhuti Bhattacharya. Both Dr. Pantula and Dr. Swallow have provided constant support to me and I really admire their abilities to take care of every student in the department. I have always admired Dr. B. B. for his passion towards teaching, his knowledge of everything and his kind nature. Because of Dr. B.B., measure theory will always be associated with some happy memories. I would also like to take this opportunity to thank Dr. G.G. Agarwal, my professor at Lucknow University, without whose guidance and help, I would not have come to the United States.

I want to thank Adrian Blue for being so helpful and Terry Byron for answering my almost always silly computing questions. Just the thought of having Terry around for all those last minute technical glitches right before the exams has always been such a huge

relief.

My friends in Raleigh have been a constant reminder of the fact that friendship brings so much comfort in times of joy as well as hardships. I would like to thank my friends Pralay, Harsha, Adi, Kristen, Alvin, Marti, Jimmy Doi, Amna and Athar for their endless support in this endeavor and for making my stay in Raleigh a memorable one. I would specially like to thank Aparna for being the nicest and the most loving roommate and friend one can ever have.

Now I want to thank someone very special in my life, even though I know that I can not thank him enough. I met Kartik, my wonderful husband and my best friend, here in Raleigh almost four and half years ago. Since then he and my in-laws have been a great support for me. I am lucky to be a part of this wonderful family. In last four years, Kartik has been there with me in my toughest times and every time managed to bring me out of it, with his affection and encouraging words. I could not have done this without his incredible support.

I have been fortunate enough to have the world's most loving and wonderful parents. My mummy, the Late Mrs. Maya Rani Gupta, was the coolest and the sweetest mother one can ask for. She was my best friend, my biggest critic, my teacher and my greatest emotional support and to this date she continues to be. I am very lucky to have a father like my babu ji Mr. Bal Kishan Gupta for not only being the greatest father he is but also for telling me again and again that he loves me the most in this whole world.

Now I want to thank those two people in my life who literally taught me how to dream. My mama ji Mr. Mukesh Gupta and my mami ji Mrs. Madhuri Gupta have been my mentors for as long as I can remember. They are the ones who made sure that I got everything that I needed to fulfill my dreams. My mama ji is the one who believed in me when even I was not sure of myself. My mami ji is the one who always kept telling me that anything is possible with sheer determination, persistence and hard work. Every time I look at them, I am amazed to see their selfless nature and their teachings made me what I am today. It is not possible for me to express my gratitude towards them in words.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Preface

Nonlinear mixed effects (NLME) framework is widely used in modeling repeated measurements data, where measurements are obtained for a number of individuals under varying experimental conditions. Nonlinear mixed effects models incorporate population (fixed) as well as individual specific (random) characteristics and hence enable to make inferences for both random and fixed effects. The concept of a nonlinear mixed effects model was first introduced in Sheiner et al. (1972) to analyze the data pooled over all individuals and since then, there has been a great deal of further research (Davidian and Giltinan, 1995) in population pharmacokinetics/pharmacodynamics (PK/PD) using nonlinear mixed effects models. Within the framework of nonlinear mixed effects (NLME) models, much of the interest is focused on representing the mean function (or mean trajectory) describing the dynamic relationship between the response and explanatory variables (such as time), by a system of ordinary differential equations (ODEs) whose parameters describe the different characteristics of the underlying population. A system of ODEs provides an attractive modeling tool to describe a dynamic process, where the interest is focused on modeling the rate of change over time rather than the static average value of the response variable, e.g., as in PK/PD models, viral dynamics etc. In real applications, it turns out that there are very few cases where it is actually possible to derive the closed form expression for the exact solution, for a well-posed differential equation problem. The absence of a closed form analytical solution for the system of ODEs makes parameter estimation in such models, challenging and computationally demanding. The objective of this research is to propose methodologies that approximate the mean function and estimate the parameters involved in it, when an exact analytical form of the mean function is not available.

In Chapter 2, we propose two methods based on likelihood approximation, for non-

linear fixed effects models and compare the performance of these two methods *Bayesian Euler's Approximation Method (BEAM)* and the *Splines Euler's Approximation Method (SEAM)* with other established methods in the literature by using a real data analysis and a simulation study motivated by the real data analysis. In this case, the data sets corresponding to several individuals/subjects were treated as separate data sets.

In Chapter 3, we extend the *BEAM* approach to the nonlinear mixed effects framework, where population consists of several individuals and the individual specific characteristics are incorporated within the model framework. A real data analysis and the corresponding simulation study were performed to compare the performance of BEAM with an established approach in the literature (Lindstrom and Bates, 1990). We implemented *BEAM* to a real data example where no closed form analytical solution exist for the system of ODEs, without assuming any assumption.

In Chapter 4, an additional simulation study was performed to illustrate the application of BEAM to the multivariate data where data consists of $(2 \times 1)$ observation vectors.

In Chapter 5, we summarized the results obtained in Chapters 2, 3 and 4 along with the advantages and disadvantages of the proposed methodologies. Finally we concluded this thesis with a brief discussion about some possible issues for future research.

# Chapter 2

# Statistical Inference for Non-linear Fixed Effects Models

## 2.1 Introduction

In the field of biomedical applications, data usually consists of repeated measurements on individuals observed under varying experimental conditions. For example, in pharmacokinetics, several blood samples are taken on participating individuals over a period of time, following the administration of a drug. These individuals can be considered as a random sample drawn from a population of interest. More often, the relationship between the measured response and the varying experimental conditions is nonlinear and involves unknown parameters of interest. The model is then fitted to data sets from different individuals, where the main interest is to make inferences about population characteristics and in special cases, about individual characteristics which requires a mixed effects modeling framework. However, in this chapter we will treat the data set for each individual as a separate data set and therefore the scope of this chapter is restricted to nonlinear fixed effects models.

Within the the framework of nonlinear models (NLM), much of the interest is focused on representing the mean function (or mean trajectory), describing the dynamic relationship between the response and explanatory variables (such as time), by a system of ordinary differential equations (ODEs) whose parameters describe the different characteristics of the underlying population. A system of ODEs provides an attractive modeling tool to describe dynamic process, where the interest is focused on modeling the rate of change over time rather than the static average value of the response variable. For example in Ho et al. (1995), HIV viral load data was analyzed using a system of

ODEs and their results led to the conclusion that HIV virus has a high rate of replication. In another example, a system of nonlinear ODEs was used to describe the temporal expectation of virus and infected cell densities after initiation of anti-retroviral treatment (Perelson et al., 1996). In the case of a HIV study, parameters involved in differential equations, can characterize rates of production, infection, death of immune system cells and viral production and clearance (Ding and Wu, 1999).

It is well-known that when a closed form analytic solution is available for the system of ODEs, the parameters can be estimated using standard statistical packages, e.g., `R, SAS, WinBUGS` etc. For example, in Han et al. (2002), parameters involved in a system of ODEs were estimated using an analytical solution of the ODEs. The closed form analytical solution was obtained by assuming that the virus dynamics are in steady state prior to the initiation of the anti-retroviral therapy. However, in practice it turns out that such steady state assumptions may not hold and thus there are very few cases where it is actually possible to derive the closed form expression for the exact solution for a well-posed system of ODEs. The parameter estimation problems for such models become challenging and computationally demanding in the absence of any analytically closed form solution for the system of ODEs.

The objective of this research is to develop computationally efficient methods to obtain statistical inference for parameters of a NLM that involves a system of ODEs, in the absence of an analytical solution. In Section 2.2, we describe the nonlinear statistical models and provide a brief review of the associated numerical methods. In Section 2.3, we present the two proposed methods based on Euler's approximation; (i) *Bayesian Euler's approximation method (BEAM)* in Section 2.3.1 and (ii) *splines Euler's approximation method (SEAM)* in Section 2.3.2. We then illustrate our methods in Section 2.4 by applying it to the data on growth colonies of paramecium aurelium. A simulation study motivated by the previous application is then presented in Section 2.5. Finally, in Section 2.6, we provide some general conclusions and directions for future research.

## 2.2 Nonlinear models involving ODEs

Let $y_j$ denote the $j$th observed response, measured at time point $t_j$, for $j = 1, 2, ..., n$ individuals. To keep our description simple, we considered time as the only dynamic explanatory variable in the model. However, methodologies proposed in this chapter can be extended to a more general case with multiple dynamic covariates. The statistical model can be written as,

$$y_j = \mu(t_j, \boldsymbol{\theta}) + \epsilon_j, \quad j = 1, 2, ..., n. \tag{2.1}$$

In equation (2.1), $\mu$ is the mean function describing the average dynamics of the response, and depends on a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ of $p$ regression parameters.

In the context of many biological applications (e.g., PK/PD or PBPK models), $\mu$ can be defined as the solution of a system of ODEs given by

$$\frac{d\boldsymbol{\nu}}{dt} = \mathbf{g}(t, \boldsymbol{\nu}(t, \boldsymbol{\theta})) \quad \text{for } t \neq t_0 \tag{2.2}$$

$$\text{and} \quad \boldsymbol{\nu}(t_0, \boldsymbol{\theta}) = \boldsymbol{\nu}_0(\boldsymbol{\theta}) \tag{2.3}$$

where $\boldsymbol{\nu}(\cdot) = (\nu_1(\cdot), \ldots, \nu_q(\cdot))^T$ represents the underlying vector of dynamics and $\boldsymbol{\nu}_0(\cdot)$ provides a set of known initial conditions (often free of $\boldsymbol{\theta}$). The $q$-vector valued function $\mathbf{g}(\cdot) = (g_1(\cdot), \ldots, g_q(\cdot))^T$ that describes the dynamics is completely known up to the unknown parameter $\boldsymbol{\theta}$. Notice that (2.2) can equivalently be expressed with a set of $q$ ODEs, $\frac{d\nu_k}{dt} = g_k(t, \boldsymbol{\nu}(t, \boldsymbol{\theta}))$ for $k = 1, \ldots, q$.

The mean function, $\mu(\cdot)$ is related to $\boldsymbol{\nu}$ by a completely known function $H \colon \mathbb{R}^q \to \mathbb{R}$ by $\mu(\cdot) = H(\boldsymbol{\nu}(\cdot))$. In this chapter, for simplicity, we use single-compartmental system, i.e., $q = 1$, for all our illustrations but methods proposed in this chapter can be applied to the general case of $q \geq 1$ (Chapter 3). The random errors $\epsilon_j$'s correspond to the measurement uncertainties associated with the observed response at different time points. These random errors are assumed to be identically, independently distributed (*iid*) with

zero mean and constant variance across all measurements, i.e.,

$$E(\epsilon_j) = 0 \text{ and } Var(\epsilon_j) = \sigma^2, \quad \text{for} \quad j = 1, 2, ..., n. \tag{2.4}$$

The *iid* assumption for the errors is clearly restrictive and in Section 2.6 we discuss how our methods can be extended to the case when we allow the variance of errors to change with time, i.e., $Var(\epsilon_j) = \sigma^2(t_j, \boldsymbol{\eta})$ with unknown parameter $\boldsymbol{\eta}$.

The objective is to estimate the parameter vector $\boldsymbol{\theta}$ and the variance parameter, $\sigma^2$. The parameter estimates can be obtained by numerical methods, using packages like `nlm` in `R` or `proc nlin` in `SAS` if an analytic closed form expression were available for the mean function $\mu(\cdot)$. However, as we discussed earlier, in most cases, such an analytic solution for the system of ODEs either require restrictive assumptions or simply not available. The lack of a closed form expression for the mean function makes the parameter estimation problem challenging and this is the focus of our research work.

The usual approach to overcome this problem, is to solve the system of ODEs numerically by using popular ODE solvers at a *known* set of values of the parameter $\boldsymbol{\theta}$. However the success of most of these numerical approximation methods depends on a "good" choice of a starting value for $\boldsymbol{\theta}$ and some characteristics of the system (e.g., the steepness etc.). A "bad" starting value often leads to an unstable solution and creates numerical problems for the optimization method that is followed by these ODE solvers. The popular `odesolve` package in `R` provides an interface to the Fortran ODE solver `lsoda` (Petzold, 1987). The numerical solution obtained from these ODE solvers are then used to obtain parameter estimates either by a Bayesian method (Gelman et al., 1996; Wakefield, 1996; Lunn et al., 2002; Putter et al., 2002) or by the maximum likelihood method (Davidian and Giltinan, 1995; Racine-Poon and Wakefield, 1998). Even though ODE solvers are widely used for estimating parameters in PK/PD modeling, it may be difficult to implement or lack control of essential numerical subroutines required to obtain the desired numerical solution for ODE. Moreover these numerical methods also turn out to be unstable especially in case of censored or missing data (Putter et al., 2002). Apart from such numerical instabilities, these methods are computationally intensive and

iterative procedures in which the system of ODEs must be solved at each time point for each individual and this becomes more complicated in the case of multi-compartmental problems with censored or missing data.

An alternative approach, known as the "Integrated Data (ID)" method for parameter estimation for models described by a system of ODEs, was proposed by Holte et al. (2003). The idea behind the ID method is to simplify a nonlinear regression problem by transforming a system of ODEs into a system of integral equations and fitting a linear regression model with "covariates" as the approximate integrals in these equations. However, the ID method requires a set of dense measurements taken from each compartment represented in the ODE system, which may be difficult or costly, if not impossible, especially when the system consists of multi-compartments.

In this chapter, we propose two alternative methodologies to resolve the numerical problems related to parameter estimation for the situation when there is no closed form solution available for the system of ODEs.

The first approach will be termed as the "*Bayesian Euler's Approximation Method (BEAM)*" which is built on the existing Bayesian framework for parameter estimation in PK/PD modeling (Gelman et al., 1996; Lunn et al., 2002; Han et al., 2002; Putter et al., 2002; Wakefield, 1996; Huang et al., 2004) using the Euler's method of solving a system of ODEs and thus providing an *analytic* closed form approximation for the likelihood function. The advantages of BEAM are in providing a closed form analytic approximation for the likelihood as well as its ability to handle missing or censored data by using data augmentation methods (Schafer, 1997). It also has the flexibility to handle sparse and/or unbalanced data. The availability of entire posterior distributions for the unknown parameter $(\boldsymbol{\theta}, \sigma)$ also makes it straightforward to draw statistical inferences.

The second approach will be termed as the "*Splines Euler's Approximation Method (SEAM)*", that uses a suitable class of interpolating spline functions (Wahba, 1990) to pre-process the data before applying the Euler's method to approximate the likelihood function. The advantages of SEAM are in relaxing the distributional assumptions for the errors and a huge computational efficiency over the competing methods.

## 2.3 Likelihood approximation using the Euler's method

Many different numerical approximation methods are available for computing approximate solutions to a system of ODEs with a given set of initial conditions such as the generic problem given by (2.2) and (2.3).

According to the Lambert (1991), a numerical approximation method is basically a prescription for replacing the system of ODEs by a system of linear algebraic equations that can be solved on a computer using software written in a standard programming language. A detailed discussion of these methods can be found in the literature (Shampine, 1994; Lambert, 1991; Atkinson, 1978). All the numerical approximation methods involve discretizing the time points by an amount $h$ known as the "step size," which is the distance between two consecutive time points. This step size may or may not be the same for all consecutive time points, but for our description we assume $h$ to be constant over the range of time points, i.e. we assume that $t_k^0 = t_0 + hk$ for $k = 0, 1, 2, \ldots$, represent the discretized time points. It is easy to see that the solution to the system of ODEs in equation (2.2) can be expressed as

$$\boldsymbol{\nu}(t, \boldsymbol{\theta}) = \int_{t_0}^{t} \mathbf{g}(s, \boldsymbol{\nu}(s, \boldsymbol{\theta}))ds + \boldsymbol{\nu}_0(\boldsymbol{\theta}) \tag{2.5}$$

which suggests the approximation, as $h \to 0$,

$$\boldsymbol{\nu}(t + h, \boldsymbol{\theta}) - \boldsymbol{\nu}(t, \boldsymbol{\theta}) = \int_{t}^{t+h} \mathbf{g}(s, \boldsymbol{\nu}(s, \boldsymbol{\theta}))ds \approx h\mathbf{g}(t, \boldsymbol{\nu}(t, \boldsymbol{\theta})) \tag{2.6}$$

and hence an approximation for $\mu(t, \boldsymbol{\theta}) = H(\boldsymbol{\nu}(t, \boldsymbol{\theta}))$, where $H$ is a completely known continuous function. Thus, using (2.6) we can obtain a recurrence relation to approximate the mean function. We now describe a method based on (2.6) to approximate the likelihood that arise from the model given by equations (2.1-2.4). Let $t_1 < t_2 < \cdots < t_n$ denote the observed time points in the data set and we observe the response values $\{Y_j = Y(t_j) : j = 1, 2, \ldots, n\}$. As the observed time points can be unevenly distributed, we first consider a discretization by $N$ fixed time points $t_0 = t_1^0 < t_2^0 < ... < t_N^0$ such that

$t^0_{k+1} - t^0_k = h$ for $k = 1, 2, ..., (N-1)$. In order to cover the range of observed time points we choose the maximum value for these fixed time points to be larger than $t_n$. In other words, we assume $t^0_N > t_n$. The choice of $h$ (and hence that of $N$) will depend on the sample size $n$. Letting $\tilde{\boldsymbol{\nu}}_k \equiv \tilde{\boldsymbol{\nu}}(t^0_k, \boldsymbol{\theta})$ and $\tilde{\mu}_k \equiv \tilde{\mu}(t^0_k, \boldsymbol{\theta}) = H(\tilde{\boldsymbol{\nu}}_k)$ for $k = 1, 2, \ldots, N-1$, we can write

$$\begin{aligned} \tilde{\boldsymbol{\nu}}_{k+1} &= \tilde{\boldsymbol{\nu}}_k + h\mathbf{g}(t^0_k, \boldsymbol{\nu}_k) \\ \tilde{\mu}_k &= H(\tilde{\boldsymbol{\nu}}_k), \end{aligned} \tag{2.7}$$

with initial condition, $\tilde{\boldsymbol{\nu}}_1 = \boldsymbol{\nu}_0(\boldsymbol{\theta})$. This simple approximation defined by equation (2.7), forms the basis of our analytical approximation. Now to define $\tilde{\mu}(t, \boldsymbol{\theta})$ for any value of $t \in [t^0_1, t^0_N]$, we use a linear interpolation. More precisely, given a time point $t \in [t^0_1, t^0_N]$, we define the labels,

$$L(t) = \sum_{k=1}^{N} I(t^0_k \leq t). \tag{2.8}$$

Notice that for any $t \in [t^0_1, t^0_N]$, the function $L(t)$ takes the values in the range $\{1, \ldots, N\}$ and determines that how many $t^0_k$'s are less than $t$ which in turn provides the lower limit of the interval that contains $t$, i.e., $t^0_{L(t)} \leq t < t^0_{L(t)+1}$. The value of approximate mean function at $t \in [t^0_1, t^0_N]$ is then given by,

$$\tilde{\mu}(t, \boldsymbol{\theta}) \equiv \tilde{\mu}_h(t, \boldsymbol{\theta}) = \tilde{\mu}_{L(t)} + \frac{t - t^0_{L(t)}}{t^0_{L(t)+1} - t^0_{L(t)}}(\tilde{\mu}_{L(t)+1} - \tilde{\mu}_{L(t)}), \tag{2.9}$$

where $L(t)$ is defined in (2.8) and $\tilde{\mu}_k$'s are defined in (2.7). Thus we can approximate the true likelihood function of $(\boldsymbol{\theta}, \sigma)$ using the $\tilde{\mu}(t, \boldsymbol{\theta})$ function. Notice that $\tilde{\mu}_h(t, \boldsymbol{\theta}) \to \mu(t, \boldsymbol{\theta})$ as $h \to 0$ (Lambert, 1991). In fact, $\tilde{\mu}_h(t, \boldsymbol{\theta}) = \mu(t, \boldsymbol{\theta}) + o(h)$ if we use the above method known as the "naive" Euler's method. If we use the "improved" Euler's method (see Appendix A), then $\tilde{\mu}_h(t, \boldsymbol{\theta}) = \mu(t, \boldsymbol{\theta}) + o(h^2)$ and more generally Runge-Kutta method yields $\tilde{\mu}_h(t, \boldsymbol{\theta}) = \mu(t, \boldsymbol{\theta}) + o(h^4)$. Thus, it follows that as $h \to 0$, the likelihood function based on true mean function $\mu(t, \boldsymbol{\theta})$ can be well approximated by the likelihood function

based on the approximating $\tilde{\mu}(t, \boldsymbol{\theta})$ function as defined above.
We now use (2.7) and (2.9) as the basis to propose two methods which can be further improved using other numerical recipies as described in Appendix A at the cost of computational time.

### 2.3.1 The Bayesian Euler's Approximation Method (BEAM)

In order to construct a likelihood based on (2.9) we assume that the errors are iid and normally distributed with mean zero and variance $\sigma^2$. Given the observed data $\mathcal{D} = \{(Y_j, t_j) : j = 1, 2, \ldots, n\}$ the model described by equations (2.1-2.4) can now be approximated by the following hierarchical model:

$$y_j|(\boldsymbol{\theta}, \sigma^2) \overset{indep}{\sim} \mathrm{N}(\tilde{\mu}_j(\boldsymbol{\theta}), \sigma^2) \text{ for } j = 1, 2, ..., n \tag{2.10}$$

where $\tilde{\mu}_j(\boldsymbol{\theta}) = \tilde{\mu}(t_j, \boldsymbol{\theta})$ as defined in (2.9). For practical applications, the $y_j$'s may be the log-transformed (or more generally Box-Cox transformed) values depending on whether normality or log-normality is the more appropriate assumption for the data (Lunn et al., 2002).

Second and the final stage of this hierarchical model consists of the specifying prior distributions for parameters as follows:

$$\boldsymbol{\theta}|\sigma^{-2} \sim \mathrm{MVN}_p(\boldsymbol{\theta}_0, H_0) \text{ and } \sigma^{-2} \sim \mathrm{G}(a_0, b_0), \tag{2.11}$$

where $\mathrm{MVN}_p(\boldsymbol{\theta}_0, H_0)$ denotes a multivariate normal distribution with mean $\boldsymbol{\theta}_0$ and variance matrix $H_0$ and $\mathrm{G}(a_0, b_0)$ denotes a gamma distribution with mean $a_0 b_0$. The values of $a_0,\ b_0,\ \boldsymbol{\theta}_0,\ H_0$, are assumed to be known, which are used to elicit prior information when available. In the lack of prior information, we choose these known quantities to reflect on prior ignorance by choosing these values to yield vague priors (i.e., priors with large variances). For a detailed discussion about the choice of prior distribution, see Natarajan and Kass (2000). The joint posterior distribution for parameters $\boldsymbol{\theta}$ and $\sigma^{-2}$

based on the model (2.10-2.11) can be written as:

$$p(\boldsymbol{\theta}, \sigma^{-2}|\mathcal{D}) \propto p(\mathbf{Y}|\boldsymbol{\theta}, \sigma^{-2})p(\boldsymbol{\theta})p(\sigma^{-2}). \tag{2.12}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)$. Clearly the above posterior density is analytically intractable as it is highly nonlinear in $\boldsymbol{\theta}$. In order to use sampling based methods, such as Markov chain Monte Carlo (Robert and Casella, 2005) we obtain the full conditionals of $\boldsymbol{\theta}$ and $\sigma^{-2}$, which are given by,

$$\begin{aligned} p(\boldsymbol{\theta}|\sigma^{-2}, \mathcal{D}) &\propto p(\mathbf{Y}|\boldsymbol{\theta}, \sigma^{-2})p(\boldsymbol{\theta}) & (2.13)\\ \text{and } p(\sigma^{-2}|\boldsymbol{\theta}, \mathcal{D}) &\propto p(\mathbf{Y}|\boldsymbol{\theta}, \sigma^{-2})p(\sigma^{-2}). & (2.14) \end{aligned}$$

If the right hand side of above equations have densities of the standard form then we can use Gibbs sampling (Geman and Geman, 1984) to simulate values from the posterior distribution of $(\boldsymbol{\theta}, \sigma^{-2})$. For instance a standard form is available for the full conditional of $\sigma^{-2}$:

$$\sigma^{-2}|\boldsymbol{\theta}, \mathcal{D} \propto \mathrm{G}\left\{a_0 + \frac{n}{2}, \left(\frac{1}{b_0} + \frac{1}{2}\sum_{j=1}^{n}(y_j - \tilde{\mu}_j(\boldsymbol{\theta}))^2\right)^{-1}\right\}, \tag{2.15}$$

and therefore we can easily draw samples from the full conditional of $\sigma^{-2}$. Since we do not have a standard form for the conditional distribution of $\boldsymbol{\theta}$, we can use the Metropolis-Hastings algorithm (Hastings, 1970) to draw samples. Though we used `WinBUGS` to generate approximate samples from posterior distribution of $(\boldsymbol{\theta}, \sigma^{-2})$, here we give a brief outline of the iterative MCMC algorithm suitable for BEAM.

1. Initialize the iteration of the chain at $l = 0$ and start with some initial values, $S^{(0)} = (\sigma^{-2(0)}, \boldsymbol{\theta}^{(0)})$

2. Obtain $S^{(l)}$ from $S^{(l-1)}$ in two steps:

   (a) Draw $\sigma^{-2(l)} \sim \pi(\sigma^{-2}|\boldsymbol{\theta}^{(l-1)}, \boldsymbol{Y})$ using (2.15) and

   (b) For $\boldsymbol{\theta}^{(l)}$, generate a new value $\boldsymbol{\phi}$ from a symmetric proposal density $q(\boldsymbol{\phi}|\boldsymbol{\theta}^{(l-1)})$.

Evaluate the acceptance probability of the move, given by,

$$\alpha(\boldsymbol{\phi}|\boldsymbol{\theta}^{(l-1)}) = min\left\{1, \frac{\pi(\boldsymbol{\phi}|\sigma^{-2(l)}, \mathbf{Y})}{\pi(\boldsymbol{\theta}^{(l-1)}|\sigma^{-2(l)}, \mathbf{Y})}\right\}$$

Also, independently sample a $u$ from the uniform $(0, 1)$ and if $u \leq \alpha(\phi|\boldsymbol{\theta}^{(l-1)})$ the move is accepted else stay at $\boldsymbol{\theta}^{(l-1)}$. In other words, if the move is accepted then $\boldsymbol{\theta}^{(l)} = \boldsymbol{\phi}$ otherwise $\boldsymbol{\theta}^{(l)} = \boldsymbol{\theta}^{(l-1)}$.

3. Move from chain $l$ to $l+1$ using step (2) and repeat until the Markov chain $\{S^{(l)}, l = 1, 2, \ldots\}$ converges (to $p(\boldsymbol{\theta}, \sigma^{-2}|\mathcal{D})$).

In practice, `WinBUGS` uses a Metropolis algorithm based on a normal proposal distribution with the mean as the current value of the parameter and variance determined by tuning over the first 4000 iterations to achieve an acceptance rate between 20% and 40%. In order to diagnose convergence of algorithms, we used graphical techniques such as history plots (available in `WinBUGS`) of the values of the multiple chains for each parameter. Based on these plots we decided upon an initial number of burn-in iterations (e.g., at least 4000) followed by say $B = 2000$ samples per chain drawn afterwards. We performed MCMC sampling based on three parallel chains, therefore all the posterior summaries are based on a total of $3B = 6000$ samples. Mean of posterior distribution for each of unknown parameters was taken to be the posterior estimate of that unknown parameter along with a 95% posterior interval formed by 2.5% and 97.5% posterior percentiles.

### 2.3.2 The Splines Euler's Approximation Method (SEAM)

Interpolation is a method of constructing a smooth curve from a discrete set of known data points, which is a specific case of curve fitting, in which the function must go exactly through the data points. Spline interpolation is a form of interpolation where the interpolant is a special type of piecewise polynomial function called a spline. Spline interpolation is preferred over regular polynomial interpolation because the interpolation error can be made small even when using low degree polynomials for the spline. The motiva-

tion behind the second approach, "Splines Euler's Approximation Method (SEAM)" is to obtain a smooth estimate of the mean using interpolating spline by regressing observed response on observed time points. Given the observed data $\mathcal{D}$ (as defined in previous section), we fit a cubic spline interpolation passing through each of the time points $t_j$'s. More specifically for each interval $[t_j, t_{j+1}]$, there is a separate cubic polynomial each with its own coefficients:

$$S_j(t) = a_j(t-t_j)^3 + b_j(t-t_j)^2 + c_j(t-t_j) + d_j \text{ for } t \in [t_j, t_{j+1}]$$

together, these segments constitute the spline $S(t)$ which must satisfy the following conditions:

(i) Piecewise continuous: $S_j(t_j) = y_j,\ S_j(t_{j+1}) = y_{j+1}$ and

(ii) First and second derivatives continuous: $S'_{j-1}(t_j) = S'_j(t_j)$ and $S''_{j-1}(t_j) = S''_j(t_j)$.

A detailed discussion about splines and interpolation can be found in literature (Wahba, 1990).

In order to construct the approximating mean function $\tilde{\mu}(t, \boldsymbol{\theta})$, we use the following steps:

1. Fit an interpolating spline $\hat{Y}(t)$ by regressing observed data $Y_j$'s on observed time points $t_j$'s for $j = 1, 2, ..., n$ and obtain the predicted values $\hat{Y}_k = \hat{Y}(t_k^0)$, corresponding to each of the fixed time points $t_k^0 = t_0 + hk$, for $k = 1, \ldots, N$. This creates a pseudo-data set $\{(\hat{Y}_k, t_k^0) : k = 1, 2, \ldots, N\}$ suitable for Euler's approximation.

2. Next, use Euler's method (2.7) to construct an approximate mean function $\tilde{\mu}(t_k^0, \boldsymbol{\theta})$.

3. Finally, obtain the least square estimate $\hat{\boldsymbol{\theta}}$ by minimizing

$$SS(\boldsymbol{\theta}) = \sum_{k=1}^{N} (\hat{Y}_k - \tilde{\mu}(t_k^0, \boldsymbol{\theta}))^2, \tag{2.16}$$

and then obtain $\hat{\sigma^2} = \frac{1}{N-p} \sum_{k=1}^{N} (\hat{Y}_k - \tilde{\mu}(t_k^0, \hat{\boldsymbol{\theta}}))^2$.

The first advantage of this approach lies in the fact that no parametric distribution assumption is necessary for the errors. Another advantage comes from the use of splines for curve fitting, since splines can be used to fit data observed over sparsely and unevenly spaced time points, the $SS(\boldsymbol{\theta})$ defined by (2.16) becomes a smooth function of $\boldsymbol{\theta}$ facilitating the minimization problem in (2.16). Moreover, this approach is computationally very efficient and easy to implement requiring no iterative procedure. Therefore, SEAM provides an attractive tool in comparison of other computationally intensive procedures.

## 2.4 Analysis of Growth Colonies of *Paramecium Aurelium*

Diggle (1990) presents a data set that describes the growth of three closed colonies of *paramecium aurelium* in a nutritive medium on a 19 days period. For a detailed description of this experiment, we encourage readers to refer to Diggle (1990, p. 8). One of the main goals of this experiment was to develop a dynamic model for the growth count (say $x_j$'s) of paramecium aurelium, as a function of time $t$. The data is assumed to follow a log-normal distribution, with

$$\begin{aligned} y_j = \log\{x_j\} &= \log\{\nu(t_j, \boldsymbol{\theta})\} + \epsilon_j \\ &= \mu(t_j, \boldsymbol{\theta}) + \epsilon_j, \qquad \epsilon_j \sim \mathrm{N}(0, \sigma^2) \end{aligned} \tag{2.17}$$

where $x_j$ is the observed growth count at time point $t_j$ (measured in days) and $\mu(t_j, \boldsymbol{\theta}) = \log\{\nu(t_j, \boldsymbol{\theta})\}$.

Next, it is assumed that $\nu(t)$ follows the standard 2 parameter logistic growth curve described by the non-linear differential equation

$$\begin{aligned} \frac{d\nu}{dt} &= g(t, \nu(t, \boldsymbol{\theta})) \\ &\equiv \nu(\theta_1 - \theta_2\nu), \quad \text{and } \nu(0) = y_0 = 2. \end{aligned} \tag{2.18}$$

Equivalently, we can express the equation(2.18) in terms of $\mu(\cdot)$:

$$\begin{aligned} \frac{d\mu}{dt} &= g^*(t, \mu(t, \boldsymbol{\theta}))) \\ &\equiv \theta_1 - \theta_2 e^{\mu} \text{ and } \mu(0) = \log(2). \end{aligned} \tag{2.19}$$

In a logistic growth curve model, $\theta_1$ represents the growth per capita, $\theta_1/\theta_2$ measures the carrying capacity of the population, and $y_0$ represents the initial size of the population.

There are three data sets consisting of individual counts, corresponding to three replications of the same experiment and here we analyzed all three data sets separately. Although we proposed our estimation methods for situations where an analytical closed form solution is not available for the system of differential equations, here an analytically closed form solution is actually available. The reason behind the choice of this simple growth curve model is that it will allow us to compare the performance of the estimation based on our proposed approaches to that based on the ideal nonlinear regression approach, which requires a closed form mean function be available. A detailed description of nonlinear regression techniques can be found in a classic book by Davidian and Giltinan (1995). We will also perform ID method (Holte et al., 2003) on these three data sets.

To approximate the likelihood using BEAM, we chose $N = 19$ and for SEAM we chose $N = 40$. These choices of the tuning parameter $N$ (or equivalently, $h = \frac{t_N^0 - t_1^0}{N}$) are not based on any analytical work, rather the choices are mainly driven by computational convenience. In general, the finer the grid points (with large $N$ and hence small $h$) is chosen, the better the likelihood will be approximated. Alternatively, one may also use improved numerical approximation methods (such as those described in the Appendix A) at the cost of computing time.

To implement the BEAM on this log transformed data, we followed the same steps as described in Section 2.3.1, coupled with equations (2.17) and (2.19). For the MCMC runs required within BEAM, we generated samples based on three parallel Markov chains with an initial burn-in of 4000 iterations followed by 2000 post-burn-in samples per

chain, giving us a total of 6000 approximate samples from the posterior distribution of $(\boldsymbol{\theta}, \sigma)$, where samples for $\sigma$ were observed by calculating $\sigma = \frac{1}{\sqrt{\sigma^{-2}}}$. Convergence of the chains was diagnosed visually by inspecting simultaneous trace and acf plots of the values of all three chains, for each parameter. Plots showing a good mixing of chains were a reasonable indication for convergence. The software `WinBUGS` freely available at the website `http://www.mrc-bsu.cam.ac.uk/bugs/` was used to perform all of the computations for this data analysis. The following values were used to elicit priors: $\boldsymbol{\theta}_0 = (0,0)^T, H_0 = 10\mathrm{I}_{2\times 2}$ for the prior on $\boldsymbol{\theta}$ and $a_0 = b_0 = 0.01$ for the prior on $\sigma^{-2}$. Finally, the MC estimate of the mean and standard deviation of the posterior distributions for these parameters were used as point estimates and standard errors, respectively.

In a similar approach, to implement SEAM to fit the model to these data sets we followed the same steps as described in Section 2.3.2. The function "`interpSpline`" in `R` was used to obtain the spline interpolation followed by the use of the `optim` function available in `R` to perform the minimization of $SS(\boldsymbol{\theta})$ function in (2.16). Finally, to fit the NLM, we used `nlm` function in `R` to obtain parameter estimates and associated standard errors (SE).

Results from all four estimation approaches are presented in Table 2.1. In Table 2.1, we presented estimates and standard errors for parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)$ and $\sigma$, corresponding to nonlinear least square (NLS) method, BEAM, SEAM and ID method. Estimated standard errors for the parameter $\sigma$ are not presented in the table, because of its unavailability with the statistical software used to minimize the likelihood functions in "NLS", "SEAM" and "ID" methods. Additionally, in order to compare the performance of these estimation approaches, we plotted estimated mean function corresponding to four approaches along with the observed data points in Figure 2.1. From Figure 2.1, all four approaches seem to perform very similarly in capturing the trajectory of mean function reasonably well. It can also be observed from Figure 2.1 that both BEAM and SEAM provide data fits close to the fits obtained by using the NLS method, which uses the exact analytical form of the mean function. It can be observed from Table 2.1, that

Table 2.1: Parameter estimates and standard errors (SE) based on the logistic growth model for colonies of the bacteria paramecium aurelium using NLS, BEAM and SEAM methods to fit three data sets.

| Data Set | Estimates | Method | | | |
|---|---|---|---|---|---|
| | | NLS | BEAM | SEAM | ID |
| I | $\hat{\theta}_1$ | 0.789 | 0.760 | 0.783 | 0.767 |
| | ESE | 0.025 | 0.029 | 0.044 | 0.015 |
| | $\hat{\theta}_2(*10^{-3})$ | 1.446 | 1.470 | 1.464 | 1.369 |
| | ESE | 0.120 | 0.152 | 0.238 | 0.055 |
| | $\hat{\sigma}$ | 0.218 | 0.266 | 0.233 | 0.187 |
| | ESE | - | 0.049 | - | - |
| II | $\hat{\theta}_1$ | 0.837 | 0.803 | 0.827 | 0.792 |
| | ESE | 0.025 | 0.034 | 0.049 | 0.015 |
| | $\hat{\theta}_2(*10^{-3})$ | 1.672 | 1.678 | 1.685 | 1.548 |
| | ESE | 0.126 | 0.175 | 0.265 | 0.057 |
| | $\hat{\sigma}$ | 0.201 | 0.266 | 0.207 | 0.164 |
| | ESE | - | 0.049 | - | - |
| III | $\hat{\theta}_1$ | 0.892 | 0.857 | 0.875 | 0.856 |
| | ESE | 0.018 | 0.025 | 0.051 | 0.011 |
| | $\hat{\theta}_2(*10^{-3})$ | 1.594 | 1.596 | 1.579 | 1.489 |
| | ESE | 0.078 | 0.118 | 0.246 | 0.063 |
| | $\hat{\sigma}$ | 0.132 | 0.201 | 0.137 | 0.135 |
| | ESE | - | 0.037 | - | - |

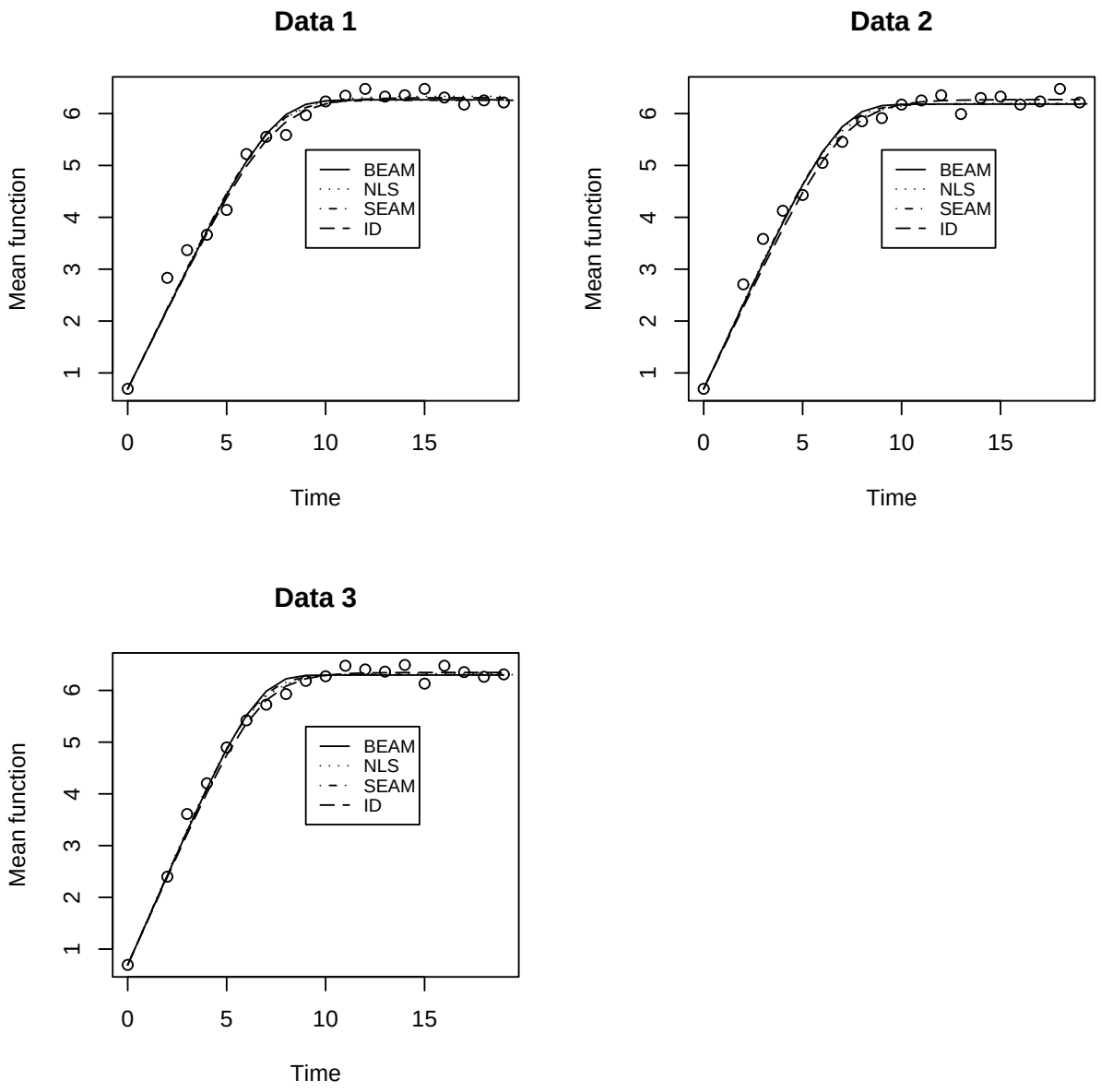


Figure 2.1: Plots of observations on growth colonies of paramecium aurelium (in log scale) and the estimated mean trajectories obtained by NLS, BEAM and SEAM for each of the three data sets.

estimated standard errors (ESE) for parameters $\theta_1$ and $\theta_2$ are smaller for ID method than remaining other three approaches which may appear to suggest that ID method provides more precise estimates than other approaches. However this situation seems to arise from the estimation approach for $\sigma$ by ID method. ID method estimates the population variance beforehand and then plugs in that estimated variance in the model. This estimated variance is then treated as a known parameter while estimating rest of the parameters in the model. Here an estimate is being considered as a known value for rest of the estimation procedure and therefore uncertainty associated with this estimate is not being accounted. Therefore, this plug-in variance estimator leads to the underestimation of variance of $\theta_1$ and $\theta_2$ both. Mathematically, estimated variance of $\hat{\theta}_i$ can be expressed as

$$V(\hat{\theta}_i) = E[V(\hat{\theta}_i|\hat{\sigma})] + V[E(\hat{\theta}_i|\hat{\sigma})], \quad i = 1, 2 \tag{2.20}$$

Now under the ID method, since $\hat{\sigma}$ is treated as known while estimating $\boldsymbol{\theta}$, the second term on the right hand side of equation (2.20) becomes zero. This leads to the underestimation of $V(\hat{\theta}_i)$ for $i = 1, 2$. It can be noticed in next section that the simulation standard error for parameters, in Table 2.2 are in accordance for all four methods. We also used BEAM to calculate the point estimates based on the posterior distribution of the carrying capacity per capita $(\theta_1/\theta_2)$ for all three data sets, along with the corresponding 95% posterior credible interval. These point estimates were obtained by calculating the mean of ratio of posterior samples for $\theta_1$ and $\theta_2$. For the data set (1), the estimated carrying capacity is 523.91 and the corresponding 95% posterior credible interval is (445.80, 606.92). For the data set (2), the estimated carrying capacity is 526.88 and the corresponding 95% posterior credible interval is (448.19, 612.83). Similarly, for the data set (3), the estimated carrying capacity is 533.74 and the corresponding 95% posterior credible interval is (453.38, 616.70).

## 2.5 Simulation Study

A simulation study, motivated by the above real data analysis was carried out to compare the performance of the two proposed methods, BEAM and SEAM, to the NLS procedure, in terms of estimation accuracy and efficiency. The true values of parameters for data generation are chosen based on the estimates obtained for the real data application. For the simulation study, data is generated using the model (2.19) with $\mu(t, \boldsymbol{\theta})$ given by the closed form solution:

$$\begin{aligned} \mu(t, \boldsymbol{\theta}) &= \log(\nu(t, \boldsymbol{\theta})) \\ &= \log(\theta_1) + \mu(0) + t\theta_1 - \log[\theta_2 e^{\mu(0)}(e^{t\theta_1} - 1) + \theta_1] \end{aligned} \tag{2.21}$$

We chose the time points as given in the real data set and simulated data sets using equation (2.17) and (2.21) with true values of the parameter set at $\theta_1 = 0.8$, $\theta_2 = 0.0015$ and $\sigma = 0.25$.

We chose the sample size same as the real data set, i.e., $n = 19$ and replicated the data generations for 1000 Monte Carlo runs. To fit the model by BEAM we chose $N = n$ and the same prior distribution as used in the real data analysis and used same number of burn-in and MCMC samples for each of the 1000 data sets as was used in the real data application. Similarly for SEAM we used $N = 40$ to fit the models to each of the simulated data sets.

A summarization of the comparative study of the four procedures based on this simulation study are given in Table 2.2 and Figure 2.2. In Table 2.2, we summarize our finding in terms of (i) the bias, which is the difference between the MC mean of the point estimates and the true value of a parameter; (ii) the estimated standard error (ESE), which is the MC mean of the standard errors of the parameter estimates, (iii) the Monte Carlo simulation standard error (MCSE), which is the standard deviation of the estimates and (iv) the mean square error (MSE) obtained as Bias$^2$+ MCSE$^2$.

From Table 2.2, it is evident that all four methods performed equally well in terms of bias, standard errors and MSE's. For a better understanding of MC distribution of the

Table 2.2: Simulation Results for the logistic growth model for colonies of paramecium aurelium using NLS, BEAM, SEAM and ID Methods with 1000 MC Runs.

| Parameters | Estimates | Method | | | |
|---|---|---|---|---|---|
| | | NLS | BEAM | SEAM | ID |
| $\theta_1$ | Bias | 0.002 | -0.014 | -0.008 | -0.007 |
| | MCSE | 0.029 | 0.028 | 0.029 | 0.035 |
| | ESE | 0.029 | 0.029 | 0.233 | 0.017 |
| | MSE | 0.001 | 0.001 | 0.001 | 0.001 |
| $\theta_2(*10^{-3})$ | Bias | 0.005 | 0.036 | 0.017 | -0.024 |
| | MCSE | 0.146 | 0.149 | 0.149 | 0.152 |
| | ESE | 0.140 | 0.152 | 0.235 | 0.063 |
| | MSE | 0.021 | 0.024 | 0.022 | 0.024 |
| $\sigma$ | Bias | -0.005 | 0.014 | -0.017 | -0.002 |
| | MCSE | 0.042 | 0.044 | 0.044 | 0.045 |
| | ESE | - | 0.048 | - | - |
| | MSE | 0.0018 | 0.0021 | 0.0022 | 0.0021 |

parameter estimates in comparing the BEAM and SEAM with the NLS and ID method, we present box plots of the estimates obtained by each of the four methods in Figure 2.2. The horizontal solid line in each case represents the true value of the parameter. Figure 2.2 reveals that although BEAM and SEAM tend to underestimate $\theta_1$, the inter-quartile range of the estimates from all four methods contain the true value of $\theta_1$. Similarly for $\theta_2$, we observe from Figure 2.2 and Table 2.2 that in this case as well, all four methods perform almost identically. For $\sigma$, Figure 2.2 apparently indicates that BEAM tends to overestimate and SEAM tends to underestimate the true value but none of these biases are statistically significant. Finally, in terms of comparing the MSEs (see Table 2.2) obtained by these four methods, we find that, as expected NLS has the minimum MSE compared to the two proposed methods and ID method, but the gain is very nominal considering the fact that NLS uses exact analytical form of the mean function. In practice, when

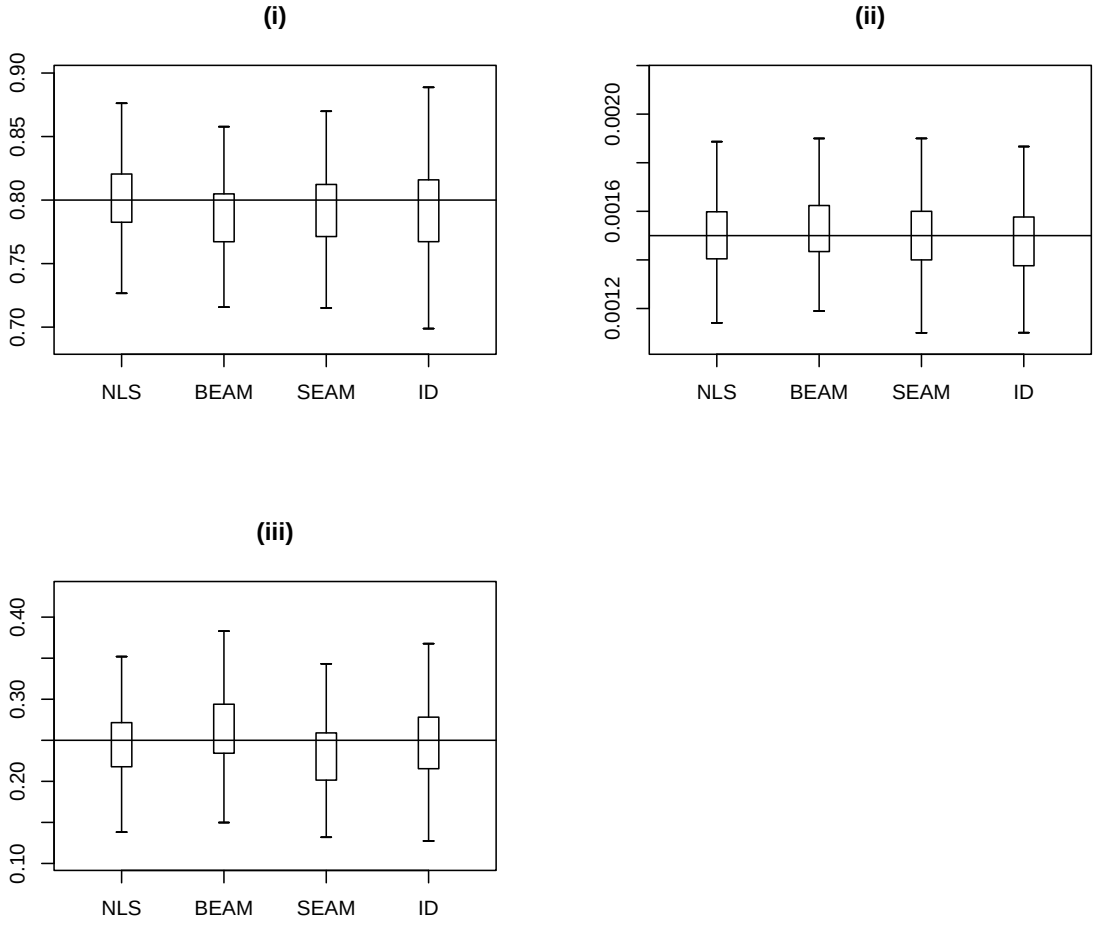


Figure 2.2: Box plots of point estimates: (i) $\hat{\theta}_1$'s (ii) $\hat{\theta}_2$'s (iii) $\hat{\sigma}$'s based on 1000 simulated data sets. (The horizontal solid line in each case represents the true value of the parameters.)

an analytically closed form for the mean function is not available, NLS is not applicable, but BEAM and SEAM will still work.

## 2.6 Discussion

The main objective of the data analysis (in Section 2.4) and simulation study (in Section 4) was to compare the performance of BEAM and SEAM with NLS for situations where a closed form analytical solution for system of differential equations is available. Results of data analysis and simulation study suggest that both of these methods provide results very close to the results obtained by the NLS method and therefore the Euler's approximation to the mean function described in Section (2.4) works quite accurately. Also the strikingly similar values of the MSE's of the parameters suggest that the proposed methods are as efficient as the NLS method when an analytic solution is available. Though for the model considered in this chapter, ID method also provided similar results, it requires certain assumptions to be satisfied. For example, this method requires that the maximum interval length between two consecutive time points should be $\mathrm{O}(n^{-1})$, which may or may not be true, specially for large longitudinal studies.

In this chapter we assumed that random errors are identically, independently distributed but methodologies presented here are not restricted to this assumption. This assumption can be relaxed by using a generalized nonlinear modeling framework (Davidian and Giltinan, 1995) with $\tilde{\mu}(t, \boldsymbol{\theta})$ as the mean function and $\sigma^2(t, \boldsymbol{\eta})$ as the variance function, which is assumed to be a known function up to the unknown parameter $\boldsymbol{\eta}$. In order to implement BEAM with this generalized nonlinear modeling framework we need to use suitable priors for the parameter $\boldsymbol{\eta}$ of the variance function. For the SEAM we have to replace the $SS(\boldsymbol{\theta})$ in (2.16) by a weighted least square criteria, where say the weights can be chosen to be inversely proportional to the variance function.

Further, in most of biomedical applications, data constitutes of several individuals and modeling of such data involves population specific parameters as well as individual-specific parameters, and that requires a mixed effects modeling framework. Future work

for this research consists of extending the proposed methodologies to the mixed effects model framework where data are subject to missingness and censoring.

One of the main advantages of BEAM and SEAM is that these do not require any restrictive assumptions other than those typically considered in nonlinear modeling. The proposed likelihood approximation method also provides a closed form approximation of the mean function, $\tilde{\mu}(t, \boldsymbol{\theta})$. Therefore these methods can be used to estimate the mean function at any time point lying within the close vicinity of the observed time range. This is a huge advantage as it avoids evaluating the numerical solution of the mean function at the parameter estimate again and again for interpolation/extrapolation. Because of the Bayesian framework, one of the key advantages of BEAM also lies in its ability to handle missing data that is very common in longitudinal studies. The availability of posterior distributions for the unknown parameters, also makes it straightforward to draw statistical inferences. At the same time advantage of SEAM comes from not only from its accuracy of estimation and weaker distributional assumptions but also from its computational convenience as compared to BEAM. Although BEAM provides estimates that are not only accurate but applicable with missing or censored data, there is no denying that this is a computationally intensive procedure. In comparison to BEAM, SEAM takes much less computation time, but SEAM is limited to handling only uncensored data. At the end, we will leave the choice between BEAM and SEAM up to readers, as both have their pros and cons.

For our proposed methods, we used the "naive" Euler's approximation method in both cases. We chose Euler's approach just for the sake of simplicity and also because it provided reasonable estimates for parameters in our simulation studies. However, other improved numerical methods (see Appendix A) can also be implemented within proposed methods, though that will mostly likely increase the computational time.

# Chapter 3

# Bayesian Inference in Non-linear Mixed Effect Models involving ODEs

## 3.1 Introduction

Nonlinear mixed effects (NLME) models are widely used in practice involving data sets, where repeated measurements are obtained for a number of individuals under varying experimental conditions. Nonlinear mixed effects models incorporate population level (fixed) effects as well as individual specific (random) characteristics and hence enable to make inferences for both random and fixed effects. The conceptual framework of a nonlinear mixed effects model was first introduced by Sheiner et al. (1972) to analyze the data pooled over all individuals and since then, there has been an explosion of further research (Davidian and Giltinan, 1995) with application to population pharmacokinetics/pharmacodynamics (PK/PD) models and physiologically based pharmacokinetic (PBPK) models based on nonlinear mixed effects modeling framework. Within the framework of NLME models, much of the interest is focused on representing the mean function (or mean trajectory), describing the dynamic relationship between the response and explanatory variables (such as time), by a system of ordinary differential equations (ODEs) whose parameters describe the different characteristics of the underlying population. A system of ODEs provides an attractive modeling tool to describe dynamic processes, where the interest is focused on modeling the rate of change over time rather than the static average value of the response variable, e.g., as in PK/PD models, viral dynamics etc. As an illustrative example, consider a very commonly used system of differential equations (Perelson et al., 1996; Han et al., 2002; Holte et al., 2003), which is used to describe the temporal expectation of virus and infected cell densities after initiation of

anti-retroviral treatment given by:

$$\begin{aligned} \frac{d\nu_1}{dt} &= -\delta\nu_1 + kT_0\nu_2, \quad \nu_1(0) = T_0^*, \\ \frac{d\nu_2}{dt} &= -c\nu_2, \quad \nu_2(0) = V_{I_0}, \\ \frac{d\nu_3}{dt} &= N_{fv}\delta\nu_1 - c\nu_3, \quad \nu_3(0) = V_{NI_0}. \end{aligned} \tag{3.1}$$

In this model, $\nu_1(t)$, $\nu_2(t)$ and $\nu_3(t)$ represent the density of infected cells at time $t$, the density of infectious virus at time $t$ and the density of non-infectious virus at time $t$ respectively. $k$ is the infectivity constant, $T_0$ is the density of un-infected cells at the initiation of treatment, $\delta$ is the death rate for infected cells, $c$ is the clearance rate for free virus, and $N_{fv}$ is the number of free virions produced per infected cell in its lifetime. $T_0^*$, $V_{I_0}$, and $V_{NI_0}$ are the known initial values (at $t$=0) for $\nu_1(t)$, $\nu_2(t)$ and $\nu_3(t)$ respectively. If unknown, these initial values can be treated as parameters in the model and can be estimated along with other parameters involved in the model. Assuming that the virus dynamics are in quasi-steady state prior to the initiation of the treatment i.e., $N_{fv}kT_0 = c$, Perelson et al. (1996) solved this system of ODEs for total viral density $V(t) = \nu_2(t) + \nu_3(t)$ at time $t$ and the solution is given by

$$V(t) = V_0\exp(-ct) + \frac{cV_0}{c-\delta}\left[\frac{c}{c-\delta}\{\exp(-\delta t) - \exp(-ct)\} - \delta t\exp(-ct)\right], \tag{3.2}$$

where $V_0 = V_{I_0} + V_{NI_0}$. However, in practice it turns out that this steady state assumption usually holds only during a short period after the initiation of treatment (Wu, 2005) and thus there are very few cases where it is actually possible to derive the closed form expression for the exact solution for a well-posed differential equation problem. The absence of a closed form analytical solution for the system of ODEs makes parameter estimation in such models, challenging and computationally demanding.

In this chapter, we aim to extend the *Bayesian Euler's approximation method (BEAM)* proposed in the previous chapter, to the nonlinear mixed effects framework involving a system of ODEs, for which an analytical solution is not available. In Section 3.2, we

describe a statistical framework for nonlinear mixed effects models followed by the description of *BEAM* for mixed effects models in Section 3.3. We then illustrate the method in Section 3.4 by applying it to a data on growth colonies of paramecium aurelium and a simulation study motivated by the real data analysis is then presented in Section 3.5, followed by the application of BEAM to the motivating example described by equation (1), in Section 3.6. Finally, in Section 3.7, we provide some general conclusions and directions for future research.

## 3.2 Nonlinear mixed effects models involving ODEs

The problem in hand can be described as follows. Let $y_{ij}$ denotes the $j$th observed response, for the $i$th individual measured at timepoint $t_{ij}$, for $i = 1, 2, ..., m$ and $j = 1, 2, ..., n_i$. For example, in the pharmacokinetic settings, $t_{ij}$ can be the time associated with the $j$th drug concentration for subject $i$. To keep our description simple, we considered time as the only dynamic explanatory variable in the model. However, methodologies proposed in this chapter can be extended to a more general case with multiple dynamic covariates. A statistical model can be written as,

$$y_{ij} = \mu(t_{ij}, \boldsymbol{\theta}_i) + \epsilon_{ij}, \quad \text{for } i = 1, 2, \ldots m \text{ and } j = 1, 2, ..., n_i \tag{3.3}$$

In equation (3.3), $\mu$ is the mean function describing the within-individual behavior, which depends on a vector of $p$ parameters, $\boldsymbol{\theta}_i$, specific to individual $i$. The random effects $\boldsymbol{\theta}_i$'s are assumed to arise from a common distribution with mean $\boldsymbol{\theta}$ and variance $\Sigma_\theta$. More specifically we can write

$$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{e}_i; \;\; E(\boldsymbol{e}_i) = \mathbf{0}, \;\; V(\boldsymbol{e}_i) = \Sigma_\theta \tag{3.4}$$

For example, in case of the model described by equation (3.1), the parameter vector in the model is $\boldsymbol{\theta}_i = (c_i, \delta_i, k_i, T_{0i}, N_{fvi})$. Some of these parameters may have known values or may be constant for all individuals and in such a case $\boldsymbol{\theta}_i$ will be the sub-

vector of remaining unknown individual specific parameters. For an individual $i$, the intra-individual error $\epsilon_{ij}$ corresponds to the measurement uncertainty associated with the observed response at time point $t_{ij}$. These random errors are assumed to be independently distributed with zero mean and constant variance across all measurements, i.e.,

$$E(\epsilon_{ij}) = 0 \text{ and } Var(\epsilon_{ij}) = \sigma^2 \text{ for } i = 1, 2, ..., m \text{ and } j = 1, 2, ..., n_i \tag{3.5}$$

For simplicity, errors are assumed to be iid though this assumption is clearly restrictive. However, the method described in this chapter can be extended to the case when we allow variance of errors to be of the form, $V(\boldsymbol{\epsilon}_i) = \sigma^2 D_i$ where $D_i$ is a $n_i \times n_i$ positive definite matrix, that may or may not depend on additional parameters and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, ..., \epsilon_{in_i})^T$. We will discuss this aspect briefly in Section 3.7. Further $\boldsymbol{e}_i$ and $\boldsymbol{\epsilon}_i$ are assumed to be independent. Then the likelihood function for this modeling framework assuming normality of the responses and random effects can be written as ,

$$L(\boldsymbol{\theta}, \Sigma_\theta, \sigma) \propto \prod_{i=1}^{m} \int \prod_{j=1}^{n_i} \frac{1}{\sigma} e^{-\frac{1}{2\sigma^2}(y_{ij} - \mu(t_{ij}, \boldsymbol{\theta}_i))^2} |\Sigma_\theta|^{-1/2} e^{-\frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\theta})^T \Sigma_\theta^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\theta})} d\boldsymbol{\theta}_i$$

$$= \prod_{i=1}^{m} \left(\frac{1}{\sigma^2}\right)^{\frac{n_i}{2}} |\Sigma_\theta|^{-1/2} \int e^{-\frac{1}{2\sigma^2} \sum_{j=1}^{n_i} (y_{ij} - \mu(t_{ij}, \boldsymbol{\theta}_i))^2 - \frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\theta})^T \Sigma_\theta^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\theta})} d\boldsymbol{\theta}_i \tag{3.6}$$

Ideally, the parameters involved in the NLME model can be estimated by maximizing the likelihood function, given by equation (3.6). If $\mu$ is a linear function in terms of parameters $\boldsymbol{\theta}_i$, the intergral in equation (3.6) can be evaluated to obtain an analytic expression. However, more often in case of NLME models, $\mu$ is a nonlinear function of parameters $\boldsymbol{\theta}_i$, making it impossible to obtain an analytic expression for the integral described in equation (3.6) and therefore classical approach such as maximum likelihood method for parameter estimation becomes analytically intractable. A common approach to handle the integral (with respect to $\boldsymbol{\theta}_i$) in (3.6) involves linearization of the nonlinear model by using either Taylor's series expansion (Beal and Sheiner, 1982; Lindstrom and Bates, 1990) or by applying Laplace's approximation to the likelihood (Wolfinger, 1993)

and then estimating parameters from the resulting approximated likelihood functions. A detailed description of these procedures is presented in Davidian and Giltinan (1990, p.151). A common feature of these approximation procedures is the availability of a closed form expression for the mean function $\mu(\cdot, \boldsymbol{\theta})$. In the context of many biological applications (e.g., PK/PD models) $\mu$ is obtained as the solution to a system of ODEs given by,

$$\frac{d(\boldsymbol{\nu}(t,\boldsymbol{\theta}))}{dt} = \mathbf{g}(t, \boldsymbol{\nu}(t,\boldsymbol{\theta})) \quad \forall\, \boldsymbol{\theta} \text{ and } t \neq t_0 \tag{3.7}$$

$$\text{and} \quad \boldsymbol{\nu}(t_0,\boldsymbol{\theta}) = \boldsymbol{\nu}_0(\boldsymbol{\theta}) \tag{3.8}$$

where $\boldsymbol{\nu}(\cdot) = (\nu_1(\cdot), \ldots, \nu_q(\cdot))^T$ represents the underlying vector of dynamics and $\boldsymbol{\nu}_0(\cdot)$ provides a set of known initial conditions (often free of $\boldsymbol{\theta}$). The $q$-vector valued function $\mathbf{g}(\cdot) = (g_1(\cdot), \ldots, g_q(\cdot))^T$ that describes the dynamics, is completely known up to the unknown parameter $\boldsymbol{\theta}$. Notice that (3.7) can equivalently be expressed with a set of $q$ ODEs, $\frac{d\nu_k}{dt} = g_k(t, \boldsymbol{\nu}(t,\boldsymbol{\theta}))$ for $k = 1, \ldots, q$. The mean function, $\mu(\cdot)$ is related to $\boldsymbol{\nu}$ by a completely known function $H\colon \mathbb{R}^q \to \mathbb{R}$ by $\mu(\cdot) = H(\boldsymbol{\nu}(\cdot))$. For instance, in the case of the model described by equation (3.1), the mean function $\mu(t, \boldsymbol{\theta}_i) = log(\nu_2(t, \boldsymbol{\theta}_i) + \nu_3(t, \boldsymbol{\theta}_i))$.

In the absence of a closed form solution to the system described by equation (3.7), the usual approach to overcome this problem, is to solve the system of ODEs numerically by using popular ODE solvers at a *given* set of values of the random effects $\boldsymbol{\theta}_i$'s and use those numerical solutions in the estimation procedure (Davidian and Giltinan, 1995). These numerical solutions obtained from the ODE solvers can also be used in nonlinear hierarchical Bayesian framework (Gelman et al., 1996; Wakefield, 1996; Lunn et al., 2002; Putter et al., 2002), to estimate the parameters involved in the model. The Fortran ODE solver `lsoda` (Petzold, 1987) and the `odesolve` package in `R` that provides an interface to the Fortran ODE solver are commonly used for this purpose. However, the success of most of these numerical approximation methods depends on a "good" choice of a starting value for $\boldsymbol{\theta}_i$'s and some characteristics of the system (e.g., the steepness). A "bad" starting value often leads to an unstable solution and creates numerical problems

for the optimization method that is followed by these ODE solvers.

Even though ODE solvers are widely used for estimating parameters in PK/PD modeling, it may be difficult to implement or lack control of essential numerical subroutines required to obtain the desired numerical solution for ODE and sometimes also turn out to be unstable especially in case of censored or missing data (Putter et al., 2002). Apart from such numerical instabilities, these methods are computationally intensive iterative procedure in which the system of ODEs must be solved at each time point for each individual and this becomes more complicated in the case of multi-compartmental problems with censored or missing data. For situations where the use of ODE solvers is combined with approximated likelihood using Taylor's series expansion (Lindstrom and Bates, 1990) or using Laplace's approximation (Wolfinger, 1993), the system of ODEs as well as the derivatives of the mean function (obtained by solving the system of differential equations), need to be evaluated for all random effects parameters, at each time point. The extent of computational burden increases with the increase in number of random effects parameters.

The objective of this research is to estimate the parameters involved in the NLME model (3.3-3.5) by providing a closed form approximation to the mean function described by the system of ordinary differential equations. In this research, we are mainly interested in the estimates of the "mean" of the random effects ($\boldsymbol{\theta}$), intra-individual variability ($\sigma^2$) and the variability associated with the random effects ($\Sigma_\theta$).

In this chapter, we extend the Bayesian Euler's approximation method (BEAM), earlier proposed for nonlinear fixed effects models (Chapter 2), to the nonlinear mixed effect framework. This approach combines the existing Bayesian framework for parameter estimation in PK/PD modeling (Gelman et al., 1996; Lunn et al., 2002; Han et al., 2002; Putter et al., 2002; Wakefield, 1996; Huang et al., 2004) with a numerical approximation method, providing an analytical closed form approximation for the system of ordinary differential equations. The advantages of this method lies in providing a closed form approximation for the solution of system of ordinary differential equations thus removing the need to repeat the use of numerical methods to solve the system of differential equa-

tions for every random effects parameter separately. This is a great advantage because not only this saves a lot of computational difficulty but also the approximated mean function can be incorporated in nonlinear Bayesian hierarchical modeling to obtain parameter estimates, without imposing any restrictive condition about the mean function. Due to its Bayesian framework this approach has the ability to handle the missing data by using data augmentation methods (Schafer, 1997) and the flexibility to handle sparse and unbalanced data by utilizing information across individuals sampled from the population. The availability of posterior distributions for the unknown parameters involved in the model, also makes it straightforward to draw statistical inferences. This method involves a numerical approximation method, viz. naive Euler's approximation method that provides reasonably good results in this case but there are more accurate numerical approximation methods, which can be used in place of Euler's method, without any loss of generality.

## 3.3 The Bayesian Euler's Approximation Method

Many different numerical approximation methods are available for computing approximate solutions to system of ODEs with a given initial condition such as the problem given by (3.7) and a detailed discussion of these methods can be found in the literature (Shampine, 1994; Lambert, 1991; Atkinson, 1978). The algorithm for *BEAM* is based on the Euler's approximation method. All the numerical approximation methods involve discretizing the time points by an amount $h$ known as the "step size," which is the distance between two consecutive time points. This step size may or may not be same for all consecutive time points, but for our description we assume $h$ to be constant over the range of time points, i.e., we assume that $t_k = t_0 + kh$ for $k = 0, 1, 2, ....$ Mostly, a uniform step size is used to simplify programming. It is easy to see that the solution to the system of ODEs given by equation (3.7) can be expressed as

$$\boldsymbol{\nu}(t, \boldsymbol{\theta}) = \int_{t_0}^{t} \mathbf{g}(s, \boldsymbol{\nu}(s, \boldsymbol{\theta}))ds + \boldsymbol{\nu}_0(\boldsymbol{\theta}) \tag{3.9}$$

which suggests the approximation, as $h \to 0$,

$$\boldsymbol{\nu}(t+h,\boldsymbol{\theta}) - \boldsymbol{\nu}(t,\boldsymbol{\theta}) = \int_t^{t+h} \mathbf{g}(s,\boldsymbol{\nu}(s,\boldsymbol{\theta}))ds \approx h\mathbf{g}(t,\boldsymbol{\nu}(t,\boldsymbol{\theta})) \tag{3.10}$$

and hence an approximation for $\mu(t,\boldsymbol{\theta}) = H(\boldsymbol{\nu}(t,\boldsymbol{\theta}))$ for all values of the parameter $\boldsymbol{\theta}$, where $H$ is a completely known continuous function. Thus, using (3.10) we can obtain a recurrence relation to approximate the mean function.

We now describe a method based on (3.10) to approximate the likelihood that arise from the model given by equations (3.3-3.7) and the observed data $\mathcal{D} = \{(Y_{ij}, t_{ij}) : i = 1,2,\ldots,m \text{ and } j = 1,2,\ldots,n_i\}$. Here $t_{i1} < t_{i2} < \cdots < t_{in_i}$, denote the observed time points for the $i^{th}$ individual in the data set and $\{Y_{ij} = Y(t_{ij}) : i = 1,2,\ldots,m \text{ and } j = 1,2,\ldots,n_i\}$ denote the response values. As the observed time points can be unevenly distributed we first consider a discretization by $N$ fixed time points $t_0 = t_1^0 < t_2^0 < ... < t_N^0$ such that $t_{k+1}^0 - t_k^0 = h$ for $k = 1,2,...,(N-1)$. In order to cover the range of observed time points we choose the maximum value for these fixed time points such that $t_N^0 \geq \max_{1\leq i\leq m} t_{in_i}$. The choice of $N$ (and hence that of $h$) will depend on $n_i$. Letting $\tilde{\boldsymbol{\nu}}_k(\boldsymbol{\theta}) \equiv \tilde{\boldsymbol{\nu}}(t_k^0,\boldsymbol{\theta})$ and $\tilde{\mu}_k(\boldsymbol{\theta}) \equiv \tilde{\mu}(t_k^0,\boldsymbol{\theta}) = H(\tilde{\boldsymbol{\nu}_k}(\boldsymbol{\theta}))$ for $k = 1,2,\ldots,N-1$, we can write

$$\tilde{\boldsymbol{\nu}}_{k+1}(\boldsymbol{\theta}) = \tilde{\boldsymbol{\nu}}_k(\boldsymbol{\theta}) + h\mathbf{g}(t_k^0, \boldsymbol{\nu}_k(\boldsymbol{\theta})) \tag{3.11}$$

$$\tilde{\mu}_k(\boldsymbol{\theta}) = H(\tilde{\boldsymbol{\nu}}_k(\boldsymbol{\theta})) \tag{3.12}$$

with initial conditions $\tilde{\boldsymbol{\nu}}_1(\boldsymbol{\theta}) = \boldsymbol{\nu}_0(\boldsymbol{\theta})$. This simple approximation defined by equations (3.11) and (3.12) is known as the naive Euler's approximation and forms the basis for most of the numerical approximation methods available in the literature. A brief description of some refined and numerically more accurate approximation methods that are used in practice is given in Appendix A. We now use (3.11) as the basis to a method which can be further improved using other numerical recipies at the cost of computational time. Now to define $\tilde{\mu}(t,\boldsymbol{\theta}_i)$ for any value of $t \in [t_1^0, t_N^0]$, we use a linear interpolation. More

precisely, given a time point $t \in [t_1^0, t_N^0]$, we define the labels,

$$L(t) = \sum_{k=1}^{N} I(t_k^0 \leq t). \tag{3.13}$$

Notice that for any time point $t \in [t_1^0, t_N^0]$, the function $L(t)$ takes the values in the range $\{1, \ldots, N\}$ and determines that how many $t_k^0$'s are less than the time point $t$ which in turn provides the lower limit of the interval that contains $t$, i.e., $t_{L(t)}^0 \leq t < t_{L(t)+1}^0$. The value of approximate mean function at any time point $t \in [t_1^0, t_N^0]$ is then given by,

$$\tilde{\mu}(t, \boldsymbol{\theta}) \equiv \tilde{\mu}_h(t, \boldsymbol{\theta}) = \tilde{\mu}_{L(t)}(\boldsymbol{\theta}) + \frac{t - t_{L(t)}^0}{t_{L(t)+1}^0 - t_{L(t)}^0}(\tilde{\mu}_{L(t)+1}(\boldsymbol{\theta}) - \tilde{\mu}_{L(t)}(\boldsymbol{\theta})) \tag{3.14}$$

where $L(t)$ is defined in (3.13) and $\tilde{\mu}_k(\boldsymbol{\theta})$'s are obtained from equation (3.11). It should be noticed that this approximation of mean function described by the equation (3.14) holds for all values of $\boldsymbol{\theta}$, therefore the approximation for an individual specific mean function can be obtained by substituting the individual specific parameter vector in the equation (3.14). Hence, we can approximate the true likelihood function of $(\boldsymbol{\theta}, \Sigma_\theta, \sigma)$ using the $\tilde{\mu}_h(t, \boldsymbol{\theta}_i)$ function in equations (3.3-3.5).

Notice that $\tilde{\mu}_h(t, \boldsymbol{\theta}) \to \mu(t, \boldsymbol{\theta})$ as $h \to 0$ (Lambert, 1991; p.151). In fact, $\tilde{\mu}_h(t, \boldsymbol{\theta}) = \mu(t, \boldsymbol{\theta}) + o(h)$ if we use the above method known as the "naive" Euler's method (eq 3.11). If we use the "improved" Euler's method (see Appendix A), then $\tilde{\mu}_h(t, \boldsymbol{\theta}) = \mu(t, \boldsymbol{\theta}) + o(h^2)$ and more generally Runge-Kutta method yields $\tilde{\mu}_h(t, \boldsymbol{\theta}) = \mu(t, \boldsymbol{\theta}) + o(h^4)$. Thus, it follows that as $h \to 0$, the likelihood function based on true mean function $\mu(t, \boldsymbol{\theta})$ can be well approximated by the likelihood function based on the approximating $\tilde{\mu}(t, \boldsymbol{\theta})$ function as defined by equation (3.14). Finally the approximated mean function is incorporated in the Bayesian framework to obtain estimates for parameters involved in the model. By using this approximated mean function (3.14), the first stage of Bayesian hierarchical model can be written as,

$$y_{ij} | (\boldsymbol{\theta}_i, \sigma^2) \overset{indep}{\sim} N(\tilde{\mu}(t_{ij}, \boldsymbol{\theta}_i), \sigma^2); \quad i = 1, 2, ..., m \text{ and } j = 1, 2, ..., n_i, \tag{3.15}$$

where $\tilde{\mu}(t, \boldsymbol{\theta})$ is as defined in equation (3.14). For practical applications, the $y_{ij}$'s may be the log-transformed (or more generally Box-Cox transformed) values depending on whether normality or log-normality is the more appropriate assumption for the data (Lunn et al., 2002).

The second stage of the Bayesian hierarchical model in this case, involves making distributional assumptions regarding the individual specific random effects $\boldsymbol{\theta}_i$:

$$\boldsymbol{\theta}_i|(\boldsymbol{\theta}, \Sigma_\theta) \sim \text{MVN}_p(\boldsymbol{\theta}, \Sigma_\theta), \quad i = 1, 2, ..., m \tag{3.16}$$

where $\text{MVN}_p(.,.)$ represents a $p$-dimensional multivariate normal distribution with mean $\boldsymbol{\theta}$ which is a vector of $p$ fixed effect parameters and variance-covariance matrix $\Sigma_\theta$. The third and the final stage of this hierarchical model consists of the prior distributions given as,

$$\boldsymbol{\theta}|(\Sigma_\theta^{-1}, \sigma^{-2}) \sim \text{MVN}_p(\boldsymbol{\theta}_0, H_0); \;\; \Sigma_\theta^{-1}|\sigma^{-2} \sim \text{W}_p(R_0, \rho_0); \;\; \sigma^{-2} \sim \text{G}(a_0, b_0) \tag{3.17}$$

where $\text{G}(\cdot, \cdot)$ represents the gamma distribution with shape parameter $a_0$, scale parameter $b_0$ with mean $a_0 b_0$, $\text{W}_p(\cdot, \cdot)$ denotes a $p$-dimensional Wishart distribution with a positive definite scale matrix $R_0$ and degrees of freedom $\rho_0$ with mean $\rho_0 R_0$. The values of $a_0,\ b_0,\ \boldsymbol{\theta}_0,\ H_0,\ R_0,\ \rho_0$ are assumed to be known, and are used to elicit prior information when available. In the lack of prior information, we choose these known quantities to reflect on prior ignorance by imposing priors with large variances. When the number of subjects is not too small, the prior distribution of $\boldsymbol{\theta}$ will have relatively little influence on the posterior distribution of $\boldsymbol{\theta}$. A detailed discussion about the choice of prior values are given in Natarajan and Kass (2000). From the literature (Gelfand and Smith, 1990; Wakefield et al., 1994; Wakefield, 1996; Huang et al., 2004), it has been established that for model described by equations (3.15-3.17), the full conditional distributions for the parameters $\sigma^2, \boldsymbol{\theta}$ and $\Sigma_\theta$ are given by,

$$\sigma^{-2}|(\boldsymbol{\theta}, \Sigma_\theta, \boldsymbol{\Theta}, \mathcal{D}) \sim \text{G}\left(a_0 + \frac{\sum_{i=1}^m n_i}{2}, A^{-1}\right) \tag{3.18}$$

$$\boldsymbol{\theta}|(\sigma^{-2}, \Sigma_\theta^{-1}, \boldsymbol{\Theta}, \mathcal{D}) \sim \text{MVN}_p(B^{-1}C, B^{-1}) \tag{3.19}$$

$$\Sigma_\theta^{-1}|(\boldsymbol{\theta}, \sigma^{-2}, \boldsymbol{\Theta}, \mathcal{D}) \sim \text{W}_p(D^{-1}, m + \rho_0) \tag{3.20}$$

where $A = b_0^{-1} + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}[y_{ij} - \tilde{\mu}(t_{ij}, \boldsymbol{\theta}_i)]$, $B = m\Sigma_\theta^{-1} + H_0^{-1}$, $C = \Sigma_\theta^{-1}\sum_{i=1}^{m}\boldsymbol{\theta}_i + H_0^{-1}\boldsymbol{\theta}_0$, $D = R_0^{-1} + \sum_{i=1}^{m}(\boldsymbol{\theta}_i - \boldsymbol{\theta})(\boldsymbol{\theta}_i - \boldsymbol{\theta})^T$ and $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_i : \ i = 1, \ldots, m\}$ denote the set of all individual specific random effects.

The full conditional distribution of each $\boldsymbol{\theta}_i$, given the remaining parameters and the data, is proportional to

$$exp\left\{\frac{1}{2\sigma^2}\sum_{j=1}^{n_i}[y_{ij} - \tilde{\mu}_{ij}(\boldsymbol{\theta}_i)]^2 - \frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\theta})\Sigma_\theta^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\theta})^T\right\} \tag{3.21}$$

This framework allows us to use Gibbs sampling to draw samples for $\sigma^{-2}, \boldsymbol{\theta}$ and $\Sigma_\theta^{-1}$, from the posterior distributions, while we need to use Metropolis-Hastings algorithm to draw samples for $\boldsymbol{\theta}_i$. Though we used WinBUGS to draw samples from posterior distributions, here we give a brief outline of the iterative MCMC algorithm (Huang et al., 2004) for our models.

1. Initialize the iteration of the chain at $l = 0$ and start with some initial values,

$$S^{(0)} = (\sigma^{-2(0)}, \boldsymbol{\theta}^{(0)}, \Sigma_\theta^{-1(0)}, \boldsymbol{\Theta}^{(0)})^T$$

2. Obtain $S^{(l)}$ from $S^{(l-1)}$ in the following way:

   (a) Draw $\sigma^{-2(l)} \sim \pi(\sigma^{-2}|(\boldsymbol{\theta}^{(l-1)}, \Sigma_\theta^{-1(l-1)}, \boldsymbol{\Theta}^{(l-1)}, \mathcal{D}))$ using equation (3.18).

   (b) Draw $\boldsymbol{\theta}^{(l)} \sim \pi(\boldsymbol{\theta}|(\sigma^{-2(l)}, \Sigma_\theta^{-1(l-1)}, \boldsymbol{\Theta}^{(l-1)}, \mathcal{D}))$ using equation (3.19).

   (c) Draw $\Sigma_\theta^{-1(l)} \sim \pi(\Sigma_\theta^{-1}|(\boldsymbol{\theta}^{(l)}, \sigma^{-2(l)}, \boldsymbol{\Theta}^{(l-1)}, \mathcal{D}))$ using equation (3.20) and

   (d) For $\boldsymbol{\theta}_i^{(l)}$, generate a new value $\boldsymbol{\phi}$ from a symmetric proposal density $q(\boldsymbol{\phi}|\boldsymbol{\theta}_i^{(l-1)})$.

Evaluate the acceptance probability of the move, given by,

$$\alpha(\boldsymbol{\phi}|\boldsymbol{\theta}_i^{(l-1)}) = min\left\{1, \frac{\pi(\boldsymbol{\phi}|\sigma^{-2(l)}, \boldsymbol{\theta}^{(l)}, \Sigma_\theta^{-1(l)}, \boldsymbol{\Theta}_{\{i\}}^{(l-1)}, \mathcal{D})}{\pi(\boldsymbol{\theta}_i^{(l)}|\sigma^{-2(l)}, \boldsymbol{\theta}^{(l)}, \Sigma_\theta^{-1(l)}, \boldsymbol{\Theta}_{\{i\}}^{(l-1)}, \mathcal{D})}\right\}$$

Also, independently sample a $u$ from the uniform $(0, 1)$ and if $u \leq \alpha(\boldsymbol{\phi}|\boldsymbol{\theta}_i^{(l-1)})$ the move is accepted else stay at $\boldsymbol{\theta}_i^{(l-1)}$. In other words, if the move is accepted then $\boldsymbol{\theta}_i^{(l)} = \boldsymbol{\phi}$ and otherwise $\boldsymbol{\theta}_i^{(l)} = \boldsymbol{\theta}_i^{(l-1)}$. Set $\boldsymbol{\Theta}^{(l)} = (\boldsymbol{\theta}_1^{(l)}, \ldots, \boldsymbol{\theta}_m^{(l)})$.

3. Move from chain $l$ to $l+1$ using step (2) and repeat until the Markov chain $\{S^{(l)},\ l = 1, 2, \ldots\}$ converges to the true conditional posterior distribution of $\theta_i$ (3.21).

In practice, `WinBUGS` uses a Metropolis algorithm based on a normal proposal distribution with mean as the current value of the parameter and variance determined by tuning over the first 4000 iterations to achieve an acceptance rate between 20% and 40%. In order to diagnose convergence of algorithms, we used graphical techniques such as history plots (available in `WinBUGS`) of the values of the multiple chains for each parameter. Based on these plots we decided upon an initial number of burn-in iterations (e.g., at least 4000) followed by say $B = 2000$ samples per chain drawn afterwards. We performed MCMC sampling based on three parallel chains, therefore all the posterior summaries are based on a total of $3B = 6000$ samples. Mean of posterior distribution for each of unknown parameters was taken to be the posterior estimate of that unknown parameter along with a 95% posterior interval formed by 2.5% and 97.5% posterior percentiles.

## 3.4 Analysis of Growth Colonies of *Paramecium Aurelium*

In this section we will illustrate the application of BEAM to the data set used in Chapter 2 (Section 2.4), in a mixed effects set up. For a detailed description of this experiment, we encourage readers to read the book by Diggle (1990). This data set describes the growth of three closed colonies of *paramecium aurelium* in a nutritive medium on a 19

days period and one of the main goals of this experiment was to develop a dynamic model for the growth count (say $x_{ij}$'s) of paramecium aurelium, as a function of time $t$. Here, the three closed colonies of *paramecium aurelium* are considered as three subjects in nonlinear mixed effects set up. The data is assumed to follow a log-normal distribution and is described as,

$$\begin{aligned} y_{ij} = log\{x_{ij}\} &= log\{\nu(t_{ij}, \boldsymbol{\theta}_i)\} + \epsilon_{ij}, \quad \text{for } i = 1, 2, 3 \text{ and } j = 1, 2, \ldots, 19 \\ &= \mu(t_{ij}, \boldsymbol{\theta}_i) + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0, \sigma^2) \end{aligned} \tag{3.22}$$

where $x_{ij}$ is the observed growth count corresponding to the $i^{th}$ subject, at the time point $t_{ij}$ (measured in days) and $\mu(t_{ij}, \boldsymbol{\theta}_i) = \log\{\nu(t_{ij}, \boldsymbol{\theta}_i)\}$. Next, it is assumed that for all individuals $i = 1, \ldots, m$, $\nu(t, \boldsymbol{\theta})$ follows the standard 2 parameter logistic growth curve described by the non-linear differential equation

$$\begin{aligned} \frac{d\nu(t, \boldsymbol{\theta})}{dt} &= g(t, \nu(t, \boldsymbol{\theta})) \\ &= \nu(t, \boldsymbol{\theta})(\theta_1 - \theta_2 \nu(t, \boldsymbol{\theta})), \quad \forall\ t \text{ and } \nu(t_1, \boldsymbol{\theta}) = y_0 = 2 \end{aligned} \tag{3.23}$$

where $t_1 = t_0 = 0$ is the initial time point for all three subjects. Equivalently, we can express the equation(3.23) in terms of $\mu(\cdot)$:

$$\begin{aligned} \frac{d\mu(t, \boldsymbol{\theta})}{dt} &= g^*(t, \mu(t, \boldsymbol{\theta})) \\ &\equiv \theta_1 - \theta_2 \exp(\mu(t, \boldsymbol{\theta})), \quad \forall\ t \text{ and } \mu(t_0, \boldsymbol{\theta}) = log(2) \end{aligned} \tag{3.24}$$

Although we proposed our estimation methods for situations where an analytical closed form solution is not available for the system of differential equations, here an analytically closed form solution is actually available. This choice of the simple growth curve model allows us to compare the performance of the estimates based on the proposed approach, to the estimates obtained by the iterative generalized least square method (GLS) proposed by Lindstrom and Bates (1990). This approach is based on conditional first order linearization and requires a closed form mean function be available. For a detailed de-

scription of this iterative GLS approach involving a closed form nonlinear mean function and mixed effects framework, one may read the article by Lindstrom and Bates (1990) or a classic book by Davidian and Giltinan (1995). For this data analysis and the simulation studies presented in next section, we used `nlme` function in `R` to obtain parameter estimates using the iterative GLS approach which is based on the analytical solution given by

$$\begin{aligned} \mu(t,\boldsymbol{\theta}) &= \log(\nu(t,\boldsymbol{\theta})) \\ &= \log(\theta_1) + \mu(0) + t\theta_1 - \log[\theta_2 e^{\mu(0)}(e^{t\theta_1} - 1) + \theta_1] \end{aligned} \tag{3.25}$$

To approximate the likelihood using BEAM, we chose $N = 19$. This choice of the tuning parameter $N$ (or equivalently, $h = \frac{t_N^0 - t_1^0}{N}$) is not based on any analytical work, rather the choice is mainly driven by computational convenience. In general, the finer the grid points (with large $N$ and hence small $h$) is chosen, the better the likelihood will be approximated. Alternatively, one may also use improved numerical approximation methods (such as those described in the Appendix A) at the cost of computing time. To implement the BEAM on this log transformed data, we followed exact same steps as described in Section 3.3, coupled with equations (3.22) and (3.24). Assuming that the random effects involved in the model $\boldsymbol{\theta}_i = (\theta_{1i}, \theta_{2i})$ follows log normal distribution or equivalently assuming $\boldsymbol{\theta}_i^* = \log(\boldsymbol{\theta}_i) = (\log(\theta_{1i}), \log(\theta_{2i}))^T$ normally distributed with mean vector $\boldsymbol{\theta}^* = (\theta_1^*, \theta_2^*)^T$, the Bayesian framework assumed for parameter estimation can be written as,

$$\begin{aligned} \boldsymbol{\theta}_i^* &\sim \text{MVN}_2(\boldsymbol{\theta}^*, \boldsymbol{\Sigma}_{\theta^*}) \\ \boldsymbol{\theta}^* &\sim \text{MVN}_2(\boldsymbol{\theta}_0, H_0) \\ \boldsymbol{\Sigma}_{\theta^*}^{-1} &\sim \text{W}_2(R_0, \rho_0) \text{ and } \sigma^{-2} \sim \text{G}(a_0, b_0) \end{aligned} \tag{3.26}$$

Where $\boldsymbol{\Sigma}_{\theta^*}$ is the $(2 \times 2)$ variance-covariance matrix associated with the random effects vector $\boldsymbol{\theta}_i^*$. The values for hyper-parameters are chosen as follows,

$$a_0 = 0.01, b_0 = 100, \rho_0 = 2, \boldsymbol{\theta}_0 = (0,0)^T, H_0 = \text{diag}(10,10), R_0 = \text{diag}(0.1, 0.1)$$

For the MCMC runs required within the BEAM, we generated samples based on three parallel Markov chains with an initial burn-in of 4000 iterations followed by 2000 post-burn-in samples per chain, giving us a total of 6000 approximate samples from the posterior distributions of $\theta^*$, $\sigma^{-2}$, and $\Sigma_{\theta^*}^{-1}$. Samples for $\boldsymbol{\Sigma}_{\theta^*}$ can be obtained by $\boldsymbol{\Sigma}_{\theta^*} = (\boldsymbol{\Sigma}_{\theta^*}^{-1})^{-1}$ and similarly samples for $\sigma^2$ can be obtained by $\sigma^2 = (\sigma^{-2})^{-1}$. Convergence of the chains was diagnosed visually by inspecting simultaneous trace and acf plots of the values of all three chains, for each parameter. Plots showing a good mixing of chains were a reasonable indication for convergence. The software `WinBUGS` freely available at the website `http://www.mrc-bsu.cam.ac.uk/bugs/` was used to perform all of the computations for this data analysis. Finally, the MC estimate of the mean and standard deviation of the posterior distributions for these parameters were used as point estimates and standard errors, respectively. Results from BEAM and NLME estimation approaches are presented in Table 3.1. For NLME estimation approach, estimated standard errors (SE) of the variance parameters are not given in the table due to the unavailability of these estimates by using the `NLME` function in `R`. It can be observed from Table 3.1 that estimated standard errors for $\theta_1^*$ and $\theta_2^*$ using NLME are smaller than those obtained by using BEAM. This difference in estimated standard errors may be attributed to the fact that the `NLME` function uses iterative generalized least-square method based on the standard asymptotic theory, while the estimates from BEAM are based on finite sample calculations. We investigate this aspect again in the next section in the context of results from the simulation study. For the real data analysis, the estimated covariance between the random effects was found to be approximately zero and therefore we only reported the squared root of the variance estimates of random effects $\theta_1^*$ and $\theta_2^*$, denoted by $\sqrt{\Sigma_{\theta_1^*}}$ and $\sqrt{\Sigma_{\theta_2^*}}$ respectively . Additionally, in order to compare the performance of these estimation approaches, we plotted estimated mean function corresponding to

Table 3.1: Results of estimation of parameters in the logistic growth model to the data set on growth colonies of the bacteria Paramecium Aurelium using NLME and BEAM.

| Parameters | Method | |
|---|---|---|
| | NLME | BEAM |
| $\theta_1^*$ | -0.177 | -0.217 |
| SE | 0.029 | 0.182 |
| $\theta_2^*$ | -6.459 | -6.463 |
| SE | 0.043 | 0.209 |
| $\sigma$ | 0.185 | 0.223 |
| SE | - | 0.022 |
| $\sqrt{\Sigma_{\theta_1^*}}$ | 0.042 | 0.210 |
| SE | - | 0.236 |
| $\sqrt{\Sigma_{\theta_2^*}}$ | 0.022 | 0.223 |
| SE | - | 0.287 |

BEAM and NLME approaches along with the observed data points corresponding to three subjects, in Figure 3.1.

From Figure 3.1, both approaches seem to perform very similarly in capturing the trajectory of mean function reasonably well. It can also be observed from Figure 3.1 that BEAM provides data fits close to the fits obtained by using the nonlinear likelihood estimation approach that uses the exact form analytical solution of the mean function. Because of its Bayesian framework, BEAM provides mean estimates along with its posterior distribution and Figure 3.1 shows that 95% confidence band (posterior credible interval) for estimated mean trajectory by using BEAM, covers the observed data points corresponding to all three subjects. This 95% posterior band was formed by calculating 2.5% and 97.5% posterior percentiles of the function $\tilde{\mu}(t, \theta)$ at each observed time point.

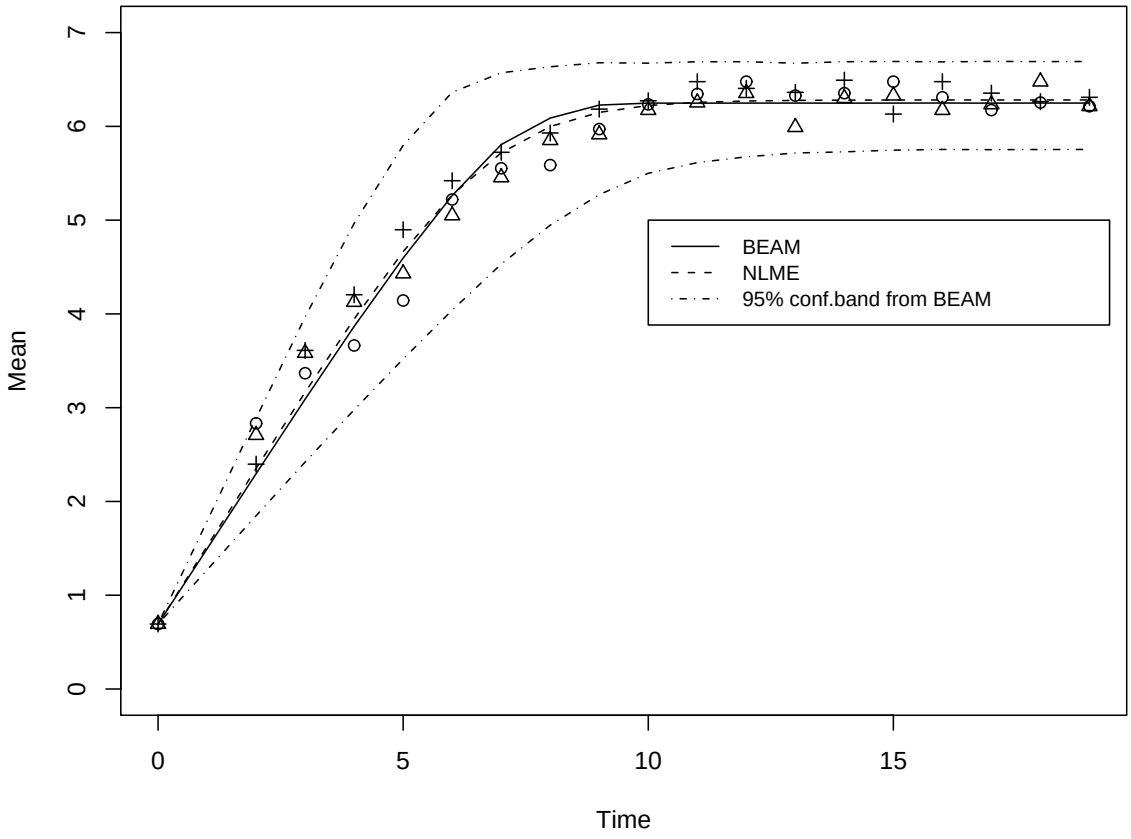


Figure 3.1: Plots of observations on growth colonies of paramecium aurelium (in log scale) and the estimated mean trajectories obtained by NLME and BEAM along with 95% posterior confidence band from BEAM.

## 3.5 Simulation Study

A simulation study, motivated by the above real data analysis was carried out to compare the performance of the proposed BEAM to the NLME procedure, in terms of estimation accuracy and efficiency. For the simulation study, data is generated using the model (3.24) with $\mu(t, \boldsymbol{\theta}_i)$ given by the closed form solution defined in equation (3.25). We

chose the observed time points in the simulated data set same as given in the real data set and hence the sample size same as the real data set, i.e., $n_i = 19$ for $i = 1, 2, 3$. True values of the parameter were set at $\boldsymbol{\theta}^* = (\log(0.8), \log(0.0015))$, $\Sigma_{\theta^*} = 0.1\mathrm{I}_{2\times 2}$ and $\sigma = 0.25$. These true values of parameters for data generation are chosen based on the estimates obtained for the real data application. To fit the model by BEAM, we chose same number of fixed time points as observed time points and the same prior distributions as used in the real data analysis. This simulation setup was replicated for 500 Monte Carlo runs. For the MCMC runs required within BEAM, we generated samples based on three parallel Markov chains with an initial burn-in of 4000 iterations followed by 2000 post-burn-in samples per chain, giving us a total of 6000 approximate samples from the posterior distributions of $\theta_1^*$, $\theta_2^*$, $\sigma$, $\Sigma_{\theta_1^*}$ and $\Sigma_{\theta_2^*}$. Initial simulation runs showed that the covariance between random effects is close to zero and hence we did not include results corresponding to that in the table. A summarization of the comparative study of BEAM and NLME approach, based on this simulation are given in Table 3.2 and Figure 3.2. In Table 3.2, we summarize our finding in terms of (i) the bias, which is the difference between the MC mean of the point estimates and the true value of a parameter; (ii) the estimated standard error (ESE), which is the the MC mean of the standard errors of the parameter estimates, (iii) the Monte Carlo simulation standard error (MCSE), which is the standard deviation of the estimates and (iv) the mean square error (MSE) obtained as $\text{Bias}^2 + \text{MCSE}^2$.

From Table 3.2, it seems that both BEAM and NLME performed almost same in terms of MSE's. It can be noticed that for NLME, the ESEs for fixed effects are smaller than the MCSEs. This supports the conclusion drawn in the previous section that NLME underestimates the standard errors associated with the parameter estimates. On the other hand in case of BEAM, we observe that the corresponding ESEs are larger than the corresponding MCSEs and this overestimation can be result of finite sample uncertainty with only $m = 3$ subjects.

For a better understanding of the MC distribution of the parameter estimates in comparing the BEAM with the NLME method, we present box plots of the estimates

Table 3.2: Simulation Results for the logistic growth model for colonies of paramecium aurelium (on log scale) using NLME and BEAM, with 500 MC Runs.

| Parameters | Estimates | Method | |
|---|---|---|---|
| | | NLME | BEAM |
| $\theta_1^*$ | Bias | 0.010 | -0.035 |
| | MCSE | 0.136 | 0.136 |
| | ESE | 0.121 | 0.186 |
| | MSE | 0.0180 | 0.0200 |
| $\theta_2^*$ | Bias | 0.015 | 0.016 |
| | MCSE | 0.145 | 0.147 |
| | ESE | 0.129 | 0.202 |
| | MSE | 0.0210 | 0.0220 |
| $\sigma$ | Bias | -0.002 | 0.009 |
| | MCSE | 0.019 | 0.021 |
| | ESE | - | 0.020 |
| | MSE | 0.0004 | 0.0005 |
| $\sqrt{\Sigma_{\theta_1^*}}$ | Bias | -0.050 | 0.014 |
| | MCSE | 0.098 | 0.126 |
| | ESE | - | 0.189 |
| | MSE | 0.0120 | 0.0120 |
| $\sqrt{\Sigma_{\theta_2^*}}$ | Bias | -0.051 | 0.012 |
| | MCSE | 0.111 | 0.133 |
| | ESE | - | 0.214 |
| | MSE | 0.0150 | 0.0140 |

obtained by both methods in Figure 3.2. The horizontal solid line in each case represents the true value of the parameter. Figure 3.2 (i) reveals that the inter-quartile range of the $\theta_1^*$ estimates from both BEAM and NLME contain the true value of parameters. Similarly for $\theta_2^*$, we observe from Figure 3.2 (ii) that both methods perform almost identically. For

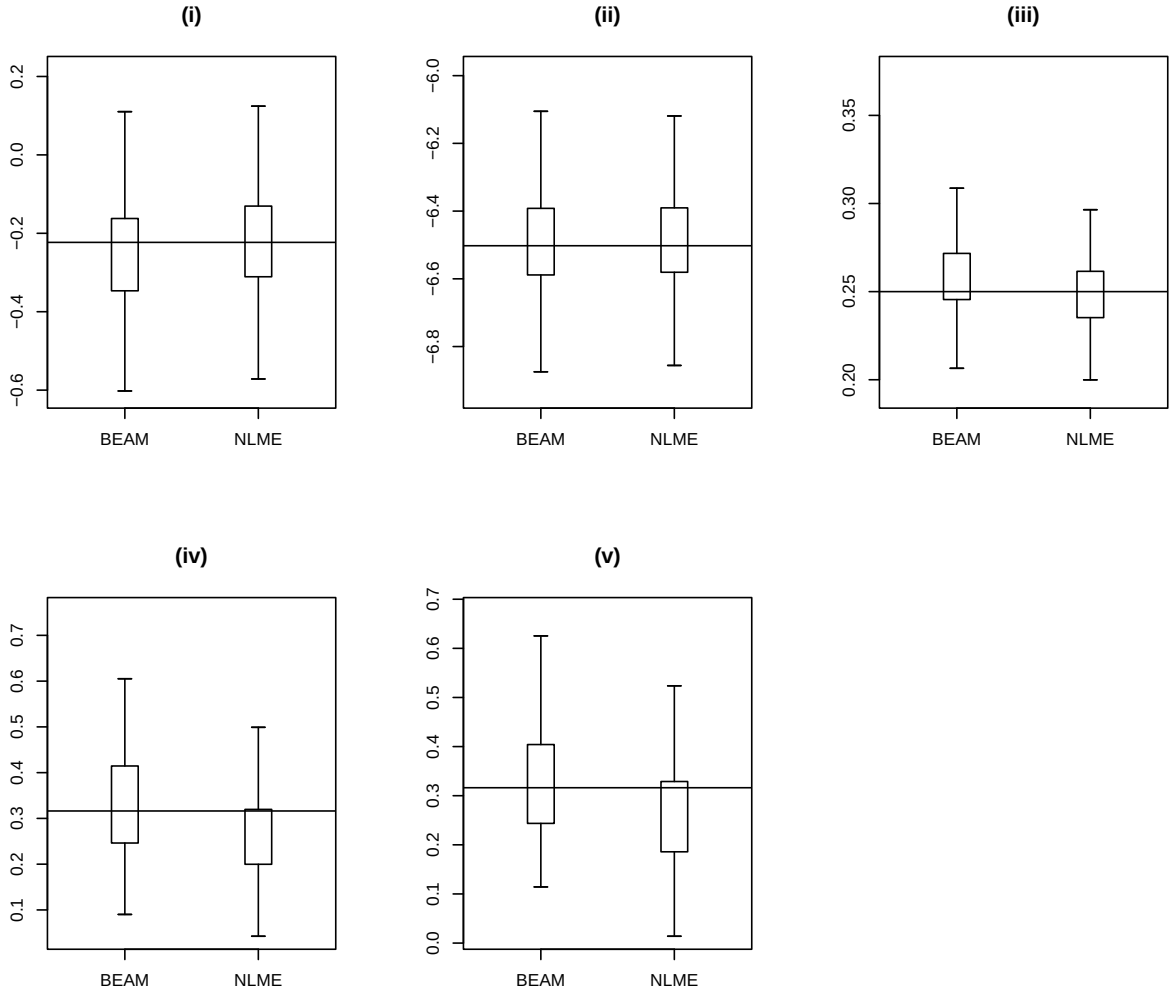


Figure 3.2: Box plots of point estimates: (i) $\hat{\theta}_1^*$'s (ii) $\hat{\theta}_2^*$'s (iii) $\hat{\sigma}$'s (iv) $\sqrt{\hat{\Sigma}_{\theta_1^*}}$'s (v) $\sqrt{\hat{\Sigma}_{\theta_2^*}}$'s based on 500 simulated data sets. (The horizontal solid line in each case represents the true value of the parameters.)

$\sigma$, Figure 3.2 (iii) indicates that BEAM tends to slightly overestimate whereas NLME tends to underestimate the true value, but the inter-quartile range of the estimates from both BEAM and NLME contain the true value of parameters in this case as well. It can also be noticed that in terms of MSE, NLME is slightly more efficient (with smaller

MSE) than BEAM for estimating $\theta_1^*$, $\theta_2^*$ and $\sigma$. However, the gain is very nominal considering the fact that NLME uses exact analytical form of the mean function. On the other hand, box plots for estimated inter-individual variance parameters $\sqrt{\Sigma_{\theta_1^*}}$ and $\sqrt{\Sigma_{\theta_2^*}}$ show that the BEAM seems to perform better than NLME (Figure 3.5 (iv) and (v)), where NLME tends to underestimate the inter-individual variance parameters. This advantage of BEAM is significant considering the fact that NLME uses exact analytical form of the mean function and in practice, when an analytically closed form for the mean function is not available, NLME is not applicable, but BEAM will still work.

## 3.6 HIV Model Revisited

In this section we revisit the HIV model described by equation (3.1) and apply the BEAM to estimate the parameters involved in this model. Han et al. (2002) presented the data set consisting of observations for total viral density measured at $0, 2, 4, 6, 12, 18, 24, 30, 36, 42, 48$ hours, and at days $3, 4, 5, 6$. These observations were measured on five individuals and the data follows a log-normal distribution with mean function described by the system of ODEs (3.1). As described in the Section 3.1, the total viral density ($V(t)$) can be expressed as a sum of the density of infectious virus $\nu_2(t)$ and the density of non-infectious virus, $\nu_3(t)$ at time $t$. The statistical model for the log transformed data can be written as,

$$\begin{aligned} \log(y_{ij}) &= \mu(t_{ij}, \boldsymbol{\theta}_i) + \epsilon_{ij}, \quad \text{for } i = 1, 2, \ldots 5 \text{ and } j = 1, 2, ..., 15 \\ &= \log(V(t_{ij}, \boldsymbol{\theta}_i)) + \epsilon_{ij}, \end{aligned} \tag{3.27}$$

where $V(t, \boldsymbol{\theta}_i) = \nu_2(t, \boldsymbol{\theta}_i) + \nu_3(t, \boldsymbol{\theta}_i)$ for all time points in the data set. We are assuming $c_i$ and $\delta_i$ as random effects denoted by the parameter vector $\boldsymbol{\theta}_i = (c_i, \delta_i)$ and rest of the parameters $(N_{fv}, k, T_0)$ involved in the system of ODEs are assumed to be known. Assuming that the virus dynamics are in quasi-steady state prior to the initiation of the treatment i.e., $N_{fv}kT_0 = c$, Perelson et al. (1996) solved this system of ODEs for total

viral density $V(t) = \nu_2(t) + \nu_3(t)$ at time $t$ and the solution is given by,

$$V(t) = V_0 \exp(-ct) + \frac{cV_0}{c-\delta}\left[\frac{c}{c-\delta}\{\exp(-\delta t) - \exp(-ct)\} - \delta t \exp(-ct)\right], \quad (3.28)$$

where $V_0 = V_{I_0} + V_{NI_0}$. Han et al. (2002) used this closed form expression of the total viral density to estimate the parameters involved in the model. However, in practice it turns out that this steady state assumption usually holds only during a short period after the initiation of treatment (Wu, 2005) and in the absence of this assumption, the system of ODEs described by equation (3.1) does not have the closed form solution given by equation (3.28). For our model fitting, we do not assume the steady state assumption and therefore the closed form mean function given by equation (3.28) is not being used by the Bayesian Euler's approximation method.

To implement BEAM for estimating parameters, we followed the exact same steps as described in Section 3.3, coupled with equations (3.27) and the log-transformed ODEs given by,

$$\begin{aligned}
\frac{d\log(\nu_1)}{dt} &= -\delta + kT_0\frac{\nu_2}{\nu_1}, \quad \log(\nu_1(0)) = \log(T_0^*), \\
\frac{d\log(\nu_2)}{dt} &= -c, \quad \log(\nu_2(0)) = \log(V_{I_0}), \\
\frac{d\log(\nu_3)}{dt} &= N_{fv}\delta\frac{\nu_1}{\nu_3} - c, \quad \log(\nu_3(0)) = log(V_{NI_0}).
\end{aligned} \quad (3.29)$$

Further we assumed that the random effects vector $\boldsymbol{\theta}_i$ also follows a log-normal distribution with the restriction $(c_i > \delta_i)$ and the Bayesian framework for the parameter estimation is as follows.

$$\begin{aligned}
\boldsymbol{\theta}_i^* = (\log(c_i), \log(\delta_i))^T &\sim \text{MVN}_2(\boldsymbol{\theta}^*, \boldsymbol{\Sigma}_{\theta^*}) \\
\boldsymbol{\theta}^* = (\log(c), \log(\delta))^T &\sim \text{MVN}_2(\boldsymbol{\theta}_0, H_0) \\
\boldsymbol{\Sigma}_{\theta^*}^{-1} &\sim \text{W}_2(R_0, \rho_0) \text{ and } \sigma^{-2} \sim \text{G}(a_0, b_0)
\end{aligned} \quad (3.30)$$

Where $\boldsymbol{\Sigma}_{\theta^*}^{-1}$ is the $(2 \times 2)$ inverse variance-covariance matrix associated with the random

effects vector $\boldsymbol{\theta}_i^*$. The values for hyper-parameters are chosen as follows,

$$a_0 = 4.5, b_0 = 9, \rho_0 = 3, \boldsymbol{\theta}_0 = (1.1, -1.0)^T, H_0 = \text{diag}(10, 10), R_0 = \text{diag}(2.5, 2.5)$$

For this study, the initial values for the system of ODEs were assumed to be known and are given by $\log(T_0^*) = 8.006$, $\log(V_{I_0}) = 12.101$, $\log(V_{NI_0}) = 11$ and to avoid the problem of identifiability we assumed $kT_0 = 0.015$. We also estimated parameters involved in this model by using informative priors and the values for hyper-parameters in case of informative prior are chosen as follows,

$$a_0 = 4.5, b_0 = 9, \rho_0 = 3, \boldsymbol{\theta}_0 = (1.1, -1.0)^T, H_0 = \text{diag}(0.1, 0.01), R_0 = \text{diag}(2.5, 2.5)$$

These values for prior parameters and other known parameters were chosen based on the priors used in Han et al. (2002) and Holte et al. (2003), for the same HIV experiment. Estimation results from BEAM along with the estimates presented in Han et al. (2002) for the same data set, using NLME estimation approach (Lindstrom and Bates, 1990) and the Bayesian hierarchical modeling with mean function given by equation (3.28) are given in the Table 3.3. The Bayesian Euler's approximation method was used with

Table 3.3: Results of estimation of parameters in the HIV model using BEAM, NLME and Bayesian hierarchical modeling involving a closed form expression for the mean function.

| Parameters | Steady State Assumption | | No Assumption | |
|---|---|---|---|---|
| | NLME | Han et al. (2002) | BEAM | |
| | | Info. Prior | Noninfo. Prior | Info. Prior |
| $\theta_1^*$ | 0.935 | 1.01 | 0.643 | 0.983 |
| SE | 0.282 | 0.237 | 0.519 | 0.264 |
| $\theta_2^*$ | -0.730 | -0.946 | -0.868 | -0.980 |
| SE | 0.175 | 0.094 | 0.632 | 0.097 |

15000 MCMC samples followed by a burn-in of 10000 iterations using a single chain and

therefore all posterior estimates are thus based on 15000 MCMC samples. Here, a large single chain was used rather than multiple parallel chains for MCMC sampling to save the computational time.

Table 3.3 shows that BEAM provides estimates for the log of the clearance rate of free virus ($\theta_1^*$) and for the log of the death rate for infected cells ($\theta_2^*$) without assuming the steady state assumption and therefore estimates the parameters, in the absence of a closed form solution for the mean function. It should be noticed that BEAM can also be used in the presence of a closed form solution for the system of ODEs, as illustrated in the Section 3.4 and therefore provides an estimation framework for both situations. On the other hand, results reported in Han et al. (2002) holds only for the model assuming the closed form mean function described by equation (3.28).

Although BEAM in combination with non-informative priors tends to provide estimates with larger standard errors (SE) as compared to the other two approaches, it should be noticed that in case of the NLME estimation SE's reported in the Table 3.3 are the asymptotic standard errors and in case of Bayesian estimation, all the estimation summaries are calculated based on $150,000$ samples and relatively informative priors. With informative priors, BEAM gives parameter estimates and estimated standard errors similar to the estimates obtained by `NLME` and Bayesian hierarchical estimation method that assume steady-state assumption. These results may suggest that for this data set, steady state assumption seems to be a valid assumption. BEAM is advantageous because estimates from BEAM will remain applicable even for situations for which the above mentioned assumption may not hold and hence can be used to test the assumption.

We also estimated $\sigma$ and the variance components associated with the random effects and their estimates using BEAM are as follows: $\hat{\sigma} = 0.824$ with SE=0.069, $\sqrt{\hat{\Sigma}_{\theta_1^*}} = 1.088$ with SE=0.293 and $\sqrt{\hat{\Sigma}_{\theta_1^*}} = 1.030$ with SE=0.366.

## 3.7 Discussion

In this chapter, we extended the *Bayesian Euler's Approximation Method (BEAM)* to the nonlinear mixed effects (NLME) framework. We first illustrated the application of BEAM in a model when there is a closed form mean function available, through a real data analysis and a simulation study. The model with the known closed form mean function was chosen to compare the performance of BEAM with the iterative GLS estimation approach (Lindstrom and Bates, 1990), implemented in `nlme` function in `R`. For the real data analysis and the corresponding simulation study, we compared the results from BEAM to those from the iterative GLS estimation approach described in Lindstrom and Bates (1990) and found that the both approaches provide very similar estimation of the mean function (see Figure 3.1) and also parameter estimates of the fixed effects and variance components with very similar MSE's (Table 3.2). From Figure 3.2 we also observed that for fixed effects parameters ($\boldsymbol{\theta}^*$), MC distributions of estimates from both NLME and BEAM look similar however NLME seems to underestimate variance parameters considerably. This is a significant advantage for BEAM mainly because NLME uses exact analytical solution for the mean function as compared to BEAM which uses only the system of ODEs to approximate the mean function.

We also applied BEAM to estimate the parameters involved in the HIV model (3.1), when there is no closed form solution available i.e. when the steady state assumption described in Section 3.1 does not hold.

Because of the Bayesian framework, one of the key advantages of BEAM also lies in its ability to handle missing data that is very common in longitudinal studies. We will briefly discuss this in the next chapter. The availability of posterior distributions for the unknown parameters, also makes it straightforward to draw statistical inferences. Although BEAM provides estimates that are not only accurate but applicable with missing or censored data, there is no denying that this is a computationally intensive procedure. However, the computational intensive nature of BEAM should not be an issue with its applications to the real data sets, considering its several other advantages.

In this chapter we assumed that random errors are identically, independently dis-

tributed but methodologies presented here are not restricted to this assumption. This assumption can be relaxed by using a generalized nonlinear modeling framework with $\tilde{\mu}(t, \boldsymbol{\theta}_i)$ as the mean function and $\sigma^2 D_i(t, \boldsymbol{\eta})$ as the variance covariance matrix, which is assumed to be a known matrix up to the unknown parameter $\boldsymbol{\eta}$. A detailed description of the available estimation approaches for this generalized model is presented in (Davidian and Giltinan, 1995). In order to implement BEAM with this generalized nonlinear modeling framework we just need to use a suitable prior for the parameter $\boldsymbol{\eta}$ in the Bayesian model setup. The Bayesian Euler's approimation method provides a closed form approximation of the mean function, $\tilde{\mu}(t, \boldsymbol{\theta})$, without imposing any restrictive assumptions regarding the system of ODEs. Therefore, due to the availability of a closed form approximated mean function, this method can be used to estimate the mean function at any time point lying within the close vicinity of the observed time range. This is a huge advantage as it avoids evaluating the numerical solution of the mean function at the parameter estimate again and again for interpolation/extrapolation.

# Chapter 4

# An Extension of BEAM to the Mutivariate Response NLME Models

## 4.1 NLME Models with Multivariate Responses

In the previous chapter we illustrated the application of BEAM to the NLME models where for every individual, only one observation is available at each time point. In other words we considered models where data consists of measurements taken on one response variable. For example in Han et al. (2002), data consists of observations corresponding to the total viral density which is the sum of density of infectious and non-infectious virus described by equation 3.1. In this chapter we consider the NLME models which involves data observed for more than one type of response on each individual (Davidian and Giltinan, 1995, p. 110). The model setup for this case can be described as follows.

$$\mathbf{y}_{ij} = \boldsymbol{\mu}(t_{ij}, \boldsymbol{\theta}_i) + \boldsymbol{\epsilon}_{ij}, \quad \text{for } i = 1, 2, \ldots m \text{ and } j = 1, 2, ..., n_i, \tag{4.1}$$

where $\mathbf{y}_{ij} = (y_{ij1}, \ldots, y_{ijl})^T$ is the $(l \times 1)$ response vector with elements corresponding to the different response variables, observed at the same time point $t_{ij}$ for the $i^{th}$ individual. $\boldsymbol{\mu}(\cdot) = (\mu_1(\cdot), \ldots, \mu_l(\cdot))^T$ represents the mean vector such that $\mu_l(\cdot, \cdot)$ is the mean function corresponding to the response variable $y_{ijl}$ for all $i = 1, 2, \ldots m$ and $j = 1, 2, ..., n_i$ and depends on a vector of $p$ parameters, $\boldsymbol{\theta}_i$, specific to individual $i$. Next, the random effects $\boldsymbol{\theta}_i$'s are assumed to arise from a common distribution with mean $\boldsymbol{\theta}$ and variance $\Sigma_\theta$. More specifically we can write

$$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{e}_i; \;\; E(\boldsymbol{e}_i) = \mathbf{0}, \;\; V(\boldsymbol{e}_i) = \Sigma_\theta \tag{4.2}$$

For an individual $i$, the intra-individual error vector $\boldsymbol{\epsilon}_{ij} = (\epsilon_{ij1}, \ldots, \epsilon_{ijl})^T$ corresponds to the measurement uncertainty associated with the observed response vector $\mathbf{y}_{ij}$ at time point $t_{ij}$. The random error vector $\boldsymbol{\epsilon}_{ij}$ is assumed to be independently distributed with mean vector $\mathbf{0}$ and variance-covariance matrix $\Sigma$, i.e.,

$$E(\boldsymbol{\epsilon}_{ij}) = \mathbf{0}_{(l\times 1)} \text{ and } V(\boldsymbol{\epsilon}_{ij}) = \Sigma_{(l\times l)} \text{ for } i = 1, 2, ..., m \text{ and } j = 1, 2, ..., n_i \tag{4.3}$$

As described in Chapter 3 (Section 3.2), the mean vector $\boldsymbol{\mu}$ can be defined as a solution to a given system of ODEs as follows,

$$\frac{d(\boldsymbol{\nu}(t,\boldsymbol{\theta}))}{dt} = \mathbf{g}(t, \boldsymbol{\nu}(t,\boldsymbol{\theta})) \quad \forall\, \boldsymbol{\theta} \text{ and } t \neq t_0 \tag{4.4}$$

$$\text{and} \quad \boldsymbol{\nu}(t_0,\boldsymbol{\theta}) = \boldsymbol{\nu}_0(\boldsymbol{\theta}) \tag{4.5}$$

and the mean function $\boldsymbol{\mu}(\cdot)$ is related to $\boldsymbol{\nu}$ by a completely known function $H\colon \mathbb{R}^q \to \mathbb{R}^q$ by $\boldsymbol{\mu}(\cdot) = H(\boldsymbol{\nu}(\cdot))$. It can be noticed here that we are assuming the availability of the data corresponding to all states of the system of ODEs. Once again the objective is to estimate the parameters $(\boldsymbol{\theta}, \Sigma, \Sigma_\theta)$, when no exact form solution is available for $\boldsymbol{\mu}(\cdot)$. In the next section we perform a simulation study based on the first two differential equations involved in the model (3.1), to illustrate the application of BEAM in this case.

## 4.2 HIV Model Simulation Study for Multivariate Data

This simulation study is motivated from the HIV experiment described in Perelson et al. (1996) that involved the administration of a drug (in this case, a protease inhibitor) to five HIV-infected patients. Measurements corresponding to the plasma HIV RNA concentration were taken at time points, $0, 2, 4, 6, 12, 18,\ 24, 30, 36, 42, 48$ hours, and at days $3, 4, 5, 6$. In the experiment data was observed corresponding to the total viral density, which is the sum of the density of infectious virus $\nu_2(t)$ and the density of

non-infectious virus, $\nu_3(t)$ at time $t$. Assuming that viral dynamics is in the quasi-steady state prior to the initiation of treatment, the analytic solution of this system of ODE is given by equation (3.2). However, this simulation study was performed to illustrate the application of BEAM for a multivariate response and for that purpose, we considered differential equations corresponding to the density of infected cells, $\nu_1(t)$ at time $t$ and corresponding to the density of infectious virus, $\nu_2(t)$ at time $t$ only and generated the data corresponding to both differential equations. This data follows a log-normal distribution and therefore we used the logarithm of the data as our response and the transformed differential equations are given as,

$$\begin{aligned} \frac{d\{\log(\nu_1)\}}{dt} &= -\delta + kT_0\frac{\nu_2}{\nu_1}, \quad \log\{\nu_1(0)\} = \log(T_0^*) && (4.6)\\ \frac{d\{\log(\nu_2)\}}{dt} &= -c, \qquad\qquad\quad \log\{\nu_2(0)\} = \log(V_{I_0}) && (4.7) \end{aligned}$$

To keep this example simple, we assumed the parameters c and $\delta$ to be the random effects, and the remaining parameters were assumed to be known. Also to avoid the problem of identifiability in equation (3.1), we assumed $kT_0$ as one parameter and its value was considered to be known. This simulation study was performed with the data for the log of infected cells ($y_{ij1}$), and for the log of infectious virus ($y_{ij2}$) simulated as,

$$\log(\mathbf{y}_{ij}) = \boldsymbol{\mu}(t_{ij}, \boldsymbol{\theta}_i) + \boldsymbol{\epsilon}_{ij}, \quad \boldsymbol{\epsilon}_{ij} \sim \mathrm{MVN}_2(\mathbf{0}, \mathrm{I}_2) \tag{4.8}$$

where

$$\begin{aligned} \log(\mathbf{y}_{ij}) &= (\log(y_{ij1}),\ \log(y_{ij2}))^T, \\ \boldsymbol{\mu}(\cdot) &= (\log(\nu_1),\ \log(\nu_2))^T, \\ \boldsymbol{\epsilon}_{ij} &= (\epsilon_{ij1}, \epsilon_{ij2})^T, \end{aligned} \tag{4.9}$$

for $i = 1, 2, \ldots, 20$ individuals and $j = 1, 2, ..., 15$ time points.

The analytical expressions for the mean functions as given in Holte et al. (2003), were

used for the data generation purposes and are given as following,

$$\log(\nu_1(t)) = \log\left(\frac{\{kT_0V_{I_0} + T_0^*(c-\delta)\} * e^{-\delta t} - kT_0V_{I_0}e^{-ct}}{c-\delta}\right) \quad (4.10)$$

$$\log(\nu_2(t)) = \log(V_{I_0}e^{-ct}) \quad (4.11)$$

The initial values for the differential equations are given by $\log(\nu_1(0)) = \log(3000)$ and $\log(\nu_2(0)) = \log(100000)$. The Bayesian framework assumed for estimation of parameters $\log(c_i)$ and $\log(\delta_i)$ with the restriction $(c_i > \delta_i)$, is as follows,

$$\begin{aligned} \boldsymbol{\theta}_i^* = (\log(c_i), \log(\delta_i))^T &\sim \text{MVN}_2(\boldsymbol{\theta}^*, \boldsymbol{\Sigma}_{\theta^*}) \\ \boldsymbol{\theta}^* = (\log(c), \log(\delta))^T &\sim \text{MVN}_2(\boldsymbol{\theta}_0, H_0) \\ \boldsymbol{\Sigma}_{\theta^*}^{-1} &\sim \text{W}_2(R_0, \rho_0) \text{ and } \sigma^{-2} \sim \text{G}(a_0, b_0) \end{aligned} \quad (4.12)$$

Where $\boldsymbol{\Sigma}_{\theta^*}^{-1}$ is the $(2 \times 2)$ inverse variance-covariance matrix associated with the random effects vector $\boldsymbol{\theta}_i^*$. The values for hyper-parameters are chosen as follows,

$$a_0 = 0.01, b_0 = 0.01, \rho_0 = 2, \boldsymbol{\theta}_0 = (1.1, -1.0)^T, H_0 = \text{diag}(10, 10), R_0 = \text{diag}(0.1, 0.1)$$

We assumed $kT_0 = 0.015$ for the data generation and simulated the data corresponding to 20 individuals. For the data generation process, we used $\boldsymbol{\theta}^{*T} = (1.1, -1.0)$ and $\boldsymbol{\Sigma} = 0.2\text{I}_2$. These parameter values were chosen based on estimates from the similar models, available in the literature (Han et al., 2002; Huang et al., 2004). As described above, we performed the simulation using BEAM, with 1000 replicates and with number of fixed time points $N = 15$. Results from this simulation study are presented in the Table 4.1.

The Bayesian Euler's approximation method was used with 2000 MCMC samples followed by a burn-in of 4000 iterations using three parallel chains. All posterior estimates are thus based on 6000 MCMC samples generated for each of the 1000 MC runs. In this case BEAM provides unbiased parameter estimates for the parameters $\theta_1^*$ and $\sigma$ (p-values=0.246 and 0.146 respectively), while the bias in case of the parameters $\theta_2^*$, $\sqrt{\Sigma_{\theta_1^*}}$ and $\sqrt{\Sigma_{\theta_2^*}}$ was found to be significant (p-values $< 0.0001$). Table 4.1 lists the bias

Table 4.1: Simulation Results for HIV model using BEAM with 1000 MC Runs.

| Parameters | Estimates | Method |
|---|---|---|
| | | BEAM |
| $\theta_1^*$ | Bias | 0.001 |
| | MCSE | 0.046 |
| | ESE | 0.050 |
| $\theta_2^*$ | Bias | 0.085 |
| | MCSE | 0.070 |
| | ESE | 0.069 |
| $\sigma$ | Bias | 0.001 |
| | MCSE | 0.030 |
| | ESE | 0.029 |
| $\sqrt{\Sigma_{\theta_1^*}}$ | Bias | 0.014 |
| | MCSE | 0.034 |
| | ESE | 0.031 |
| $\sqrt{\Sigma_{\theta_2^*}}$ | Bias | 0.020 |
| | MCSE | 0.045 |
| | ESE | 0.039 |

of estimates along with estimated standard errors (ESE) and Monte Carlo simulation standard errors (MCSE). Box plots of estimates from this simulation study are presented in Figure 4.1.

The horizontal line in each box plot represents the true value of the respective parameter, used for data generation. It can be observed from these box plots that for all parameters except $\hat{\theta}_2^*$, interquartile range of estimates cover the true value of corresponding parameters and even in that case, the distribution range of estimates obtained from BEAM covers the true value. We also estimated the posterior credible intervals for all parameters by computing the 2.5% and 97.5% percentiles of the posterior distributions of all parameters and calculated coverage probability based on above defined 95% posterior

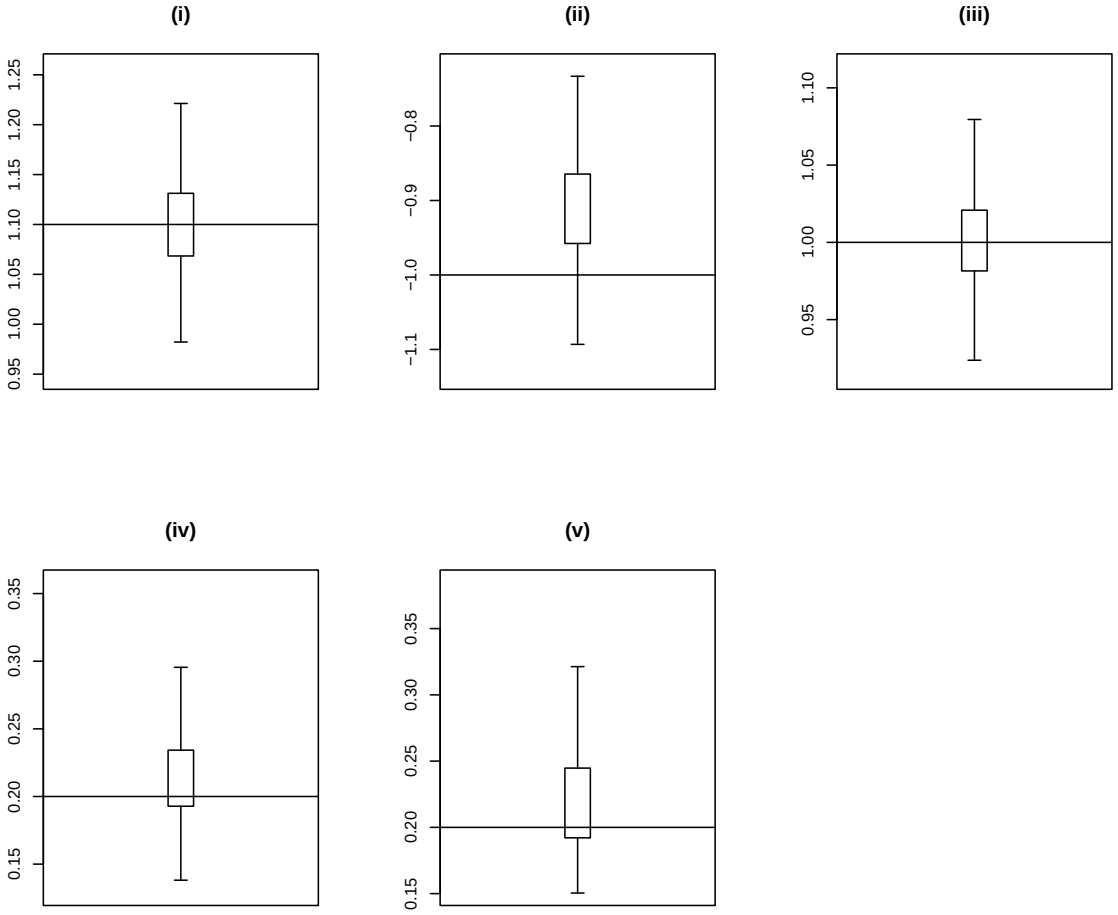


Figure 4.1: Box plots of point estimates: (i) $\hat{\theta}_1^*$'s (ii) $\hat{\theta}_2^*$'s (iii) $\hat{\sigma}$'s (iv) $\sqrt{\hat{\Sigma}_{\theta_1^*}}$'s (v) $\sqrt{\hat{\Sigma}_{\theta_2^*}}$'s based on 1000 simulated data sets for HIV model. (The horizontal solid line in each case represents the true value of the parameters.)

credible intervals. This probability gives us the percentage of posterior credible intervals of parameter estimates that contain the true value of the corresponding parameters. Based on this simulation study, coverage probability for all parameter estimates is of about 96.2% for all parameters.

## 4.3 Discussion

The objective of the simulation study described in this chapter was to present an extension of the BEAM to estimate the parameters in the NLME models where multiple responses are available for each individual. For simplicity we assumed a case study where observations were available for each response type at the same time points for every individual. However, BEAM can be used without loss of generality, with models where different time points were considered for different individuals as well as for different response types. Results from the simulation study presented in the previous section suggests that BEAM seems to perform well in terms of unbiasedness for $\theta_1^*$ and $\sigma$ and provide reasonably good estimates in terms of coverage probabilities for all parameters. Even though BEAM seems to slightly over estimate the variability associated with the random effects, Figure 4.2 (iv) and (v) suggests the availability of the reasonable posterior interval estimates using BEAM. However, further investigation is needed to explain the over estimation of $\theta_2^*$ in the simulation study.

# Chapter 5

# Conclusion and Future Work

The objective of this research was to propose estimation techniques that provide estimates of the parameters involved in the mean function described by a system of ODEs when there is no closed form solution available for the system of ODEs. We proposed two methodologies namely "Bayesian Euler's approximation method (BEAM)" and "Splines Euler's approximation method (SEAM)". In Chapter 2, we proposed the algorithms for these two methods, for fixed effects models and illustrated the use of two approaches by applying BEAM and SEAM to the real data (Section 2.3.1) and by means of a simulation study (Section 2.3.2) based on the real data. We also compared results from these two approaches with the established iterative GLS approach (Lindstrom and Bates, 1990) which is implemented in the "`nlme`" package of `R` and with "Integrated data (ID)" approach (Holte et al., 2003). Our results showed that both BEAM and SEAM provide competitive estimates for parameters as compared to the NLME and ID estimation approaches.

For the illustration purpose, we used a model for which the mean function has a closed form expression available or equivalently a closed form solution for the system of ODEs is achievable. This choice of model enabled us to compare results from the proposed two estimation approaches with the NLME estimation approach which requires a closed form expression for the mean function. Competitive results from the proposed estimation approaches, with respect to NLME estimation approach, show the usefulness of proposed methodologies. Since for most of the real data applications, it is impossible to derive a closed form solution of the system of differential equations and therefore for all such situations NLME is not applicable while BEAM and SEAM still provide approximately unbiased and efficient estimates.

Next in Chapter 3, we extended the BEAM to the nonlinear mixed effects modeling framework where mean function for each individual depends on the individual specific pa-

rameters (random effects). This class of models is commonly used in HIV data modeling, PK/PD modeling and other dynamic mean modeling and more often the mean function involved in these models is defined as the solution of a given system of ODEs. As in case of fixed effects models, for situation where there is no closed form solution available for the system of ODEs, parameter estimation becomes challenging and that was the focus of our research. For the mixed effect framework, the algorithm of BEAM is very similar to the one used in fixed effects framework in terms of mean function approximation. However, for the NLME models, we estimated individual specific mean functions by substituting the values of random effects parameters for all individuals, in the expression for approximated mean function given by equations 3.14 and incorporated these individual specific approximated mean functions into the Bayesian nonlinear hierarchical modeling framework to obtain the posterior summaries of the parameters of interest.

In Chapter 3, first we used the same data as used in Chapter 2 but unlike Chapter 2 where we treated the data corresponding to different subjects as separate data sets, in this case we combined all subjects and considered the combined data set as the single dataset. Both NLME estimation technique and BEAM were applied to this real data set under the mixed effects framework and a simulation study was also performed to compare the performance of these two estimation approaches. We did not compare BEAM with ID method in this case since it was not obvious to us about how to extend ID method to the mixed effects models. Results again showed that BEAM appears to provide similar estimates as compared to the NLME approach and this is more evident when we look at the plots of approximated mean function provided by both NLME and BEAM (Figure 3.4).

In Section 3.6 we illustrated the application of BEAM to the HIV model described in Section 3.1, for the situation when no closed form expression for the mean function is available. This real data analysis was significant because it achieves our objective of estimating parameters involved in the given system of ODEs, for which an analytical closed form solution is unavailable.

In Chapter 4, we performed a simulation study to illustrate the application of BEAM

to the case when more than one observation is available at each observed time point. For this purpose we used the first two differential equations of the HIV model described by equations 3.1.

An extension of the SEAM to the NLME modeling is an issue for future research although we did some preliminary investigation into this matter. We combined the individual estimates obtained by applying SEAM to the real data corresponding to the paramecium growth model as described in Chapter 3 (Section 3.4), by using the " standard two stage method" (Davidian and Giltinan, 1995) and results showed that SEAM also provided similar estimates as NLME and BEAM. The parameter estimates obtained by using SEAM for nonlinear mixed effects model along with the results from BEAM and NLME approach (presented in the Table 3.1) are summarized in the Table 5.1. Even

Table 5.1: Results of estimation of parameters in the logistic growth model to the data set on growth colonies of the bacteria Paramecium Aurelium using NLME, BEAM and SEAM.

| Parameters | Method | | |
|---|---|---|---|
| | NLME | BEAM | SEAM |
| $\theta_1^*$ | -0.177 | -0.217 | -0.189 |
| ESE | 0.029 | 0.182 | 0.006 |
| $\theta_2^*$ | -6.459 | -6.463 | -6.455 |
| ESE | 0.043 | 0.209 | 036 |
| $\sigma$ | 0.185 | 0.223 | 0.192 |
| ESE | - | 0.022 | - |

though these estimates look very similar to the results we obtain by using NLME and BEAM, a more formal approach to implement SEAM with in the classical nonlinear mixed effects modeling framework is desirable. However, results from the preliminary investigation (Table 5.1) definitely give hope for the possible good results.

At this stage, we will summarize a list of potential statistical questions that have the ability to motivate the future research in this field. Future research may include:

- Extension of SEAM to the mixed effects models to obtain parameter estimates.
- Illustration of the application of BEAM with missing or censored data.
- Derivation of a formal approach to determine the number of fixed time points in the proposed approaches.
- A formal proof for the consistency of the estimates obtained by BEAM and SEAM.

One of the main advantages of both BEAM and SEAM is that, these methods provide estimates for the parameters involved in the system of ODEs, without imposing any restrictive assumptions regarding the dynamics of the data set. The proposed likelihood approximation method also provides a closed form approximation of the mean function, $\tilde{\mu}(t, \boldsymbol{\theta})$. Therefore these methods can be used to estimate the mean function at any time point lying within the close vicinity of the observed time range. This is a huge advantage as it avoids evaluating the numerical solution of the mean function at the parameter estimate again and again for interpolation/extrapolation.

Because of the Bayesian framework, one of the key advantages of BEAM also lies in its ability to handle missing or censored data that is very common in longitudinal studies. Suppose that the observation corresponding to the $i^{th}$ individual, at the time point $t_{ij}$ is missing, then this missing observation can be generated using (3.15). In case of the censored data, a suitable truncated distribution can be used to replace the censored observations. Next, this imputed data can be included in the Gibbs sampling algorithm described by equations (2.13-2.15) in case of fixed effects models and by equations (3.18-3.21) in case of the mixed effects models. A detailed discussion of the problem of censored data with in the nonlinear modeling framework and the Bayesian approach to tackle this problem is presented in Banks et al. (2005). The availability of posterior distributions for the unknown parameters, also makes it straightforward to draw statistical inferences.

At the same time advantage of SEAM comes not only from its accuracy of estimation and weaker distributional assumptions but also from its computational convenience as compared to BEAM. Although BEAM provides estimates that are not only accurate but

applicable with missing or censored data, there is no denying that this is a computationally intensive procedure. In comparison to BEAM, SEAM takes much less computation time, but SEAM is limited to handling only completely observed data and hence restricting its application in comparison to BEAM. Finally we conclude that for mixed effects models we proposed BEAM as the approach to estimate the parameters in the absence of a closed form mean function however, for the fixed effects models, the choice between BEAM and SEAM is up to the readers to decide.

# Bibliography


Atkinson, K. (1978). *An Introduction to Numerical Analysis*. New York: Wiley.

Banks, H. T., Grove, S., Hu, S. and Ma, Y. (2005). A hierarchical bayesian approach for parameter estimation in hiv models. *Inverse Problems* **21**, 1803–1822.

Beal, S. L. and Sheiner, L. B. (1982). Estimating population kinetics. *CRC Critical Reviews in Biomedical Engineering* **8**, 195–222.

Butcher, J. C. (2003). *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons.

Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. London:Chapman & Hall.

Diggle, P. J. (1990). *Time Series A Biostatistical Introduction*. Oxford Science publications.

Ding, A. A. and Wu, H. (1999). Relationships between antiviral treatment effects and biphasic viral decay rates in modelling hiv dynamics. *Mathematical Biosciences* **160**, 63–82.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **91**, 1400–1412.

Gelman, A., Bois, F. and Jing, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association* **85**, 398–409.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions and the bayesian resolution of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Han, C., Chaloner, K. and Perelson, A. S. (2002). Bayesian analysis of a population hiv dynamic model. *Case Sudies in Bayesian Statistics* **6**, 223–237.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika* **57**, 97–109.

Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M. and Markowitz, M. (1995). Rapid turnover of plasma virions and cd4 lymphocytes in hiv-1 infection. *Nature* **373**, 123–126.

Holte, S. E., Cornelisse, P., Heagerty, P. and Self, S. (2003). An alternative to nonlinear least-squares regression for estimating parameters in ordinary differential equations models (personal communication).

Huang, Y., Liu, D. and Wu, H. (2004). Hierarchical bayesian methods for estimation of parameters in a longitudinal hiv dynamic system. *Submitted to Biometrics*.

Lambert, J. D. (1991). *Numerical Methods for Ordinary Differential Systems: The Initial Value Problem*. Chichester: John Wiley.

Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**, 673–687.

Lunn, D. J., Best, N., Thomas, A., Wakefield, J. and Spiegelhalter, D. (2002). Bayesian analysis of population pk/pd models: General concepts and software. *Journal of Pharmacokinetics and Pharmacodynamics* **29**, 271–307.

Natarajan, R. and Kass, R. (2000). Reference bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association* **95**, 227–237.

Perelson, A. S., Neumann, A. V., Markowitz, M., Leonard, J. M. and Ho, D. H. (1996). Hiv-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time.. *Science* **271**, 1582–1587.

Petzold, L. R. (1987). Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *Journal of Scientific and Statistical Computing* **4**, 136–148.

Putter, H., Heisterkamp, S. H., Lange, J. M. A. and Wolf, F. d. (2002). A bayesian approach to parameter estimation in hiv dynamical models. *Statistics in Medicine* **21**, 2199–2214.

Racine-Poon, A. and Wakefield, J. (1998). Statistical methods for population pharmacokinetic modelling. *Statistical Methods in Medical Research* **7**, 63–84.

Robert, C. P. and Casella, G. (2005). *Monte Carlo Statistical Methods*. Springer.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC.

Shampine, L. F. (1994). *Numerical Solutions of Ordinary Differential Equations*. New York:Chapman & Hall.

Sheiner, L. B., Rosenberg, B. and Melmon, K. L. (1972). Modeling of individual pharmacokinetics for computer-aided drug dosage. *Computers and Biomedical Research* **5**, 441–459.

Wahba, G. (1990). *Splines for Observational Data*. Society for Industrial and Applied Mathematics.

Wakefield, J. C. (1996). The bayesian analysis to population pharmacokinetic models. *Journal of the Amreican Statistical Association* **91**, 62–75.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A. and Gelfand, A. E. (1994). Bayesian analysis of linear and nonlinear population models using the gibbs sampler. *Applied Statistics* **43**, 201–221.

Wolfinger, R. (1993). Laplace's approximation for nonlinear models. *Biometrika* **80**, 791–795.

Wu, H. (2005). Statistical methods for hiv dynamic studies in aids clinical trials. *Statistical Methods in Medical Research* **14**, 171–192.

# Appendix

# Appendix A

# Numerical methods to solve a system of ODEs

To describe few popular approximation methods, for simplicity, we assume $\mu \equiv \nu$ with $q = 1$.

1. *Naive Euler's method:* This method is based on the following sequence of solutions,

$$\mu(t_n) = \mu_n \approx \mu_{n-1} + g(t_{n-1}, \mu_{n-1})h \tag{A.1}$$

   The order of accuracy for this method is $o(h)$. Recall that the order of accuracy of a method is the order of accuracy with which the solution function is approximated.

   *An approximation to a solution of a system of ODEs is said to be $r^{th}$ order accurate and is denoted by $o(h^r)$, if the term corresponding to $h^r$ in the Taylor expansion of the solution is correctly reproduced.*

2. *Improved Euler's method:* This second algorithm improves the naive Euler's method by modifying the algorithm as follows:

$$\mu(t_n) \approx \mu_n = \mu_{n-1} + \frac{1}{2}[g(t_{n-1}, \mu_{n-1}) + g(t_n, \mu_{n-1} + g(t_{n-1}, \mu_{n-1})h)]h \tag{A.2}$$

   The order of accuracy for the improved Euler's method is $o(h^2)$.

3. *Runge-Kutta method:* This algorithm uses the Simpson's rule:

$$\int_{t_n}^{t_n+h} g(t,\mu(t))dt \approx \frac{h}{6}[g(t_n,\mu(t_n)) + 4g(t_n+\frac{h}{2},\mu(t_n+\frac{h}{2})) + g(t_n+h,\mu(t_n+h))]$$

The final algorithm, after approximating $\mu(t_n)$, $\mu(t_n+\frac{h}{2})$ and $\mu(t_n+h)$ is given by following sequence of steps:

$$\begin{aligned} K_{n,1} &= g(t_n,\mu_n) \\ K_{n,2} &= g(t_n+\frac{h}{2},\mu_n+\frac{h}{2}K_{n,1}) \\ K_{n,3} &= g(t_n+\frac{h}{2},\mu_n+\frac{h}{2}K_{n,2}) \\ K_{n,4} &= g(t_n+h,\mu_n+hK_{n,3}) \\ \mu_{n+1} &= \mu_n+\frac{h}{6}[K_{n,1}+2K_{n,2}+2K_{n.3}+K_{n,4}] \end{aligned} \tag{A.3}$$

This method has an accuracy of $o(h^4)$ which is much better compared to previous two approximation methods, but is also more computationally intensive than the previous two methods.

There is a huge literature on numerical methods to solve ODEs and an extensive review of these numerical methods can be found in a classic book by Butcher (2003).