Centrum voor Wiskunde en Informatica

REPORT *RAPPORT*

Rates of Convergence and Asymptotic Normality of Curve Estimators for Ergodic Diffusion Processes

J.H. van Zanten

Probability, Networks and Algorithms (PNA)

PNA-R0011 October 31, 2000

Report PNA-R0011
ISSN 1386-3711

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

Rates of Convergence and Asymptotic Normality of Curve Estimators for Ergodic Diffusion Processes

Harry van Zanten
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
hvz@cwi.nl

ABSTRACT
For ergodic diffusion processes, we study kernel-type estimators for the invariant density, its derivatives and the drift function. We determine rates of convergence and find the joint asymptotic distribution of the estimators at different points.

2000 Mathematics Subject Classification: 60J60, 62G20.
Keywords and Phrases: Ergodic diffusions, nonparametric estimation, rate of convergence, asymptotic normality.
Note: Work carried out under project PNA3.3 'Stochastic Processes and Applications'.

1 Introduction

We consider diffusion processes that solve a stochastic differential equation of the form

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t, \tag{1.1}$$

where W is a standard Brownian motion and b and σ are certain measurable functions. More precisely, we consider a solution X of equation (1.1) that has the ergodic property with invariant measure μ, meaning that the law of large numbers holds, i.e. that

$$\frac{1}{t}\int_0^t g(X_s)\,ds \stackrel{\text{as}}{\to} \int g\,d\mu \tag{1.2}$$

for every $g \in L^1(\mu)$ and that we have the weak convergence $X_t \rightsquigarrow \mu$ as $t \to \infty$. In the next section we state precise conditions on the coefficients b and σ under which solutions of (1.1) have the ergodic property and we recall the relation between the functions b and σ and the density f of the invariant measure μ.

The invariant density f is the main object we are interested in. Our aim is to find the asymptotic properties of nonparametric estimators for the derivatives of the function f. Based on the observation of a trajectory $\{X_s : s \leq t\}$ of the diffusion, we can estimate f itself by the kernel estimator $\hat{f}_{t,h}$ defined by

$$\hat{f}_{t,h}(x) = \frac{1}{ht} \int_0^t K\left(\frac{x - X_s}{h}\right) ds,$$

where K is some appropriate kernel function and $h > 0$ is a bandwidth. Obvious estimators for the derivatives $f^{(m)}$ of f are then obtained by differentiating this expression. If the kernel K has an m-th derivative $K^{(m)}$, we define

$$\hat{f}_{t,h}^{(m)}(x) = \frac{1}{h^{m+1}t} \int_0^t K^{(m)}\left(\frac{x - X_s}{h}\right) ds. \tag{1.3}$$

Kernel estimators of this type have been studied by several authors. Banon (1978) considered estimators for the density f and its derivative f', i.e. the cases $m = 0$ and $m = 1$. In Banon's paper, conditions were given for pointwise mean-square consistency. For $m = 0$, uniform consistency was investigated by Nguyen (1978). Kutoyants (1998) showed that the rate of convergence of the kernel estimator $\hat{f}_{t,h_t}$ for f is $\sqrt{t}$ (independent of the bandwidth h_t) and he proved pointwise asymptotic normality and efficiency. Van Zanten (2000b) showed that this asymptotic normality is in fact uniform in the variable x. More precisely, for every compact subinterval I of the state space of the diffusion we have the weak convergence

$$\sqrt{t}(\hat{f}_{t,h_t} - f) \rightsquigarrow G \tag{1.4}$$

in the space $\ell^\infty(I)$ of bounded functions on I, provided that the bandwidths h_t converge to 0 at the right speed. Here G denotes a certain Gaussian random map in the space $\ell^\infty(I)$. Uniform convergence of the estimators for the derivatives of f was also considered in Van Zanten (2000b). It has been shown there that for compact intervals I

$$\sup_{x \in I} \left| \hat{f}_{t,h_t}^{(m)}(x) - f^{(m)}(x) \right| = o_P\left(\frac{1}{h_t^m \sqrt{t}}\right)$$

for $m \in \mathbb{N} = \{1, 2, \ldots\}$. The exact rate of convergence for $m \in \mathbb{N}$ still remained unknown, but it was conjectured to be $(th_t^{2m-1})^{1/2}$. The purpose of the present paper is to prove that the rate is indeed equal to $(th_t^{2m-1})^{1/2}$ and to find the weak limit of the normalized difference

$$\sqrt{th_t^{2m-1}}\big[\hat{f}_{t,h_t}^{(m)} - f^{(m)}\big].$$

It is not immediately obvious that we should have the rate $(th_t^{2m-1})^{1/2}$. Comparing our problem with the problem of density estimation for i.i.d. random variables would give a different guess. In the i.i.d. case, kernel estimators for the m-th derivative of the density have a rate $(nh_n^{2m+1})^{1/2}$, where n is the number of observations and h_n is the bandwidth. But such a straightforward comparison with the i.i.d. case can not be made. It fails already in the case $m = 0$, where we have a parametric rate for ergodic diffusion models (see (1.4)). On the other hand, there are continuous-time models for which kernel estimators for the derivative of the density also have the parametric rate $t^{1/2}$ (see Lucas (1998)). The results of this paper show that this is not the case for ergodic diffusions.

While density estimators behave very differently in i.i.d. models and ergodic diffusion models, it is known that kernel estimators for nonparametric regression-type problems behave similarly for both types of observations. Some heuristics then leads to the correct guess regarding the asymptotic behaviour of kernel estimators for the derivative f' of the invariant density: For the drift function b of the SDE (1.1) it holds that

$$E\left(X_{t+\Delta} - X_t \,|\, X_t = x\right) = \Delta b(x) + o(\Delta)$$

as $\Delta \to 0$. So for small $\Delta > 0$, the function b is approximately equal to the regression function of $(X_{t+\Delta} - X_t)/\Delta$ on X_t. This observation has led Pham Dinh Tuan (1981) to construct a nonparametric estimator for b by mimicking the construction of the well-known Nadaraya-Watson estimator. For i.i.d. observations, the Nadaraya-Watson estimator has the rate of convergence $(nh_n)^{1/2}$, where n is the number of observations and h_n is the bandwidth. Moreover, at different points of the real line the estimators are asymptotically independent (see Schuster (1972) for both assertions). Pham Dinh Tuan (1981) found the same asymptotic behaviour for the nonparametric estimator of b, with the number n replaced by the length t of the observation interval. If σ is differentiable, the relation between the invariant density f and the drift and diffusion functions b and σ is given by (3.1), as is explained below. Since we already know that kernel estimators for f have the rate $t^{1/2}$, the observations we just made about kernel-type estimators for b lead us to guess that kernel estimators for f' have the rate $(th_t)^{1/2}$ and that estimators at different points are asymptotically independent. Theorem 3.1 below confirms this guess.

The structure of this paper is as follows. In the next section we give a precise description of the ergodic diffusion models that we consider. We give conditions on the functions b and σ and recall the form of the invariant density. In section 3 the main results are presented. For every $m \in \mathbb{N}$, theorem 3.1 gives the rate of convergence and the joint asymptotic distribution of the estimator (1.3) at different points. Corollary 3.2 concerns a nonparametric drift estimator that was proposed by Banon (1978). It turns out that it has exactly the same asymptotic behaviour as the estimator of Pham Dinh Tuan (1981) that we discussed in the preceding paragraph. The proof of theorem 3.1 is deferred to section 4.

2 Model assumptions

We consider the stochastic differential equation (1.1), where W is a standard Brownian motion and b and σ are measurable functions. We assume that the SDE has a unique strong solution for every initial condition (see for instance Karatzas and Shreve (1991) for conditions in terms of the coefficients b and σ). By (l, r) we denote the (possibly unbounded) state space of the diffusion. More precisely, we assume that if the law of the initial random variable X_0 is concentrated on (l, r), then the whole process X takes values in this interval. Typically, this is ensured by the condition that $\sigma(x) > 0$ for all $x \in (l, r)$ and

$$\begin{aligned} \sigma(l) = 0, \quad b(l) > 0 \quad \text{if } -\infty < l, \\ \sigma(r) = 0, \quad b(r) < 0 \quad \text{if } r < \infty \end{aligned}$$

(see Gihman and Skorohod (1972), theorem 2, p. 149). To avoid technical difficulties, we assume that both b and σ are continuous on the state space (l,r) and that $\sigma > 0$ on (l,r).

Now fix a point $x_0 \in (l,r)$. Recall that the derivative of the scale function associated to the stochastic differential equation (1.1) is the function s on (l,r) defined by

$$s(x) = \exp\left[-2\int_{x_0}^{x} \frac{b(y)}{\sigma^2(y)}\,dy\right]. \tag{2.1}$$

It is assumed that

$$s(l) = s(r) = \infty \quad \text{and} \quad D = \int_l^r \frac{1}{\sigma^2(x)s(x)}\,dx < \infty. \tag{2.2}$$

The probability measure μ on (l,r) is defined by $\mu(dx) = f(x)\,dx$, where

$$f(x) = \frac{1}{D\sigma^2(x)s(x)}. \tag{2.3}$$

The distribution function of the measure μ is denoted by F. It is well-known (see e.g. Gihman and Skorohod (1972) or Skorokhod (1989)) that condition (2.2) implies that the solution X of (1.1) is ergodic in the sense that the law of large numbers holds, i.e. that (1.2) holds for every $g \in L^1(\mu)$. Moreover, the solution that satisfies the initial condition $\mathcal{L}(X_0) = \mu$ is stationary. Throughout the paper, the symbol X denotes this stationary, ergodic solution of the stochastic differential equation (1.1). We call μ the invariant measure of the process, and f and F the invariant density and distribution function, respectively.

3 Main results

First we formulate two important integrability conditions that we shall use in the sequel.

$$C_1: \qquad \int \sigma^2\,d\mu < \infty \ \text{ and there exists an } \varepsilon > 0 \text{ such that}$$
$$\int |b|^{1+\varepsilon}\,d\mu < \infty \quad \text{and} \quad \int |x|^{1+\varepsilon}\mu(dx) < \infty.$$

$$C_2: \qquad \int \left[\frac{F(1-F)}{\sigma f}\right]^2 d\mu < \infty.$$

Very often, linear growth conditions are imposed on the functions b and σ. In that case, condition C_1 reduces to the requirement that the invariant measure μ has a finite second moment. Condition C_2 is a technical condition that ensures the integrability of certain functions appearing in the proof of theorem 3.1. It is in fact also equivalent to the finiteness of the covariance function of the Gaussian random map G in (1.4) (see Van Zanten (2000b)).

Throughout the paper, the kernel function K is assumed to be a symmetric probability density with compact support. As usual, we denote by $N_d(0,\Sigma)$ the d-dimensional normal distribution with mean vector 0 and covariance matrix Σ. The following theorem is the main result of the paper. We will prove it in the next section.

Theorem 3.1. *Let $m \in \mathbb{N}$ be given. Suppose that conditions C_1 and C_2 hold, that f is $m+2$ times continuously differentiable and that K is m times continuously differentiable. Furthermore, suppose that $th_t^{2m+3} \to 0$ as $t \to \infty$. Then for all distinct $x_1, \ldots, x_d \in (l, r)$ we have*

$$\sqrt{th_t^{2m-1}} \begin{bmatrix} \hat{f}_{t,h_t}^{(m)}(x_1) - f^{(m)}(x_1) \\ \vdots \\ \hat{f}_{t,h_t}^{(m)}(x_d) - f^{(m)}(x_d) \end{bmatrix} \rightsquigarrow N_d(0, \Sigma)$$

as $t \to \infty$, where $\Sigma = diag(\Sigma_1, \ldots, \Sigma_d)$ and

$$\Sigma_i = 4\frac{f(x_i)}{\sigma^2(x_i)} \int_{\mathbb{R}} \left[K^{(m-1)}(w)\right]^2 dw$$

for $i = 1, \ldots, d$.

Once we have this result we can use the delta-method to find the limiting distribution of estimators that are smooth functions of the kernel estimators. An interesting case is the nonparametric drift estimator that was proposed by Banon (1978). Suppose that the function σ is known and continuously differentiable on (l, r). Then definitions (2.1) and (2.3) give the relation

$$b(x) = \tfrac{1}{2}\sigma^2(x)\frac{f'(x)}{f(x)} + \sigma(x)\sigma'(x) \tag{3.1}$$

for every $x \in (l, r)$. An obvious nonparametric estimator for the function b is obtained by replacing f' and f in this expression by their kernel estimators. This leads to the estimator

$$\hat{b}_{t,h_t} = \tfrac{1}{2}\sigma^2\frac{\hat{f}_{t,h_t}^{(1)}}{\hat{f}_{t,h_t}} + \sigma\sigma'. \tag{3.2}$$

The delta-method gives us the following corollary of theorem 3.1.

Corollary 3.2. *Suppose that conditions C_1 and C_2 hold, that f is 3 times continuously differentiable and that K is continuously differentiable. Furthermore, suppose that $th_t \to \infty$ and $th_t^4 \to 0$ as $t \to \infty$. Then for all distinct $x_1, \ldots, x_d \in (l, r)$ we have*

$$\sqrt{th_t} \begin{bmatrix} \hat{b}_{t,h_t}(x_1) - b(x_1) \\ \vdots \\ \hat{b}_{t,h_t}(x_d) - b(x_d) \end{bmatrix} \rightsquigarrow N_d(0, \Gamma)$$

as $t \to \infty$, where $\Gamma = diag(\Gamma_1, \ldots, \Gamma_d)$ and

$$\Gamma_i = \frac{\sigma^2(x_i)}{f(x_i)} \int_{\mathbb{R}} K^2(w)\, dw$$

for $i = 1, \ldots, d$.

Proof. First we introduce some notation. Let the vectors S and T be given by

$$S = \Big[f(x_1), \dots, f(x_d)\Big]^\top, \quad T = \Big[f'(x_1), \dots, f'(x_d)\Big]^\top$$

and let the random vectors S_t and T_t be the corresponding estimators

$$S_t = \Big[\hat{f}_{t,h_t}(x_1), \dots, \hat{f}_{t,h_t}(x_d)\Big]^\top, \quad T_t = \Big[\hat{f}^{(1)}_{t,h_t}(x_1), \dots, \hat{f}^{(1)}_{t,h_t}(x_d)\Big]^\top.$$

Now consider the map $\phi : (0,\infty)^d \times \mathbb{R}^d \to \mathbb{R}^d$ given by $\phi = (\phi_1, \dots, \phi_d)^\top$ and

$$\phi_i(y, z) = \tfrac{1}{2}\sigma^2(x_i)\frac{z_i}{y_i} + \sigma(x_i)\sigma'(x_i).$$

Then by relations (3.1) and (3.2) it holds that

$$\sqrt{th_t}\begin{bmatrix} \hat{b}_{t,h_t}(x_1) - b(x_1) \\ \vdots \\ \hat{b}_{t,h_t}(x_d) - b(x_d) \end{bmatrix} = \sqrt{th_t}\Big[\phi(S_t, T_t) - \phi(S, T)\Big].$$

Under the conditions of the corollary, the estimator $\hat{f}_{t,h_t}$ of f has a rate of convergence $\sqrt{t}$. More precisely, the random vector $\sqrt{t}(S_t - S)$ converges weakly to some Gaussian random vector (see Van Zanten (2000b), theorem 3.4). Using theorem 3.1 we thus find that

$$\sqrt{th_t}\Big[(S_t, T_t) - (S, T)\Big] \rightsquigarrow \Big(0, V\Big)$$

in $\mathbb{R}^d \times \mathbb{R}^d$, where V has a $N_d(0, \Sigma)$-distribution with $\Sigma = \mathrm{diag}(\Sigma_1, \dots, \Sigma_d)$ and

$$\Sigma_i = 4\frac{f(x_i)}{\sigma^2(x_i)}\int_{\mathbb{R}} K^2(w)\, dw.$$

The map ϕ is differentiable on its domain, which contains the point (S, T). So by the delta-method (see Van der Vaart (1998), theorem 3.1)

$$\sqrt{th_t}\Big[\phi(S_t, T_t) - \phi(S, T)\Big] \rightsquigarrow$$
$$\phi'(S, T)\begin{pmatrix} 0 \\ V \end{pmatrix} = \begin{bmatrix} \frac{\partial\phi_1}{\partial z_1}(S, T) & \dots & \frac{\partial\phi_1}{\partial z_d}(S, T) \\ \vdots & \ddots & \vdots \\ \frac{\partial\phi_d}{\partial z_1}(S, T) & \dots & \frac{\partial\phi_d}{\partial z_d}(S, T) \end{bmatrix} V. \tag{3.3}$$

The matrix of partial derivatives on the right hand side is easily seen to be equal to

$$\Delta = \tfrac{1}{2}\mathrm{diag}\left[\frac{\sigma^2(x_1)}{f(x_1)}, \dots, \frac{\sigma^2(x_d)}{f(x_d)}\right].$$

The right hand side of (3.3) therefore has a $N_d(0, \Gamma)$ distribution, where $\Gamma = \Delta\Sigma\Delta$. This is precisely the matrix that is described in the statement of the corollary. □

4 Proof of the main theorem

We prove theorem 3.1 in two steps, see sections 4.1 and 4.2 below. In section 4.1 we first write the difference $\hat{f}^{(m)}_{t,h_t}(x) - f^{(m)}(x)$ as the sum of stochastic integrals and an error term R_t that is of order

$$R_t = O(h_t^2) + O_P\left(\frac{1}{th_t^{m-1}}\right) + O_P\left(\frac{1}{\sqrt{t}}\right)$$

for $t \to \infty$. Since we assume that $\sqrt{th_t^{2m+3}} \to 0$, this implies that

$$\sqrt{th_t^{2m-1}}R_t \xrightarrow{\text{P}} 0$$

as $t \to \infty$. So in section 4.1 the problem of proving theorem 3.1 is reduced to a central limit problem for certain stochastic integrals. In section 4.2 we study the asymptotic behaviour of these integrals.

4.1 Step one

In this section we prove that up to terms of lower order, the difference $\hat{f}^{(m)}_{t,h_t} - f^{(m)}$ can be written as a sum of stochastic integrals. The exact formulation is as follows.

Theorem 4.1. *Let $m \in \mathbb{N}$ be given. Suppose that C_2 holds, that f is $m+2$ times continuously differentiable, K is m times continuously differentiable and $h_t \to 0$. Then for every $x \in (l, r)$*

$$\hat{f}^{(m)}_{t,h}(x) - f^{(m)}(x) = \sum_{n=1}^{m} \frac{1}{t} \int_0^t A_{n,t,x}(X_s)\, dW_s + R_t, \tag{4.1}$$

with

$$A_{n,t,x}(y) = \frac{2}{h_t^n \sigma(y)} K^{(n-1)}\left(\frac{x-y}{h_t}\right)$$

and

$$R_t = O(h_t^2) + O_P\left(\frac{1}{th_t^{m-1}}\right) + O_P\left(\frac{1}{\sqrt{t}}\right).$$

as $t \to \infty$.

The remainder term R_t also depends on m and x. However, since that is not relevant for the rest of the paper, we do not make it explicit in the notation.

Proof. Let $x \in (l, r)$ be fixed. The proof of the theorem is broken down in several steps. The first step is to write

$$\hat{f}^{(m)}_{t,h_t}(x) - f^{(m)}(x) = \hat{f}^{(m)}_{t,h_t}(x) - E\,\hat{f}^{(m)}_{t,h_t}(x) + R_{1,t},$$

where

$$R_{1,t} = E\,\hat{f}^{(m)}_{t,h_t}(x) - f^{(m)}(x).$$

As was noted in Van Zanten (2000b), the bias term $R_{1,t}$ can be handled in the same manner as the bias of a kernel estimator for the density of i.i.d. observations. Using repeated partial integration and Taylor's formula, one finds that under the conditions of the theorem, $R_{1,t} = O(h_t^2)$ (see Van Zanten (2000b), lemma 4.2), which explains the term $O(h_t^2)$ in the statement of the theorem.

Next, observe that by stationarity of the process X we have the relation

$$\hat{f}^{(m)}_{t,h_t}(x) - E\,\hat{f}^{(m)}_{t,h_t}(x) = \frac{1}{t}\int_0^t \phi_{m,t,x}(X_s)\,ds,$$

where

$$\phi_{m,t,x}(\cdot) = \frac{1}{h_t^{m+1}}K^{(m)}\left(\frac{x-\cdot}{h_t}\right) - \int_l^r \frac{1}{h_t^{m+1}}K^{(m)}\left(\frac{x-y}{h_t}\right) f(y)\,dy.$$

By the generalized Itô formula (see Karatzas and Shreve (1991)) and the form (2.3) of the invariant density f it holds for every function $\phi \in L^1(\mu)$ that

$$\int_0^t \phi(X_s)\,ds = \psi(X_t) - \psi(X_0) - \int_0^t \sigma(X_s)\psi'(X_s)\,dW_s,$$

where ψ is a function on (l,r) whose derivative ψ' is given by

$$\psi'(x) = \frac{2}{f(x)\sigma^2(x)}\int_l^x \phi(y)f(y)\,dy.$$

Taking $\phi = \phi_{m,t,x}$ we find that

$$\hat{f}^{(m)}_{t,h_t}(x) - E\,\hat{f}^{(m)}_{t,h_t}(x) = \frac{1}{t}\Big[\psi_{m,t,x}(X_t) - \psi_{m,t,x}(X_0)\Big] - \frac{1}{t}\int_0^t \sigma(X_s)\psi'_{m,t,x}(X_s)\,dW_s,$$

where

$$\psi_{m,t,x}(y) = \int_{x_0}^y \psi'_{m,t,x}(z)\,dz$$

and

$$\psi'_{m,t,x}(y) = \frac{2}{f(y)\sigma^2(y)}\int_l^y \phi_{m,t,x}(z)f(z)\,dz.$$

The next step is to find a useful expression for the function $\psi'_{m,t,x}$.

Say that the support of K is contained in the compact interval J and let the bandwidth $h_t > 0$ be small enough to ensure that $l, r \notin \{x\} - h_t J$. In that case, repeated partial integration gives

$$\begin{aligned}\int_l^y \frac{1}{h_t^{m+1}}K^{(m)}\left(\frac{x-z}{h_t}\right) f(z)\,dz = \\ \int_l^y \frac{1}{h_t}K\left(\frac{x-z}{h_t}\right) f^{(m)}(z)\,dz \\ - \frac{1}{h_t}K\left(\frac{x-y}{h_t}\right) f(y) - \cdots - \frac{1}{h_t^m}K^{(m-1)}\left(\frac{x-y}{h_t}\right) f(y)\end{aligned}$$

for every $y \in (l,r)$. Consequently, we find that

$$\int_l^y \phi_{m,t,x}(z) f(z)\,dz =$$
$$\int_l^r \frac{1}{h_t} K\left(\frac{x-z}{h_t}\right) f^{(m)}(z) \left(1_{(z,r)}(y) - F(y)\right)\,dz$$
$$- \frac{1}{h_t} K\left(\frac{x-y}{h_t}\right) f(y) - \cdots - \frac{1}{h_t^m} K^{(m-1)}\left(\frac{x-y}{h_t}\right) f(y).$$

It follows that

$$\psi'_{m,t,x}(y) = \frac{1}{\sigma(y)}\left[B_{m,t,x}(y) - \sum_{n=1}^m A_{n,t,x}(y)\right],$$

where the functions $A_{n,t,x}$ are defined as in the statement of the theorem and

$$B_{m,t,x}(y) = \int_l^r \frac{1}{h_t} K\left(\frac{x-z}{h_t}\right) f^{(m)}(z) \lambda_z(y)\,dz,$$

with

$$\lambda_z(y) = 2\frac{1_{(z,r)}(y) - F(y)}{f(y)\sigma(y)}.$$

Hence, we get

$$\hat{f}^{(m)}_{t,h}(x) - E\,\hat{f}^{(m)}_{t,h}(x) = \sum_{n=1}^m \frac{1}{t}\int_0^t A_{n,t,x}(X_s)\,dW_s + R_{2,t} - R_{3,t},$$

where

$$R_{2,t} = \frac{1}{t}\Big[\psi_{m,t,x}(X_t) - \psi_{m,t,x}(X_0)\Big]$$

and

$$R_{3,t} = \frac{1}{t}\int_0^t B_{m,t,x}(X_s)\,dW_s.$$

To finish the proof of theorem 4.1 we will show that

$$R_{2,t} = O_P\left(\frac{1}{t h_t^{m-1}}\right) \tag{4.2}$$

and

$$R_{3,t} = O_P\left(\frac{1}{\sqrt{t}}\right). \tag{4.3}$$

For the proof of (4.2), observe that there exists a constant $C > 0$ such that we have $|B_{m,t,x}(y)| \leq C/(f(y)\sigma(y))$ for every y. It follows that

$$|\psi'_{m,t,x}(y)| \leq \frac{C}{f(y)\sigma^2(y)} + \sum_{n=1}^m \frac{A_{n,t,x}(y)}{\sigma(y)}.$$

By a change of variables, it is easily seen that for h_t small enough it holds that

$$\left|\int_{x_0}^{z} \frac{A_{n,t,x}(y)}{\sigma(y)}\,dy\right| \le \int_J \frac{2}{h_t^{n-1}} K^{(n-1)}(w) \frac{1}{\sigma^2(x-h_t w)}\,dw \le \frac{D_n}{h_t^{n-1}}$$

for some appropriate constant D_n. Hence, with $D = m\max\{D_1,\dots,D_m\}$, we have

$$|\psi_{m,t,x}(z)| = \left|\int_{x_0}^{z} \psi'_{m,t,x}(y)\,dy\right| \le C\left|\int_{x_0}^{z} \frac{1}{f(y)\sigma^2(y)}\,dy\right| + D\frac{1}{h_t^{m-1}}$$

for h_t small enough. In view of the stationarity of the process X this implies relation (4.2).

To prove (4.3) let

$$\varepsilon_{m,t,x}(y) = \int_l^r \frac{1}{h_t} K\left(\frac{x-z}{h_t}\right) f^{(m)}(z)\lambda_z(y)\,dz - f^{(m)}(x)\lambda_x(y)$$

so that

$$B_{m,t,x}(y) = f^{(m)}(x)\lambda_x(y) + \varepsilon_{m,t,x}(y).$$

Under condition C_2, the function λ_x is μ-square integrable. So by the ergodic property (1.2) and the central limit theorem for martingales we have the weak convergence

$$\frac{1}{\sqrt{t}}\int_0^t f^{(m)}(x)\lambda_x(X_s)\,dW_s \rightsquigarrow [f^{(m)}(x)]^2 N\big(0, \|\lambda_x\|^2_{L^2(\mu)}\big).$$

To finish the proof of (4.3) it thus suffices to show that

$$\frac{1}{\sqrt{t}}\int_0^t \varepsilon_{m,t,x}(X_s)\,dW_s \xrightarrow{\mathrm{P}} 0. \tag{4.4}$$

By a change of variables, we see that for h_t small enough we can write

$$\varepsilon_{m,t,x}(y) = \int_J K(w)\Big[f^{(m)}(x-h_t w)\lambda_{x-h_t w}(y) - f^{(m)}(x)\lambda_x(y)\Big]\,dw.$$

To shorten the notation somewhat, we write $(\cdots)$ in place of the large expression $f^{(m)}(x - h_t w)\lambda_{x-h_t w}(y) - f^{(m)}(x)\lambda_x(y)$. Using Jensen's inequality and Fubini's theorem we get

$$\begin{aligned} E\left[\frac{1}{\sqrt{t}}\int_0^t \varepsilon_{m,t,x}(X_s)\,dW_s\right]^2 &= E\,\frac{1}{t}\int_0^t \varepsilon^2_{m,t,x}(X_s)\,ds \\ &= \int_l^r \varepsilon^2_{m,t,x}(y) f(y)\,dy \\ &= \int_l^r \left[\int_J K(w)(\cdots)\,dw\right]^2 f(y)\,dy \\ &\le \int_l^r \int_J K(w)(\cdots)^2\,dw\, f(y)\,dy \\ &= \int_J K(w)\int_l^r (\cdots)^2 f(y)\,dy\,dw. \end{aligned} \tag{4.5}$$

Straightforward calculus shows that the assumption that f is $m+1$ times continuously differentiable and condition C_2 imply that there exists a constant $C>0$ (not depending on t) such that

$$\int_l^r (\cdots)^2 f(y)\,dy \le Ch_t(|w|+w^2). \tag{4.6}$$

Indeed, the condition on f implies that

$$(\cdots) = f^{(m)}(x)\Big(\lambda_{x-h_tw}(y)-\lambda_x(y)\Big)+\lambda_{x-h_tw}(y)\int_x^{x-h_tw} f^{(m+1)}(v)\,dv.$$

Square this identity, recall the definition of the functions λ_z and integrate to find (4.6). It follows from (4.5) and (4.6) that

$$E\left[\frac{1}{\sqrt{t}}\int_0^t \varepsilon_{m,t,x}(X_s)\,dW_s\right]^2 \le Ch_t\int_J(|w|+w^2)K(w)\,dw \to 0$$

as $t\to\infty$. Hence, we have established (4.4) and the proof of theorem 4.1 is finished. □

4.2 Step two

Let us now finish the proof of theorem 3.1. Using theorem 4.1 we see that for every $x\in(l,r)$

$$\begin{aligned}&\sqrt{th_t^{2m-1}}\left[\hat f^{(m)}_{t,h_t}(x)-f^{(m)}(x)\right]=\\&\sqrt{\frac{h_t^{2m-1}}{t}}\int_0^t A_{m,t,x}(X_s)\,dW_s+\sum_{n=1}^{m-1}h_t^{m-n}\sqrt{\frac{h_t^{2n-1}}{t}}\int_0^t A_{n,t,x}(X_s)\,dW_s\\&+O_P\left(\sqrt{\frac{h_t}{t}}\right)+O_P\left(\sqrt{h_t^{2m-1}}\right)+O\left(\sqrt{th_t^{2m+3}}\right).\end{aligned}$$

Since we assume that $th_t^{2m+3}\to 0$ as $t\to\infty$, theorem 3.1 follows from the following lemma.

Lemma 4.2. *Let $n\in\mathbb{N}$ be given. Suppose that C_1 holds, that K is $n-1$ times continuously differentiable and that $h_t\to 0$. Then for all distinct $x_1,\ldots,x_d\in(l,r)$ we have*

$$\begin{bmatrix}\sqrt{\frac{h_t^{2n-1}}{t}}\int_0^t A_{n,t,x_1}(X_s)\,dW_s\\ \vdots\\ \sqrt{\frac{h_t^{2n-1}}{t}}\int_0^t A_{n,t,x_d}(X_s)\,dW_s\end{bmatrix}\rightsquigarrow N_d(0,\Sigma)$$

as $t\to\infty$, where $\Sigma=diag(\Sigma_1,\ldots,\Sigma_d)$ and

$$\Sigma_i = 4\frac{f(x_i)}{\sigma^2(x_i)}\int_{\mathbb{R}}\left[K^{(n-1)}(w)\right]^2\,dw$$

for $i=1,\ldots,d$.

Proof. We denote by f_t the empirical density of the process X. So for $x \in (l,r)$ and $t > 0$ we put $f_t(x) = 2L_t(x)/t\sigma^2(x)$, where $\{L_t(x) : t \geq 0, x \in (l,r)\}$ denotes the semimartingale local time of the process X. Then for $t > 0$ we have the relation

$$\frac{1}{t}\int_0^t g(X_s)\,ds = \int_l^r g(x)f_t(x)\,dx$$

for every $g \in L^1(\mu)$ (see Karatzas and Shreve (1991)). For $i,j \in \{1,\ldots,d\}$ it follows that

$$\begin{aligned}
&\frac{h_t^{2n-1}}{t}\int_0^t A_{n,t,x_i}(X_s)A_{n,t,x_j}(X_s)\,ds \\
&\quad = \frac{4}{t}\int_0^t \frac{1}{h_t}K^{(n-1)}\left(\frac{x_i - X_s}{h_t}\right)K^{(n-1)}\left(\frac{x_j - X_s}{h_t}\right)\frac{1}{\sigma^2(X_s)}\,ds \\
&\quad = 4\int_l^r \frac{1}{h_t}K^{(n-1)}\left(\frac{x_i - z}{h_t}\right)K^{(n-1)}\left(\frac{x_j - z}{h_t}\right)\frac{f_t(z)}{\sigma^2(z)}\,dz.
\end{aligned}$$

Let the support of K be contained in the compact interval J. Then if $i \neq j$, the neighborhoods $x_i - h_tJ$ and $x_j - h_tJ$ are disjoint for h_t small enough, so the last integral vanishes as $t \to \infty$. If $i = j$, a change of variables shows that for h_t small enough, the last integral is equal to

$$\begin{aligned}
&4\int_J \left[K^{(n-1)}(w)\right]^2 \frac{f_t(x_i - h_tw)}{\sigma^2(x_i - h_tw)}\,dw = \\
&\quad 4\frac{f(x_i)}{\sigma^2(x_i)}\int_{\mathbb{R}}\left[K^{(n-1)}(w)\right]^2\,dw \\
&\quad + 4\int_J \left[K^{(n-1)}(w)\right]^2\left[\frac{f_t(x_i - h_tw)}{\sigma^2(x_i - h_tw)} - \frac{f(x_i)}{\sigma^2(x_i)}\right]dw.
\end{aligned} \tag{4.7}$$

Under condition C_1 we have the uniform convergence

$$\sup_{x\in I}|f_t(x) - f(x)| \xrightarrow{\mathrm{P}} 0$$

for every compact interval $I \subseteq (l,r)$ (see Van Zanten (2000a), theorem 7 and the remarks following theorem 3.1 of Van Zanten (2000b)). Using also the fact that f and σ are continuous we see that

$$\sup_{w\in J}\left|\frac{f_t(x_i - h_tw)}{\sigma^2(x_i - h_tw)} - \frac{f(x_i)}{\sigma^2(x_i)}\right| \xrightarrow{\mathrm{P}} 0$$

as $t \to \infty$. It follows that the second integral on the right hand side of (4.7) vanishes as $t \to \infty$. All together we have found that

$$\frac{h_t^{2n-1}}{t}\int_0^t A_{n,t,x_i}(X_s)A_{n,t,x_j}(X_s)\,ds \xrightarrow{\mathrm{P}}
\begin{cases} 4\frac{f(x_i)}{\sigma^2(x_i)}\int_{\mathbb{R}}\left[K^{(n-1)}(w)\right]^2 dw & , \text{ if } i = j, \\ 0 & , \text{ if } i \neq j. \end{cases}$$

The assertion of the lemma now follows from the central limit theorem for martingales. □

References

Banon, G. (1978). Nonparametric identification for diffusion processes. *SIAM J. Control Optim.* **16**, 380–395.

Gihman, I.I. and Skorohod, A.V. (1972). *Stochastic differential equations.* Springer.

Karatzas, I. and Shreve, S.E. (1991). *Brownian motion and stochastic calculus*, 2nd edition. Springer.

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

Lucas, A. (1998). Can we estimate the density's derivative with suroptimal rate? *Stat. Inference Stoch. Process.* **1**, 29–41.

Nguyen, H.T. (1979). Density estimation in a continuous-time stationary Markov process. *Ann. Statist.* **7**, 341–348.

Nishiyama, Y. (2000). *Entropy methods for martingales.* CWI Tract **128**, CWI Amsterdam.

Pham Dinh Tuan (1981). Nonparametric estimation of the drift coefficient in the diffusion equation. *Math. Operationsforsch. Statist. Ser. Statist.* **12**, 61–73.

Schuster, E.F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.* **43**, 84–88.

Skorokhod, A.V. (1989). *Asymptotic methods in the theory of stochastic differential equations.* AMS.

Van der Vaart, A.W. (1998). *Asymptotic statistics.* Cambridge University Press.

Van Zanten, J.H. (2000a). On the uniform convergence of the empirical density of an ergodic diffusion. *Stat. Inference Stoch. Process.*, to appear. See `http://www.cwi.nl/~hvz`.

Van Zanten, J.H. (2000b). Uniform convergence of curve estimators for ergodic diffusion processes. CWI report PNA-R0006, CWI Amsterdam. See `http://www.cwi.nl/~hvz`.