Aust. N. Z. J. Stat. **49**(1), 2007, 93–107 doi: 10.1111/j.1467-842X.2006.00466.x

ON LIKELIHOOD ESTIMATION FOR DISCRETELY OBSERVED MARKOV JUMP PROCESSES

Dominique Dehay[1]* and Jian-Feng Yao[2]

Institut de Recherche Mathématique de Rennes, France

Summary

The parameter estimation problem for a Markov jump process sampled at equidistant time points is considered here. Unlike the diffusion case where a closed form of the likelihood function is usually unavailable, here an explicit expansion of the likelihood function of the sampled chain is provided. Under suitable ergodicity conditions on the jump process, the consistency and the asymptotic normality of the likelihood estimator are established as the observation period tends to infinity. Simulation experiments are conducted to demonstrate the computational facility of the method.

Key words: discrete observations; likelihood estimator; Markov jump process.

1. Introduction

Consider a Markov jump process $X=(X_t)_{\{t\geq 0\}}$ with a countable state space E. The process is defined by a Markov transition kernel $p\equiv\{p(x,y)\colon (x,y)\in E^2\}$ satisfying $p(x,x)=0$ for all $x\in E$, and an intensity function $\lambda(\cdot)$ defined on E. The defining functions $\lambda(\cdot)$ and $p(\cdot\,,\cdot)$ of the process are assumed to depend on a parameter $\theta\in\Theta\subset\mathbb{R}^s$, for some integer $s>0$. The aim of this paper is to estimate this parameter θ from a regularly time-spaced observation of the process, that is from the sampled chain $Z\equiv(Z_n)_{\{n\in\mathbb{N}\}}=(X_{n\delta})_{\{n\in\mathbb{N}\}}$. The sampling step is usually known in practice; then, without loss of generality, the assumption $\delta=1$ is made to simplify the notation.

The likelihood estimation theory based on a continuous time observation of a Markov jump process $X=(X_t)_{\{t\geq 0\}}$ is classical and well-known; see, for example, Billingsley (1961a) and Jacobsen (1982). However, estimation theory for discrete observations has been developed more recently. In particular, intensive research has appeared in the last decade for the estimation of a discretely-observed diffusion process; see for example Mishra & Prakasa Rao (2001), Jacod (2007), Aït-Sahalia (2002), Bibby *et al.* (2007). The main difficulty comes from the fact that the likelihood function cannot usually be obtained in a closed form. Consequently, direct maximum likelihood estimation is not available.

To the best of the authors' knowledge, there are few results for a discretely-observed Markov jump process. The parameter estimation problem for a discretely-observed birth

*Author to whom correspondence should be addressed.

[1]UFR Sciences Sociales, Université de Rennes 2 Haute Bretagne, Place du recteur H. Le Moal, 35043 Rennes, France.
e-mail: dominique.dehay@uhb.fr.

[2]IRMAR, Université de Rennes 1, 35042 Rennes cedex, France.

process and a birth-and-death process was considered by Keiding (1974, 1975). These situations, however, are simpler as the maximum likelihood estimator is explicitly given by a formula. Note that, in the case of a birth-and-death process, Keiding also assumed some auxiliary information in addition to the sampled observations.

A closely related work is Bladt & Sørensen (2005), where the authors proposed two different methods for estimating the intensity matrix (also called the infinitesimal matrix) of a general Markov jump process with finite state space: namely, a maximum likelihood estimator and a Bayesian estimator obtained by a Markov Chain Monte Carlo procedure. A different approach is proposed here in a parametric framework, where the parameters can be different from the terms of the intensity matrix. Furthermore, strong consistency and asymptotic normality are established under fairly general assumptions. Another difference from the cited reference is that the method presented here applies to any discrete state space (not necessarily finite).

The method presented here is based on an explicit expansion 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. This boundedness condition, however, excludes the models studied by Keiding (1974, 1975).

The explicit formula for the transition matrix is now introduced. Indeed, this formula is well-known; see, for example, Jensen (1953) and Ethier & Kurtz (1986). Let Q be the intensity 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, and $\theta \in \Theta$ is fixed. Assume that $\widetilde{\lambda}(\theta) = \sup_{x\in E} \lambda(\theta; x) \in (0, \infty)$. Then a new kernel can be defined:

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

It follows that

$$q = \exp(Q) = e^{-\widetilde{\lambda}} \exp(\widetilde{\lambda} I + Q) = e^{-\widetilde{\lambda}} \exp(\widetilde{\lambda}\widetilde{p}),$$

or explicitly

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

where $\widetilde{p}^k(\theta; x, y)$ denotes the (x, y)-term of the kth power of the kernel matrix $\widetilde{p}(\theta; \cdot)$. From a numerical point of view, accurate approximations of the transition kernel $q(\theta;\ x,\ y)$ can be computed since, in this formula, the $\widetilde{p}^k(\theta; x, y)$s are bounded and the series converges exponentially fast. It is worth noting that (2) is still valid when $\widetilde{\lambda}(\theta)$ is replaced by any finite value greater than $\widetilde{\lambda}(\theta)$, and $\widetilde{p}(\theta; x, y)$ by the corresponding value obtained by (1).

As the state space is not necessarily finite, the likelihood estimation theory is not straightforward. Traditional important references in parametric estimation problems for Markov chains are the papers of Billingsley (1961a,b) where many precise results, including asymptotic normality as well as limiting distributions of chi-square statistics, are given. Consistency, however, in the strict sense, i.e. convergence of the likelihood estimator to the true parameter, was not established there. Moreover, the regularity conditions used by Billingsley (1961a, Condition 1.1), e.g. the existence of third-order partial derivatives, are typically much more than necessary for consistency. Another important reference is the book of Dacunha-Castelle & Duflo (1986, Chapter 4), where a detailed study of the likelihood estimator was carried out. In particular, the consistency problem was solved by introducing a suitable continuity condition with respect to the parameter, and an identifiability condition.

The consistency and the asymptotic normality of the likelihood estimator are established in Section 2, following the method developed by Dacunha-Castelle & Duflo (1986) is essentially followed. Although the results are quite standard, several technical improvements are introduced using recent results from the stability theory of Markov chains as developed by Duflo (1997) and by Meyn & Tweedie (1993). For the reader's convenience, the main steps of the proofs are postponed until Section 5. In Section 3 two specific examples are considered (a finite state space model and the M/M/1 queue), and a Lyapunov drift criterion is given for the recurrence of the sampled chain (Subsection 3.3). In Section 4 the paper is completed with simulation experiments for the two models studied in the previous section, proving the computational feasibility of the likelihood estimator method.

2. Convergence of the Likelihood estimator

The observation chain Z is a Markov chain with transition kernel $q(\theta;\ x,\ y) = \mathrm{P}(X_1 = y|X_0 = x)$. Therefore, conditionally on $Z_0 = X_0 = z$, the log-likelihood of $(Z_1, \dots, Z_n)$ is

$$l_n(\theta) = \sum_{k=1}^{n} \log q(\theta; Z_{k-1}, Z_k) = \sum_{x,y\in E} N_n(x, y) \log q(\theta; x, y), \tag{3}$$

where $N_n(x, y) = \sum_{k=1}^{n} \mathbb{I}_{\{(Z_{k-1},Z_k)=(x,y)\}}$, where $\mathbb{I}_A$ denotes the indicator function of set A. The likelihood estimator is then defined as

$$\widehat{\theta}_n = \arg \max_{\theta\in\Theta} l_n(\theta).$$

For simplicity, the state space E is embedded in $\mathbb{R}^k$ with the Euclidean norm denoted by $|\cdot|$. The true value of the parameter is denoted by α. Therefore $\mathrm{P}_{\alpha,x}$ stands for the probability distribution of the chain $(Z_n)_{\{n\in\mathbb{N}\}} = (X_n)_{\{n\in\mathbb{N}\}}$ under the true model with the initial condition $Z_0 = x$. Moreover, the notation $\xrightarrow{a.s.}$ means an almost sure convergence under $\mathrm{P}_{\alpha,x}$ for every x.

As for the recurrence of the Markov chain $(Z_n)_{\{n\in\mathbb{N}\}}$, the following condition will be assumed throughout the paper.

Assumptions (R).

(R1) Under the true model α, the Markov chain $(Z_n)_{\{n\in\mathbb{N}\}}$ has an unique invariant probability measure μ_α having a moment of order a, for some $a \geq 1$, i.e. $\sum_{x\in E} |x|^a \mu_\alpha(x) < \infty$.

(R2) For any μ_α-integrable function $\phi : E \to \mathbb{R}$, the following strong law of large numbers (SLLN) holds:

$$\frac{1}{n}\sum_{i=1}^{n} \phi(Z_i) \xrightarrow{a.s.} \sum_{x\in E} \phi(x)\mu_\alpha(x), \qquad n \to \infty.$$

Note that a standard way to ensure such kinds of recurrence is to use a drift condition with the Lyapunov function $V(x) = |x|^a$, together with some continuity condition on the transition kernel. In Section 3.3 an explicit example of this approach will be given. It is also noted that a similar approach was employed by Yao (2000) to derive the asymptotic properties of the least-squares estimator for a stable nonlinear autoregressive process.

2.1. Strong consistency

Define the estimating function (or contrast)

$$U_n(\theta) = l_n(\alpha) - l_n(\theta),$$

so that the likelihood estimator $\widehat{\theta}_n$ minimizes $U_n(\cdot)$. Let $g(\theta; x, y) = \log q(\alpha; x, y) - \log q(\theta; x, y)$. Then the Kullback–Leibler information between the probability measures $q(\alpha; x, \cdot)$ and $q(\theta; x, \cdot)$ is given by

$$K(\alpha, \theta; x) = \mathrm{E}_\alpha(g(\theta; Z_0, Z_1) \mid Z_0 = x) = \sum_{y \in E} g(\theta; x, y)\, q(\alpha; x, y).$$

It follows that $(U_n(\theta))_{\{n \in \mathbb{N}\}}$ is a sub-martingale having the decomposition

$$U_n(\theta) = \sum_{i=1}^{n} (g(\theta; Z_{i-1}, Z_i) - K(\alpha, \theta; Z_{i-1})) + \sum_{i=1}^{n} K(\alpha, \theta; Z_{i-1}). \tag{4}$$

To go further, the following assumptions are required. First, call a continuity modulus any increasing function $G(\cdot)$ defined on $[0, \infty)$ satisfying $\lim_{v \to 0} G(v) = G(0) = 0$. Throughout the paper, C will denote a generic constant.

Assumptions (S).

(S1) The parameter space Θ is a compact subset of $\mathbb{R}^s$.

(S2) For all θ, $p(\theta; \cdot)$ is an irreducible kernel and $\lambda(\theta; \cdot)$ a positive function.

(S3) (i) $|\log q\ (\alpha; x, y)| \le C\ (1 + |x|^{a/2} + |y|^{a/2})$ for all $x, y \in E$;

(ii) there exists a continuity modulus $G(\cdot)$ such that, for all $x, y \in E$ and $\theta, \theta' \in \Theta$,

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

Condition (S1) is standard. Condition (S2) is also basic; it implies that $q(\theta; x, y) > 0$ for every (θ, x, y). Condition (S3) together with the ergodicity assumption (R2) guarantee a SLLN for functions like $|g(\cdot)|$ or $|g(\cdot)|^2$. Dacunha-Castelle & Duflo (1986, Theorem 3.2.8) assumed that the family of functions $\{\theta \mapsto g(\theta; x, y)\colon x, y \in E\}$ is equicontinuous to get the consistency of the maximum likelihood estimator of the parameter of a Markov chain. Here condition (S3)(ii) is much weaker when the state space E is unbounded.

The following proposition determines the limit of the estimating function $U_n(\cdot)$.

Proposition 1. *Under assumptions (R) and (S),*

$$\frac{1}{n} U_n(\theta) \xrightarrow{a.s.} k(\theta) = \sum_{x \in E} K(\alpha, \theta; x)\, \mu_\alpha(x) \tag{5}$$

as $n \to \infty$, for any fixed $\theta \in \Theta$. Moreover the limit function $k(\cdot)$ is continuous in θ and non-negative.

By definition, $k(\alpha) = 0$, so the true value α is a global minimum of $k(\cdot)$. Whether or not it is the unique minimum depends on the identifiability of the model. A natural condition in the current context is the following condition.

Assumption (D).

(D) For any $\theta \neq \alpha$, $q(\theta;\cdot) \neq q(\alpha;\cdot)$.

Therefore, under (D), $k(\theta) = 0$ if and only if $\theta = \alpha$. The model is said to be identifiable at α. However, it is not trivial to check this identifiability condition in practice. This problem is not yet solved in full generality. For finite state space Markov jump processes, Bladt & Sørensen (2005) give a detailed discussion on this subject. Nevertheless, in the case of a M/M/1 queue discussed below, this identifiability problem can be solved.

Theorem 1. *Let assumptions (R), (S) and (D) hold. Then the likelihood parameter estimator $\widehat{\theta}_n$ is strongly consistent, i.e. $\widehat{\theta}_n \xrightarrow{a.s.} \alpha$ as $n \to \infty$.*

2.2. Asymptotic normality

For asymptotic normality of $\widehat{\theta}_n$, some additional conditions are typically required on the second order differentiability of the estimating function $U_n(\cdot)$. In the following, partial derivatives of a function $\phi(\cdot)$ of θ are denoted $D_i\phi = \partial\phi/\partial\theta_i$, $D^2_{ij}\phi = \partial^2\phi/(\partial\theta_i\partial\theta_j)$.

Assumptions (N). Assume that α is an interior point of Θ, and there is a neighborhood $\mathcal{V}_\alpha$ of α where, for any $(x, y) \in E^2$, the function $\theta \mapsto g(\theta; x, y)$ is twice continuously differentiable and satisfies, for all $i, j = 1, \dots, s$, the following conditions:

(N1) (i) $\max\left\{|D_i \log q(\alpha; x, y)|, \left|D^2_{ij} \log q(\alpha; x, y)\right|\right\} \le C\left(1 + |x|^{a/2} + |y|^{a/2}\right)$;

(ii) there exists a continuity modulus σ_{ij} such that, for $\theta \in \mathcal{V}_\alpha$, $(x, y) \in E^2$,

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

(N2) for every $x \in E$, the family of transition kernels $\{q(\theta; x, \cdot)\colon \theta \in \mathcal{V}_\alpha\}$ is regular at α, in the sense that

(i) $\displaystyle\sum_{y\in E} (D_i \log q(\alpha; x, y))\, q(\alpha, x, y) = 0$;

(ii)
$$I_{ij}(\alpha; x) = \sum_{y\in E} (D_i \log q(\alpha; x, y))(D_j \log q(\alpha; x, y))\, q(\alpha; x, y) = -\sum_{y\in E}\left(D^2_{ij} \log q(\alpha; x, y)\right) q(\alpha; x, y).$$

Condition (N1) is similar to (S3), guaranteeing a SLLN for functions involving first or second order derivatives of $g(\cdot)$. Condition (N2) is a natural extension of classical regularity conditions for independent and identically distributed samples to the present Markov chain case (see e.g. Dacunha-Castelle & Duflo, 1986, Section 4.4). The matrix $I(\alpha; x) = [I_{ij}(\alpha; x)]_{i,j=1,\dots,s}$ is the Fisher information matrix at α associated with the family of distributions $\{q(\theta; x, \cdot) : \theta \in \mathcal{V}_\alpha\}$. Moreover, it will be seen below that, under (N2), the matrix

$$I(\alpha) = \sum_{x \in E} I(\alpha; x)\, \mu_\alpha(x),$$

is well-defined. It is usually called the (asymptotic) Fisher information matrix of the Markov chain $(Z_n)_{\{n \in \mathbb{N}\}}$.

Theorem 2. *Let assumptions (R), (S) and (N) hold, and the matrix $I(\alpha)$ be invertible. Then $\sqrt{n}(\widehat{\theta}_n - \alpha)$ converges in distribution to the zero-mean s-dimensional Gaussian distribution with covariance matrix $I(\alpha)^{-1}$, as $n \to \infty$, for every weakly consistent estimator $\widetilde{\theta}_n$ of α.*

3. Examples and remarks

3.1. Jump process on a finite state space

When E is finite, assumptions (S) and (N) become much simpler. Consider first (S). When the jump kernels $p(\theta;\cdot)$ are irreducible, $q(\theta; x, y) > 0$ for every (θ, x, y). Therefore the chain $(Z_n)_{\{n \in \mathbb{N}\}}$ is irreducible and positive recurrent.

Note, however, that if the maps $\theta \mapsto p(\theta; x, y)$ and $\theta \mapsto \lambda(\theta; x)$ are continuous, then the same holds for the map $\theta \mapsto \log q(\theta; x, y)$ since the series on the right hand side of (2) converges uniformly on the compact set $\Theta \subset \mathbb{R}^s$. As the state space is finite, there exists a continuity modulus $G(\cdot)$ such that, for all $(x, y) \in E^2$ and $(\theta, \theta') \in \Theta^2$,

$$|\log q(\theta; x, y) - \log q(\theta'; x, y)| \le G(|\theta - \theta'|).$$

Thus (S3) holds. The following corollary is obtained as a consequence of Theorem 1.

Corollary 1. *Let $X = (X_t)_{\{t \ge 0\}}$ be a Markov jump process with a finite state space such that*

(i) *the parameter space $\Theta \subset \mathbb{R}^s$ is compact;*
(ii) *for all θ, $p(\theta;\cdot)$ is an irreducible kernel and $\lambda(\theta;\cdot)$ is a positive function;*
(iii) *for all $x, y \in E$, the functions $\theta \mapsto p(\theta; x, y)$ and $\theta \mapsto \lambda(\theta; x)$ are continuous;*
(iv) *the model is identifiable at α: for $\theta \ne \alpha$, $q(\theta;\cdot) \ne q(\alpha;\cdot)$.*

The likelihood estimator $\widehat{\theta}_n$ is then strongly consistent.

Conditions (N1)(i), (N1)(ii) and (N2) are also automatically satisfied whenever the state space is finite, and the following corollary of Theorem 2 is obtained.

Corollary 2. *In addition to the assumptions of Corollary 1, it is assumed that the true value α of the parameter is an interior point of Θ and there is a neighborhood $\mathcal{V}_\alpha$ of α where $\theta \mapsto p(\theta; x, y)$ is twice-continuously differentiable for all $x, y \in E$. Assume also that the associated Fisher information matrix $I(\alpha)$ is invertible. Then $\sqrt{n}(\widehat{\theta}_n - \alpha)$ converges in distribution to the zero-mean s-dimensional Gaussian distribution with covariance matrix $I(\alpha)^{-1}$.*

3.2. Case of a M/M/1 queue

A M/M/1 queue process is a Markov jump process defined on the state space $E = \mathbb{N}$ and its transition kernel is characterized by the inter-arrival rate $\beta > 0$ and the service rate $\delta > 0$.

The parameter of the process is the pair $\theta = (\beta, \delta)$. The transition kernels $p(\theta; \cdot)$ and $q(\theta; \cdot)$ can be obtained in a straightforward way; see e.g. Asmussen (2003, Section III.8).

Furthermore, under the condition $\rho = \beta/\delta < 1$, it is well-known that the Markov process M/M/1, as well as the sampled chain $(Z_n)_{\{n\in\mathbb{N}\}}$, are irreducible and positive recurrent. Moreover, the invariant probability distribution is the geometric distribution $\mu_\theta(x) = (1-\rho)\rho^x,\ x \in \mathbb{N}$. In particular, all moments of μ are finite. Hence the sampled chain satisfies the recurrence condition (S2) for any exponent $a \geq 1$.

The problem of the identifiability of the $M/M/1$ queue model can be solved in the following way. Assume that $\theta \neq \theta'$ where $\theta' = (\beta', \delta')$ and $\rho' = \beta'/\delta' < 1$. If $\rho \neq \rho'$, the corresponding invariant probability measures μ_θ and $\mu_{\theta'}$ are distinct, and so are the transition probability kernels $q(\theta; \cdot)$ and $q(\theta'; \cdot)$. On the other hand, assume that $\rho = \rho'$ so that $\beta\delta \neq \beta'\delta'$. Thanks to Takács (1962) (see Asmussen, 2003, Chapter III Formula 8.8), the transition probability $q(\theta; 0, 0)$ can be expressed as

$$q(\theta; 0, 0) = (1-\rho) + \frac{\rho}{\pi}\int_{-\pi}^{\pi} \frac{\exp\left\{-\frac{\beta\delta}{\sqrt{\rho}}\left|1-\sqrt{\rho}\,e^{iu}\right|^2\right\}}{\left|1-\sqrt{\rho}\,e^{iu}\right|^2}\sin^2 u\,du.$$

As a function of $\beta\delta$, this expression is strictly decreasing for $0 < \rho < 1$ fixed. Hence, $q(\theta; 0, 0) \neq q(\theta'; 0, 0)$ and the model is identifiable at $\theta = (\beta, \delta)$.

By elementary but cumbersome computations on the transition probabilities, the consistency and the asymptotic normality of the likelihood estimator can be obtained by adapting Theorems 1 and 2 to the present case.

3.3. A Lyapunov method to check the recurrence condition

It is assumed in this subsection that the state space $E \subset \mathbb{R}^k$ is countable. In considering the recurrence condition (R) for the sampled chain, it may be difficult to carry out a direct analysis as has been done above for the M/M/1 queue. Here a method is proposed based on a Lyapunov drift condition in terms of $p(\alpha; \cdot)$.

Proposition 2. *Assume that*

(i) *for all* x, $0 < c \leq \lambda(\alpha; x) \leq d < \infty$ *for some constants* c *and* d;

(ii) *the jump kernel* $p = p(\alpha; \cdot)$ *is irreducible, satisfying the Lyapunov drift condition*

$$pV(x) \ \leq\ \beta V(x) + \gamma,$$

with $V(x) = |x|^a$ *and constants* $a \geq 1$, $0 \leq \beta < 1$ *and* $\gamma \geq 0$.

Then, under the true model α, *the sampled chain* Z *satisfies the recurrence condition (R) with index* a.

Proof. First it is proven that, since the jump kernel $p(\alpha; \cdot)$ satisfies a Lyapunov drift condition, then the same holds for the kernel $q(\alpha; \cdot)$ of the sampled chain Z.

Assume first that the intensity function $\lambda(\alpha; \cdot)$ is constant. Then the drift condition in (ii) implies that $p^k V \leq \beta^k\ V + \gamma'$ for any k, where $\gamma' = \gamma/(1-\beta)$. Hence the transition kernel $q = q(\alpha; \cdot)$ of the sampled chain Z satisfies

$$qV(x) \ =\ e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\ p^k V(x) \ \leq\ \beta' V(x) + \gamma',$$

where $\beta' = e^{-\lambda(1-\beta)}$ and $0 < \beta' < 1$.

In the general case, where the intensity function is not necessarily constant, from (i) it can be seen that $0 < c \le \widetilde{\lambda}(\alpha) = \sup_{x\in E} \lambda(\alpha; x) \le d < \infty$. Hence, (1) and (ii) imply that

$$\begin{aligned}\widetilde{p}V(x) &= \left(1 - \frac{\lambda(\alpha; x)}{\widetilde{\lambda}(\alpha)}\right)V(x) + \frac{\lambda(\alpha; x)}{\widetilde{\lambda}(\alpha)}\, pV(x) \\ &\le \left(1 - \frac{c}{\widetilde{\lambda}(\alpha)}(1-\beta)\right)V(x) + \gamma\,.\end{aligned}$$

According to the inequalities $0 \le 1 - c\,\widetilde{\lambda}(\alpha)^{-1}(1-\beta) < 1$, and from the previous computations, it is readily deduced that the kernel $q = q(\alpha;\,\cdot)$ satisfies the Lyapunov drift criterion

$$qV(x) \;\le\; \beta'' V(x) + \gamma'',$$

for some $0 \le \beta'' < 1$ and $\gamma'' \ge 0$.

On the other hand, it is known that the state space E is countable and the jump kernel $p(\alpha;\,\cdot)$ is irreducible. Thus the $q(\alpha;\, x,\, y)$s are positive and the sampled chain Z is irreducible and aperiodic. With the Lyapunov drift criterion, it is deduced that Z has an invariant probability and satisfies the ergodicity condition (R) (see Meyn & Tweedie, 1993).

4. Simulation experiments

In all the experiments reported below, the nonlinear optimization program OPTIM (Scilab, `http://www.scilab.org/`) is used to compute the likelihood estimator. As the likelihood function relies on the expansion (2), there is not an explicit expression for its gradient. Therefore, an approximation of the gradient is computed by a finite-difference scheme using the SCILAB function DIFFERENCE.

4.1. A three-states example

Let the state space be $E = \{1, 2, 3\}$. The parameters of the process are $\theta = (\lambda_1, \lambda_2, \lambda_3, a_1, a_2, a_3)$, where the a_ks define the transition matrix

$$p = \begin{pmatrix} 0 & a_1 & 1-a_1 \\ 1-a_2 & 0 & a_2 \\ a_3 & 1-a_3 & 0 \end{pmatrix}.$$

The true values of the parameters are set to be $\alpha = (3,\ 1,\ 2,\ 0.3,\ 0.6,\ 0.8)$ leading to the following (true) transition matrices

$$p_\alpha = \begin{pmatrix} 0 & 0.3 & 0.7 \\ 0.4 & 0 & 0.6 \\ 0.8 & 0.2 & 0 \end{pmatrix}, \qquad q_\alpha = \begin{pmatrix} 0.2795 & 0.3196 & 0.4009 \\ 0.1995 & 0.5012 & 0.2993 \\ 0.2808 & 0.2857 & 0.4335 \end{pmatrix}.$$

Here, the matrix q_α is computed via (2).

For three different time durations $T = 100$, 500 and 2000, 100 independent paths of the jump process X were simulated on the interval $[0, T]$. The sampled chains Z were then extracted at times 1,2,..., $n = T$. For each sampled chain, the likelihood estimator $\widehat{\theta_n}$ was computed. An estimate $q_{\widehat{\theta_n}}$ of the transition matrix q_α follows from (2). Moreover, the

TABLE 1

Averaged bias and standard deviation of the likelihood estimates of the parameters $\widehat{\theta}_n$ of the transition matrix $q_{\widehat{\theta}_n}$ and of the Kullback distance $k(\widehat{\theta}_n)$ from 100 runs of sampled chains of length 100, 500 and 2000, respectively. Note that $(\lambda_1, \lambda_2, \lambda_3, a_1, a_2, a_3) = (3, 1, 2, 0.3, 0.6, 0.8)$.

	$n = T = 100$		$n = T = 500$		$n = T = 2000$	
Bias	average	std. dev.	average	std. dev.	average	std. dev.
$\widehat{\lambda}_1 - \lambda_1$	−0.2995	1.3093	0.6723	1.2295	0.0331	0.9724
$\widehat{\lambda}_2 - \lambda_1$	0.1042	0.4007	0.1164	0.2208	−0.0012	0.0586
$\widehat{\lambda}_3 - \lambda_3$	−0.2027	0.8840	0.3555	0.9201	0.0682	0.5626
$\widehat{a}_1 - a_1$	0.0417	0.3021	0.0143	0.1908	0.0009	0.0917
$\widehat{a}_2 - a_2$	−0.0750	0.3730	−0.1552	0.3128	0.0335	0.1716
$\widehat{a}_3 - a_3$	−0.1174	0.2652	0.0183	0.1525	−0.0256	0.0886
$(q_{\widehat{\theta}_n} - q_\alpha)_{11}$	0.0223	0.0701	0.0019	0.0288	0.0060	0.0180
$(q_{\widehat{\theta}_n} - q_\alpha)_{21}$	0.0054	0.0497	0.0196	0.0240	−0.0048	0.0135
$(q_{\widehat{\theta}_n} - q_\alpha)_{31}$	−0.0240	0.0597	−0.0022	0.0286	−0.0047	0.0113
$(q_{\widehat{\theta}_n} - q_\alpha)_{12}$	−0.0030	0.0780	0.0128	0.0389	−0.0008	0.0195
$(q_{\widehat{\theta}_n} - q_\alpha)_{22}$	−0.0024	0.0786	−0.0124	0.0418	0.0014	0.0151
$(q_{\widehat{\theta}_n} - q_\alpha)_{32}$	0.0039	0.0705	0.0090	0.0343	0.0042	0.0154
$(q_{\widehat{\theta}_n} - q_\alpha)_{13}$	−0.0193	0.0814	−0.0148	0.0353	−0.0051	0.0220
$(q_{\widehat{\theta}_n} - q_\alpha)_{23}$	−0.0030	0.0645	−0.0072	0.0362	0.0034	0.0172
$(q_{\widehat{\theta}_n} - q_\alpha)_{33}$	0.0200	0.0664	−0.0068	0.0363	0.0005	0.0181
$k(\widehat{\theta}_n)$	0.0230	0.0143	0.0059	0.0034	0.0013	0.0008

convergence of $\widehat{\theta}_n$ can also be accessed by the evaluation (5) of the Kullback deviation $k(\widehat{\theta}_n)$ of $q_{\widehat{\theta}_n}$ from q_α.

Table 1 displays the mean and standard deviation statistics for the bias $\widehat{\theta}_n - \alpha$, $q_{\widehat{\theta}_n} - q_\alpha$ and for $k(\widehat{\theta}_n)$. As the sample size n increases, the standard deviations of all estimates decrease as expected. More interestingly, the corresponding mean Kullback distance, 0.0230, 0.0059 and 0.0013, decreases significantly. For the parameters λ_ks and a_ks, although the bias of the estimates remain small, some deviations are quite large, especially for the biggest λ_1. This is not surprising since the sojourn time at this state is the lowest, so that this state is the least favored by the discrete time sampling scheme. Finally, the estimation method in this case takes about 0.6 second (CPU time) on a DELL Latitude 600 notebook (running Linux).

In Table 2, the estimations obtained are displayed using 20 runs from the three different lengths. The estimates are very close to those obtained with 100 runs.

In Table 3, the estimates obtained are displayed using a single run for the three different sample lengths. These values give an idea about the fluctuations of the estimates, which are significantly bigger for the parameters (λ_j, a_j), especially λ_1, than for the Kullback distance or the transition probabilities q_{ij}. This is due to the small values taken by the gradients for the (λ_j, a_j)s.

4.2. An infinite state space case: the M/M/1 queue

This section considers the M/M/1 queue discussed in Section 3.2. As the state space is infinite, a difficulty occurs in the computation of the log likelihood function $l_n(\cdot)$. More specifically, for the use of (2) for a given set of parameters θ, the infinite transition matrix $q(\theta;\, \cdot)$ needs to be truncated. A practical rule is proposed as follows. It is noted from (3)

TABLE 2

Averaged bias and standard deviation of the likelihood estimates of the parameters $\widehat{\theta}_n$ of the transition matrix $q_{\widehat{\theta}_n}$ and of the Kullback distance $k(\widehat{\theta}_n)$ from 20 runs of sampled chains of length 100, 500 and 2000, respectively. Note that $(\lambda_1, \lambda_2, \lambda_3, a_1, a_2, a_3) = (3, 1, 2, 0.3, 0.6, 0.8)$.

	$n = T = 100$		$n = T = 500$		$n = T = 2000$	
Bias	average	std. dev.	average	std. dev.	average	std. dev.
$\widehat{\lambda}_1 - \lambda_1$	−0.2225	1.5192	1.0565	1.4022	−0.2218	0.8351
$\widehat{\lambda}_2 - \lambda_1$	0.0194	0.4553	0.0644	0.1302	0.0051	0.0420
$\widehat{\lambda}_3 - \lambda_3$	0.0043	1.2105	0.6403	1.1169	0.0091	0.4438
$\widehat{a}_1 - a_1$	0.0109	0.2840	−0.0085	0.1697	−0.0051	0.0852
$\widehat{a}_2 - a_2$	0.0116	0.3305	−0.0973	0.2905	0.0873	0.1699
$\widehat{a}_3 - a_3$	−0.0806	0.2291	0.0548	0.1274	−0.0566	0.1013
$(q_{\widehat{\theta}_n} - q_\alpha)_{11}$	0.0462	0.0957	0.0010	0.0236	0.0123	0.0193
$(q_{\widehat{\theta}_n} - q_\alpha)_{21}$	−0.0105	0.0410	0.0119	0.0200	−0.0062	0.0159
$(q_{\widehat{\theta}_n} - q_\alpha)_{31}$	−0.0183	0.0673	0.0010	0.0279	−0.0046	0.0113
$(q_{\widehat{\theta}_n} - q_\alpha)_{12}$	−0.0187	0.0765	0.0094	0.0319	−0.0042	0.0195
$(q_{\widehat{\theta}_n} - q_\alpha)_{22}$	0.0330	0.0860	−0.0056	0.0311	−0.0008	0.0116
$(q_{\widehat{\theta}_n} - q_\alpha)_{32}$	−0.0029	0.0591	0.0028	0.0309	0.0093	0.0168
$(q_{\widehat{\theta}_n} - q_\alpha)_{13}$	−0.0274	0.0788	−0.0105	0.0311	−0.0080	0.0218
$(q_{\widehat{\theta}_n} - q_\alpha)_{23}$	−0.0224	0.0751	−0.0063	0.0273	0.0071	0.0166
$(q_{\widehat{\theta}_n} - q_\alpha)_{33}$	0.0213	0.0818	−0.0038	0.0344	−0.0046	0.0198
$k(\widehat{\theta}_n)$	0.0260	0.0151	0.0038	0.0017	0.0015	0.0007

TABLE 3

Bias of the likelihood estimates of the parameters $\widehat{\theta}_n$ of the transition matrix $q_{\widehat{\theta}_n}$ and of the Kullback distance $k(\widehat{\theta}_n)$ from one single run of sampled chains of length 100, 500 and 2000, respectively. Note that $(\lambda_1, \lambda_2, \lambda_3, a_1, a_2, a_3) = (3, 1, 2, 0.3, 0.6, 0.8)$.

Bias	$n = T = 100$	$n = T = 500$	$n = T = 2000$
$\widehat{\lambda}_1 - \lambda_1$	−1.0061	−0.3369	−0.3580
$\widehat{\lambda}_2 - \lambda_2$	0.0322	−0.0246	0.0065
$\widehat{\lambda}_3 - \lambda_3$	−0.0002	0.4623	−0.3845
$\widehat{a}_1 - a_1$	0.0208	−0.2993	0.0571
$\widehat{a}_2 - a_2$	0.3986	−0.3278	−0.0628
$\widehat{a}_3 - a_3$	−0.0747	−0.2402	0.0204
$(q_{\widehat{\theta}_n} - q_\alpha)_{11}$	0.0572	0.0377	0.0072
$(q_{\widehat{\theta}_n} - q_\alpha)_{21}$	−0.0227	0.0439	0.0083
$(q_{\widehat{\theta}_n} - q_\alpha)_{31}$	0.0283	0.0041	0.0023
$(q_{\widehat{\theta}_n} - q_\alpha)_{12}$	−0.0226	−0.0468	−0.0075
$(q_{\widehat{\theta}_n} - q_\alpha)_{22}$	−0.0146	−0.0061	−0.0089
$(q_{\widehat{\theta}_n} - q_\alpha)_{32}$	0.0029	−0.0541	−0.0309
$(q_{\widehat{\theta}_n} - q_\alpha)_{13}$	−0.0346	0.0090	0.0002
$(q_{\widehat{\theta}_n} - q_\alpha)_{23}$	0.0374	−0.0378	0.0005
$(q_{\widehat{\theta}_n} - q_\alpha)_{33}$	−0.0312	−0.0582	0.0285
$k(\widehat{\theta}_n)$	0.0042	0.0074	0.0011

that $l_n(\theta)$ depends only on those transition probabilities $q(\theta;\, x,\, y)$ for which the statistics $N_n(x,\, y)$ are non-null. Therefore, given a sampled chain $(Z_1, \ldots, Z_n)$ and letting K be the maximum value of the observed values (Z_j), all matrices involved in (2) are truncated at the order $4K \times 4K$ to get a correct approximation of the transition probabilities $\{q(\theta;\, x,\, y)\colon 0 \le x,\, y \le K\}$.

As in the previous example, for a given set of parameters $(\beta,\, \delta)$, 100 independent paths of the M/M/1 queue were simulated on $[0,\, T]$ with $T = 100$. Sampled chains Z were then extracted at times $1, 2, \ldots, n = T$.

In these experiments, the intention was to analyze the effect of the queue intensities on the variation of $\widehat{\theta}_n$. Therefore, three situations were considered, with true parameter values $(\beta,\, \delta) = (0.1,\, 0.2)$, $5 \times (0.1,\, 0.2)$ and $25 \times (0.1,\, 0.2)$, respectively, corresponding to increasing queue intensities. Also the likelihood estimators $\widehat{\theta}_{n,c} = (\widehat{\beta}_{n,c},\, \widehat{\delta}_{n,c})$ were computed based on the whole continuous paths for time varying in $[0,\, T]$. These estimators serve as a benchmark for comparison.

Table 4 displays the mean and standard deviation statistics for the bias $(\widehat{\beta}_n,\, \widehat{\delta}_n) - (\beta,\, \delta)$ and $(\widehat{\beta}_{n,c},\, \widehat{\delta}_{n,c}) - (\beta,\, \delta)$. While the estimator $\widehat{\theta}_{n,c}$ from continuous observations barely varies for the three degrees of the queue intensity, the estimator $\widehat{\theta}_n$ from discrete observations seems to have a high bias and standard deviation in the highest intensity case $(\beta,\, \delta) = (2.5,\, 5)$. This is, in fact, a scaling effect as confirmed by the box plots of the relative biases $(\widehat{\beta}_n - \beta)/\beta$ and $(\widehat{\delta}_n - \delta)/\delta$ displayed in Figure 1. It can be seen that these relative biases remain quite stable

TABLE 4

Averaged bias and standard deviation of the discrete-time likelihood estimates $(\widehat{\beta}_n, \widehat{\delta}_n)$, of the continuous-time likelihood estimates $(\widehat{\beta}_{n,c}, \widehat{\delta}_{n,c})$ from 100 independent runs on [0, 100] and for three queue intensities $(\beta, \delta) = (0.1, 0.2)$, (0.5, 1) and (2.5,5), respectively.

	$\{(\beta, \delta) = (0.1, 0.2)$		$\{(\beta, \delta) = (0.5, 1)$		$\{(\beta, \delta) = (2.5, 5)$	
Bias	average	std. dev.	average	std. dev.	average	std. dev.
$\widehat{\beta}_n - \beta$	−0.0088	0.0301	0.0011	0.1045	0.1159	0.8961
$\widehat{\delta}_n - \delta$	0.0614	0.1132	0.0793	0.2546	0.3533	1.8595
$\widehat{\beta}_{n,c} - \beta$	0.0155	0.0486	0.0118	0.0947	−0.0301	0.2028
$\widehat{\delta}_{n,c} - \delta$	0.0163	0.1110	0.0072	0.2014	0.0249	0.4192

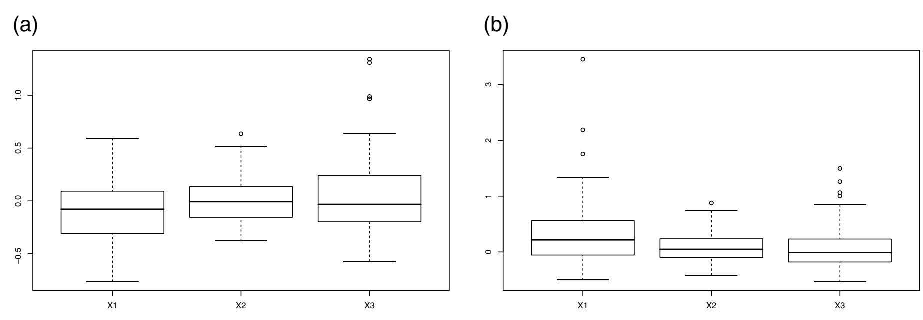

Figure 1. Box plots of the relative biases $(\widehat{\beta}_n - \beta)/\beta$ and $(\widehat{\delta}_n - \delta)/\delta$, in figure (a) and (b) respectively, from 100 independent simulations. Column `X1`: $(\beta, \delta) = (0.1, 0.2)$; Column `X2`: $(\beta, \delta) = (0.5, 1)$; Column `X3`: $(\beta, \delta) = (2.5, 5)$.

across the different levels of the queue intensity. Finally, the estimation method in this case takes about 0.8 second (CPU time) on a DELL Latitude 600 notebook (running Linux).

5. Proofs

Proof of Proposition 1. Denote the first sum in (4) by M_n and the second one by A_n. Then $(M_n)_{\{n\in\mathbb{N}\}}$ is a martingale and $(A_n)_{\{n\in\mathbb{N}\}}$ is an increasing sequence of positive random variables. Under the assumptions, $K(\theta;\, x,\, y) \geq 0$ and

$$\begin{aligned}\sum_{x\in E} K(\alpha,\theta;x)\,\mu_\alpha(x) &\leq \sum_{x\in E}\sum_{y\in E} |g(\theta;x,y)|\,q(\alpha;x,y)\mu_\alpha(x)\\ &\leq G(\theta)\sum_{x\in E}\sum_{y\in E}\big(1+|x|^{a/2}+|y|^{a/2}\big)\,q(\alpha;x,y)\mu_\alpha(x)\\ &\leq G(\theta)\left(1+\sum_{x\in E}|x|^{a/2}\mu_\alpha(x)+\sum_{y\in E}|y|^{a/2}\mu_\alpha(y)\right)\\ &< \infty.\end{aligned}$$

Then, thanks to the SLLN (R2), $n^{-1}\, A_n$ converges almost surely to $k(\theta)$ as $n \to \infty$.

Furthermore, as $g^2(\theta;\, x,\, y) \leq C\ (1+|x|^a+|y|^a)$, the sequence $(M_n)_{\{n\in\mathbb{N}\}}$ is a square-integrable martingale with the increasing process

$$\langle M\rangle_n = \sum_{i=1}^{n} \mathrm{E}_\alpha\big((g(\theta;X_i,X_{i-1}) - K(\alpha,\theta;X_{i-1}))^2 \mid X_{i-1}\big).$$

Again the SLLN (R2) applies, showing that $n^{-1}\langle M\rangle_n$ converges almost surely to a non-negative constant $c(\alpha,\theta,g)$ as $n\to\infty$. Using the standard convergence theorem for martingales (e.g. Duflo, 1997, Theorem 1.3.15), this implies that $n^{-1}M_n$ converges almost surely to 0.

Finally note that, under (S3)(ii), the function $k(\cdot)$ is continuous in θ. Indeed,

$$\begin{aligned}|k(\theta)-k(\theta')| &\leq \sum_{x\in E}|K(\alpha,\theta;x)-K(\alpha,\theta';x)|\,\mu_\alpha(x)\\ &\leq G(|\theta-\theta'|)\left(1+\sum_{x\in E}|x|^{a/2}\mu_\alpha(x)+\sum_{y\in E}|y|^{a/2}\mu_\alpha(y)\right)\\ &\leq c\,G(|\theta-\theta'|),\end{aligned}$$

for some $c > 0$. Then the proposition is proved.

Proof of Theorem 1. According to the general criterion for consistency given by van de Vaart (1998, Theorem 5.7), the following conditions need to be checked.

$$\text{(C1)}\qquad \sup_{\theta\in\Theta}\big|n^{-1}U_n(\theta)-k(\theta)\big|\xrightarrow{a.s.}0\ ,$$

$$\text{(C2)}\qquad \inf_{\theta:|\theta-\alpha|\geq\varepsilon} k(\theta) > 0 \quad \text{for every } \varepsilon > 0\ .$$

Thanks to Proposition 1 and the assumption (D), the function $k(\cdot)$ is continuous and positive on the compact set $\{\theta \in \Theta\colon |\theta - \alpha| \geq \varepsilon\}$. Then (C2) holds.

For (C1), let $W_n(\cdot)$ denote the uniform continuity modulus of $n^{-1}U_n(\cdot)$, i.e.

$$W_n(\eta) = \sup_{|\theta-\theta'|\leq\eta} n^{-1}|U_n(\theta) - U_n(\theta')|, \qquad \eta > 0.$$

Therefore, it is enough to find a deterministic sequence $(u_j)_{\{j>0\}}$ decreasing to 0, such that, for every $j > 0$,

$$\limsup_{n\to\infty} W_n(1/j) \leq u_j \qquad \text{a.s.} \tag{6}$$

Indeed, since the parameter space Θ is compact, for any fixed j it can be recovered with a finite number of balls of radius $1/j$ centered at points, say ξ_ℓs. Then for any $\theta \in \Theta$,

$$|n^{-1}U_n(\theta) - k(\theta)| \;\leq\; W_n(1/j) + \max_{\ell} |n^{-1}U_n(\xi_\ell) - k(\xi_\ell)| + [k](1/j),$$

where $[k](\cdot)$ denotes the uniform continuity modulus of the function $k(\cdot)$ which is continuous on the compact set Θ. Letting $n \to \infty$, it holds

$$\begin{aligned}\limsup_{n\to\infty} \sup_{\theta\in\Theta} \left|n^{-1}U_n(\theta) - k(\theta)\right| &\leq \limsup_{n\to\infty} W_n(1/j) + [k](1/j)\\ &\leq u_j + [k](1/j)\ , \quad \text{a.s.}\end{aligned}$$

Condition (C1) then follows by letting $j \to \infty$.

To end the proof, a sequence $(u_j)_{\{j>0\}}$ is constructed which satisfies (6). By (S3)(ii),

$$\begin{aligned}|U_n(\theta) - U_n(\theta')| &\leq \sum_{i=1}^{n} |\log q(\theta; X_{i-1}, X_i) - \log q(\theta'; X_{i-1}, X_i)|\\ &\leq G(|\theta - \theta'|) \sum_{i=1}^{n} \big(1 + |X_{i-1}|^{a/2} + |X_i|^{a/2}\big).\end{aligned}$$

Therefore,

$$\begin{aligned}\limsup_{n} W_n(1/j) &\leq G(1/j) \limsup_{n} \frac{1}{n} \sum_{i=1}^{n} \big(1 + |X_{i-1}|^{a/2} + |X_i|^{a/2}\big)\\ &= c\, G(1/j) \qquad \text{a.s.}\ ,\end{aligned}$$

where c is a positive constant. Thus $u_j = c\, G(1/j)$ can be taken and the proof is complete.

Proof of Theorem 2. Only the main lines are sketched here as the proof is quite standard; see, for example, Yao (2000, Theorem 2) where one can find more details.

First, the following consequences of the assumptions are outlined. Since Θ is compact, one has for all $i, j, \theta \in V$ and $x, y \in E$

$$\begin{aligned}
\left|D_{ij}^2 \log q(\theta; x, y)\right| &\le C\left(1 + |x|^{a/2} + |y|^{a/2}\right),\\
|D_i \log q(\theta; x, y) - D_i \log q(\alpha; x, y)| &\le C|\theta - \alpha|\left(1 + |x|^{a/2} + |y|^{a/2}\right),\\
|D_i \log q(\theta; x, y)| &\le C\left(1 + |x|^{a/2} + |y|^{a/2}\right),\\
|I_{ij}(\alpha; x)| &\le C(1 + |x|^a).
\end{aligned}$$

Second, the gradient vector of U_n at α satisfies

$$D_i U_n(\alpha) = -\sum_{k=1}^{n} D_i \log q(\alpha; X_{k-1}, X_k).$$

Thus, $(DU_n(\alpha))_{\{n\in\mathbb{N}\}}$ is a square-integrable s-dimensional martingale according to (N2) and the computations above. Its increasing process is

$$\langle DU_n(\alpha)\rangle_n = \sum_{1\le k\le n} I(\alpha; X_{k-1}),$$

and $n^{-1}\langle DU_n(\alpha)\rangle_n$ converges almost surely to $I(\alpha)$ as $n \to \infty$. It can be checked that the Lindberg condition is also satisfied. Hence, by the Lindberg–Feller theorem (see Duflo, 1997, Corollary 2.1.10), $n^{-1/2}DU_n(\alpha)$ converges in distribution to the zero-mean s-dimensional Gaussian distribution with covariance matrix $I(\alpha)$ for every $x \in E$. Furthermore the second-order derivatives of $U_n(\cdot)$ at α are

$$D_{ij}^2 U_n(\alpha) = -\sum_{k=1}^{n} D_{ij}^2 \log q(\alpha; X_{k-1}, X_k).$$

One can prove that

(i) almost surely, $n^{-1}D_{ij}^2 U_n(\alpha)$ converges to $I_{ij}(\alpha)$ thanks to a martingale decomposition analogous to the one used for $U_n(\theta)$ in the proof of Proposition 1 and by taking (N2) into account;

(ii) almost surely,

$$\limsup_{n\to\infty} \sup_{\theta:|\theta-\alpha|\le\varepsilon} \frac{1}{n}\left|D_{ij}^2 U_n(\theta) - D_{ij}^2 U_n(\alpha)\right| \le \psi(\varepsilon),$$

where ψ is some deterministic function satisfying $\lim_{\varepsilon\to 0} \psi(\varepsilon) = 0$.

Then the central limit result follows from a Taylor expansion of $DU_n(\widehat{\theta}_n)$ for $\widehat{\theta}_n$ near α.

References

AÏT-SAHALIA, Y. (2002). Maximum likelihood estimation for discretely sampled diffusions: a closed-form approximation approach. *Econometrica* **70**, 223–262.

ASMUSSEN, S. (2003). *Applied Probability and Queues*. New York: Springer-Verlag.

BLADT, M. & SØRENSEN, M. (2005). Statistical inference for discretely observed Markov jump processes. *J. R. Statist. Soc. B* **67**, 395–410.

BIBBY, B.M., JACOBSEN, M. & SØRENSEN, M. (2007). Estimating functions for discretely sampled diffusion-type models. In Aït-Sahalia, Y. & Hansen, L.P., editors, *Handbook of Financial Econometrics*. Amsterdam: North Holland.

BILLINGSLEY, P. (1961a). *Statistical Inference for Markov Processes*. Chicago: The University of Chicago Press.

BILLINGSLEY, P. (1961b). Statistical methods in Markov chains. *Ann. Math. Stat.* **32**, 12–40.

DACUNHA-CASTELLE, D. & DUFLO, M. (1986). Probability and Statistics. Volume II. New York: Springer-Verlag.

DUFLO, M. (1997). Random Iterative Models. Berlin: Springer-Verlag.

ETHIER, S.N. & KURTZ, T.G. (1986). *Markov Processes, Characterization and Convergence*. NewYork: John Wiley & Sons.

JACOBSEN, M. (1982). *Statistical Analysis of Counting Processes. Lecture Notes in Statistics 12*. New York: Springer-Verlag.

JACOD, J. (2007). Inference for Stochastic Processes. In Aït-Sahalia, Y. & Hansen, L.P., editors, *Handbook of Financial Econometrics*. Amsterdam: North Holland. Preprint PMA-683. Available at http://www.proba.jussieu.fr/mathdoc/preprints/index.html#2001.

JENSEN, A. (1953). Markoff chains as an aid in the study of Markoff processes. *Skand. Aktuarietidskr* **36**, 87–91.

KEIDING, N. (1974). Estimation in the birth process. *Biometrika* **61**, 71–80.

KEIDING, N. (1975). Maximum likelihood estimation in the birth-and-death process. *Ann. Statistics* **3**, 363–372.

MEYN, S. & TWEEDIE, R. (1993). *Markov Chains and Stochastic Stability*. Berlin: Springer-Verlag.

MISHRA, M.N. & PRAKASA RAO, B.L.S. (2001). Asymptotic minimax estimation in nonlinear stochastic differential equations from discrete observations. *Statistics &Decision* **19**, 121–136.

TAKÁCS, L. (1962). *Introduction to the Theory of Queues*. Oxford: Oxford University Press.

VAN DE VAART, A.W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

YAO, J.-F. (2000). On least squares estimation for stable nonlinear autoregressive processes, *Ann. Inst. Math. Statist.* **52**, 316–331.