Available online at www.sciencedirect.com
SCIENCE @ DIRECT®
C. R. Acad. Sci. Paris, Ser. I 342 (2006) 341–344
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistics

On likelihood estimation for a discretely observed jump process

Dominique Dehay, Jian-feng Yao

IRMAR, campus de Beaulieu, 35042 Rennes cedex, France

Received 25 May 2005; accepted after revision 20 December 2005

Available online 20 January 2006

Presented by Paul Deheuvels

Abstract

We consider the parameter estimation problem for a Markov jump process sampled at periodic epochs with a constant step. Unlike the diffusion case where a closed form of the likelihood function is usually unavailable, we provide here an explicit expression of the likelihood function of the sampled chain. Moreover under suitable ergodicity condition on the jump process, we establish the consistency and the asymptotic normality of the likelihood estimator as the observation period tends to infinity. ***To cite this article: D. Dehay, J.-f. Yao, C. R. Acad. Sci. Paris, Ser. I 342 (2006).***

Résumé

Sur l'estimation d'un processus de sauts discretisé. Soit un processus de sauts markovien observé en des temps discrets. À l'aide d'une formule explicite de la vraisemblance de la chaîne observée, nous proposons une théorie asymptotique de l'estimateur de vraisemblance. ***Pour citer cet article : D. Dehay, J.-f. Yao, C. R. Acad. Sci. Paris, Ser. I 342 (2006).***

1. Introduction

Consider a Markov jump process $X = (X_t)_{t\geqslant 0}$ with a countable state space E. We are given a Markov transition kernel $p = \{p(x, y),\ (x, y) \in E^2\}$ satisfying $p(x, x) = 0$ for all $x \in E$, and an intensity function $\lambda(\cdot)$ defined on E. The characteristic data $\lambda(\cdot)$ and $p(\cdot\,,\,\cdot)$ of the process are assumed to depend on a parameter $\theta \in \Theta \subset \mathbb{R}^s$. The aim of the Note is to estimate this parameter θ from regularly spaced observations of the process, that is from the *sampled chain* $Z = (Z_n)_{n\geqslant 0} := (X_{n\delta})_{n\geqslant 0}$. The sampling step is usually known in practice, then without loss of generality we will assume $\delta = 1$.

The likelihood estimation theory based on a continuous time observation of a Markov jump process $X = (X_t)_{t\geqslant 0}$ is classical and well-known. On the other hand, estimation theory from discrete observations has been developed more recently. In particular the estimation for a discretely observed diffusion process has been subject to an intensive research in the last decade [9,1,3]. One of the difficulties here is that usually the likelihood function cannot be obtained in a closed form. Consequently the direct maximum likelihood estimation is not available.

E-mail addresses: dominique.dehay@uhb.fr (D. Dehay), jian-feng.yao@univ-rennes1.fr (J.-f. Yao).

1631-073X/$ – see front matter © 2006 Académie des sciences. Published by Elsevier SAS. All rights reserved.
doi:10.1016/j.crma.2005.12.025

To the best of our knowledge, there are few results for a discretely observed Markov jump process. The parameter estimation problem for a discretely observed birth process and a birth-and-death process was considered by Keiding [7,8]. However, both situations are simpler as the maximum likelihood estimator is explicitly given by a formula. Note also that in the case of a birth-and-death process, the author assumed also the existence of some auxiliary statistics in addition of the sampled observations. For a general Markov jump process with finite state space, Bladt and Sørensen proposed in a recent work [2] two independent methods for estimating the intensity matrix (infinitesimal matrix).

In this Note we propose a different approach based on an explicit formula for the transition matrix of the sampled chain Z assuming that the intensity function $\lambda(\cdot)$ is bounded above and away from 0. This makes the likelihood estimation feasible. From a numerical point of view, accurate approximations of the transition matrix can be easily computed as the series converges exponentially fast.

Now we introduce that explicit formula for the transition matrix. Let Q be the intensity matrix (infinitesimal matrix) of the process X,

$$Q(\theta; x, y) = -\lambda(\theta; x)\delta_x(y) + \lambda(\theta; x)p(\theta; x, y),$$

where δ_x is the Dirac function at x. Assume that $\tilde{\lambda} := \sup_{x\in E} \lambda(\theta; x) \in (0, \infty)$ and define a new kernel

$$\tilde{p}(\theta; x, y) := \left(1 - \frac{\lambda(\theta; x)}{\tilde{\lambda}}\right)\delta_x(y) + \frac{\lambda(\theta; x)}{\tilde{\lambda}} p(\theta; x, y). \tag{1}$$

Then using matrix exponential calculus, it is easy to see that we have $q = \exp(Q) = \mathrm{e}^{-\tilde{\lambda}} \exp(\tilde{\lambda} I + Q) = \mathrm{e}^{-\tilde{\lambda}} \exp(\tilde{\lambda}\tilde{p})$, or explicitly

$$q(\theta; x, y) = \mathrm{e}^{-\tilde{\lambda}} \sum_{k=0}^{\infty} \frac{\tilde{\lambda}^k}{k!}\, \tilde{p}^k(\theta; x, y), \quad (x, y) \in E^2, \tag{2}$$

with $\tilde{p}^k$ the k-th power of the kernel $\tilde{p}$. It then becomes easy to compute the transition kernel $q(\theta; x, y)$ since in this formula the $\tilde{p}^k(\theta; x, y)$'s are bounded and the series converges exponentially fast.

As the state space is unbounded, the likelihood estimation theory is not straightforward. Traditionally important references in parametric estimation problem for Markov chains are Billingsley's papers [4,5] where many precise results, including the asymptotic normality as well as the limiting distributions of khi-square statistics are given. However the consistency in the strict sense, i.e. convergence of the likelihood estimator to the true parameter, was not established there. Moreover, Condition 1.1 used in [4] followed Cramér's analysis of the likelihood estimator. These regularity conditions, e.g. existence of third-order partial derivatives, are typically much more than necessary for the consistency. Another important reference is Dacunha-Castelle and Duflo's book [6] (Chapter 4) where a detailed study of the likelihood estimator was carried out. In particular the consistency problem was solved there by introducing a suitable continuity condition (w.r.t. the parameters) and an identifiability condition. Our approach follows the method developed in [6]. Although the results are quite standard, we have introduced several technical improvements using recent results from the stability theory of Markov chains.

2. Convergence of the likelihood estimator

The observation chain Z is a Markov chain with transition kernel $q(\theta; x, y)$. Let be the counting statistics $N_n(x, y) = \sum_{k=1}^{n} \mathbb{1}_{(Z_{k-1}, Z_k)=(x,y)}$. Therefore, conditionally to $Z_0 = X_0 = z$, the log-likelihood of $(Z_1, \ldots, Z_n)$ and the likelihood estimator are

$$L_n(\theta) = \sum_{k=1}^{n} \log q(\theta; Z_{k-1}, Z_k) = \sum_{x,y\in E} N_n(x, y) \log q(\theta; x, y), \quad \hat{\theta}_n := \arg\max_{\theta\in\Theta} L_n(\theta). \tag{3}$$

Before starting a mathematical analysis of this estimator, let us indicate how to compute it in practice. First, we may use some a priori bounds for the parameters or a preliminary estimator to set up the compact parameter space Θ. Next, we can use standard gradient descent method (from some numerical optimization toolbox) to find $\hat{\theta}_n$ with the aid of Eqs. (2) and (3).

The true value of the parameter is denoted by α. As for the recurrence of the Markov chain $(X_n)_{n\geqslant 0}$, we will assume throughout the Note the following condition

[R]: (i) *For some* $a \geqslant 1$, *under the true model* α, *the chain* (X_n) *has an unique invariant probability measure* μ_α *having a moment of order* a, *i.e.* $\sum_{x\in E} |x|^a \mu_\alpha(x) < \infty$.
(ii) *For any* μ_α*-integrable function* $\phi : E \to \mathbb{R}$, *the following Strong law of large numbers* (*SLLN*). *Holds we have for any initial condition* $X_0 = x$, $\frac{1}{n}\sum_{i=1}^n \phi(X_i) \xrightarrow{a.s.} \sum_{x\in E} \phi(x)\mu_\alpha(x)$.

It is worth noticing that a standard way to ensure such type of recurrence is to use a drift condition with the Lyapunov function $V(x) = |x|^a$, together with some continuity of the transition kernel.

2.1. Strong consistency

First call a *continuity modulus* any increasing function G defined on $[0,\infty)$ satisfying $\lim_{\theta\to 0} G(\theta) = G(0) = 0$. Throughout the article C will denote a generic constant.

Assumptions [**S**]:

(1) The parameter space Θ is a compact subset of $\mathbb{R}^s$.
(2) For all θ, $p(\theta;\cdot)$ is a irreducible kernel and $\lambda(\theta;\cdot)$ a positive function.
(3) (a) $|\log q(\alpha; x, y)| \leqslant C(1 + |x|^{a/2} + |y|^{a/2})$.
   (b) There exists a continuity modulus G such that for all $(x, y) \in E^2$ and $(\theta,\theta') \in \Theta^2$,

$$\bigl|\log q(\theta; x, y) - \log q(\theta'; x, y)\bigr| \leqslant G\bigl(|\theta-\theta'|\bigr)\bigl(1 + |x|^{a/2} + |y|^{a/2}\bigr).$$

Condition [S]-(1) is standard. Condition [S]-(2) is also basic; it implies in particular that $q(\theta; x, y) > 0$ for every (θ, x, y). Conditions [S]-(3) are continuity assumptions on $g = \log q$ together with its boundedness by a polynomial at infinity. Furthermore, these properties together with the ergodicity assumption (2) guarantee a SLLN for functions like $|g|$ or $|g|^2$. With respect to Condition [S]-(3.b), it is worth noticing that in Theorem 4.4.21 in [6] the authors assume that the family $\{\theta \mapsto g(\theta; , x, y)\}$ indexed by $(x, y) \in E^2$ is equicontinuous to get the consistency of the estimator. So Condition [S]-(3.b) is much weaker when the state space E is unbounded.

To ensure the fact that the true value α is the unique global maximum of the limiting function of L_n, we need the following identifiability condition

[D]: for any $\theta \neq \alpha$, $\mu_\alpha\{x:\ q(\theta; x, y) \neq q(\alpha; x, y),\ \text{for some } y \in E\} > 0$.

Theorem 2.1. *Assume that Conditions* [R], [S] *and* [D] *hold. Then the likelihood parameter estimator* $\hat\theta_n$ *is strongly consistent, i.e. for all* $x \in E$, $\mathbb{P}_{\alpha,x}$*-almost surely,* $\hat\theta_n \to \alpha$.

2.2. Asymptotic normality

For asymptotic normality of $\hat\theta_n$, we typically need some additional conditions on second order differentiability of the process (L_n). In the following partial derivatives of a function $\phi(\theta)$ are denoted $D_i\phi := \partial\phi/\partial\theta_i$, $D^2_{ij}\phi := \partial^2\phi/(\partial\theta_i\partial\theta_j)$.

Assumption [**N**]: Assume that α is an interior point of Θ, and there is a neighbourhood V of α where for any $(x, y) \in E^2$, the function $\theta \mapsto g(\theta;\ x, y)$ is twice continuously differentiable such that for all $i, j = 1,\dots,s$ we have

(1) (a) $|D_i \log q(\alpha;\ x, y)| \vee |D^2_{ij} \log q(\alpha;\ x, y)| \leqslant C(1 + |x|^{a/2} + |y|^{a/2})$;
   (b) there exists a continuity modulus σ_{ij} such that for $\theta \in V$, $(x, y) \in E^2$

$$\bigl|D^2_{ij} \log q(\theta;\ x, y) - D^2_{ij} \log q(\alpha;\ x, y)\bigr| \leqslant \sigma_{ij}\bigl(|\theta-\alpha|\bigr)\bigl(1 + |x|^{a/2} + |y|^{a/2}\bigr);$$

(2) for all $x \in E$, the family of transition kernels $\{q(\theta; x, \cdot);\ \theta \in V\}$ is *regular at* α in the sense that
   (a) $\sum_{y\in E}[D_i \log q(\alpha; x, y)]\, q(\alpha, x, y) = 0$.
   (b) $I_{ij}(x;\alpha) := \sum_{y\in E}[D_i \log q(\alpha;\ x, y)][D_j \log q(\alpha;\ x, y)]\, q(\alpha; x, y) = -\sum_{y\in E}[D^2_{ij} \log q(\alpha;\ x, y)]q(\alpha; x, y)$.

The assumption [N]-(2) is a natural extension of classical regularity conditions for i.i.d. samples to the present Markov chain case (see e.g. [6], §4.4). The matrix $I(\alpha;x) := [I_{ij}(\alpha;x)]$ is the Fisher information matrix at α associated to the family of distributions $\{q(\theta;x,\cdot)\colon\ \theta \in V\}$. Moreover, we will see below that under this assumption the matrix $I(\alpha) := \sum_{x\in E} I(x;\alpha)\mu_\alpha(x)$, is well-defined and usually called the (*asymptotic*) *Fisher information matrix* of the Markov chain (X_n).

Theorem 2.2. *Assume that Conditions* [R], [S] *and* [N] *hold and the matrix* $I(\alpha)$ *is invertible. Then for any weakly consistent estimator* $\hat\theta_n$, $\sqrt{n}\,(\hat\theta_n - \alpha)$ *converges in distribution to the zero-mean* s*-dimensional Gaussian distribution with covariance matrix* $I(\alpha)^{-1}$.

References

[1] Y. Aït-Sahalia, Maximum likelihood estimation for discretely sampled diffusions: a closed-form approximation approach, Econometrica 70 (2002) 223–262.
[2] M. Bladt, M. Sørensen, Statistical inference for discretely observed Markov jump processes, J.R.S.S (B) 67 (3) (2005) 395–410.
[3] B.M. Bibby, M. Jacobsen, M. Sørensen, Estimation functions for discretely sampled diffusion-type models, in: Y. Aït-Sahalia, L.P. Hansen (Eds.), Handbook of Financial Econometrics, North-Holland, Amsterdam, in press.
[4] P. Billingsley, Statistical Inference for Markov Processes, The University of Chicago Press, Chicago, 1961.
[5] P. Billingsley, Statistical methods in Markov chains, Ann. Math. Statist. 32 (1961) 12–40.
[6] D. Dacunha-Castelle, M. Duflo, Probability and Statistics, vol. II, Springer-Verlag, New York, 1986.
[7] N. Keiding, Estimation in the birth process, Biometrika 61 (1974) 71–80.
[8] N. Keiding, Maximum likelihood estimation in the birth-and-death process, Ann. Statist. 3 (1975) 363–372.
[9] M. Kessler, M. Sørensen, Estimating equations based on eigenfunctions for discretely observed diffusion processes, Bernoulli 5 (1999) 299–314.