Bernoulli **12**(3), 2006, 383–401

Parametric inference for discretely observed non-ergodic diffusions

JEAN JACOD

UFR de Mathématiques, Université Pierre et Marie Curie, 4 place Jussieu, 75252 Paris, France. E–mail: jj@ccr.jussieu.fr

We consider a multidimensional diffusion process X whose drift and diffusion coefficients depend respectively on a parameter λ and θ. This process is observed at $n+1$ equally spaced times $0, \Delta_n, 2\Delta_n, \ldots, n\Delta_n$, and $T_n = n\Delta_n$ denotes the length of the 'observation window'. We are interested in estimating λ and/or θ. Under suitable smoothness and identifiability conditions, we exhibit estimators $\hat{\lambda}_n$ and $\hat{\theta}_n$, such that the variables $\sqrt{n}(\hat{\theta}_n - \theta)$ and $\sqrt{T_n}(\hat{\lambda}_n - \lambda)$ are tight for $\Delta_n \to 0$ and $T_n \to \infty$. When λ is known, we can even drop the assumption that $T_n \to \infty$. These results hold without any kind of ergodicity or even recurrence assumption on the diffusion process.

Keywords: non-ergodic diffusion processes; parametric inference for diffusions

1. Introduction

In this paper we consider a d-dimensional diffusion process X whose drift (diffusion) coefficient depends on a (possibly multidimensional) unknown parameter λ (θ). That is, it solves the equation

$$dX_t = a(\lambda, X_t)dt + \sigma(\theta, X_t)dW_t, \qquad X_0 = x_0. \tag{1.1}$$

This process is observed at n regularly spaced times $0, \Delta_n, 2\Delta_n, \ldots, n\Delta_n$, and $T_n = n\Delta_n$ denotes the length of the 'observation window'. There are some smoothness and boundedness assumptions on the coefficients a and σ, but *neither ergodicity nor even recurrence* is assumed. Our aims are as follows:

1. If λ is known, to construct estimators for θ converging at rate $\sqrt{n}$ as $n \to \infty$ (meaning that the sequence $\sqrt{n}(\hat{\theta}_n - \theta)$ is tight, or bounded in probability), for $\Delta_n \to 0$, and regardless of the behaviour of T_n.
2. If λ is unknown, to construct estimators for θ converging at $\sqrt{n}$, and simultaneously estimators for λ converging at rate $\sqrt{T_n}$, for $\Delta_n \to 0$ and $T_n \to \infty$.

When λ is known and T_n does not depend on n (that is, we observe the diffusion at times iT/n for $i = 0, \ldots, n$ on a fixed interval $[0, T]$), this is a rather old result: see, for example, Dohnal (1987) or Genon-Catalot and Jacod (1993). When $T_n \to \infty$ and when the diffusion is ergodic under the true value of the parameters, this is also a known result – see, for example, Yoshida (1992), Kessler (1997), Kessler and Sørensen (1999), Aït-Sahalia (2002), Prakasa Rao (1999a, 1999b) and Kutoyants (2004) – and indeed in this case one

1350–7265

does not need Δ_n to go to 0. On the other hand, in the non-ergodic situations and when $T_n \to \infty$, there are so far very few results and most are in very specific cases: see Basawa and Scott (1983), and Prakasa Rao (1999a, 1999b) for a review of known results.

This paper is thus the first to provide estimators which work, at the above-prescribed rate or better, and without assuming ergodicity or recurrence. The estimators are explicit, although they are based upon moments of the diffusion which are usually not 'explicitly' known as functions of the parameters; in principle, it is always possible to use some sort of Monte Carlo (see Pedersen 1995) or other technique to approximate these moments.

Let us end this introductory section with some comments about the nature and the limitations of the forthcoming results.

1. When the diffusion is ergodic, the variables $\sqrt{n}(\hat{\theta}_n - \theta)$ and $\sqrt{T_n}(\hat{\lambda}_n - \lambda)$ converge to centred non-vanishing Gaussian vectors, and in particular the rates are 'efficient'. When the diffusion is non-ergodic, we do not know whether these variables converge in law, and if they do the limit could be 0, which means that the rate are actually 'larger' than $\sqrt{n}$ or $\sqrt{T_n}$.
2. For λ the rate $\sqrt{T_n}$ is not generally efficient: take for example an Ornstein–Uhlenbeck process, for which there are estimators for the drift converging at rate T_n in the null recurrent case and exponential in T_n in the transient case.
3. However, it is quite likely (but remains to be proved) that the rate $\sqrt{n}$ for θ is efficient, and even that $\sqrt{n}(\hat{\theta}_n - \theta)$ converges in law to some non-degenerate random vector, in the non-ergodic case as well. This is what happens, for example, for the diffusion coefficient of an Ornstein–Uhlenbeck process, regardless of the ergodicity or non-ergodicity.
4. Of course, we also need some identifiability assumptions for the parameters. These identifiability assumptions seem to be weak enough when the coefficients are bounded. Otherwise those assumptions are quite strong, and for example rule out the Ornstein–Uhlenbeck model. So for unbounded coefficients the results are far from satisfactory: some more precise comments about this question are made after the statement of the results.

The paper is organized as follows. In Section 2 we construct the estimators and state our results. Section 3 is devoted to a number of technical preliminaries, and the proofs of the results are in Section 4.

2. The results

Let us first give a precise statement of the setting and assumptions. In (1.1), W is an m-dimensional standard Wiener process, and the coefficients a and σ have the relevant dimensions (d for a, and dm for σ). The initial value $x_0 \in \mathbb{R}^d$ is known. As for the parameters, we have three cases:

1. The parameter λ is known, while the parameter θ is unknown and belongs to some compact convex domain Θ of $\mathbb{R}^q$; we then write $a(x)$ instead of $a(\lambda, x)$.

2. The parameter θ is known, while the parameter λ is unknown and belongs to some compact convex domain Λ of $\mathbb{R}^r$; we then write $\sigma(x)$ instead of $\sigma(\theta, x)$.
3. Both parameters λ and θ are unknown and belong to compact convex domains Λ and Θ of $\mathbb{R}^r$ and $\mathbb{R}^q$, respectively.

We systematically use vector or matrix notation. If f is an $\mathbb{R}^n$-valued function on $\Lambda \times \Theta \times \mathbb{R}^d$, we denote by $\nabla^i_\lambda \nabla^j_\theta \nabla^k_x f$ the $n \times ir \times jq \times kd$-dimensional array of partial derivatives of order i, j and k of the components of f with respect to the components of λ, θ and x, respectively; the partial derivative of f with respect to the kth component of λ is denoted by $\partial_{\lambda_k} f$, and similarly for the other variables. In case 1 (2) derivatives with respect to λ (θ) are irrelevant, and subsequently we arbitrarily set them to 0. We also denote by $\|y\|$ the Euclidian norm of y in whichever space it lies. The transpose of a vector or matrix y is $y^\star$, and we define the *diffusion coefficient* to be $c = \sigma\sigma^\star$.

In the following, if we are in case 1 (2) the derivatives with respect to λ (θ) are irrelevant. We state our smoothness assumption in case 3 only, the adaptation to cases 1 and 2 being straightforward.

Assumption (H) *Smoothness. The function a (σ) is three times differentiable in λ (θ). The functions $\nabla^j_\lambda a$ and $\nabla^j_\theta \sigma$, for $j = 0, 1, 2, 3$, are three times differentiable in x. Further:*

(a) *the functions $\nabla^j_\lambda \nabla^k_x a$ and $\nabla^j_\theta \nabla^k_x \sigma$, for $j = 0, 1, 2, 3$ and $k = 1, 2, 3$, are bounded by a constant;*
(b) *we have*

$$j = 0, 1, 2, 3 \Rightarrow |\nabla^j_\lambda a(\lambda, x)| \leqslant A(x), \; |\nabla^j_\theta \sigma(\theta, x)| \leqslant A(x) \tag{2.1}$$

for some C^∞ function $A : \mathbb{R}^d \to [1, \infty)$, whose derivatives of any order $n \geqslant 1$ are bounded, and such that $A(x) \leqslant C(1 + \|x\|)$.

Therefore the coefficients and their derivatives with respect to the parameters are of linear growth, uniformly in λ and θ (remember that Λ and Θ are compact). Note that (2.1) is always satisfied with $A(x) = \sqrt{1 + \|x\|^2}$, and a particularly important case is when one may take $A(x) = 1$ for all x (we refer to this as the 'bounded' case).

Assumption (H) yields the result that (1.1) admits a unique strong solution, which we denote by X for the 'true values' of the parameters which themselves are denoted by λ and θ. Other values of the parameters are denoted by u and v, respectively. We wish to estimate λ and/or θ, upon observing the values $X_{i\Delta_n}$ for $i = 1, \ldots, n$ at stage n, for some time-lag Δ_n which goes to 0 as $n \to \infty$. For this, we additionally need some identifiability assumptions. For each parameter we have two such assumptions: a 'global' one which ensures that asymptotically efficient estimators do exist (which amounts to slightly more than just saying that for distinct values of the parameters the processes have different laws); and a 'local' one which says that the model is not 'flat' at the true value of the parameter and thus accounts for the rate of convergence of sequences of estimators. Both these conditions are stated in terms of λ and θ and of the 'true' process X.

Assumption (I1-λ) *Global identifiability for λ. We have, for all ε > 0,*

$$\lim_{\eta\to 0}\limsup_n \mathbb{P}\left(\inf_{u:\|u-\lambda\|>\varepsilon}\frac{1}{T_n}\int_0^{T_n}\frac{\|a(u, X_s)-a(\lambda, X_s)\|^2}{A(X_s)^4}\,\mathrm{d}s \leqslant \eta\right)=0. \tag{2.2}$$

Assumption (I1-θ) *Global identifiability for θ. We have, for all ε > 0,*

$$\lim_{\eta\to 0}\limsup_n \mathbb{P}\left(\inf_{v:\|v-\theta\|>\varepsilon}\frac{1}{T_n}\int_0^{T_n}\frac{\|c(v, X_s)-c(\theta, X_s)\|^2}{A(X_s)^6}\,\mathrm{d}s \leqslant \eta\right)=0. \tag{2.3}$$

Let a^i and c^{ij} denote the components of a and c, and consider $\nabla_\lambda a^i$ and $\nabla_\theta c^{ij}$ as column vectors, so that, for example, $\nabla_\theta c(\theta, x)^\star y = (\nabla_\theta c^{ij}(\theta, x)^\star y)_{1\leqslant i,j\leqslant d}$ is a $d\times d$ matrix if $y\in\mathbb{R}^q$.

Assumption (I2-λ) *Local identifiability for λ.* We have

$$\lim_{\varepsilon\to 0}\limsup_n \mathbb{P}\left(\inf_{y\in\mathbb{R}^r:\|y\|=1}\frac{1}{T_n}\int_0^{T_n}\frac{\|\nabla_\lambda a^i(\lambda, X_s)^\star y\|^2}{A(X_s)^4}\,\mathrm{d}s \leqslant \varepsilon\right)=0. \tag{2.4}$$

Assumption (I2-θ) *Local identifiability for θ. We have*

$$\lim_{\varepsilon\to 0}\limsup_n \mathbb{P}\left(\inf_{y\in\mathbb{R}^q:\|y\|=1}\frac{1}{T_n}\int_0^{T_n}\frac{\|\nabla_\theta c^{ij}(\theta, X_s)^\star y\|^2}{A(X_s)^6}\,\mathrm{d}s \leqslant \varepsilon\right)=0. \tag{2.5}$$

We can now construct our estimators. Note first that, at stage n, we observe $X_{i\Delta_n}$ for $i=0, 1, \ldots, n$ or, equivalently, the increments or the 'normalized' increments:

$$\chi_i^n = X_{i\Delta_n}-X_{(i-1)\Delta_n}, \qquad \chi_i'^n=\frac{\chi_i^n}{\sqrt{\Delta_n}}. \tag{2.6}$$

Next, we denote by $X^{u,v,x}$ the solution to (1.1) when the starting point is x and the parameter values are u and v, and we set

$$\phi_n(u, v, x)=\mathbb{E}(X^{u,v,x}_{\Delta_n}-x), \qquad \phi_n'(u, v, x)=\mathbb{E}\Big((X^{u,v,x}_{\Delta_n}-x)(X^{u,v,x}_{\Delta_n}-x)^\star\Big), \tag{2.7}$$

which are (in principle) known functions of (u, v, x). Note that ϕ_n is an $\mathbb{R}^d$-valued function, while ϕ_n' takes its values in the set of $d\times d$ symmetric non-negative matrices.

The estimators will be minimizers of suitable contrast functions, and for the sake of clarity we single out the three cases.

In case 1 (when λ is known), we set

$$U_n(v)=\sum_{i=1}^n\frac{1}{A(X_{(i-1)\Delta_n})^6}\,\|\chi_i^n\chi_i^{n\star}-\phi_n'(\lambda, v, X_{(i-1)\Delta_n})\|^2. \tag{2.8}$$

As we shall see later, $v\mapsto U_n(v)$ is continuous, so it has a minimum on the compact set Θ, and due to the measurable selection theorem we can find a measurable (with respect to the observed σ-field at stage n) variable $\hat\theta_n$ satisfying

$$U_n(\hat{\theta}_n) = \inf_{v \in \Theta} U_n(v). \tag{2.9}$$

Theorem 2.1. *Assume that we are in case 1 and that Assumptions* (H), (I1-θ) *and* (I2-θ) *hold. Then if θ is in the interior of Θ, and if $\Delta_n \to 0$, the estimators $\hat{\theta}_n$ are $\sqrt{n}$-consistent, in the sense that the sequence $\sqrt{n}(\hat{\theta}_n - \theta)$ is tight.*

In case 2 (when θ is known), we set

$$U_n(u) = \sum_{i=1}^{n} \frac{1}{A(X_{(i-1)\Delta_n})^4} \phi_n(u, \theta, X_{(i-1)\Delta_n})^{\star} \big(\phi_n(u, \theta, X_{(i-1)\Delta_n}) - 2\chi_i^n\big) \tag{2.10}$$

and, as above, $\hat{\lambda}_n$ denotes a measurable variable such that

$$U_n(\lambda_n) = \inf_{u \in \Lambda} U_n(u). \tag{2.11}$$

Theorem 2.2. *Assume that we are in case 2 and that* (H), (I1-λ) *and* (I2-λ) *hold. Then if λ is in the interior of Λ, and if $\Delta_n \to 0$ and $T_n \to \infty$, the estimators $\hat{\lambda}_n$ are $\sqrt{T_n}$-consistent, in the sense that the sequence $\sqrt{T_n}(\hat{\lambda}_n - \lambda)$ is tight.*

Finally, in case 3 we set

$$\begin{aligned} U_n(u, v) = &\sum_{i=1}^{n} \frac{1}{A(X_{(i-1)\Delta_n})^6} \phi_n(u, v, X_{(i-1)\Delta_n})^{\star} \big(\phi_n(u, v, X_{(i-1)\Delta_n}) - 2\chi_i^n\big) \\ &+ \sum_{i=1}^{n} \frac{1}{A(X_{(i-1)\Delta_n})^4} \|\chi_i^n \chi_i^{n\star} - \phi_n'(u, v, X_{(i-1)\Delta_n})\|^2 \end{aligned} \tag{2.12}$$

and, as above, we denote by $(\hat{\lambda}_n, \hat{\theta}_n)$ a measurable variable such that

$$U_n(\hat{\lambda}_n, \hat{\theta}_n) = \inf_{(u,v) \in \Lambda \times \Theta} U_n(u, v). \tag{2.13}$$

Theorem 2.3. *Assume that we are in case 3 and that* (H), (I1-λ), (I2-λ), (I1-θ) *and* (I2-θ) *hold. Then if α and θ are in the interiors of Λ and Θ, and if $\Delta_n \to 0$ and $T_n \to \infty$, the estimators $\hat{\lambda}_n$ and $\hat{\theta}_n$ are respectively $\sqrt{T_n}$-consistent and $\sqrt{n}$-consistent.*

Some comments on our assumptions are in order. First, (H) is a standard and probably reasonable hypothesis. One might perhaps ask for a smaller order of differentiability, but we do need *some* differentiability.

Things are different for the identifiability assumptions. To understand the strength of these assumptions, let us consider an example concerned with linear dependence in the parameters. Let us suppose that λ and θ are one-dimensional (i.e. $r = q = 1$), and that the coefficients a and σ have the form

$$a(\lambda, x) = \lambda \alpha(x), \qquad \sigma(\theta, x) = \sqrt{\theta}\sigma'(x). \tag{2.14}$$

Then Λ in cases 2 and 3, and Θ in cases 1 and 3, are bounded closed intervals of $\mathbb{R}$ and $\mathbb{R}_+$

respectively, and $c(\theta, x) = \theta\gamma(x)$ where $\gamma = \sigma'\sigma'^{\star}$. Assumption (H) reduces to the fact the functions α and σ' are three times differentiable, with all partial derivatives of order 1, 2, 3 being bounded. The 'bounded case' corresponds to the additional assumption that α and σ' (hence γ as well) are themselves bounded.

The identifiability assumptions take a simple form here. Let us introduce the random variables

$$Z_n = \frac{1}{T_n}\int_0^{T_n} \frac{\|\alpha(X_s)\|^2}{A(X_s)^4}\,\mathrm{d}s, \qquad Z'_n = \frac{1}{T_n}\int_0^{T_n} \frac{\|\gamma(X_s)\|^2}{A(X_s)^6}\,\mathrm{d}s. \tag{2.15}$$

Then the two assumptions (I1-λ) and (I2-λ) on the one hand, and (I1-θ) and (I2-θ) on the other hand, respectively reduce to the following two hypotheses:

Assumption (I3-$\boldsymbol{\lambda}$). *The sequence* $(1/Z_n)$ *is tight.*

Assumption (I3-$\boldsymbol{\theta}$). *The sequence* $(1/Z'_n)$ *is tight.*

In the 'bounded case' these are quite weak, because we can take $A(x) = 1$ identically. They are satisfied for example when, respectively, $\inf_x\|\alpha(x)\| > 0$ and $inf_x\|\gamma(x)\| > 0$, and also in many other situations.

In the unbounded case, things are quite different. Of course when $T_n = T$ does not depend on n, (I3-λ) is irrelevant, and (I3-θ) is automatically satisfied (unless the diffusion coefficient is identically 0). But when $T_n \to \infty$, the process X typically spends more and more time far away from the origin, and $A(x)^p$ behaves like $\|x\|^p$ for large $\|x\|$, so Z_n and/or Z'_n have a tendency to decrease as n increases. In the genuine linear growth case, $\|\alpha(x)\|$ behaves more or less like $|x|$ or $A(x)$, while $\|\gamma(x)\|$ behaves like $|x|^2$ or $A(x)^2$, so in fact in (2.15) we have the wrong powers in the denominator: we would have more reasonable conditions if $A(X_s)$ appeared with the power 2 instead of 4 for Z_n and 4 instead of 6 for Z'_n. Note, however, that even these 'more reasonable' conditions are not fulfilled by the Ornstein–Uhlenbeck process, which is a model linear in the two parameters if we write it as $\mathrm{d}X_t = \lambda X_t \mathrm{d}t + \sqrt{\theta}\,\mathrm{d}W_t$.

Of course, even in the unbounded case, the ergodicity of X implies the identifiability assumptions (the ergodic theorem implies that both Z_n and Z'_n converge to positive limits, execpt in some trivial degenerate cases). But then the results are already well established in the literature.

3. Preliminaries

We will present a unified proof for the three cases. For this, we set $\kappa = \kappa' = 1$ in case 3, $\kappa = 1$ and $\kappa' = 0$ in case 2, and $\kappa = 0$ and $\kappa' = 1$ in case 1. In all cases the contrast U_n can be written as

$$U_n(u, v) = \sum_{i=1}^{n} \frac{\kappa}{A(X_{(i-1)\Delta_n})^4} \phi_n(u, v, X_{(i-1)\Delta_n})^\star \left(\phi_n(u, v, X_{(i-1)\Delta_n}) - 2\chi_i^n\right)$$

$$+ \sum_{i=1}^{n} \frac{\kappa'}{A(X_{(i-1)\Delta_n})^6} \|\chi_i^n \chi_i^{n\star} - \phi'_n(u, v, X_{(i-1)\Delta_n})\|^2. \tag{3.1}$$

In case 1 (2) we take $u = \lambda$ ($v = \theta$). Recall that A satisfies (2.1).

3.1. On the diffusion process

First, we need some (classical) results on (1.1): we refer the reader to Revuz and Yor (1991). We start with the standard Wiener space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ endowed with the (m-dimensional) canonical process W. To simplify the notation, we write $w = (u, v)$ for a pair of parameters, and $\beta = (\lambda, \theta)$ for the true value (derivatives with respect to u and v are still denoted ∇_λ and ∇_θ). Recall that for any $w = (u, v) \in \Lambda \times \Theta$ and $x \in \mathbb{R}^d$ we denote by $X^{w,x}$ the solution to the equation starting at x and with parameters u and v.

Let us introduce the following auxiliary functions (recall that Λ and Θ are compact):

$$\alpha(x) = 1 + \sup_{u \in \Lambda} \sum_{k=0}^{3} \|\nabla_\lambda^k a(u, x)\|, \qquad \gamma(x) = 1 + \sup_{v \in \Theta} \sum_{k=0}^{3} \|\nabla_\theta^k \sigma(v, x)\|. \tag{3.2}$$

Observe that

$$X_t^{w,x} - x - \sigma(v, x)W_t - a(u, x)t$$

$$= \int_0^t (a(u, X_s^{w,x}) - a(u, x))\mathrm{d}s + \int_0^t (\sigma(v, X_s^{w,x}) - \sigma(v, x))\,\mathrm{d}W_s.$$

Then by (H) and Gronwall's lemma, we obtain for $p \geqslant 1$ and $t \in (0, 1]$ and for a constant C_p which changes from line to line and may depend on p, but not on t, x, u or v:

$$\mathbb{E}\left(\sup_{s \leqslant t} \|X_s^{w,x} - x\|^p\right) \leqslant C_p A(x)^p t^{p/2}. \tag{3.3}$$

$$\mathbb{E}\left(\|\mathbb{X}_t^{w,x} - x - \sigma(v, x)W_t\|^p\right) \leqslant C_p A(x)^p t^p, \tag{3.4}$$

$$\|\mathbb{E}\left(X_t^{w,x} - x - a(u, x)t\right)\| \leqslant CA(x)t^{3/2}. \tag{3.5}$$

Next, by classical differentiation properties for stochastic differential equations (see, for example, Protter 1990), (H) implies that $w \mapsto X^{w,x}$ is differentiable in L^2, with derivatives $\nabla_\lambda X^{w,x}$ and $\nabla_\theta X^{w,x}$ being the unique solutions of the following linear equations (in straightforward matrix notation):

$$\nabla_\lambda X_t^{w,x} = \int_0^t (\nabla_\lambda a(u, X_s^{w,x}) + \nabla_x a(u, X_s^{w,x})\nabla_\lambda X_s^{w,x})\mathrm{d}s + \int_0^t \nabla_x \sigma(v, X_s^{w,x})\nabla_\lambda X_s^{w,x}\mathrm{d}W_s,$$

$$\nabla_\theta X_t^{w,x} = \int_0^t \nabla_x a(u, X_s^{w,x})\nabla_\theta X_s^{w,x}\mathrm{d}s + \int_0^t (\nabla_x \sigma(v, X_s^{w,x})\nabla_\theta X_s^{w,x} + \nabla_\theta \sigma(v, X_s^{w,x}))\mathrm{d}W_s.$$

Then Gronwall's lemma, (H) and (3.3) together yield, for $t \in (0, 1]$:

$$\begin{aligned} &\mathbb{E}(\|\nabla_\lambda X_t^{w,x}\|^p) \leqslant C_p A(x)^p t^p, \qquad \mathbb{E}(\|\nabla_\theta X_t^{w,x}\|^p) \leqslant C_p A(x)^p t^{p/2}, \\ &\mathbb{E}\big(\|\nabla_\lambda X_t^{w,x} - \nabla_\lambda a(u, x)t\|^p\big) \leqslant C_p A(x)^p t^{3p/2}, \\ &\mathbb{E}\big(\|\nabla_\theta X_t^{w,x} - \nabla_\theta \sigma(v, x) W_t\|^p\big) \leqslant C_p A(x)^p t^p, \qquad \|\mathbb{E}(\nabla_\theta X_t^{w,x})\| \leqslant CA(x) t^{3/2}. \end{aligned} \tag{3.6}$$

In a similar way one can differentiate a second time, to obtain

$$\begin{aligned} \nabla_\lambda^2 X_t^{w,x} = &\int_0^t (\nabla_\lambda^2 a(u, X_s^{w,x}) + 2\nabla_\lambda \nabla_x a(u, X_s^{w,x})\nabla_\lambda X_s^{w,x})\mathrm{d}s \\ &+ \int_0^t (\nabla_x^2 a(u, X_s^{w,x})(\nabla_\lambda X_s^{w,x})^2 + \nabla_x a(u, X_s^{w,x})\nabla_\lambda^2 X_s^{w,x})\mathrm{d}s \\ &+ \int_0^t (\nabla_x^2 \sigma(v, X_s^{w,x})(\nabla_\lambda X_s^{w,x})^2 + \nabla_x \sigma(v, X_s^{w,x})\nabla_\lambda^2 X_s^{w,x})\,\mathrm{d}W_s, \\ \nabla_\lambda \nabla_\theta X_t^{w,x} = &\int_0^t \big(\nabla_x^2 a(u, X_s^{w,x})\nabla_\lambda X_s^{w,x}\nabla_\theta X_s^{w,x} + \nabla_x a(u, X_s^{w,x})\nabla_\lambda \nabla_\theta X_s^{w,x}\big)\mathrm{d}s \\ &+ \int_0^t \nabla_\lambda \nabla_x a(u, X_s^{w,x})\nabla_\theta X_s^{w,x}\mathrm{d}s + \int_0^t \nabla_x^2 \sigma(v, X_s^{w,x})\nabla_\lambda X_s^{w,x}\nabla_\theta X_s^{w,x}\mathrm{d}W_s \\ &+ \int_0^t \big(\nabla_x \nabla_\theta \sigma(v, X_s^{w,x})\nabla_\lambda X_s^{w,x} + \nabla_x \sigma(v, X_s^{w,x})\nabla_\lambda \nabla_\theta X_s^{w,x}\big)\,\mathrm{d}W_s, \\ \nabla_\theta^2 X_t^{w,x} = &\int_0^t \big(\nabla_x^2 a(u, X_s^{w,x})(\nabla_\theta X_s^{w,x})^2 + \nabla_x a(u, X_s^{w,x})\nabla_\theta^2 X_s^{w,x}\big)\mathrm{d}s \\ &+ \int_0^t \big(\nabla_x^2 \sigma(v, X_s^{w,x})(\nabla_\theta X_s^{w,x})^2 + 2\nabla_x \nabla_\theta \sigma(v, X_s^{w,x})\nabla_\theta X_s^{w,x}\big)\,\mathrm{d}W_s \\ &+ \int_0^t \big(\nabla_x \sigma(v, X_s^{w,x})\nabla_\theta^2 X_s^{w,x} + \nabla_\theta^2 \sigma(v, X_s^{w,x})\big)\,\mathrm{d}W_s. \end{aligned}$$

We then deduce that

$$\begin{aligned} &\mathbb{E}(\|\nabla_\lambda^2 X_t^{w,x}\|^p) \leqslant C_p A(x)^{2p} t^p, \qquad \mathbb{E}(\|\nabla_\lambda \nabla_\theta X_t^{w,x}\|^p) \leqslant C_p A(x)^{2p} t^{3p/2}, \\ &\mathbb{E}(\|\nabla_\theta^2 X_t^{w,x}\|^p) \leqslant C_p A(x)^{2p} t^{p/2}, \qquad \|\mathbb{E}(\nabla_\theta^2 X_t^{w,x})\| \leqslant CA(x)^2 t^{3/2}. \end{aligned} \tag{3.7}$$

We can even differentiate a third time: this gives similar formulae, from which one obtains

$$\begin{aligned}
&\mathbb{E}(\|\nabla_\lambda^3 X_t^{w,x}\|^p) \leqslant C_p A(x)^{3p} t^p, \\
&\mathbb{E}(\|\nabla_\lambda^2 \nabla_\theta X_t^{w,x}\|^p) + \mathbb{E}(\|\nabla_\lambda \nabla_\theta^2 X_t^{w,x}\|^p) \leqslant C_p A(x)^{3p} t^{3p/2}, \qquad (3.8)\\
&\mathbb{E}(\|\nabla_\theta^3 X_t^{w,x}\|^p) \leqslant C_p A(x)^{3p} t^{p/2}, \qquad \|\mathbb{E}(\nabla_\theta^3 X_t^{w,x})\| \leqslant CA(x)^3 t^{3/2}.
\end{aligned}$$

3.2. Estimates on moments

We now turn our attention to moment estimates, such as ϕ_n and ϕ'_n in (2.7). It is easier to consider normalized moments $\Phi_n = \phi_n/\sqrt{\Delta_n}$ and $\Phi'_n = \phi'_n/\Delta_n$, whose components are given by

$$\Phi_n^i(w, x) = \frac{1}{\sqrt{\Delta_n}} \mathbb{E}(X^{w,x,i} - x^i), \qquad \Phi_n^{ij}(w, x) = \frac{1}{\Delta_n} \mathbb{E}((X^{w,x,i} - x^i)(X^{w,x,j} - x^j))$$

(here and below, $X^{w,x,i}$ and x^i denote the ith component of $X^{w,x}$ and $x \in \mathbb{R}^d$). We also need a subfamily of fourth-order moments, namely the matrix $\Psi_n(w, x)$ whose entries are

$$\Psi_n^{ij}(w, x) = \frac{1}{\Delta_n^2} \mathbb{E}((X^{w,x,i} - x^i)^2 (X^{w,x,j} - x^j)^2).$$

First, (3.3) and (3.5) lead to

$$\|\Phi_n(w, x)\| \leqslant CA(x)\sqrt{\Delta_n}, \qquad \|\Phi'_n(w, x)\| \leqslant CA(x)^2, \qquad \|\Psi_n(w, x)\| \leqslant CA(x)^4, \qquad (3.9)$$

and (3.4) yields, after some calculation,

$$|\Psi_n^{ij}(w, x) - c^{ii}(v, x)c^{jj}(v, x) - 2(c^{ij}(v, x))^2| \leqslant CA(x)^4 \sqrt{\Delta_n}. \qquad (3.10)$$

Next, the results of the previous subsection show that $\Phi_n(w, x)$ and $\Phi'_n(w, x)$ are three times differentiable in w, with (∇_μ, $\nabla_{\mu'}$ and $\nabla_{\mu''}$ denoting either ∇_λ or ∇_θ):

$$\begin{aligned}
\nabla_\mu \Phi_n(w, x) &= \frac{1}{\sqrt{\Delta_n}} \mathbb{E}(\nabla_\mu X_{\Delta_n}^{w,x}), \\
\nabla_\mu \nabla_{\mu'} \Phi_n(w, x) &= \frac{1}{\sqrt{\Delta_n}} \mathbb{E}(\nabla_\mu \nabla_{\mu'} X_{\Delta_n}^{w,x}), \\
\nabla_\mu \nabla_{\mu'} \nabla_{\mu''} \Phi_n(w, x) &= \frac{1}{\sqrt{\Delta_n}} \mathbb{E}(\nabla_\mu \nabla_{\mu'} \nabla_{\mu''} X_{\Delta_n}^{w,x}), \\
\nabla_\mu \Phi'^{ij}_n(w, x) &= \frac{1}{\Delta_n} \mathbb{E}\Big((X_{\Delta_n}^{w,x,i} - x^i)\nabla_\mu X_{\Delta_n}^{w,x,j} + (X_{\Delta_n}^{w,x,j} - x^j)\nabla_\mu X_{\Delta_n}^{w,x,i}\Big), \\
\nabla_\mu \nabla_{\mu'} \Phi'^{ij}_n(w, x) &= \frac{1}{\Delta_n} \mathbb{E}((X_{\Delta_n}^{w,x,i} - x^i)\nabla_\mu \nabla_{\mu'} X_{\Delta_n}^{w,x,j} + (X_{\Delta_n}^{w,x,j} - x^j)\nabla_\mu \nabla_{\mu'} X_{\Delta_n}^{w,x,i} \\
&\quad + \nabla_\mu X_{\Delta_n}^{w,x,i} \nabla_{\mu'} X_{\Delta_n}^{w,x,j} + \nabla_{\mu'} X_{\Delta_n}^{w,x,i} \nabla_\mu X_{\Delta_n}^{w,x,j}),
\end{aligned}$$

$$
\begin{aligned}
\nabla_\mu\nabla_{\mu'}\nabla_{\mu''}\Phi'^{ij}_n(w, x) = \frac{1}{\Delta_n}\,\mathbb{E}((X^{w,x,i}_{\Delta_n} - x^i)\nabla_\mu\nabla_{\mu'}\nabla_{\mu''}X^{w,x,j}_{\Delta_n} \\
&+ (X^{w,x,j}_{\Delta_n} - x^j)\nabla_\mu\nabla_{\mu'}\nabla_{\mu''}X^{w,x,i}_{\Delta_n} + \nabla_{\mu''}X^{w,x,i}_{\Delta_n}\nabla_\mu\nabla_{\mu'}X^{w,x,j}_{\Delta_n} + \nabla_{\mu'}X^{w,x,i}_{\Delta_n}\nabla_\mu\nabla_{\mu''}X^{w,x,j}_{\Delta_n} \\
&+ \nabla_\mu X^{w,x,i}_{\Delta_n}\nabla_{\mu'}\nabla_{\mu''}X^{w,x,j}_{\Delta_n} + \nabla_\mu\nabla_{\mu'}X^{w,x,i}_{\Delta_n}\nabla_{\mu''}X^{w,x,j}_{\Delta_n} \\
&+ \nabla_\mu\nabla_{\mu''}X^{w,x,i}_{\Delta_n}\nabla_{\mu'}X^{w,x,j}_{\Delta_n} + \nabla_{\mu'}\nabla_{\mu''}X^{w,x,i}_{\Delta_n}\nabla_\mu X^{w,x,j}_{\Delta_n}).
\end{aligned}
$$

Then (3.3), (3.6), (3.7) and (3.8) yield (suppressing in some cases the arguments w and x):

$$
\left.
\begin{array}{ll}
\|\nabla_\lambda\Phi_n\| \leqslant CA\sqrt{\Delta_n}, & \|\nabla_\theta\Phi_n\| \leqslant CA\Delta_n, \\
\|\nabla^2_\lambda\Phi_n\| \leqslant CA^2\sqrt{\Delta_n}, & \|\nabla_\lambda\nabla_\theta\Phi_n\| + \|\nabla^2_\theta\Phi_n\| \leqslant CA^2\Delta_n, \\
\|\nabla^3_\lambda\Phi_n\| \leqslant CA^3\sqrt{\Delta_n}, & \|\nabla^2_\lambda\nabla_\theta\Phi_n\| + \|\nabla_\lambda\nabla^2_\theta\Phi_n\| + \|\nabla^3_\theta\Phi_1\| \leqslant CA^3\Delta_n, \\
\multicolumn{2}{l}{\|\Phi_n(w, x) - a(u, x)\sqrt{\Delta_n}\| + \|\nabla^k_\lambda\Phi_n(w, x) - \nabla_\lambda a(u, x)\sqrt{\Delta_n}\| \leqslant CA(x)\Delta_n,}
\end{array}
\right\} \tag{3.11}
$$

$$
\left.
\begin{array}{ll}
\|\nabla_\theta\Phi'_n\| \leqslant CA^2, & \|\nabla_\lambda\Phi'_n\| \leqslant CA^2\sqrt{\Delta_n}, \\
\|\nabla^2_\theta\Phi'_n\| \leqslant CA^3, & \|\nabla^2_\lambda\Phi'_n\| + \|\nabla_\lambda\nabla_\theta\Phi'_n\| \leqslant CA^3\sqrt{\Delta_n}, \\
\|\nabla^3_\theta\Phi'_n\| \leqslant CA^4, & \|\nabla^3_\lambda\Phi'_n\| + \|\nabla^2_\lambda\nabla_\theta\Phi'_n\| + \|\nabla_\lambda\nabla^2_\theta\Phi'_n\| \leqslant CA^4\sqrt{\Delta_n}, \\
\multicolumn{2}{l}{\|\Phi_n(w, x) - c(v, x)\| + \|\nabla_\theta\Phi_n(w, x) - \nabla_\theta c(v, x)\| \leqslant CA(x)^2\sqrt{\Delta_n}.}
\end{array}
\right\} \tag{3.12}
$$

3.3. Contrast

Multiplying $U_n(u, v) = U_n(w)$ by a positive number and adding to it another number, both possibly depending on n and ω and on the true value $\beta = (\lambda, \theta)$ of the parameter (but not on w), does not change the estimators $\hat{\lambda}_n$ and $\hat{\theta}_n$. So instead of (3.1) we can use the following definition:

$$
U_n(w) = \sum_{i=1}^n \zeta^n_i(w),
$$

where (with the notation $X^n_i = X_{(i-1)\Delta_n}$; recall that χ'^n_i is given by (2.6))

$$
\begin{aligned}
\zeta^n_i(w) = {}& \frac{\kappa}{n\Delta_n A(X^n_i)^4}\sum_{j=1}^d \left(\Phi^j_n(w, X^n_i)(\Phi^j_n(w, X^n_i) - 2\chi'^{n,j}_i) + \Phi^j_n(\beta, X^n_i)^2\right) \\
&+ \frac{\kappa'}{nA(X^n_i)^6}\sum_{j,k=1}^d \left((\chi'^{n,j}_i\chi'^{n,k}_i)^2 + \Phi'^{jk}_n(w, X^n_i)^2 - 2\Phi'^{jk}_n(w, X^n_i)\chi'^{n,j}_i\chi'^{n,k}_i - 2\Phi^{jk}_n(\beta, X^n_i)^2\right).
\end{aligned}
$$

Observe that ζ^n_i is three times differentiable, with

$$\nabla_\mu \zeta_i^n(w) = \frac{2\kappa}{n\Delta_n A(X_i^n)^4} \sum_{j=1}^d \Big(\nabla_\mu \Phi_n^j(w, X_i^n)(\Phi_n^j(w, X_i^n) - \chi_i^{\prime n,j})\Big)$$

$$+ \frac{2\kappa'}{nA(X_i^n)^6} \sum_{j,k=1}^d \Big(\nabla_\mu \Phi_n^{\prime jk}(w, X_i^n)(\Phi_n^{\prime jk}(w, X_i^n) - \chi_i^{\prime n,j}\chi_i^{n,k})\Big),$$

$$\nabla_\mu \nabla_{\mu'} \zeta_i^n(w) = \frac{2\kappa}{n\Delta_n A(X_i^n)^4} \sum_{j=1}^d (\nabla_\mu \nabla_{\mu'} \Phi_n^j(w, X_i^n)(\Phi_n^j(w, X_i^n) - \chi_i^{\prime n,j})$$

$$+ \nabla_\mu \Phi_n^j(w, X_i^n) \nabla_{\mu'} \Phi_n^j(w, X_i^n))$$

$$+ \frac{2\kappa'}{nA(X_i^n)^4} \sum_{j,k=1}^d (\nabla_\mu \nabla_{\mu'} \Phi_n^{\prime jk}(w, X_i^n)(\Phi_n^{\prime jk}(w, X_i^n) - \chi_i^{\prime n,j}\chi_i^{n,k})$$

$$+ \nabla_\mu \Phi_n^{\prime jk}(w, X_i^n) \nabla_{\mu'} \Phi_n^{\prime jk}(w, X_i^n)),$$

$$\nabla_\mu \nabla_{\mu'} \nabla_{\mu''} \zeta_i^n(w) = \frac{2\kappa}{n\Delta_n A(X_i^n)^4} \sum_{j=1}^d (\nabla_\mu \nabla_{\mu'} \nabla_{\mu''} \Phi_n^j(w, X_i^n)(\Phi_n^j(w, X_i^n) - \chi_i^{\prime n,j})$$

$$+ \nabla_\mu \Phi_n^j(w, X_i^n) \nabla_{\mu'} \nabla_{\mu''} \Phi_n^j(w, X_i^n) + \nabla_{\mu'} \Phi_n^j(w, X_i^n) \nabla_\mu \nabla_{\mu''} \Phi_n^j(w, X_i^n)$$

$$+ \nabla_{\mu''} \Phi_n^j(w, X_i^n) \nabla_\mu \nabla_{\mu'} \Phi_n^j(w, X_i^n))$$

$$+ \frac{2\kappa'}{nA(X_i^n)^6} \sum_{j,k=1}^d (\nabla_\mu \nabla_{\mu'} \nabla_{\mu''} \Phi_n^{\prime jk}(w, X_i^n)(\Phi_n^{\prime jk}(w, X_i^n) - \chi_i^{\prime n,j}\chi_i^{\prime n,k})$$

$$+ \nabla_\mu \Phi_n^{\prime jk}(w, X_i^n) \nabla_{\mu'} \nabla_{\mu''} \Phi_n^{\prime jk}(w, X_i^n) + \nabla_{\mu'} \Phi_n^{\prime jk}(w, X_i^n) \nabla_\mu \nabla_{\mu''} \Phi_n^{\prime jk}(w, X_i^n)$$

$$+ \nabla_{\mu''} \Phi_n^{\prime jk}(w, X_i^n) \nabla_\mu \nabla_{\mu'} \Phi_n^{\prime jk}(w, X_i^n)).$$

Then, combining (3.3), (3.9), (3.11) and (3.12) with the previous equalities, and recalling that $A \geqslant 1$, we obtain for any $j \in \mathbb{N}^\star$, and if $\mathcal{F}_i^n = \mathcal{F}_{i\Delta_n}$:

$$k = 0, 1, 2, 3 \Rightarrow \mathbb{E}(\|\nabla_\lambda^k \zeta_i^n(w)\|^j | \mathcal{F}_{i-1}^n) \leqslant \frac{C_j}{n^j}\left(\frac{\kappa}{\Delta_n^{j/2}} + 1\right), \tag{3.13}$$

$$k = 1, 2, 3,\ l = 0, \ldots, k-3 \Rightarrow \mathbb{E}(\|\nabla_\lambda^l \nabla_\theta^k \zeta_i^n(w)\|^j | \mathcal{F}_{i-1}^n) \leqslant \frac{C_j}{n^j}. \tag{3.14}$$

$$\|\mathbb{E}(\nabla_\lambda^2 \zeta_i^n(w) | \mathcal{F}_{i-1}^n)\| \leqslant \frac{C}{n}, \qquad \|\mathbb{E}(\nabla_\lambda \nabla_\theta \zeta_i^n(w) | \mathcal{F}_{i-1}^n)\| \leqslant \frac{C\sqrt{\Delta_n}}{n}. \tag{3.15}$$

Finally, we set

$$F(w, x) = \frac{\kappa}{A(x)^4} \, \|a(u, x) - a(\lambda, x)\|^2 + \frac{\kappa'}{A(x)^6} \, \|c(v, x) - c(\theta, x)\|^2. \tag{3.16}$$

This function is three times differentiable, and we have

$$\left.\begin{aligned} \partial^2_{\lambda_j\lambda_k} F(\beta, x) &= \frac{2\kappa}{B(x)} \sum_{i=1}^{r} \partial_{\lambda_j} a^i(\lambda, x) \partial_{\lambda_k} a^i(\lambda, x), \\ \partial^2_{\theta_j\theta_k} F(\beta, x) &= \frac{2\kappa'}{C(x)} \sum_{i,l=1}^{q} \partial_{\theta_j} c^{il}(\theta, x) \partial_{\theta_k} c^{il}(\theta, x). \end{aligned}\right\} \tag{3.17}$$

Therefore (3.10, (3.11) and (3.12) yield

$$\left| \mathbb{E}\big(\zeta_i^n(w) | \mathcal{F}_{i-1}^n\big) - \frac{1}{n} \, F(w, X_i^n) \right| \leqslant C \frac{\sqrt{\Delta_n}}{n}, \tag{3.18}$$

$$\left\| \mathbb{E}\big(\nabla_\lambda^2 \zeta_i^n(\beta) | \mathcal{F}_{i-1}^n\big) - \frac{2\kappa}{n} \nabla_\lambda^2 F(\beta, X_i^n) \right\| \leqslant C \, \frac{\sqrt{\Delta_n}}{n} \Big(\kappa + \sqrt{\Delta_n}\Big), \tag{3.19}$$

$$\left\| \mathbb{E}\big(\nabla_\theta^2 \zeta_i^n(\beta) | \mathcal{F}_{i-1}^n\big) - \frac{2\kappa'}{n} \, \nabla_\theta^2 F(\beta, X_i^n) \right\| \leqslant C \, \frac{\sqrt{\Delta_n}}{n}, \tag{3.20}$$

$$\| \mathbb{E}\big(\nabla_\lambda \nabla_\theta \zeta_i^n(\beta) | \mathcal{F}_{i-1}^n\big) \| \leqslant C \, \frac{\sqrt{\Delta_n}}{n}. \tag{3.21}$$

3.4. Riemann integrals

Itô's formula yields, for any real-valued C^2 function h,

$$h(X_t^{w,x}) - h(x) = \int_0^t \left(\nabla_x h(X_s^{w,x}) a(u, X_s^{w,x}) + \frac{1}{2} \, \nabla_x^2 h(X_s^{w,x}) c(v, X_s^{w,x}) \right) \mathrm{d}s$$

$$+ \int_0^t \nabla_x h(X_s^{w,x}) \sigma(v, X_s^{w,x}) \mathrm{d}W_s.$$

Then, if the two functions $A\nabla_x h$ and $A^2 \nabla_x^2 h$ are bounded, we obtain

$$\mathbb{E}(|h(X_t^{w,x}) - h(x)|) \leqslant C\sqrt{t}.$$

In this case, we have (recall $X_i^n = X_{(i-1)\Delta_n}$)

$$\mathbb{E}\left(\left| h(X_i^n) - \frac{1}{\Delta_n} \int_{(i-1)\Delta_n}^{i\Delta_n} h(X_s) \mathrm{d}s \right| \right) \leqslant C\sqrt{\Delta_n}.$$

Then we deduce that

$$\mathbb{E}\left(\left| \frac{1}{n} \sum_{i=1}^{n} h(X_i^n) - \frac{1}{T_n} \int_0^{T_n} h(X_s) \mathrm{d}s \right| \right) \leqslant C\sqrt{\Delta_n}. \tag{3.22}$$

3.5. An application of the Burkholder–Gundy inequality

Let us suppose that η_i^n are real-valued random variables, each η_i^n being $\mathcal{F}_i^n$-measurable. Assume also that for some constants γ and γ' and some integer $m \geqslant 1$, we have

$$|\mathbb{E}(\eta_i^n|\mathcal{F}_{i-1}^n)| \leqslant \frac{\gamma}{n}, \qquad \mathbb{E}((\eta_i^n)^{2m}) \leqslant \frac{\gamma'^{2m}}{n^{2m}}\left(1+\frac{\delta}{\Delta_n^m}\right), \tag{3.23}$$

where δ is either 0 or 1. Then we have the following lemma.

Lemma 3.1. *For any integer $m \geqslant 1$ there is a universal constant K_m such that, for any family (η_i^n) satisfying* (3.23)*,*

$$\mathbb{E}\left(\left(\sum_{i=1}^n \eta_i^n\right)^{2m}\right) \leqslant K_m\left(\gamma^{2m}+\gamma'^{2m}\frac{1}{n^m}\left(1+\frac{\delta}{\Delta_n^m}\right)\right). \tag{3.24}$$

Proof. Set $V_n = \sum_{i=1}^n \eta_i^n$, $\delta_i^n = \eta_i^n - \mathbb{E}(\eta_i^n|\mathcal{F}_{i-1}^n)$ and $M_j^n = \sum_{i=1}^j \delta_i^n$ for $j \in \mathbb{N}$. The first part of (3.23) gives

$$|V_n - M_n^n| \leqslant \gamma. \tag{3.25}$$

By construction the sequence $(M_j^n)_{j\in\mathbb{N}}$ is a martingale with respect to the filtration $(\mathcal{F}_j^n)_{j\in\mathbb{N}}$, so the Burkholder–Gundy inequality yields a universal constant $K'_m \geqslant 1$ such that

$$\mathbb{E}((M_n^n)^{2m}) \leqslant K'_m\mathbb{E}\left(\left(\sum_{i=1}^n(\delta_i^n)^2\right)^m\right).$$

Then the Hölder inequality tells us that $\mathbb{E}((\delta_i^n)^{2m}) \leqslant 2^{2m}\mathbb{E}((\eta_i^n)^{2m})$, and that

$$\begin{aligned}\mathbb{E}((M_n^n)^{2m}) &\leqslant K'_m n^m \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n(\delta_i^n)^{2m}\right)\\ &\leqslant K'_m 2^{2m} n^m \mathbb{E}((\eta_i^n)^{2m}) \leqslant \frac{K'_m 2^{2m}\gamma'^{2m}}{n^m}\left(1+\frac{\delta}{\Delta_n^m}\right).\end{aligned} \tag{3.26}$$

Then (3.24) readily follows from (3.25) and (3.26), upon setting $K_m = K'_m 2^{4m-1}$. □

4. Proof of the theorems

One can consider Theorem 2.2 as a particular case of Theorem 2.3, upon setting $\hat{\theta}_n = \theta$ (since θ is known). It is not quite the same for Theorem 2.1, because the hypothesis $T_n \to \infty$ is not assumed here; however, in this case we have $\kappa = 0$ and the derivatives in λ appear nowhere, so it is straightforward to check that the proof of Theorem 2.3 entails the proof of Theorem 2.1, even if $T_n \to \infty$ fails. We therefore prove only Theorem 2.3.

However, we repeatedly mention κ and κ', so the reader can easily verify that $T_n \to \infty$ is not necessary when $\kappa = 0$.

Suppose that the assumptions of Theorem 2.3 are in force. Set

$$V_n(w) = \frac{1}{T_n} \int_0^{T_n} F(w, X_s) \mathrm{d}s. \tag{4.1}$$

This is three times differentiable in w, and we have, for $\mu = \lambda$ or $\mu = \theta$,

$$\nabla_\mu^2 V_n(\beta) = \frac{1}{T_n} \int_0^{T_n} \nabla_\mu^2 F(\beta, X_s) \mathrm{d}s. \tag{4.2}$$

Observe also that, since A has bounded derivatives of all orders greater than or equal to 1 and also $A \geqslant 1$, we can deduce from (H), (2.1) and (3.16) that

$$k = 0, 1, 2, \quad \mu = \lambda, \theta \Rightarrow \|A(x)^k \nabla_x^k F(w, x)\| + \|A(x)^k \nabla_x^k \nabla_\mu^2 F(w, x)\| \leqslant C. \tag{4.3}$$

Therefore (3.22) yields, for $\mu = \lambda$ and $\mu = \theta$,

$$\left.\begin{aligned} &\mathbb{E}\left(\left|\frac{1}{n}\sum_{i=1}^n F(w, X_i^n) - V_n(w)\right|\right) \leqslant C\sqrt{\Delta_n}, \\ &\mathbb{E}\left(\left|\frac{1}{n}\sum_{i=1}^n \nabla_\mu^2 F(\beta, X_i^n) - \nabla_\mu^2 V_n(\beta)\right|\right) \leqslant C\sqrt{\Delta_n}. \end{aligned}\right\} \tag{4.4}$$

4.1. Convergence of contrasts

In view of (3.13) we have

$$\sum_{i=1}^n \mathbb{E}\big(\zeta_i^n(w)^2 \mathcal{F}_{i-1}^n\big) \leqslant C\left(\frac{\kappa}{\mathrm{T}_n} + \frac{1}{n}\right).$$

Next, (3.18) and (4.4) yield

$$\sum_{i=1}^n \mathbb{E}\big(\zeta_i^n(w) \mathcal{F}_{i-1}^n\big) - V_n(w) \xrightarrow{\mathbb{P}} 0.$$

Therefore, since $T_n \to \infty$ when $\kappa = 1$, in all cases we arrive at

$$V'_n(w) := U_n(w) - V_n(w) \xrightarrow{\mathbb{P}} 0. \tag{4.5}$$

On the other hand, if $w = (u, v) = ((u_i)_{i \leqslant r}, (v_i)_{i \leqslant q})$ and $w' = (u', v') = ((u'_i)_{i \leqslant r}, (v'_i)_{i \leqslant q})$ are two pairs of parameters, we have

$$
\begin{aligned}
\eta_i^n &:= \zeta_i^n(w') - \zeta_i^n(w) \qquad (4.6)\\
&= \sum_{k=1}^{r} \int_{u_k}^{u'_k} \partial_{\lambda_k} \zeta_i^n((u'_1, \ldots, u'_{k-1}, x, u_{k+1}, \ldots, u_r), v)\mathrm{d}x \\
&\quad + \sum_{k=1}^{q} \int_{v_k}^{v'_k} \partial_{\theta_k} \zeta_i^n(u', (v'_1, \ldots, v'_{k-1}, x, v_{k+1}, \ldots, v_q))\mathrm{d}x,
\end{aligned}
$$

if the hypercube in $\mathbb{R}^{r+q}$ having w and w' for summits is entirely contained in $\Lambda \times \Theta$. Since $\Lambda \times \Theta$ is compact and convex, we deduce that for any integer m and any pair (w, w') we have

$$
\mathbb{E}((\eta_i^n)^{2m}) \leqslant C_m \|w' - w\|^{2m} \sup_{w'' \in \Lambda \times \Theta} \mathbb{E}\big(\|\nabla_\lambda \zeta_i^n(w'')\|^{2m} + \|\nabla_\theta \zeta_i^n(w'')\|^{2m}\big) \qquad (4.7)
$$

for some constant C_m. Using (3.13) and (3.14), we see that the variables η_i^n satisfy the second part of (3.23) with $\gamma' = C\|w - w'\|$ and $\delta = \kappa$. On the other hand, (3.18) and (4.3) imply that the first part of (3.23) is also satisfied with $\gamma = \gamma'$. Hence Lemma 3.1 yields

$$
\mathbb{E}(|U_n(w) - U_n(w')|^{2m}) \leqslant C\|w' - w\|^{2m}.
$$

By (4.3) again, we also have $|V_n(w) - V_n(w')|^{2m} \leqslant C\|w' - w\|^{2m}$, so finally

$$
\mathbb{E}(|V'_n(w) - V'_n(w')|^{2m}) \leqslant C\|w' - w\|^{2m}. \qquad (4.8)
$$

Now w lies in a compact convex subspace of $\mathbb{R}^{r+q}$, so if we take m such that $2m \geqslant r + q + 1$ it is then well known (see, for example, Theorem 20 in Ibragimov and Has'sminskii 1981: 378) that (4.8) combined with (4.5) implies that the sequence $(V'_n(w) : w \in \Lambda \times \Theta)$ of processes is tight for the uniform convergence, and further satisfies

$$
M'_n := \sup_{w \in \Lambda \times \Theta} |V'_n(w)| \xrightarrow{\mathbb{P}} 0. \qquad (4.9)
$$

4.2. Convergence of derivatives of contrasts

By (3.13) and (3.14), we have

$$
\sum_{i=1}^{n} \big(\mathbb{E}(\|\nabla_\lambda^2 \zeta_i^n(w)\|^2 \mid \mathcal{F}_{i-1}^n) + \mathbb{E}(\|\nabla_\theta^2 \zeta_i^n(w)\|^2 \mid \mathcal{F}_{i-1}^n)\big) \leqslant C\left(\frac{1}{n} + \frac{\kappa}{T_n}\right),
$$

which goes to 0 in all cases. Combining this with (3.19), (3.20) and (4.4) yields

$$
\nabla_\lambda^2 U_n(\beta) - \nabla_\lambda^2 V_n(\beta) \xrightarrow{\mathbb{P}} 0, \qquad \nabla_\theta^2 U_n(\beta) - \nabla_\theta^2 V_n(\beta) \xrightarrow{\mathbb{P}} 0. \qquad (4.10)
$$

Next, combine (3.14) with (3.21) to obtain

$$
\mathbb{E}(\|\nabla_\lambda \nabla_\theta U_n(\beta)\|^2) \leqslant C\left(\frac{1}{n} + \Delta_n\right).
$$

Therefore,

$$\kappa = 1 \Rightarrow \text{the sequence } (\nabla_\lambda \nabla_\theta U_n(\beta)/\sqrt{\Delta_n}) \text{ is tight.} \tag{4.11}$$

Now, as above, we consider two pairs of parameters $w = (u, v)$ and $w' = (u', v')$. If ∇_μ and $\nabla_{\mu'}$ denote either ∇_λ or ∇_θ, we replace η_i^n in (4.6) by any component of $\eta_i^n := \nabla_\mu \nabla_{\mu'} \zeta_i^n(w') - \nabla_\mu \nabla_{\mu'} \zeta_i^n(w)$. The same argument as for (4.7) leads to

$$\mathbb{E}((\eta_i^n)^{2m}) \leq C\|w' - w\|^{2m} \sup_{w'' \in \Lambda \times \Theta} \mathbb{E}\big(\|\nabla_\lambda \nabla_\mu \nabla_{\mu'} \zeta_i^n(w'')\|^{2m} + \|\nabla_\theta \nabla_\mu \nabla_{\mu'} \zeta_i^n(w'')\|^{2m}\big).$$

Then if we use (3.13) and (3.14), we see that the variables η_i^n satisfy the second part of (3.23) with $\gamma' = C\|w - w'\|$ and with $\delta = \kappa$ when $\mu = \mu' = \lambda$, and $\delta = 0$ otherwise. On the other hand, (3.14) and (3.15) imply that the first part of (3.23) is also satisfied with $\gamma = \gamma' \sqrt{\Delta_n}$ if $\mu = \lambda$ and $\mu' = \theta$, and $\gamma' = \gamma$ otherwise. Hence Lemma 3.1 and a summation over all components of $\nabla_\mu \nabla_\mu U_n$ yield

$$\mathbb{E}(\|\nabla_\theta^2 U_n(w) - \nabla_\theta^2 U_n(w')\|^{2m}) \leq C\|w' - w\|^{2m},$$

$$\kappa = 1 \Rightarrow \mathbb{E}(\|\nabla_\lambda^2 U_n(w) - \nabla_\lambda^2 U_n(w')\|^{2m}) \leq C\|w' - w\|^{2m}\left(1 + \frac{1}{T_n^m}\right),$$

$$\kappa = 1 \Rightarrow \mathbb{E}(\|\nabla_\lambda \nabla_\theta U_n(w) - \nabla_\lambda \nabla_\theta U_n(w')\|^{2m}) \leq C\|w' - w\|^{2m}\left(1 + \frac{1}{T_n^m}\right)\Delta_n^m.$$

These properties, combined with (4.10) and (4.11), imply that the sequences of processes $G_n(w) = \nabla_\lambda^2 U_n(w)$, or $G_n(w) = \nabla_\theta^2 U_n(w)$, or $G_n(w) = \nabla_\lambda \nabla_\theta U_n(w)/\sqrt{\Delta_n}$, indexed by the $(r+q)$-dimensional parameter w in a compact convex set, are tight for the uniform convergence in the set of continuous functions over $\Lambda \times \Theta$. Hence, again using (4.10) and (4.11), we readily deduce that for any (random) sequence w_n which converges in probability to β, we have

$$\nabla_\theta^2 U_n(w_n) - \nabla_\theta^2 V_n(\beta) \xrightarrow{\mathbb{P}} 0, \tag{4.12}$$

$$\kappa = 1 \Rightarrow \nabla_\lambda^2 U_n(w_n) - \nabla_\lambda^2 V_n(\beta) \xrightarrow{\mathbb{P}} 0, \tag{4.13}$$

$$\kappa = 1 \Rightarrow \text{the sequence } (\nabla_\lambda \nabla_\theta U_n(w_n)/\sqrt{\Delta_n}) \text{ is tight in } \mathbb{R}^{r+q}. \tag{4.14}$$

Finally, we will also need a result on the first-order derivatives at the point $w = \beta$. The explicit expression $\nabla_\mu \zeta_i^n(w)$ and the definitions of Φ_n and Φ_n' give us

$$\mathbb{E}(\nabla_\lambda \zeta_i^n(\beta)|\mathcal{F}_{i-1}^n) = \mathbb{E}(\nabla_\theta \zeta_i^n(\beta)|\mathcal{F}_{i-1}^n) = 0.$$

Then, combining this with (3.13) and (3.14) yields

$$\mathbb{E}(\|\nabla_\lambda U_n(\beta)\|^2) \leq C\left(\frac{1}{n} + \frac{\kappa}{T_n}\right), \qquad \mathbb{E}(\|\nabla_\theta U_n(\beta)\|^2) \leq \frac{C}{n}.$$

Therefore we have:

$$\left.\begin{array}{l} \text{the sequence } \sqrt{n}\,\nabla_\theta U_n(\beta) \text{ is tight in } \mathbb{R}^q, \\ \kappa = 1 \Rightarrow \text{the sequence } \sqrt{T_n}\nabla_\lambda U_n(\beta) \text{ is tight in } \mathbb{R}^r. \end{array}\right\} \tag{4.15}$$

4.3. Consistency of the estimators

Now we prove the consistency of the estimators $\hat{\beta}_n = (\hat{\lambda}_n, \hat{\theta}_n)$. Observe that $V_n(\beta) = 0$. Set $C_n(\varepsilon, \eta) = \{inf_{w:\|w-\beta\|>\varepsilon} V_n(w) \geqslant \eta\}$. Comparing (3.16) and (4.1) with (2.2) and (2.3), we see that

$$\forall \varepsilon > 0, \qquad \lim_{\eta\to 0} \liminf_n \mathbb{P}(C_n(\varepsilon, \eta)) = 1. \tag{4.16}$$

On the set $C_n(\varepsilon, \eta) \cap \{M'_n < \eta/2\}$, we have $U_n(\beta) < \eta/2$, and $U_n(w) > \eta/2$ whenever $|w - \beta| > \varepsilon$. Since for ε small enough the ball $\{w : \|w - \beta\| \leqslant \varepsilon\}$ is contained in $\Lambda \times \Theta$, and since $U_n(\cdot)$ is continuous, the definition of $\hat{\beta}_n$ implies that necessarily $\|\hat{\beta}_n - \beta\| \leqslant \varepsilon$ on the set $C_n(\varepsilon, \eta) \cap \{M'_n < \eta/2\}$. Then combining (4.16) and (4.9) immediately yields that

$$\hat{\beta}_n \xrightarrow{\mathbb{P}} \beta. \tag{4.17}$$

4.4. Rate consistency of the estimators

We are now ready to prove Theorem 2.3. Recall that β is in the interior of $\Lambda \times \Theta$. So if $A'_n = \{\nabla_\lambda U_n(\hat{\beta}_n) = \nabla_\theta U_n(\hat{\beta}_n) = 0\}$, then (2.13) implies that $\mathbb{P}(A'_n) \to 1$. A Taylor expansion gives

$$\begin{aligned} -\partial_{\lambda_k} U_n(\beta) &= \sum_{l=1}^{r} \partial^2_{\lambda_k\lambda_l} U_n(w_n)(\hat{\lambda}^l_n - \lambda^l) + \sum_{l=1}^{q} \partial^2_{\lambda_k\theta_l} U_n(w_n)(\hat{\theta}^l_n - \theta^l), \\ -\partial_{\theta_k} U_n(\beta) &= \sum_{l=1}^{r} \partial^2_{\theta_k\lambda_l} U_n(w_n)(\hat{\lambda}^l_n - \lambda^l) + \sum_{l=1}^{q} \partial^2_{\theta_k\theta_l} U_n(w_n)(\hat{\theta}^l_n - \theta^l) \end{aligned} \tag{4.18}$$

on the set A'_n, where w_n is a (random) point between β and $\hat{\beta}_n$.

We will write this system of equation in another way, and for this we introduce some new notation:

$$\begin{array}{lll} G_n = \nabla^2_\lambda U_n(w_n), & G'_n = \nabla^2_\theta V_n(\beta) & (r \times r \text{ random matrices}), \\ H_n = \nabla_\lambda \nabla_\theta U_n(w_n) & & (r \times q \text{ random matrices}), \\ K_n = \nabla^2_\theta U_n(w_n), & K'_n = \nabla^2_\theta V_n(\beta) & (q \times q \text{ random matrices}), \\ R_n = \sqrt{T_n}(\hat{\lambda}_n - \lambda), & R'_n = \sqrt{T_n}\ \nabla_\lambda U_n(\beta) & (r \times 1 \text{ random matrices}), \\ S_n = \sqrt{n}(\hat{\theta}_n - \theta), & S'_n = \sqrt{n}\nabla_\theta U_n(\beta) & (q \times r \text{ random matrices}). \end{array}$$

Then we can write (4.18) as follows on the set A'_n:

$$R'_n = G_n R_n + \sqrt{\Delta_n}\, H_n S_n, \qquad S'_n = \frac{1}{\sqrt{\Delta_n}} H^\star_n R_n + K_n S_n.$$

This system of linear equations can be 'explicitly' solved as follows on the set

$$A''_n = \{\text{the matrices } G_n \text{ and } L_n := K_n - H^\star_n G_n^{-1} H_n \text{ are invertible}\}.$$

We have:

$$\left.\begin{aligned} R_n &= (G_n^{-1} + G_n^{-1} H_n L_n^{-1} H_n^\star G_n^{-1}) R'_n - \sqrt{\Delta_n} G_n^{-1} H_n L_n^{-1} S'_n \\ S_n &= L_n^{-1} S'_n - \frac{1}{\sqrt{\Delta_n}} L_n^{-1} H_n^\star G_n^{-1} R'_n \end{aligned}\right\} \text{ on } A'_n \cap A''_n. \tag{4.19}$$

At this stage, we deduce from (3.17) and (4.2) that $G'_n = \nabla^2_\lambda V_n(\beta)$ and $K'_n = \nabla^2_\theta V_n(\beta)$ are symmetric non-negative matrices, whose determinants are

$$\det(G'_n) = \inf_{y \in \mathbb{R}^r : \|y\|=1} \frac{2\kappa}{T_n} \int_0^{T_n} \frac{1}{A(X_s)^4} \sum_{i=1}^r (\nabla_\lambda a^i(\lambda, X_s) y)^2 \, ds,$$

$$\det(K'_n) = \inf_{y \in \mathbb{R}^q : \|y\|=1} \frac{2\kappa'}{T_n} \int_0^{T_n} \frac{1}{A(X_s)^6} \sum_{i,j=1}^r (\nabla_\theta c^{ij}(\lambda, X_s) y)^2 \, ds.$$

Consequently, the conditions (I2$'$-λ) and (I2$'$-θ) imply that the two sequences $(\det(G'_n))$ and $(\det(K'_n))$ of non-negative random variables, which by (2.1) are bounded, are also bounded away from 0 in probability in the sense that

$$\lim_{\varepsilon \to 0} \limsup_n \mathbb{P}(\det(G'_n) + \det(K'_n) \leqslant \varepsilon) = 0. \tag{4.20}$$

Moreover, if we apply (4.12), (4.13) and (4.14) on the one hand, and (4.15) on the other hand, we see that

$$\left.\begin{aligned} &\|G_n - G'_n\| + \|K_n - K'_n\| \xrightarrow{\mathbb{P}} 0, \\ &\text{the sequence } (H_n/\sqrt{\Delta_n}) \text{ is tight in } \mathbb{R}^{r+q}, \\ &\text{the sequences } (R'_n) \text{ and } (S'_n) \text{ are tight in } \mathbb{R}^r \text{ and } \mathbb{R}^q. \end{aligned}\right. \tag{4.21}$$

Recall that to establish Theorem 2.3 we need to show that the sequences (R_n) and (S_n) are tight in $\mathbb{R}^r$ and $\mathbb{R}^q$, respectively. By the 'subsequence principle', it is enough to prove that from any infinite subsequences of these sequences one can extract further infinite subsequences that are tight.

Therefore, take any infinite subsequence. Using (4.20) and (4.21), and also the fact that the sequences (G'_n) and (K'_n) are bounded, we can find an infinite sub-subsequence along which the random variables $(G_n, K_n, H_n/\sqrt{\Delta_n}, R'_n, S'_n)$ converge in law to some random variable (G, K, H, R', S'), where G is an $r \times r$ matrix, K is a $q \times q$ matrix, H is an $r \times q$ matrix, R' is in $\mathbb{R}^r$ and S' is in $\mathbb{R}^q$, and further G and K are bounded and $\det(G) > 0$ and $\det(K) > 0$. Up to taking a further subsequence, we can even assume (by the Skorokhod representation theorem) that this convergence holds almost surely (on an extended space).

In other words, and since $\mathbb{P}(A'_n) \to 1$, we are left to prove that if we have (non-random) elements with the relevant dimensions $(G_n, K_n, H_n, R'_n, S'_n)$ satisfying

$$\left.\begin{aligned} &G_n \to G, \qquad K_n \to K, \qquad \frac{1}{\sqrt{\Delta_n}} H_n \to H, \qquad R'_n \to R', \qquad S'_n \to S', \\ &\det(G) > 0, \qquad \det(K) > 0, \end{aligned}\right. \tag{4.22}$$

then for all n large enough the matrices G_n and $L_n = K_n - H_n^\star G_n^{-1} H_n$ are invertible, and the two vectors R_n and S_n defined by (4.19) for those ns are bounded in n.

Clearly (4.22) yields that G_n^{-1} exists for all n large enough and converges to G^{-1}, so $L_n \to K$ (because $H_n \to 0$), so L_n is also invertible for all n large enough. Then it is obvious from (4.19) and (4.22) that

$$R_n \to R = G^{-1} R', \qquad S'_n \to K^{-1} S' - K^{-1} H G^{-1} R',$$

and this completes the proof.

References

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

Basawa, I.V. and Scott, D.J. (1983) *Asymptotic Optimal Inference for Non-ergodic Models*, Lecture Notes in Statist. **17**. New York: Springer-Verlag.

Dohnal, G. (1987) On estimating the diffusion coefficient. *J. Appl. Probab.*, **24**, 105–114.

Genon-Catalot, V. and Jacod, J. (1993) On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **29**, 119–151.

Ibragimov, I.A. and Has'minskii, R.Z. (1981) *Statistical Estimation: Asymptotic Theory*. New York: Springer-Verlag.

Kessler, M. (1997) Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, **24**, 211–229.

Kessler, M. and Sørensen, M. (1999) Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, **5**, 299–314.

Kutoyants, Yu.A. (2004) *Statistical Inference for Ergodic Diffusion Processes*. London: Springer-Verlag.

Pedersen, A.R. (1995) A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.*, **22**, 55–71.

Prakasa Rao, B.L.S. (1999a) *Semimartingales and Their Statistical Inference*. Boca Raton, FL: Chapman & Hall.

Prakasa Rao, B.L.S. (1999b) *Statistical Inference for Diffusion Type Processes*. London: Arnold.

Protter, P. (1990) *Stochastic Integration and Differential Equations*. Berlin: Springer-Verlag.

Revuz, D. and Yor, M. (1991) *Continuous Martingales and Brownian Motion*. Berlin: Springer-Verlag.

Yoshida, N. (1992) Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.*, **41**, 220–242.

Received November 2004 and revised September 2005