Statistics and Probability Letters 90 (2014) 136–139

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Asymptotic normality for discretely observed Markov jump processes with an absorbing state

Alexander Kremer *, Rafael Weißbach

Chair in Statistics and Econometrics, Faculty for Economic and Social Sciences, University of Rostock, 18051 Rostock, Germany

ARTICLE INFO

Article history:
Received 23 December 2013
Accepted 14 March 2014
Available online 27 March 2014

Keywords:
Multiple Markov jump process
Discrete observations
Asymptotic normality
Parametric maximum likelihood

ABSTRACT

For a continuous-time Markov process, occasionally, only discrete-time observations are available. For a simple sample of homogeneous Markov jump processes with an absorbing state, observed each on a stochastic grid of time points, we establish asymptotic normality of the maximum likelihood estimator and close the gap in Kremer and Weißbach (2013). By showing that the solution of the Kolmogorov backward equation system is continuous differentiable, we can apply results for M-estimators.

1. Introduction

Likelihood inference for the transition intensities of a discrete-state Markov jump process, based on continuous-time observations, is well established (see e.g. Billingsley, 1961; Küchler and Sørensen, 1997, among others). However, that continuous-time phenomena can only be observed at discrete time points has lately aroused some interest. For Markov jump processes, there are few results for the discretely-monitored processes. For instance, Keiding (1974, 1975) considers the estimation problem for a discretely-monitored birth process and birth-and-death process. Bladt and Sørensen (2005) estimate the intensity matrix from a discretely-monitored multiple Markov jump process. Dehay and Yao (2007) is a closely related work; their method is based on an explicit expansion of the transition matrix of the sampled chain. Furthermore, Bladt and Sørensen (2005) and Dehay and Yao (2007) prove consistency of the maximum likelihood estimator (MLE) provided that the process is ergodic.

A contemporary application of Markov jump processes is that of credit rating trajectories (see e.g. Jarrow et al., 1997; Bladt and Sørensen, 2009). It is important to note that a rating system usually includes a default class, which is an absorbing state, disabling ergodic theory as in Bladt and Sørensen (2005) and Dehay and Yao (2007) for an asymptotic analysis. Provided that the parameters are identifiable Kremer and Weißbach (2013) establish consistency using the continuous mapping theorem. We prove asymptotic normality of the MLE using properties of differential equations, namely continuity and Lipschitz-continuity, and a theorem on M-estimators by van der Vaart (1998) and close the gap in Kremer and Weißbach (2013).

The paper is organized as follows. Section 2 describes the general model with the corresponding likelihood. We establish asymptotic normality of the maximum likelihood estimator in Section 3.

2. Notation and likelihood

Consider the homogeneous continuous-time discrete-state Markov process $\boldsymbol{X} = \{X(t),\ t \in [0, T]\}$, $\mathcal{T} := [0, T]$, $T < \infty$, with ordered states $1, \ldots, m$, defined on a compact and metrical probability space $(\Omega, \mathfrak{F}, P)$ with filtration $(\mathfrak{F}_t)_{t \in \mathcal{T}}$.

* Corresponding author. Tel.: +49 3814984428; fax: +49 3814984401.
E-mail addresses: Alexander.Kremer3@uni-rostock.de (A. Kremer), rafael.weissbach@uni-rostock.de (R. Weißbach).

http://dx.doi.org/10.1016/j.spl.2014.03.010
0167-7152/

Additionally, we assume that the states $1, \ldots, m-1$ can be reached from each other. Moreover, we assume that m is an absorbing state and can be reached from all other states. $X(0)$ has a multinomial distribution on the states $1, \ldots, m-1$. Additionally, the infinitesimal generator $\mathbf{Q} = (\theta_{hj})_{h,j=1,\ldots,m}$ with transition intensities θ_{hj} for $h \neq j$, $h = 1, \ldots, m-1$, $j = 1, \ldots, m$ and 0 for $h = m, j = 1, \ldots, m$, determines the process. Note that $\theta_{hh} := -\sum_{\substack{j=1 \\ j\neq h}}^{m} \theta_{hj}$ for $h = 1, \ldots, m-1$. Let the vector $\boldsymbol{\theta} \in \Theta = \mathbb{R}^{m^2-2m+1}$ contain all positive non-redundant items of $\mathbf{Q}$, in any order. We refer to $\boldsymbol{\theta}_0$ as the true parameter and assume that identically distributed copies $\mathbf{X}_i$, $i = 1, \ldots, n$ of $\mathbf{X}$ are independent. Quite flexibly let the process path be observed on a stochastic grid of time points $T_{i,1}, T_{i,2}, \ldots, T_{i,N_i(T)}$ for each process. As a model, we assume that there is an independent process $\mathbf{N}_i$ which has a finite expected number of jumps with rate one, e.g., the homogeneous Poisson process or Cox process. We refer to this process as the observation process. Furthermore, we assume that the observation process is non-informative and independent of the Markov process $\mathbf{X}_i$. We interpret each jump $T_{i,z}$, $z = 1, \ldots, N_i(T)$ of the observation process $\mathbf{N}_i$ as a time point at which we observe the Markov process $\mathbf{X}_i$. Not that we allow a different observation process for each Markov process. Hence, we observe processes $\mathbf{Y}_i = \left\{X_i(T_{i,z}),\ z = 1, \ldots, N_i(T)\right\}$, $i = 1, \ldots, n$. The log likelihood conditional on $T_{i,z} = t_{i,z}$, $N_i(T) = n_i$, and $X_i(0)$ is

$$\ln L(\boldsymbol{\theta}|\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_n = \mathbf{y}_n) = \sum_{i=1}^{n}\sum_{z=1}^{n_i-1} \ln p(x_i(t_{i,z}), x_i(t_{i,z+1}), t_{i,z+1} - t_{i,z}, \boldsymbol{\theta}), \tag{1}$$

where $p(x_i(t_{i,z}), x_i(t_{i,z+1}), t_{i,z+1} - t_{i,z}, \boldsymbol{\theta})$ is the probability, depending on $\boldsymbol{\theta}$, that the process in state $x_i(t_{i,z})$ at age $t_{i,z}$ is in state $x_i(t_{i,z+1})$ after time $t_{i,z+1} - t_{i,z}$. For the detailed derivation of the log likelihood we refer to Kremer and Weißbach (2013) or Bladt and Sørensen (2005). For the ease of reading, we define

$$Z_i(\boldsymbol{\theta}) := \sum_{z=1}^{N_i(T)-1} \ln p(X_i(T_{i,z}), X_i(T_{i,z+1}), T_{i,z+1} - T_{i,z}, \boldsymbol{\theta}). \tag{2}$$

3. Asymptotic normality

Our aim is to prove the asymptotic normality. One of the main requirements for the estimator is to be consistent which follows from the continuous mapping theorem (see e.g. van der Vaart and Wellner, 1996, Theorem 3.2.2) and is verified for the present model in Kremer and Weißbach (2013). In order to apply van der Vaart (1998, Theorem 5.41, p. 68) or Liese and Miescke (2008, Theorem 7.142, p. 370), we recall some properties of differential equations (see Coddington and Levison, 1955, Chapter 1, Section 7). In detail, we follow Walter (2000, Chapter III, Sections 10–13) and Amann (1995, Chapter 2, Sections 6–9).

We assume in the following that D is an open subset of the finite dimensional Banach space E. E is the space of the transition matrices without row m, stacked to vectors, and $D \subset (\epsilon_D, 1-\epsilon_D)^{m(m-1)}$, with a small $\epsilon_D \in \mathbb{R}^+$. Furthermore, let J be a compact interval of $\mathbb{R}$, in our case the time line $[\epsilon_J, T]$, for a small $\epsilon_J \in \mathbb{R}^+$. Moreover, the parameter space Λ is open in a finite dimensional Banach space, in our case $\Lambda = (\epsilon_\Lambda, \infty)^{(m-1)^2}$, for a small $\epsilon_\Lambda \in \mathbb{R}^+$. Let us consider the ordinary differential equation of the first order

$$x' = f(t, x, \lambda) \quad \text{with } x(t_0) = \xi, \tag{3}$$

where $t \in J, x \in D, \lambda \in \Lambda$. E is the image space of the mapping $f : J \times D \times \Lambda \mapsto E$, $\mathfrak{D}(f, \Lambda)$ is the domain of the mapping $u : (t, t_0, \xi, \lambda) \mapsto u(t, t_0, \xi, \lambda) \in D$, and $u(t, t_0, \xi, \lambda)$ is the solution of the initial value problem (3). It is well known that we can describe (3) as an equation including an integral, namely

$$x = \xi + \int_{t_0}^{t} f(s, x, \lambda) ds. \tag{4}$$

A unique solution u is ensured to exist (see Walter (2000, Theorem II, p. 154) or Amann (1995, Theorem 7.6, p. 110)). Furthermore, the solution is continuous in all variables. Next, we recall that the solution u is differentiable (see Amann (1995, Theorem 9.2, p. 128) or Walter (2000, Proposition X, p. 160)). Furthermore, the derivative of u with respect to D is continuous (see Walter, 2000, Proposition X, p. 160). Especially, the derivative of u with respect to Λ is continuous (see Walter, 2000, Proposition, p. 162). In detail, the derivative of u with respect to λ has the form equal to (4) (see Walter, 2000, p. 168, formula (18))

$$\frac{\partial u}{\partial \lambda} = \int_{t_0}^{t} \frac{\partial f(s, u, \lambda)}{\partial \lambda} + \frac{\partial f(s, u, \lambda)}{\partial u}\frac{\partial u}{\partial \lambda} ds.$$

Provided that f is twice differentiable, we can apply the same conclusion again, because we have a linear integral equation of form (4). This yields the following corollary (see Amann, 1995, Theorem 9.5, p. 133), which establishes that the function u