Statistics and Probability Letters 115 (2016) 60–69

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Consistent estimation of ordinary differential equation when the transformation parameter is unknown[☆]

Jie Zhou [*], Hai-Lin Feng

Department of Statistics, Xidian University, Xi'an, Shaanxi, PR China

ARTICLE INFO

Article history:
Received 12 May 2015
Received in revised form 16 February 2016
Accepted 16 February 2016
Available online 12 April 2016

Keywords:
Two-step estimation
ODE
Transformation parameter
Weak convergence

ABSTRACT

Two-step estimation of ordinary differential equation (ODE) is investigated when the transformation parameter is unknown. First we build the transformation parameter estimator by profile M-estimation. Then with a new consistency result of M-estimator, two-step estimator is proved to be consistent. Numerical studies validate the effectiveness of the proposed estimator.

1. Introduction

Consider the following ordinary differential equation

$$\dot{x}(t) = F(x(t), \theta), \quad t \in [0, 1], \tag{1.1}$$

where $x(t)$ is d-dimensional state variable and θ is p-dimensional unknown parameter. If the initial condition $x(0)$ is unknown, then it can also be included in θ. Such model has been widely used in different fields to model the dynamic phenomena. For example in biology the famous Lotka–Volterra equation is used to model the dynamic interaction between the predator species and prey species; in clinical trials of infectious disease, the relationship between the virus and immune system has been widely modeled by ODE models, see Perelson and Nelson (1999), Nowak and May (2000), Robeva et al. (2008) for details. When ODE model (1.1) is used in practice, the unknown parameter θ should be estimated from the noisy observations of $x(t)$ firstly.

However the observations Y_i's at time t_i's are often not the direct observations of $x(t_i)$ $(i = 1, \ldots, n)$ and some scale transformation should be performed firstly in order to estimate θ. For example in clinical trial of AIDS, logarithm transformation is usually applied to the measurements of load of virus when the HIV model is estimated. Let $B(\cdot|\lambda)$ denote such transformation family indexed by parameter $\lambda \in \Lambda \subseteq \mathbb{R}^d$, in which λ_i, the ith component of λ, corresponds to the transformation applied to the ith component of $x(t)$. The transformed observations are assumed to satisfy the following relationship

$$Y_{\lambda,i} \triangleq B(Y_i|\lambda) = x(t_i|\lambda) + \epsilon_{\lambda,i}, \quad i = 1, \ldots, n \tag{1.2}$$

with $E\epsilon_{\lambda,i} = 0$ and $D\epsilon_{\lambda,i} = \sigma_\lambda^2$, i.e., $x(t_i|\lambda) = EB(Y_i|\lambda)$. We assume there exists a unique $\lambda_0 \in \Lambda$ such that $x(t|\lambda_0)$ is the solution of model (1.1). We call λ_0 the true value of λ. The parameter in model (1.1) corresponding to $x(t|\lambda_0)$ is called the true

☆ Supported by NSFC (71271165).

* Corresponding author.
E-mail address: zhoujie@xidian.edu.cn (J. Zhou).

http://dx.doi.org/10.1016/j.spl.2016.02.008
0167-7152/

value of θ and is denoted by θ_0. A natural question is how to estimate the transformation parameter λ_0 given the observations Y_i's. Traditionally parameter estimation for ODE is regarded as a problem of nonlinear regression and can be carried out based on the numerical integration approaches such as Runge–Kutta algorithm. Consequently transformation parameter can be estimated by the corresponding profile likelihood approaches. Though such estimation approach is reasonable theoretically, in practice the computational loads are notoriously large and the resulted estimators often are unreliable, see Cao and Ramsay (2007), Ramsay et al. (2007) for a detailed discussion of this approach.

In this paper we study the estimation of transformation parameter under the framework of two-step estimation of ODE (1.1). Such estimation approaches were firstly proposed by Varah (1982) and recently studied by Brunel (2008), Liang and Wu (2008), Fang et al. (2011), Gugushvili and Klaassen (2012), Wu et al. (2014) among many others. We define the estimator of λ as a profile M-estimator. In order to gain some insights about such estimator, the property of consistency for ordinary M-estimator is firstly generalized to the model with unknown transformation parameter. Based on this result the consistency of the proposed profile M-estimator is established. Simulation studies show that the proposed transformation parameter estimator is efficient. A real data set from the clinical trial is examined and the results show that a better model fit can be attained by the proposed procedure compared to ordinary logarithm transformation.

This paper is organized as follows. Section 2 presents the main results. The numerical studies, including simulated data set and real data set, appear in Section 3. The detailed proofs of these results are presented in Section 4.

2. Main results

Though the state variable $x(t)$ can be d-dimensional for $d \in \mathbb{N}$, in the following we will always assume $d = 1$ for the simplicity of notation. All the results involved can be generalized to the general multivariate settings straightforwardly. Given $\lambda \in \Lambda$, denote the transformed observations by $Y = (Y_{\lambda,1}, \dots, Y_{\lambda,n})^T$ where $Y_{\lambda,i} = B(Y_i|\lambda)$ for $i = 1, \dots, n$. In order to estimate θ, the two-step approach first estimates $x(t)$ and its derivative $\dot{x}(t)$. Several choices have been proposed in this step, including spline based estimates (Varah, 1982; Brunel, 2008), kernel smoothing based estimates (Liang and Wu, 2008; Gugushvili and Klaassen, 2012; Wu et al., 2014). Here we employ the following Priestley–Chao estimate which had been proposed in Gugushvili and Klaassen (2012),

$$\hat{x}(t|\lambda) = \sum_{i=1}^{n} (t_i - t_{i-1}) \frac{1}{b} K\left(\frac{t - t_i}{b}\right) Y_{\lambda,i}, \tag{2.1}$$

where b is bandwidth, $K(\cdot)$ is the kernel function. The estimate of $\dot{x}(t)$ is assumed to be the derivative of $\hat{x}(t)$ with respect to t, i.e., $\hat{\dot{x}}(t) = \dot{\hat{x}}(t)$. In the second step the estimate of parameter θ is defined by

$$\begin{aligned} \hat{\theta}(\lambda) &= \arg\min_{\theta \in \Theta} \int_0^1 [\hat{\dot{x}}(t|\lambda) - F(\hat{x}(t|\lambda), \theta)]^2 w(t) dt \\ &= \arg\min_{\theta \in \Theta} M_n(\theta, \lambda) \end{aligned} \tag{2.2}$$

for some set $\Theta \in \mathbb{R}^p$. Here $w(t)$ is weight function with support $[\delta, 1 - \delta]$ for some $0 < \delta < 1/2$. Gugushvili and Klaassen (2012) had investigated the asymptotics of such two-step estimator when λ_0, the true value of λ, is known. Under some regular conditions they showed that $\hat{\theta}(\lambda_0)$ converges to θ_0 at the rate $\sqrt{n}$. Transformation parameter usually is unknown in practice. Therefore we substitute $\hat{\theta}(\lambda)$ into the right-hand side of (2.2) and define the estimator of λ as

$$\hat{\lambda} = \arg\min_{\lambda \in \Lambda} M_n(\hat{\theta}(\lambda), \lambda) \tag{2.3}$$

for some set $\Lambda \in \mathbb{R}$. A desired property of an estimator is the consistency. In order to show $\hat{\lambda}$ is consistent, some insights about the asymptotic properties of $\hat{\theta}(\lambda)$ are needed. To this end, let $x(t|\lambda) = EB(Y(t)|\lambda)$ and $\dot{x}(t|\lambda)$ be the derivative of $x(t|\lambda)$ with respect to t. Define

$$\begin{aligned} \theta(\lambda) &= \arg\min_{\theta \in \Theta} \int_0^1 [\dot{x}(t|\lambda) - F(x(t|\lambda), \theta)]^2 w(t) dt \\ &= \arg\min_{\theta \in \Theta} M(\theta, \lambda), \end{aligned} \tag{2.4}$$

then we have the following result.

Proposition 1. *Suppose $b \to 0$ and $nb^3/\log n \to \infty$. For given λ and $\forall \epsilon > 0$, we assume that there exists a positive constant $\tau_\lambda(\epsilon)$ such that*

$$M(\theta(\lambda), \lambda) + \tau_\lambda(\epsilon) < \inf_{\theta \in \Theta : \|\theta - \theta(\lambda)\| > \epsilon} M(\theta, \lambda), \tag{2.5}$$

then under the conditions (A1)–(A7) *given in Section 4 we have $\hat{\theta}(\lambda) \xrightarrow{p} \theta(\lambda)$.*