Computational Statistics and Data Analysis 127 (2018) 116–134

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Inference for differential equation models using relaxation via dynamical systems

Kyoungjae Lee [a], Jaeyong Lee [b,*], Sarat C. Dass [c]

[a] Department of Applied and Computational Mathematics and Statistics, The University of Notre Dame, IN, USA
[b] Department of Statistics, Seoul National University, Seoul, South Korea
[c] Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, Tronoh, Malaysia

ARTICLE INFO

Article history:
Received 12 December 2017
Received in revised form 2 April 2018
Accepted 20 May 2018
Available online 28 May 2018

Keywords:
Ordinary differential equation
Dynamic model
Runge–Kutta method
Extended Liu and West filter

ABSTRACT

Statistical regression models whose mean functions are represented by ordinary differential equations (ODEs) can be used to describe phenomena which are dynamical in nature, and which are abundant in areas such as biology, climatology and genetics. The estimation of parameters of ODE based models is essential for understanding its dynamics, but the lack of an analytical solution of the ODE makes estimating its parameter challenging. The aim of this paper is to propose a general and fast framework of statistical inference for ODE based models by relaxation of the underlying ODE system. Relaxation is achieved by a properly chosen numerical procedure, such as the Runge–Kutta, and by introducing additive Gaussian noises with small variances. Consequently, filtering methods can be applied to obtain the posterior distribution of the parameters in the Bayesian framework. The main advantage of the proposed method is computational speed. In a simulation study, the proposed method was at least 35 times faster than the other Bayesian methods investigated. Theoretical results which guarantee the convergence of the posterior of the approximated dynamical system to the posterior of true model are presented. Explicit expressions are given that relate the order and the mesh size of the Runge–Kutta procedure to the rate of convergence of the approximated posterior as a function of sample size.

1. Introduction

Many dynamical phenomena in the real world can be represented mathematically by ordinary differential equations (ODEs). Common examples include the ODE for Newton's law of cooling, Lotka–Volterra equations for predator–prey populations (Alligood et al., 1997) and the Lorenz equations for atmospheric convection (Lorenz, 1963). There are many other popular examples describing physical, chemical and biological phenomena using ODEs. Although observing data from an ODE systems is common, estimating the parameters of ODE models (ODEMs) can be challenging because of the lack of an analytical solution to the ODE. Here, we give a brief review of previous works on ODEMs.

There are several frequentist methods in the literature for parameter estimation of ODEMs. Bard (1974) used numerical integration to approximate the solution of ODEs and minimized the objective function based on a gradient method. Varah (1982) suggested a two step estimation method using cubic spline approximation. The two steps consist of estimating the regression function followed by estimating the parameters of the ODEM. Ramsay and Silverman (2005) modified the first step of Varah (1982) by adding a roughness penalty function which measures the difference between the ODE and the derivative

* Corresponding author.
E-mail address: leejyc@gmail.com (J. Lee).

https://doi.org/10.1016/j.csda.2018.05.014
0167-9473/

of the estimated mean function. Wu et al. (2012) adopted only the first step of Ramsay and Silverman (2005) and modified the second step of Varah (1982) by adopting numerical integration to approximate the derivative of the state variables. The parameter cascading method was proposed by Ramsay et al. (2007). They grouped the parameters into the regression coefficients, structural parameters, and regularization parameters. The parameters in each group are estimated in turn in a cascading fashion.

Bayesian inference of ODEMs is more challenging because naive application of Markov Chain Monte Carlo (MCMC) methods would require calculation of the numerical solution of ODE whenever parameters are sampled from the proposal distribution. Gelman et al. (1996) and Huang et al. (2006) proposed a Bayesian computation method for parameter inference of pharmacokinetic models and the longitudinal HIV dynamic system, respectively. Campbell (2007) combined the parallel tempering (Geyer, 1991) and collocation method (Ramsay et al., 2007) to get over the rough surface of the posterior, but this slows down the speed of computations significantly. Arnold et al. (2013) used particle filter framework for the inference of ODEMs with linear multistep methods for the numerical integration. Dass et al. (2017) suggested a Bayesian inference with Laplace approximation for fast computation when the dimension of the parameter space is moderate.

In this paper, we propose a Bayesian inference method for ODEMs using a relaxation technique via dynamical systems and associated dynamic models. Relaxation is achieved by a properly chosen numerical procedure, such as the Runge–Kutta, and by introducing additive Gaussian noise variables with variance tending to zero. The variance of the additive noise variables works as a measure of fidelity to the original ODEM and by letting it tend to zero, we recover the original model. The relaxation introduces inefficiency in the inference, but we gain the speed of the computation in return.

For fast computation, a filtering method is applied for inferring posterior distributions of parameters in a Bayesian framework. The relaxation technique provides a dynamical system and model for which a fast inference tool based on sequential Monte Carlo can be developed. With these sequential methods, we do not need to calculate the whole path of the numerical solution for each realization of the new parameter. It reduces the computation time significantly compared to other standard Bayesian procedures and enables us to deal with the ODEM in reasonable computing time. In Section 5.2, to emphasize its fast computation, the proposed method is compared with other methods: the parameter cascading method, the delayed rejection adaptive Metropolis algorithm and the Bayesian inferential procedure with Laplace approximation. In our simulation studies, the proposed method is approximately 9 to 46 times faster than the other methods considered.

We also derive convergence results for the approximated posteriors under suitable regularity conditions. We present a guideline for the choice of model parameters which gives a reasonable relative error rate, and provide its theoretical basis. Theoretical results which guarantee the convergence of the posterior of the approximated dynamical system to the posterior of true model are presented. Explicit expressions are given that relate the order and the mesh size of the Runge–Kutta procedure and guarantee the rate of convergence of the approximated posterior to the true posterior.

The rest of the paper is organized as follows. In Section 2, we describe a differential equation model and its corresponding relaxed dynamic model counterpart as well as prior choices. The method of posterior inference is described in Section 3. Some theoretical support for the proposed method is given in Section 4. In Section 5, we give three simulated data examples to demonstrate the speed and performance of the proposed method. A real data set, the Lynx–Hare data set, is analyzed in Section 6. The discussion is given in Section 7. The proofs of theorems are given in Appendix A.

2. Ordinary differential equation models and nonlinear dynamic models

2.1. Ordinary differential equation models

The ODEM is the regression model with regression function $x(t)$ described by an ODE. The regression function $x(t)$ is the solution of the differential equation

$$\dot{x}(t) = f(x(t), v, t; \theta), \tag{1}$$

where f is a p-dimensional smooth function, $v(t)$ is a deterministic input function, $\theta \in \Theta \subset \mathbb{R}^q$ is the unknown parameter, and $\dot{x}(t)$ denotes the first derivative of $x(t)$ with respect to time t. Since the input function $v(t)$ does not affect the general ideas of inference in this paper, it is not considered subsequently. The data are observed at n points in the time interval $t \in [0, T] \subset \mathbb{R}$, given by $0 \le t_1 \le t_2 \le \cdots \le t_n \le T$. Thus,

$$y_i = x(t_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

where y_i is a p-dimensional observation vector at time t_i, the error ϵ_i is drawn independently from the multivariate normal distribution $N_p(0, \sigma^2 I_p)$ with unknown $\sigma^2 > 0$, and $x(t_i)$ is the underlying regression function measured at time t_i.

The regression model is given by

$$\begin{aligned} y_i &= x_i + \epsilon_i, \quad i = 1, \ldots, n, \\ \dot{x}(t) &= f(x(t), t; \theta) \end{aligned} \tag{2}$$

where $x_i = x(t_i)$. The covariate x_i is determined by the initial value of x, $x_0 = x(0)$, and the parameter θ. In the rest of the paper, we call model (2) as the regression model or the true model.