A note on semiparametric estimation for ergodic diffusion processes

Yoichi Nishiyama

The Institute of Statistical Mathematics

4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan

nisiyama@ism.ac.jp

http://www.ism.ac.jp/~nisiyama/

February 28, 2009

Running title: Semiparametric estimation for diffusions

Abstract

We consider a semiparametric model for ergodic diffusion process where the drift coefficient contains an unknown finite-dimensional parameter and the diffusion coefficient is an infinite-dimensional nuisance parameter. A uniformly consistent estimator for the diffusion coefficient is given in an explicit way, and by using it we construct an asymptotically normal and efficient estimator for the drift coefficient. This paper implements a general theory established by Nishiyama (2009, *to appear in Ann. Statist.*).

Keywords. Estimating function, nuisance parameter, metric entropy, asymptotic efficiency, ergodic diffusion, discrete observation.

MSC Classification. Primary, 62F12; secondary, 62M05.

1 Introduction

This paper further develops the semiparametric estimation theory for ergodic diffusion processes initiated by Nishiyama (2009). Consider the ergodic diffusion process X on $\mathbb{R}$ which is a solution to the stochastic differential equation (SDE) given by

$$X_t = X_0 + \int_0^t S(X_s;\theta)ds + \int_0^t \sigma(X_s;h)dW_s, \tag{1}$$

where $s \rightsquigarrow W_s$ is a standard Brownian motion. Here, we consider a d-dimensional parametric family $\{S(\cdot;\theta);\theta \in \Theta\}$ for the drift coefficient indexed by a compact subset Θ of $\mathbb{R}^d$, and a possibly infinite-dimensional "parametric" family $\{\sigma^2(\cdot;h);h \in H\}$ for the diffusion coefficient indexed by a (general) totally bounded metric space (H,d_H). We denote by (θ_0,h_0) the true value of (θ,h). Our aim is to estimate θ_0 when the model is perturbed by the unknown nuisance parameter h. We refer to Nishiyama (2009) for the novelty and the importance of this problem and for a historical review; here we only state that in the case where the nuisance parameter space H is *infinite-dimensional*, the asymptotically efficient estimation for θ_0 was first established in Nishiyama (2009) which may be thought as an initial point of this research direction.

In the case where the process X is observed continuously on the time interval $[0,T]$, the diffusion coefficient σ^2 may be assumed to be known (say, σ_0^2), and the maximum likelihood estimator (MLE) for θ_0 is a solution to the estimating equation $\dot{\ell}_T(\theta) = 0$ given by

$$\dot{\ell}_T(\theta) = \frac{1}{T}\int_0^T \frac{\dot{S}(X_t;\theta)}{\sigma_0^2(X_t)}[dX_t - S(X_t;\theta)dt],$$

where $\dot{S}(\cdot;\theta)$ denotes the derivative of $S(\cdot;\theta)$ with respect to θ. In this case, the asymptotic normality of the MLE, as well as its asymptotic efficiency in the framework of Hájek-Le Cam's theory, has been already established (see Kutoyants (2004)). On the other hand, in the case where the process X is observed only at time points $0 = t_0^n < t_1^n < \cdots < t_n^n$, the diffusion coefficient has to be estimated, too. Nishiyama (2009) took an approach to use the estimating function Ψ_n, which is an approximation of $\dot{\ell}_T(\theta)$ with the true σ_0^2 replaced by the nuisance $\sigma^2(\cdot;h)$, given by

$$\Psi_n(\theta,h) = \frac{1}{t_n^n}\sum_{i=1}^n \frac{\dot{S}(X_{t_{i-1}^n};\theta)}{\sigma^2(X_{t_{i-1}^n};h)}[X_{t_i^n} - X_{t_{i-1}^n} - S(X_{t_{i-1}^n};\theta)|t_i^n - t_{i-1}^n|].$$

His assumption on the sampling scheme is that

$$\Delta_n = \max_{1\le i\le n}|t_i^n - t_{i-1}^n| = o((t_n^n)^{-1}) \quad \text{and} \quad t_n^n \to \infty,$$

as $n \to \infty$. Assuming that there exists a d_H-consistent estimator $\widehat{h}_n$ for h_0, he proved that any (approximate) solution to the estimating equation $\theta \rightsquigarrow \Psi_n(\theta,\widehat{h}_n) \approx 0$ is asymptotically normal and efficient in the sense that the asymptotic distribution coincides with that of the MLE in the continuous observation case. In order to obtain this result, he applied the weak convergence theory for ℓ^∞-valued martingales established by Nishiyama (1997, 1999, 2000) to the random fields $(\theta,h) \rightsquigarrow \sqrt{t_n^n}(\Psi_n(\theta,h) - \widetilde{\Psi}_n(\theta,h))$, where $\widetilde{\Psi}_n$ is the "compensator" of Ψ_n given by

$$\widetilde{\Psi}_n(\theta,h) = \frac{1}{t_n^n}\sum_{i=1}^n \frac{\dot{S}(X_{t_{i-1}^n};\theta)}{\sigma^2(X_{t_{i-1}^n};h)}\left[\int_{t_{i-1}^n}^{t_i^n} S(X_t;\theta_0)dt - S(X_{t_{i-1}^n};\theta)|t_i^n - t_{i-1}^n|\right],$$

in the sense that the difference $\Psi_n(\theta,h)-\widetilde{\Psi}_n(\theta,h)$ is the terminal variable of a martingale. To apply the theory, he assumed the following two conditions on (H,d_H): the Lipschitz continuity condition

$$|\sigma^2(x;h)-\sigma^2(x;h')|\leq \Lambda(x)d_H(h,h'), \tag{2}$$

where $\Lambda(x)$ is a suitable function; and the metric entropy condition

$$\int_0^1 \sqrt{\log N(H,d_H,\varepsilon)}d\varepsilon<\infty, \tag{3}$$

where $N(H,d_H,\varepsilon)$ denotes the ε-covering number of the space H with respect to the metric d_H. To check the condition (2) it is convenient to set $\sigma^2(\cdot;h)=h\in H\subset \ell^\infty(\mathbb{R})$, where $\ell^\infty(\mathbb{R})$ is the space of bounded functions on $\mathbb{R}$, and equip the space H with the uniform metric $||\cdot||_\infty$. Hereafter we adopt this parametrization and simply write like "$\sigma^2\in H$". In this case, the condition (3) is satisfied if, for example, we set H to be a class of smooth functions; see e.g. van der Vaart and Wellner (1996) and Section 3 below.

In view of this preceding result, the purpose of the current paper is to give an estimator for the diffusion coefficient $\sigma^2(x)$, which takes values in a class of smooth functions and is consistent with respect to the uniform metric. In a nonparametric setting, Hoffmann (2001) studied the rate of convergence of some estimators for the diffusion coefficient, with respect to the L_2-metric, but his results do not fit in our purpose. Nishiyama (2009) presented one example of consistent estimators by the least square method which can be applied for any metric d_H, but his estimator is the minimizer over a space of functions which is not very realistic from computational point of view. In this paper, inspired by an idea of Kristensen (2008a, b), we propose an *explicit* estimator by using a *semiparametric kernel method* which we can directly compute from the data. The idea is based on the fact that

$$\sigma^2(x)=\frac{2}{f(x)}\int_{-\infty}^x S(y;\theta)f(y)dy$$

where f is the invariant density. We substitute a consistent estimator for θ and a kernel density estimator for $f(\cdot)$ in order to obtain an estimator for $\sigma^2(\cdot)$. The present paper gives a way to implement the general theory of Nishiyama (2009).

The paper is organized as follows. The uniformly consistent estimator for σ^2 is given in Section 2, and the asymptotically efficient estimation for θ is done in Section 3.

2 Uniformly consistent estimator for diffusion coefficient

Throughout this paper, we assume the following.

C1. The solution X to the SDE (1) is a regular diffusion for which the speed measure m is finite (this implies the ergodicity), and the invariant density f is given by

$$f(x)=\frac{1}{G\sigma^2(x)}\exp\left(2\int_0^x \frac{S(y;\theta)}{\sigma^2(y)}dy\right),$$

where $G=\int_{\mathbb{R}}\frac{1}{\sigma^2(x)}\exp\left(2\int_0^x\frac{S(y;\theta)}{\sigma^2(y)}dy\right)dx<\infty$. We further suppose that f is differentiable with the derivative f' which is bounded on each compact set.

See Theorem 1.16 of Kutoyants (2004) for a sufficient condition under which the invariant density is given as above. It follows from this formula that

$$S(x;\theta) = \frac{1}{2f(x)}\frac{\partial}{\partial x}(\sigma^2(x)f(x)),$$

thus we have

$$\sigma^2(x) = \frac{2}{f(x)}\int_{-\infty}^{x} S(y;\theta)f(y)dy.$$

So we will estimate the functions $x \mapsto f(x)$ and $x \mapsto A(x)$ given by

$$A(x) = \int_{-\infty}^{x} S(y;\theta)f(y)dy.$$

It is natural to use the estimators

$$\begin{aligned}
\widehat{f}_n(x) &= \frac{1}{t_n^n}\sum_{i=1}^{n}\frac{1}{b_n}K\left(\frac{x - X_{t_{i-1}^n}}{b_n}\right)|t_i^n - t_{i-1}^n|, \\
\widehat{A}_n(x) &= \frac{1}{t_n^n}\sum_{i=1}^{n}\int_{-\infty}^{x}\frac{1}{b_n}K\left(\frac{y - X_{t_{i-1}^n}}{b_n}\right)dyS(X_{t_{i-1}^n};\widetilde{\theta}_n)|t_i^n - t_{i-1}^n|,
\end{aligned}$$

where K is a kernel function on $\mathbb{R}$ and $\widetilde{\theta}_n$ is a consistent estimator for the true θ_0. By using them, we propose the estimator

$$\widehat{\sigma}_n^2(x) = \frac{2\widehat{A}_n(x)}{\widehat{f}_n(x)} \tag{4}$$

for the true $\sigma_0^2(x)$. When $\widehat{f}_n(x)$ is zero, we set $\widehat{\sigma}_n^2(x)$ to be any fixed value. In the following, we denote by f_0 and A_0 the true functions of f and A, respectively, for the true values θ_0 and σ_0.

Kristensen (2008a, b) studied this kind of estimator in the low frequency case. See also the references therein for more information on related works. However, Kristensen's works did not reach the asymptotically optimal estimation. We consider the problem in the high frequency case and will obtain an asymptotically optimal result.

Here we collect some known results which will be used in our proofs. We denote by $l = (l_t(x) : t \in [0,\infty), x \in \mathbb{R})$ the local time of the regular diffusion X with respect to the speed measure m. For every m-integrable function g the occupation times formula says that

$$\int_0^t g(X_s)ds = \int_{\mathbb{R}} g(x)l_t(x)m(dx).$$

It was proved by van Zanten (2003) that if the speed measure is finite then

$$\sup_{x\in J}\left|\frac{l_t(x)}{t} - \frac{1}{m(\mathbb{R})}\right| \to^p 0$$

for every compact $J \subset \mathbb{R}$. We define the *local time estimator* $f^\circ = (f_t^\circ(x) : t \in [0,\infty), x \in \mathbb{R})$ for the invariant density f by

$$f_t^\circ(x) = \frac{m(\mathbb{R})f(x)l_t(x)}{t}.$$

Then, under the assumption

$$\int_{\mathbb{R}} F(x)^2(1-F(x))^2 dp(x) < \infty, \tag{5}$$

where F and p denote the invariant distribution function and the scale function respectively, $\sqrt{t}(f_t^\circ - f)$ converges weakly in $\ell^\infty(J)$ to a Gaussian random field for every compact $J \subset \mathbb{R}$. This type of theorem was first proved by Kutoyants (1998) in a special case, and it has been extended up to the present form by van der Vaart and van Zanten (2005) as a fruit of their general study on Donsker's theorems for diffusions.

Now, we shall list up some conditions on the drift and diffusion coefficients. Some of them are actually needed for Theorem 2.1 below, and the others will be used only for Theorem 3.1 which is based on Nishiyama (2009). The numbering is corresponding to that in Nishiyama (2009). We suppose that there exists a parametric family of d-dimensional vector-valued functions $\{\dot{S}(\cdot;\theta);\theta\in\Theta\}$ on $\mathbb{R}$ which satisfies the following conditions. Typically, they may be considered to be the derivatives of $S(\cdot;\theta)$ with respect to θ, that is, $\dot{S}(\cdot;\theta) = (\frac{\partial}{\partial\theta_1}S(\cdot;\theta),...,\frac{\partial}{\partial\theta_d}S(\cdot;\theta))^T$. The function Λ appearing in A1' and A3' may be chosen to be common without loss of generality. The assumption that Λ is Lipschitz continuous is necessary only for (6), but actually it is not a strong requirement.

A1'. Θ is a compact subset of $\mathbb{R}^d$. There exists a measurable function Λ on $\mathbb{R}$ such that at the true $\theta_0 \in \Theta$,

$$S(x;\theta) - S(x;\theta_0) = \dot{S}(x;\theta_0)^T(\theta-\theta_0) + \Lambda(x)\epsilon(x;\theta,\theta_0),$$

where $\sup_{x\in\mathbb{R}} |\epsilon(x;\theta,\theta_0)| = o(||\theta-\theta_0||)$ as $\theta\to\theta_0$.

A2'. There exists a constant $C>0$ such that:

$$\sup_{\theta\in\Theta} |S(x;\theta) - S(x';\theta)| \le C|x-x'|;$$

$$\sup_{\theta\in\Theta} ||\dot{S}(x;\theta) - \dot{S}(x';\theta)|| \le C|x-x'|;$$

$$\sup_{\sigma^2\in H} |\sigma^2(x) - \sigma^2(x')| \le C|x-x'|.$$

A3'. There exists a *Lipschitz continuous* function Λ on $\mathbb{R}$ such that:

$$\sup_{\theta\in\Theta} |S(x;\theta)| \le \Lambda(x); \tag{6}$$

$$\sup_{\theta\in\Theta} ||\dot{S}(x;\theta)|| \le \Lambda(x);$$

$$\inf_{\sigma^2\in H}\inf_{x\in\mathbb{R}} \sigma^2(x) > 0;$$

$$||\dot{S}(x;\theta) - \dot{S}(x;\theta)|| \le \Lambda(x)||\theta-\theta'||, \quad \forall\theta,\theta'\in\Theta.$$

A4'. $\sup_{t\in[0,\infty)} E_{\theta_0,\sigma_0^2}(\Lambda(X_t)^8 + |X_t|^4) < \infty$.

A5'. $\int_{\mathbb{R}}(\Lambda(x)^2 + |x|)f_0(x)dx < \infty$.

A6'. The matrix

$$I(\theta_0,\sigma_0^2) = \int_I \frac{\dot{S}(x;\theta_0)\dot{S}(x;\theta_0)^T}{\sigma_0^2(x)} f_0(x)dx$$

is invertible.

A8'. For every $\varepsilon > 0$

$$\inf_{\theta:\ ||\theta-\theta_0||>\varepsilon} \left|\left| \int_{\mathbb{R}} \frac{\dot{S}(x;\theta)}{\sigma_0^2(x)} [S(x;\theta_0) - S(x;\theta)] f_0(dx) \right|\right| > 0.$$

We are ready to give a uniform consistency result for the estimator $\widehat{\sigma}_n^2$ given by (4).

Theorem 2.1 *Suppose that the function $K : \mathbb{R} \to [0,\infty)$ satisfies that $\int_{\mathbb{R}} K(u)du = 1$ with a compact support, and that it differentiable with the derivative K' which is Lipschitz continuous. Let the bandwidth $b_n \downarrow 0$ satisfy that $\Delta_n^{1/2} b_n^{-3} = o(1)$ and that $b_n^{-1} = o(\sqrt{t_n^n})$. Assume C1, A2', A3', A4' and (5). For any bounded subset J of $\mathbb{R}$, we have the following.*

(i) $\sup_{x\in J} |\widehat{f}_n(x) - f_0(x)| \to^p 0$.

(ii) $\sup_{x\in J} |\widehat{A}_n(x) - A_0(x)| \to^p 0$.

(iii) $\sup_{x\in J} |\widehat{\sigma}_n^2(x) - \sigma_0^2(x)| \to^p 0$.

(iv) There exists a constant $M > 0$ such that $P(\widehat{\sigma}_n^2 \in C_M(J)) \to 1$, where $C_M(J)$ is the class of functions g on J such that

$$|||g|||_J := \sup_{x\in J} |g| + \sup_{\substack{x,y\in J\\ x\neq y}} \frac{|g(x) - g(y)|}{|x - y|} \le M.$$

Remark. As for the initial estimator $\widetilde{\theta}_n$, we suggest to use e.g. the least square estimator which is the minimizer of the function

$$\mathcal{A}_n(\theta) = \frac{1}{t_n^n} \sum_{i=1}^n |X_{t_i^n} - X_{t_{i-1}^n} - S(X_{t_{i-1}^n};\theta)|t_i^n - t_{i-1}^n||^2.$$

It is easy to prove its consistency under some mild conditions.

Remark. The assertion (iv), which is necessary for our application, is stronger than that $|||\widehat{\sigma}_n^2|||_J = O_P(1)$.

Before the proofs, we state a lemma which is well known.

Lemma 2.2 *Let X be a solution to the SDE (1) for $(\theta, h) = (\theta_0, h_0)$. Assume $|t_i^n - t_{i-1}^n| \le 1$.*

(i) For any $k \ge 2$, there exists a constant $C_k > 0$, depending only on k, such that

$$\begin{aligned} E \sup_{t\in[t_{i-1}^n, t_i^n]} |X_t - X_{t_{i-1}^n}|^k &\le C_k \sup_{s\in[0,\infty)} E\{|S(X_s;\theta_0)|^k + |\sigma(X_s;h_0)|^k\}|t_i^n - t_{i-1}^n|^{k/2} \\ &=: D_k |t_i^n - t_{i-1}^n|^{k/2}, \end{aligned}$$

provided the right hand side is finite.

(ii) For any $k \ge 2$ and any measurable function g, it holds that

$$\begin{aligned} &\sup_{t\in[t_{i-1}^n, t_i^n]} E(|X_t - X_{t_{i-1}^n}|^{k/2} |g(X_t)|) \\ &\quad \le (D_k |t_i^n - t_{i-1}^n|^{k/2})^{1/2} \sup_{s\in[0,\infty)} (E|g(X_s)|^2)^{1/2}, \end{aligned}$$

provided the right hand side is finite.

Proof. The assertion (i) is well known. (Use Hölder's inequality and Burkholder-Davis-Gundy's inequality for $\int_{t_{i-1}^n}^{t_i^n} |S(X_s;\theta_0)|ds$ and $\sup_{t\in[t_{i-1}^n,t_i^n]} \left|\int_{t_{i-1}^n}^t \sigma(X_s;h_0)dW_s\right|$.) The assertion (ii) follows from Cauchy-Schwartz's inequality and (i). □

Proof of Theorem 2.1. By assumption, there exist some constants $\dot{K}, \dot{K}' > 0$ and a bounded set $D \subset \mathbb{R}$ such that

$$|K(u_1) - K(u_2)| \le \dot{K}|u_1 - u_2|, \quad |K'(u_1) - K'(u_2)| \le \dot{K}'|u_1 - u_2|,$$

and $K(u) = K'(u) = 0$ for $u \in D^c$. We set $J_1 = [\inf J - 1, \sup J + 1]$. We shall use these notations in the proofs below. The inequality "$x \lesssim y$" for $x, y \in [0,\infty)$ means that there exists a constant $C > 0$ which is fixed throughout the proof such that $x \le Cy$.

To prove (i), put

$$\widetilde{f}_n(x) = \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \frac{1}{b_n}K\left(\frac{x - X_t}{b_n}\right)dt = \frac{1}{t_n^n}\int_0^{t_n^n} \frac{1}{b_n}K\left(\frac{x - X_t}{b_n}\right)dt,$$

$$f_n(x) = \int_{\mathbb{R}} \frac{1}{b_n}K\left(\frac{x-y}{b_n}\right)f_0(y)dy.$$

First, we have

$$|\widehat{f}_n(x) - \widetilde{f}_n(x)| \le \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \frac{1}{b_n}\dot{K}\left|\frac{X_{t_{i-1}^n} - X_t}{b_n}\right|dt,$$

thus

$$E\sup_x |\widehat{f}_n(x) - \widetilde{f}_n(x)| \lesssim \frac{\Delta_n^{1/2}}{b_n^2}.$$

So it holds that $\sup_x |\widehat{f}_n(x) - \widetilde{f}_n(x)| = O_P(\Delta_n^{1/2}b_n^{-2}) = o_P(1)$. Next, for sufficiently large n,

$$\begin{aligned}
\sup_{x\in J} |\widetilde{f}_n(x) - f_n(x)| &= \sup_{x\in J}\left|\int_{\mathbb{R}} \frac{1}{b_n}K\left(\frac{x-y}{b_n}\right)\left(\frac{l_{t_n^n}(y)m(\mathbb{R})}{t_n^n} - 1\right)f_0(y)dy\right| \\
&\le \sup_{x\in J}\int_{\mathbb{R}} \frac{1}{b_n}K\left(\frac{x-y}{b_n}\right)f_0(y)dy \cdot \sup_{y\in J_1}\left|\frac{l_{t_n^n}(y)m(\mathbb{R})}{t_n^n} - 1\right| \\
&= \sup_{x\in J}\int_{\mathbb{R}} K(u)f_0(x + b_n u)du \cdot \sup_{y\in J_1}\left|\frac{l_{t_n^n}(y)m(\mathbb{R})}{t_n^n} - 1\right| \\
&\le \sup_{x\in J_1} f_0(x) \cdot \sup_{y\in J_1}\left|\frac{l_{t_n^n}(y)m(\mathbb{R})}{t_n^n} - 1\right|, \qquad (7)
\end{aligned}$$

which converges in probability to zero by Theorem 3.2 of van Zanten (2003). Since f_0 is uniformly continuous on J_1, it is easy to see that $\sup_{x\in J}|f_n(x) - f_0(x)| \to 0$. So the proof of (i) is completed.

To prove (ii), put

$$
\begin{aligned}
A_n^1(x) &= \frac{1}{t_n^n}\sum_{i=1}^n \int_{-\infty}^x \frac{1}{b_n}K\left(\frac{y-X_{t_{i-1}^n}}{b_n}\right)dyS(X_{t_{i-1}^n};\theta_0)|t_i^n-t_{i-1}^n|,\\
A_n^2(x) &= \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n}\int_{-\infty}^x \frac{1}{b_n}K\left(\frac{y-X_{t_{i-1}^n}}{b_n}\right)dyS(X_t;\theta_0)dt,\\
A_n^3(x) &= \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n}\int_{-\infty}^x \frac{1}{b_n}K\left(\frac{y-X_t}{b_n}\right)dyS(X_t;\theta_0)dt,\\
A_n^4(x) &= \int_{-\infty}^x \frac{1}{b_n}K\left(\frac{y-z}{b_n}\right)dyS(z;\theta_0)f_0(z)dz.
\end{aligned}
$$

Notice that

$$\sup_x \int_{-\infty}^x \frac{1}{b_n}K\left(\frac{y-\xi}{b_n}\right)dy \le 1, \quad \forall \xi.$$

Hence

$$\sup_x |\widehat{A}_n(x)-A_n^1(x)| \le \frac{1}{t_n^n}\sum_{i=1}^n \Lambda(X_{t_{i-1}^n})|t_i^n-t_{i-1}^n|\cdot O(||\widetilde{\theta}_n-\theta_0||),$$

which converges in probability to zero whenever $\tilde{\theta}_n$ is a consistent estimator for θ_0. Next

$$E\sup_x |A_n^1(x)-A_n^2(x)| \le \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} E|S(X_{t_{i-1}^n};\theta_0)-S(X_t;\theta_0)|dt \lesssim \Delta_n^{1/2}.$$

Thirdly, it holds that

$$
\begin{aligned}
&\sup_x |A_n^2(x)-A_n^3(x)|\\
&\le \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n}\int_{-\infty}^x \frac{1}{b_n}\left\{1_D\left(\frac{y-X_{t_{i-1}^n}}{b_n}\right)+1_D\left(\frac{y-X_t}{b_n}\right)\right\}\dot{K}\left|\frac{X_{t_{i-1}^n}-X_t}{b_n}\right|dyS(X_t;\theta_0)dt\\
&\le \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} 2\mathrm{Leb}(D)\dot{K}\left|\frac{X_{t_{i-1}^n}-X_t}{b_n}\right|S(X_t;\theta_0)dt,
\end{aligned}
$$

where $\mathrm{Leb}(D)$ denotes the Lebesgue measure of the set D. Thus $E\sup_x |A_n^2(x)-A_n^3(x)| \lesssim \Delta_n^{1/2}/b_n$. On the other hand, by the same argument as that around (7) we can prove that $\sup_{x\in J}|A_n^3(x)-A_n^4(x)| \to^p 0$. Finally, it is easy to see that $\sup_{x\in J}|A_n^4(x)-A_0(x)| \to 0$ because $S(\cdot;\theta_0)$ and f_0 are uniformly continuous on J_1. The proof of (ii) has been finished.

Next note that

$$\sup_{x\in J}|\widehat{\sigma}_n^2(x)-\sigma_0^2(x)| \le \sup_{x\in J}\frac{|\widehat{f}_n(x)-f_0(x)|}{\widehat{f}_n(x)f_0(x)}|\widehat{A}_n(x)| + \sup_{x\in J}\frac{1}{f_0(x)}|\widehat{A}_n(x)-A_0(x)|.$$

Since $\inf_{x\in J} f_0(x) > 0$ the right hand side converges to zero in probability by (i) and (ii), so the assertion (iii) is true.

For $M_1 = 2\sup_{x\in J} A_0(x)/\inf_{x\in J} f_0(x) + 1$ it clearly holds that $P(\sup_{x\in J}\widehat{\sigma}_0^2(x) \leq M_1) \to 1$. On the other hand, we can prove that $\widehat{A}^n$ is "Lipschitz continuous in probability on J" in the sense that there exists a constant $M_2 > 0$ such that

$$P\left(\sup_{\substack{x,y\in J\\ x\neq y}} \frac{|\widehat{A}^n(x) - \widehat{A}^n(y)|}{|x-y|} \leq M_2\right) \to 1.$$

Indeed, we have

$$\begin{aligned}
&\sup_{\substack{x,y\in J\\ x\neq y}} \frac{|\widehat{A}^n(x) - \widehat{A}^n(y)|}{|x-y|} \\
\leq\ & \sup_{z\in J_1} \frac{1}{t_n^n}\sum_{i=1}^n \frac{1}{b_n}K\left(\frac{z - X_{t_{i-1}^n}}{b_n}\right)\Lambda(X_{t_{i-1}^n})|t_i^n - t_{i-1}^n| \\
\leq\ & \sup_{z\in J_1} \frac{1}{t_n^n}\int_0^{t_n^n} \frac{1}{b_n}K\left(\frac{z - X_t}{b_n}\right)\Lambda(X_{t_{i-1}^n})dt + O_P\left(\frac{\sqrt{\Delta_n}}{b_n^2}\right) \\
\leq\ & \sup_{z\in J_1} \frac{1}{t_n^n}\int_0^{t_n^n} \frac{1}{b_n}K\left(\frac{z - X_t}{b_n}\right)\Lambda(X_t)dt + O_P\left(\frac{\sqrt{\Delta_n}}{b_n}\right) + O_P\left(\frac{\sqrt{\Delta_n}}{b_n^2}\right) \qquad (8) \\
\leq\ & \sup_{z\in J_1} \frac{1}{t_n^n}\int_{\mathbb{R}} \frac{1}{b_n}K\left(\frac{z - y}{b_n}\right)\Lambda(y)l_{t_n^n}(y)m(\mathbb{R})f_0(y)dy + o_P(1) \\
\leq\ & \sup_{z\in J_1} \int_{\mathbb{R}} \frac{1}{b_n}K\left(\frac{z - y}{b_n}\right)\Lambda(y)f_0(y)dy(1 + o_P(1)) + o_P(1) \\
=\ & \sup_{z\in J_1} \int_{\mathbb{R}} K(-u)\Lambda(z + b_n u)f_0(z + b_n u)du(1 + o_P(1)) + o_P(1) \\
\leq\ & \sup_{z\in J_1} \Lambda(z)f_0(z) + o_P(1).
\end{aligned}$$

Here, we used the assumption that Λ is Lipschitz continuous in (8). On the other hand, since

$$\left|\frac{1}{\widehat{f}_n(x)} - \frac{1}{\widehat{f}_n(y)}\right| = \frac{|\widehat{f}_n(x) - \widehat{f}_n(y)|}{\widehat{f}_n(x)\widehat{f}_n(y)}$$

and $P(\inf_{x\in J}\widehat{f}_n(x) \geq \inf_{x\in J} f_0(x)/2) \to 1$, it only remains to show that $\widehat{f}_n$ is Lipschitz continuous in probability on J. To see this, it is sufficient to show that its first derivative $\widehat{f}_n'$ satisfies that $\sup_{x\in J_1}\widehat{f}_n'(x) \leq M_3 + o_P(1)$ for a constant $M_3 > 0$. We approximate $\widehat{f}_n'(x)$ by

$$\widetilde{f}_n'(x) = \frac{1}{t_n^n}\int_0^{t_n^n} \frac{1}{b_n^2}K'\left(\frac{x - X_t}{b_n}\right)dt = \int_{\mathbb{R}} \frac{1}{b_n^2}K'\left(\frac{x - y}{b_n}\right)f_{t_n^n}^{\circ}(y)dy.$$

It indeed holds that

$$\begin{aligned}
\sup_x |\widehat{f}_n'(x) - \widetilde{f}_n'(x)| &\leq \frac{1}{t_n^n}\sum_{i=1}^n \frac{\sup_{t\in[t_{i-1}^n, t_i^n]} \dot{K}'|X_{t_{i-1}^n} - X_t|}{b_n^3}|t_i^n - t_{i-1}^n| \\
&= O_P(\Delta_n^{1/2}b_n^{-3}) = o_P(1).
\end{aligned}$$

Now, it follows from Theorem 2.6 of van der Vaart and van Zanten (2005) that

$$\sqrt{t_n^n} \sup_{y\in J_1} |f_{t_n^n}^{\circ}(y) - f_0(y)| = O_P(1).$$

Since $b_n^{-1} = o(\sqrt{t_n^n})$ by assumption, we have

$$\begin{aligned}
\sup_{x\in J_1} \widetilde{f}_n'(x) &\le \sup_{x\in J_1} \left| \int_{\mathbb{R}} \frac{1}{b_n^2} K'\left(\frac{x-y}{b_n}\right) f_0(y)dy \right| + \sup_{x\in J_1} \int_{\mathbb{R}} \frac{1}{b_n} \left| K'\left(\frac{x-y}{b_n}\right) \right| dy \cdot o_P(1) \\
&= \sup_{x\in J_1} \left| \frac{d}{dx} \int_{\mathbb{R}} \frac{1}{b_n} K\left(\frac{x-y}{b_n}\right) f_0(y)dy \right| + \int_{\mathbb{R}} |K'(u)|du \cdot o_P(1) \\
&= \sup_{x\in J_1} \left| \frac{d}{dx} \int_{\mathbb{R}} K(u) f_0(x+b_n u)du \right| + o_P(1) \\
&= \sup_{x\in J_1} \left| \int_{\mathbb{R}} K(u) f_0'(x+b_n u)du \right| + o_P(1).
\end{aligned}$$

Since f_0' is assumed to be bounded on each compact set, the proof is finished. □

3 Semiparametric estimation

To proceed to the semiparametric estimation problem, let us specify the choice of the nuisance parameter space H. We take the material below from Section 2.7.1 of van der Vaart and Wellner (1996). Let I be a bounded, convex subset of $\mathbb{R}^q$. (In the current example of 1-dimensional diffusions, we are considering the case $q = 1$, but for the generality we set q to be a general positive integer.) Let $\alpha > 0$ and $M > 0$ be given, and let $\underline{\alpha}$ be the greatest integer smaller than α. For any vector $k = (k_1, ..., k_q)$ of q integers we define

$$D^k = \frac{\partial^{k.}}{\partial x_1^{k_1} \cdots \partial x_q^{k_q}}$$

where $k. = \sum_{i=1}^q k_i$. We denote by $C_M^{\alpha}(I)$ the class of functions defined on I such that

$$\max_{k. \le \underline{\alpha}} \sup_{\mathbf{x}} |D^k h(\mathbf{x})| + \max_{k. = \underline{\alpha}} \sup_{\mathbf{x},\mathbf{y}} \frac{|D^k h(\mathbf{x}) - D^k h(\mathbf{y})|}{||\mathbf{x} - \mathbf{y}||^{\alpha - \underline{\alpha}}} \le M,$$

where the sumprema are taken over all $\mathbf{x}, \mathbf{y}$ in the interior of I with $\mathbf{x} \ne \mathbf{y}$. Then, there exists a constant $K > 0$ depending only on α and q such that

$$\log N(C_M^{\alpha}(I), ||\cdot||_{\infty}, \varepsilon) \le K \text{Leb}(I^1) \left(\frac{M}{\varepsilon}\right)^{q/\alpha}$$

where $\text{Leb}(I^1)$ is the Lebesgue measure of the set $\{\mathbf{x} : ||\mathbf{x} - I|| < 1\}$. Hence the metric entropy condition (3) is satisfied if $q/(2\alpha) < 1$.

Now, let us fix a (large) compact interval $J = [J_l, J_r]$ and we assume the following conditions for the class H of functions $\sigma^2 : \mathbb{R} \to (0, \infty)$:

H1. Every $\sigma^2 \in H$ is constant on $(-\infty, J_l]$ and on $[J_r, \infty)$.

H2. There exists $M' > 0$ such that the restriction to J of every $\sigma^2 \in H$ belongs to $C_{M'}^1(J)$:

$$\sup_x \sigma(x)^2 + \sup_{x \ne y} \frac{|\sigma^2(x) - \sigma^2(y)|}{|x-y|} \le M'.$$

The assumption H1 may look artificial, but in practice we can choose an arbitrary large J. On the other hand, it should be also noted that the constant M' in H2 does not have to be known to the statisticians in the course of constructing the estimators, as it is seen below.

Based on any consistent estimator $\widetilde{\theta}_n$, let $\widehat{\sigma}_n^2(x)$ be the kernel estimator given by (4) for every $x \in J$, and extent it to all $x \in \mathbb{R}$ by defining $\widehat{\sigma}_n^2(x) = \widehat{\sigma}_n^2(J_l)$ for $x \in (-\infty, J_l]$ and $\widehat{\sigma}_n^2(x) = \widehat{\sigma}_n^2(J_r)$ for $x \in [J_r, \infty)$. Then, by Theorem 2.1 (iv), there exists a constant M'' which is specified by the true $S(\cdot;\theta_0)$, $\sigma_0(\cdot)$, the function Λ in (6) and the interval J such that the obtained estimator $\widehat{\sigma}_n^2$ takes values in the space of functions whose restrictions to J belong to $C^1_{M''}(J)$ with probability tending to 1. The constant $M' \vee M''$ is used in the proof, but the statisticians do not have to know its value in the estimating procedure. As announced in Section 1, we propose to use the estimating function

$$\Psi_n(\theta) = \frac{1}{t_n^n} \sum_{i=1}^n \frac{\dot{S}(X_{t_{i-1}^n};\theta)}{\widehat{\sigma}_n^2(X_{t_{i-1}^n})} [X_{t_i^n} - X_{t_{i-1}^n} - S(X_{t_{i-1}^n};\theta)|t_i^n - t_{i-1}^n|].$$

The following result is now immediate from Nishiyama (2009) and Theorem 2.1, because the last condition of A3 and the condition A7 in Nishiyama's paper have already been checked.

Theorem 3.1 *Assume $\Delta_n = o((t_n^n)^{-1})$ and $t_n^n \to \infty$. Assume A1' – A6', H1, H2, and all conditions in Theorem 2.1. Then, for any random sequence $\widehat{\theta}_n$ such that*

$$||\widehat{\theta}_n - \theta_0|| = o_P(1) \quad \text{and} \quad \Psi_n(\widehat{\theta}_n) = o_P((t_n^n)^{-1/2}),$$

the estimator $\widehat{\theta}_n$ is asymptotically normal and efficient:

$$\sqrt{t_n^n}(\widehat{\theta}_n - \theta_0) \to^d \mathcal{N}(0, I(\theta_0, \sigma_0^2)^{-1}).$$

When A8' is also satisfied, the assumption "$||\widehat{\theta}_n - \theta_0|| = o_P(1)$" is automatically satisfied.

References

[1] Hoffmann, M. (2001). On estimating the diffusion coefficient: parametric versus nonparametric. *Ann. Inst. H. Poincaré. –PR* **37** 339-372.

[2] Kristensen, D. (2008a). Nonparametric estimation and misspecification testing of diffusion models. *Preprint.*

[3] Kristensen, D. (2008b). Pseudo-maximum likelihood estimation in two classes of semiparametric diffusion models. *Preprint.*

[4] Kutoyants, Yu.A. (1998). Efficient density estimation for ergodic diffusion processes. *Stat. Inference Stoch. Process.* **1** 131-155.

[5] Kutoyants, Yu.A. (2004). *Statistical Inference for Ergodic Diffusion Processes.* Springer, London.

[6] Nishiyama, Y. (1997). Some central limit theorems for ℓ^∞-valued semimartingales and their applications. *Probab. Theory Relat. Fields.* **108** 459-494.

[7] Nishiyama, Y. (1999). A maximal inequality for continuous martingales and M-estimation in a Gaussian white noise model. *Ann. Statist.* **27** 675-696.

[8] Nishiyama, Y. (2000). *Entropy Methods for Martingales.* CWI Tract **128** Centrum voor Wiskunde en Informatica, Amsterdam.

[9] Nishiyama, Y. (2009). Asymptotic theory of semiparametric Z-estimators for stochastic processes, with applications to ergodic diffusions and time series. *To appear in Ann. Statist.*

[10] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag, New York.

[11] van der Vaart, A.W. and van Zanten, H. (2005). Donsker theorems for diffusions: necessary and sufficient conditions. *Ann. Probab.* **33** 1422-1451.

[12] van Zanten, H. (2003). On uniform laws of large numbers for ergodic diffusions and consisntency of estimators. *Stat. Inference Stoch. Process.* **6** 199-213.