Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

Uniform Convergence of Curve Estimators for Ergodic Diffusion Processes

J.H. van Zanten

Probability, Networks and Algorithms (PNA)

PNA-R0006 August 31, 2000

Report PNA-R0006
ISSN 1386-3711

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

Uniform Convergence of Curve Estimators for Ergodic Diffusion Processes

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

ABSTRACT
For ergodic diffusions, we consider kernel-type estimators for the invariant density, its derivatives and the drift function. Using empirical process theory for martingales, we first prove a theorem regarding the uniform weak convergence of the empirical density. This result is then used to derive uniform weak convergence for the kernel estimator of the invariant density. For kernel estimators of the derivatives of the invariant density and for a nonparametric drift estimator that was proposed by Banon [1], we give bounds for the rate at which the uniform distance between the estimator and the true curve vanishes. We also consider the problem of estimation from discrete-time observations. In that case, obvious estimators can be constructed by replacing Lebesgue integrals by Riemann sums. We show that these approximations are also uniformly consistent, provided that the bandwidths and the time between the observations are correctly balanced.

2000 Mathematics Subject Classification: 60J60, 62G05.
Keywords and Phrases: Ergodic diffusions, empirical density, nonparametric estimation, uniform convergence.
Note: Work carried out under project PNA3.3 'Stochastic Processes and Applications'.

1 Introduction

Ergodic diffusions are the continuous-time analogues of ergodic Markov chains and they occur at various places in stochastic modeling. Well-known examples are the Wright-Fisher model in genetics, the Ornstein-Uhlenbeck process in physics and stochastic volatility models and interest rate models in mathematical finance. We consider 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. By saying that a solution X of equation (1.1) has the ergodic property with invariant measure μ we mean that the law of large numbers holds, i.e. that

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

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.

In this paper we study nonparametric kernel-type estimators for the invariant density f and its derivatives and for the drift function b. Figure 1 shows a simulated sample path of an ergodic diffusion. It is a realisation of a mean reverting process X that solves the equation $dX_t = -2(X_t - 1)\,dt + dW_t$.

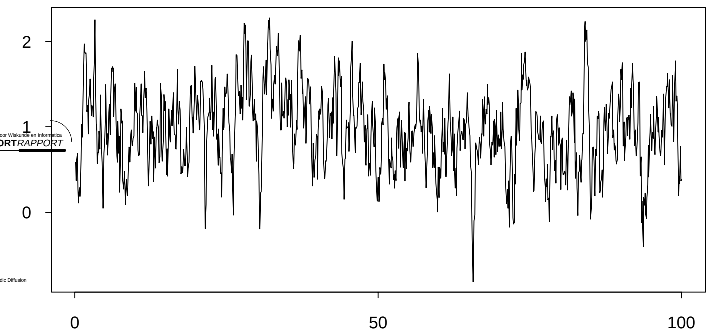

Figure 1: Simulated sample path of the solution of the stochastic differential equation $dX_t = -2(X_t - 1)\,dt + dW_t$.

In figure 2, nonparametric estimates for the invariant density f, its derivative f' and the drift function b of the process are shown, based on the observation of the sample path of figure 1.

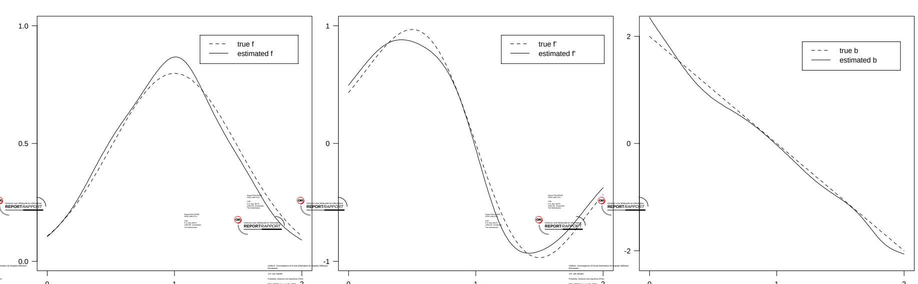

Figure 2: Nonparametric estimates for the invariant density f, its derivative f' and the drift function b. The dotted lines are the true curves.

The estimators that were used are the kernel-type estimators that we discuss in this paper. The simulations indicate that such estimators can give very decent estimates of the curves of interest. In particular, the pictures show that the estimators are not only pointwise close to the real curves, but also uniform on intervals. As a result, the estimators give a nice global picture of the real curves. The purpose of this work is to give a theoretical explanation of this fact.

Our first aim is to derive results regarding the uniform convergence of nonparametric estimators for the invariant density f and its derivatives. Based on the observation of a trajectory

$\{X_s : s \leq t\}$ of the diffusion, we can estimate f 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 for the density f and its derivative f', i.e. the cases $m = 0$ and $m = 1$, were first considered by Banon [1]. In the paper [1], conditions are given for pointwise consistency of these estimators. For $m = 0$, uniform consistency was considered in the paper [5]. Let us remark that in both papers, the authors work under the so-called condition G_2 of Rosenblatt [8]. As is shown in [1], sufficient conditions for G_2 can be formulated in terms of the functions b, σ and f, but the requirements are quite demanding. More recently, Kutoyants [4] studied the pointwise properties of the kernel estimator for the density f. He proved consistency, asymptotic normality with rate $\sqrt{t}$ and efficiency. Rather than working under Rosenblatt's condition G_2, he used the fact that for the diffusions that we consider, the empirical measure has a density.

In this paper we also exploit the properties of the empirical density f_t in order to obtain the desired results regarding uniform convergence of the kernel estimators. Our starting point is the result of [11] that states that under mild conditions, we have the convergence

$$\sup_{x \in I} |f_t(x)/f(x) - 1| \xrightarrow{\mathrm{P}} 0$$

for every compact subinterval I of the state space of the process X. Using the empirical process-type techniques of Nishiyama [6] we then derive a uniform weak convergence theorem for the normalized difference $\sqrt{t}(f_t/f - 1)$ (see theorem 3.2). With this powerful tool in hand we prove that for every $m \geq 0$

$$\sup_{x \in I} \left| \hat{f}_{t,h_t}^{(m)}(x) - f^{(m)}(x) \right| = O_P\left(\frac{1}{h_t^m \sqrt{t}}\right) \tag{1.4}$$

for every compact subinterval I of the state space of the process, provided that the bandwidths h_t converge to 0 at the right speed. For $m = 0$, the rate $t^{1/2}$ is the exact rate and we find in fact a uniform weak convergence theorem for the random map $\sqrt{t}(\hat{f}_{t,h_t} - f)$ (see theorem 3.4). For $m \geq 1$, the rate is not exact and the big O_P is actually a small o_P (see theorem 3.5). It seems that in this case, the exact rate is $(h_t^{2m-1}t)^{1/2}$, but that the convergence is not uniform at this rate. We investigate this problem in the forthcoming paper [12].

Now suppose that the diffusion coefficient σ is known and a nonparametric estimate for b is needed. Based on the relation (3.4) between the functions b, σ, f and f', Banon [1] proposed the estimator

$$\hat{b}_t = \tfrac{1}{2}\sigma^2 \frac{\hat{f}_{t,h_t}^{(1)}}{\hat{f}_{t,h_t}} + \sigma\sigma' \tag{1.5}$$

for the drift function b, provided of course that σ is differentiable. It is shown in [1] that this estimator is pointwise consistent under certain regularity conditions. By the infinite dimensional delta-method, our results for the estimators of f and f' imply that

$$\sup_{x \in I} \left| \hat{b}_t(x) - b(x) \right| = o_P \left(\frac{1}{h_t \sqrt{t}} \right) \tag{1.6}$$

for every compact subinterval I of the state space of X, provided that the bandwidths h_t converge to 0 at the right speed (see theorem 3.6).

In a real life situation, we can of course not observe a process continuously. Therefore, we also consider the case that we observe the diffusion at discrete time instants $0, \Delta, 2\Delta, \ldots, n\Delta$. In this case, nonparametric estimators can easily be constructed by replacing the Lebesgue integrals in (1.3) by the corresponding Riemann sums (see definition (3.5)). Using a maximal inequality from empirical process theory (taken from [10]), we construct an upper bound for the expected uniform distance between the continuous-time estimators (1.3) and their discrete-time approximations (see theorem 3.7). Under the assumption that $\Delta = \Delta_n \to 0$ and $n\Delta_n \to \infty$, this allows us to give results like (1.4) and (1.6) for the discrete-time estimators (see theorems 3.8 and 3.9). In particular, we show that if $\Delta_n \sim n^{-\alpha}$ for some $\alpha \in (0,1)$, then the bandwidths h_n can always be chosen in such a way that we get uniformly consistent discrete-time estimators for f and its derivatives and for b.

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 [3] 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 [2], 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. [2] or [9]) 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 $g \in L^1(\mu)$, and that $X_t \rightsquigarrow \mu$ as $t \to \infty$. 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

Let us begin with formulating two important integrability conditions that we shall use in the sequel.

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

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

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 assures the existence of the limiting covariances that we will find in the results below.

The process X is a continuous semimartingale, so its semimartingale local time $\{L_t(x) : t \geq 0, x \in (l,r)\}$ is well-defined (see for instance [3] or [7]). In particular, the empirical measure of the process has a density. Indeed, if for $t > 0$ we denote by μ_t the empirical measure

$$\mu_t(B) = \frac{1}{t}\int_0^t 1_B(X_s)\,ds,$$

then we have the relation $\mu_t(dx) = f_t(x)\,dx$, where f_t is given by

$$f_t(x) = \frac{2L_t(x)}{t\sigma^2(x)}. \tag{3.1}$$

The random function f_t is therefore called the empirical density. The ergodic property of X implies that if t grows, then the empirical density f_t behaves more and more like the invariant density f. Indeed, using (3.1), the Tanaka-Meyer formula, the ergodic property (1.2) and the law of large numbers for martingales, it is easy to prove that for every $x \in (l, r)$ we almost surely have $f_t(x)/f(x) \to 1$ as $t \to \infty$, provided that $\int |b| \, d\mu < \infty$ and $\int \sigma^2 \, d\mu < \infty$. In the paper [11] we showed that under slightly more restrictive integrability conditions, the convergence of f_t to f is in fact uniform on compact intervals. The exact formulation is as follows.

Theorem 3.1. *Suppose that C_1 holds. Then for every compact interval $I \subseteq (l, r)$ we have*

$$\sup_{x \in I} |f_t(x)/f(x) - 1| \xrightarrow{\mathrm{P}} 0$$

as $t \to \infty$.

The proof of this theorem can be found in the paper [11]. Note however that the integrability conditions of the theorem above are somewhat weaker than those needed in [11]. It is easily seen that this improvement can be achieved by reasoning slightly more careful than in section 3.1 of [11].

The following step is a convergence result for the normalized difference $\sqrt{t}(f_t/f - 1)$. It turns out that for every compact interval $I \subseteq (l, r)$, this random map converges weakly in the space $\ell^\infty(I)$ of bounded functions on I (see the book [10] for the general theory of weak convergence in such spaces). The limiting covariances are inner products of the functions λ_x given by

$$\lambda_x = 2 \frac{1_{(x,r)} - F}{\sigma f}. \tag{3.2}$$

It is easily seen that for every $x \in (l, r)$, the μ-square integrability of the function λ_x is equivalent to condition C_2.

Theorem 3.2. *Suppose that C_1 and C_2 hold and let $I \subseteq (l, r)$ be a compact interval. Then as $t \to \infty$, the random maps $\sqrt{t}(f_t/f - 1)$ converge weakly in $\ell^\infty(I)$ to a zero-mean Gaussian random map G with covariance function*

$$E\, G(x)G(y) = \langle \lambda_x, \lambda_y \rangle_{L^2(\mu)}.$$

The proof of this theorem relies on theorem 3.1 and can be found in section 4.1. An application of the continuous mapping theorem gives us the following corollary for the random maps $\sqrt{t}(f_t - f)$. Under special conditions on b and σ, a similar result was already obtained in the paper [4].

Corollary 3.3. *Suppose that C_1 and C_2 hold and let $I \subseteq (l, r)$ be a compact interval. Then as $t \to \infty$, the random maps $\sqrt{t}(f_t - f)$ converge weakly in $\ell^\infty(I)$ to a zero-mean Gaussian random map H with covariance function*

$$E\, H(x)H(y) = f(x)f(y) \langle \lambda_x, \lambda_y \rangle_{L^2(\mu)}. \tag{3.3}$$

Proof. By assumption, the function σ is continuous on (l,r). The invariant density f is therefore also continuous (see (2.3)), so $\|f\|_\infty = \sup_{x\in I}|f(x)| < \infty$. It follows in particular that the map $\phi : \ell^\infty(I) \to \ell^\infty(I)$ given by $\phi(g) = fg$ is well-defined. Clearly, $\|\phi(g_1) - \phi(g_2)\|_\infty \leq \|f\|_\infty\|g_1 - g_2\|_\infty$, so the map ϕ is continuous. By theorem 3.2 and the continuous mapping theorem we thus have

$$\sqrt{t}(f_t - f) = \phi(\sqrt{t}(f_t/f - 1)) \rightsquigarrow \phi(G) = H,$$

which completes the proof. □

The result of corollary 3.3 is interesting in itself, but is also the main ingredient in the proof of the following two results that we have obtained for the kernel estimators (1.3). The kernel K is understood to be a symmetric probability density with compact support. See section 4.2 for the proof of the theorems.

Theorem 3.4. *Suppose that C_1 and C_2 hold and let $I \subseteq (l,r)$ be a compact interval. Suppose that f is twice continuously differentiable, that the kernel K is continuous and that $h_t^2\sqrt{t} \to 0$. Then as $t \to \infty$, the random maps $\sqrt{t}(\hat{f}_{t,h_t} - f)$ converge weakly in $\ell^\infty(I)$ to a zero-mean Gaussian random map H with covariance function (3.3).*

Theorem 3.5. *Suppose that C_1 and C_2 hold and let $I \subseteq (l,r)$ be a compact interval. Suppose that f is $m+2$ times continuously differentiable, that the kernel K is m times continuously differentiable and that the bandwidths $h_t \downarrow 0$, but such that $h_t^m\sqrt{t}$ remains bounded away from 0. Then*

$$\sup_{x\in I}\left|\hat{f}^{(m)}_{t,h_t}(x) - f^{(m)}(x)\right| \xrightarrow{\mathrm{P}} 0$$

as $t \to \infty$. Moreover, if $m \geq 1$ and $h_t^{m+2}\sqrt{t} \xrightarrow{\mathrm{P}} 0$, then

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

as $t \to \infty$.

Having established these uniform convergence results for the kernel estimators for f and its derivatives, we can use the infinite dimensional delta-method to investigate functionals of the estimators. An interesting case is the estimator that Banon [1] proposed for the drift function b. Suppose that the function σ is known and continuously differentiable on (l,r). Then definitions (2.1) and (2.3) give the relation

$$b(x) = \tfrac{1}{2}\sigma^2(x)\frac{f'(x)}{f(x)} + \sigma(x)\sigma'(x) \tag{3.4}$$

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. We pick a symmetric, compactly supported, continuously differentiable probability density K, bandwidths $h_t \downarrow 0$ and we define the estimator $\hat{b}_t$ by (1.5). Using the delta-method, we obtain the following theorem (see section 4.3).

Theorem 3.6. *Suppose that C_1 and C_2 hold and let $I \subseteq (l,r)$ be a compact interval. Suppose that f is 3 times continuously differentiable, that the kernel K is 2 times continuously differentiable and that the bandwidths $h_t \downarrow 0$, but such that $h_t\sqrt{t}$ remains bounded away from 0. Then*

$$\sup_{x\in I}\left|\hat{b}_t(x) - b(x)\right| \xrightarrow{\mathrm{P}} 0$$

as $t \to \infty$. Moreover, if $h_t^3\sqrt{t} \to 0$, then

$$\sup_{x\in I}\left|\hat{b}_t(x) - b(x)\right| = o_P\left(\frac{1}{h_t\sqrt{t}}\right)$$

as $t \to \infty$.

Now suppose that we observe the diffusion X at discrete instants in time, say at times $0, \Delta, 2\Delta, \ldots, n\Delta = t$. Then obvious approximations of the kernel estimators (1.3) can be obtained by replacing the Lebesgue integrals by the corresponding Riemann-sums. For $m \geq 0$, $n \in \mathbb{N}$ and $\Delta, h > 0$ we therefore introduce the discrete-time estimators

$$\tilde{f}^{(m)}_{n,h,\Delta}(x) = \frac{1}{h^{m+1}n}\sum_{i=1}^{n} K^{(m)}\left(\frac{x - X_{(i-1)\Delta}}{h}\right). \tag{3.5}$$

In section 4.4 we prove the following estimate for the uniform distance between the continuous-time and discrete-time estimators.

Theorem 3.7. *Let a compact interval $I \subseteq (l,r)$ and $m \geq 0$ be given. Suppose that the kernel K is $m+2$ times continuously differentiable, that $\int |b|\, d\mu < \infty$ and $\int \sigma^2\, d\mu < \infty$. Then there exists a constant $C > 0$ such that*

$$E\sup_{x\in I}\left|\hat{f}^{(m)}_{n\Delta,h}(x) - \tilde{f}^{(m)}_{n,h,\Delta}(x)\right| \leq C\,\frac{\sqrt{\Delta}}{h^{m+3}}$$

for all $n \in \mathbb{N}$ and $h, \Delta > 0$ small enough.

Once we have this bound, the following two theorems follow easily from their continuous-time counter parts.

Theorem 3.8. *Suppose that C_1 and C_2 hold and let $I \subseteq (l,r)$ be a compact interval. Suppose moreover that f and K are twice continuously differentiable. Then if $n\Delta_n \to \infty$, $h_n^2\sqrt{n\Delta_n} \to 0$ and $\Delta_n\sqrt{n}/h_n^3 \to 0$ as $n \to \infty$, the random maps $\sqrt{n\Delta_n}(\tilde{f}_{n,h_n,\Delta_n} - f)$ converge weakly in $\ell^\infty(I)$ to a zero-mean Gaussian random map H with covariance function (3.3).*

Theorem 3.9. *Suppose that C_1 and C_2 hold and let $I \subseteq (l, r)$ be a compact interval. Suppose moreover that f and K are $m+2$ times continuously differentiable. Then if $n\Delta_n \to \infty$, $h_n \to 0$, $h_n^m\sqrt{n\Delta_n}$ is bounded away from zero and $\sqrt{\Delta_n}/h_n^{m+3} \to 0$ we have*

$$\sup_{x\in I}\left|\tilde{f}^{(m)}_{n,h_n,\Delta_n}(x) - f^{(m)}(x)\right| \xrightarrow{\mathrm{P}} 0$$

as $n \to \infty$. Moreover, if $m \geq 1$, $h_n^{m+2}\sqrt{n\Delta_n} \to 0$ and $\Delta_n\sqrt{n}/h_n^3 \to 0$, then

$$\sup_{x\in I}\left|\tilde{f}^{(m)}_{n,h_n,\Delta_n}(x) - f^{(m)}(x)\right| = o_P\left(\frac{1}{h_n^m\sqrt{n\Delta_n}}\right)$$

as $n \to \infty$.

Remark 3.10. It should be noted that for given Δ_n, it may be impossible to find bandwidths h_n that satisfy all the conditions of theorem 3.9. Suppose however that $\Delta_n = n^{-\alpha}$ for some $0 < \alpha < 1$ and that we want to find bandwidths h_n of the form $h_n = n^{-\beta}$, with $\beta > 0$, such that the conditions of the theorem are satisfied. In terms of α and β, the conditions for uniform consistency are $\beta < \alpha/6$ in the case $m = 0$ and $\beta < \min\{\alpha/2(m+3), 1/2m - \alpha/2m\}$ in the case $m > 0$. Note that such a β can be found for every $\alpha \in (0, 1)$. The additional conditions for the second assertion of the theorem in this case translate to $1/2(m+2)-\alpha/2(m+2) < \beta < \alpha/3-1/6$. This extra requirement can be met if and only if $(m + 5)/(2m + 7) < \alpha < 1$.

Based on the observations of X at the instants $0, \Delta_n, \ldots, n\Delta_n$, we can of course also define the estimator $\tilde{b}_n$ of b by putting

$$\tilde{b}_n(x) = \tfrac{1}{2}\sigma^2(x)\frac{\tilde{f}^{(1)}_{n,h_n,\Delta_n}(x)}{\tilde{f}_{n,h_n,\Delta_n}(x)} + \sigma(x)\sigma'(x),$$

where the h_n are suitably chosen bandwidths. As in the proof of theorem 3.6, the delta-method can be used to derive the following result from theorems 3.8 and 3.9.

Theorem 3.11. *Suppose that C_1 and C_2 hold and let $I \subseteq (l, r)$ be a compact interval. Suppose moreover that f and K are 3 times continuously differentiable. Then if $n\Delta_n \to \infty$, $h_n \to 0$, $\sqrt{\Delta_n}/h_n^4 \to 0$ and $h_n\sqrt{n\Delta_n}$ remains bounded away from 0, we have*

$$\sup_{x\in I}\left|\tilde{b}_n(x) - b(x)\right| \xrightarrow{\mathrm{P}} 0$$

as $n \to \infty$. Moreover, if $h_n^2\sqrt{n\Delta_n} \to 0$ and $\Delta_n\sqrt{n}/h_n^3 \to 0$, then

$$\sup_{x\in I}\left|\tilde{b}_t(x) - b(x)\right| = o_P\left(\frac{1}{h_n\sqrt{n\Delta_n}}\right)$$

as $t \to \infty$.

4 Proofs

4.1 Proof of theorem 3.2

We first write $f_t/f - 1$ in a convenient form. If we combine (3.1) with the Tanaka-Meyer formula for the local time $L_t(x)$ we find that for every $x \in (l, r)$

$$
\begin{aligned}
f_t(x) - f(x) = & \frac{|X_t - x| - |X_0 - x|}{t\sigma^2(x)} \\
& + \frac{1}{t\sigma^2(x)} \int_0^t \operatorname{sgn}(x - X_s)\sigma(X_s)\, dW_s \\
& + \frac{1}{t} \int_0^t \left(\frac{\operatorname{sgn}(x - X_s) b(X_s)}{\sigma^2(x)} - f(x) \right) ds.
\end{aligned} \tag{4.1}
$$

Next, we use the generalized Itô formula and definition (2.3) of the invariant density f to rewrite the Lebesgue integral in (4.1). Recall that we fixed a point $x_0 \in (l, r)$ and define

$$
\pi_x = 2\frac{1_{(x,r)} - F}{\sigma^2 f}, \quad \Pi_x(y) = \int_{x_0}^y \pi_x(z)\, dz. \tag{4.2}
$$

Note that we have the relation $\lambda_x = \sigma\pi_x$ (see definition (3.2)). From Itô's formula we get

$$
\begin{aligned}
\int_0^t \Big(\frac{\operatorname{sgn}(x - X_s) b(X_s)}{\sigma^2(x)} & - f(x) \Big)\, ds = \\
& f(x)(\Pi_x(X_t) - \Pi_x(X_0)) \\
& - \frac{|X_t - x| - |X_0 - x|}{\sigma^2(x)} \\
& - \frac{1}{\sigma^2(x)} \int_0^t \operatorname{sgn}(x - X_s)\sigma(X_s)\, dW_s \\
& - f(x) \int_0^t \lambda_x(X_s)\, dW_s.
\end{aligned} \tag{4.3}
$$

Combination of equations (4.1) and (4.3) gives

$$
\frac{f_t(x)}{f(x)} - 1 = \frac{1}{t}(\Pi_x(X_t) - \Pi_x(X_0)) - \frac{1}{t} \int_0^t \lambda_x(X_s)\, dW_s, \tag{4.4}
$$

which is the representation that we need.

We begin with considering the first term on the right hand side of (4.4). From the definitions (4.2) it is easily seen that the functions Π_x can be bounded by a function Π that does not depend on the parameter x. Indeed, we have for every $x, y \in (l, r)$

$$
|\Pi_x(y)| \le \left| \int_{x_0}^y \pi_x(z)\, dz \right| \le \left| \int_{x_0}^y 2\frac{1 - F(z)}{\sigma^2(z) f(z)}\, dz \right| + \left| \int_{x_0}^y 2\frac{F(z)}{\sigma^2(z) f(z)}\, dz \right| =: \Pi(y).
$$

Using the triangle inequality and the stationarity of the process X we then find that for every $\varepsilon > 0$

$$
\begin{aligned}
P\left(\sup_{x\in(l,r)} |\Pi_x(X_t) - \Pi_x(X_0)| > \varepsilon\sqrt{t}\right) &\le 2P\left(\sup_{x\in(l,r)} |\Pi_x(X_0)| > \varepsilon\sqrt{t}\right) \\
&\le 2P(\Pi(X_0) > \varepsilon\sqrt{t}) \to 0.
\end{aligned}
$$

This shows that

$$
\sup_{x\in(l,r)} \frac{1}{\sqrt{t}}(\Pi_x(X_t) - \Pi_x(X_0)) \xrightarrow{\mathrm{P}} 0.
$$

By Slutsky's lemma and (4.4), it thus remains to show that the random maps

$$
x \mapsto \frac{1}{\sqrt{t}} \int_0^t \lambda_x(X_s)\, dW_s \tag{4.5}
$$

converge weakly to G in $\ell^\infty(I)$. The desired finite dimensional convergence follows easily from the ergodic property (1.2) and the central limit theorem for martingales. To prove asymptotic tightness we view the random maps (4.5) as collections of endpoints of continuous martingales, so that we can apply the results of Nishiyama [6]. For every $x \in (l, r)$ and $t > 0$, define the martingale $M^{t,x}$ by

$$
M_s^{t,x} = \frac{1}{\sqrt{t}} \int_0^{st} \lambda_x(X_u)\, dW_u.
$$

Then obviously we have

$$
\frac{1}{\sqrt{t}} \int_0^t \lambda_x(X_s)\, dW_s = M_1^{t,x}
$$

for every x. An important quantity related to the family $M = \{M_1^{t,x} : x \in I\}$ is the quadratic modulus (in this case with respect to the metric $d(x, y) = \sqrt{|x - y|}$)

$$
\begin{aligned}
\|M\|_t &= \sup_{x,y\in I} \frac{\sqrt{\langle M^{t,x} - M^{t,y}\rangle_1}}{\sqrt{|x-y|}} \\
&= \sup_{x,y\in I} \frac{\sqrt{\frac{1}{t}\int_0^t (\lambda_x - \lambda_y)^2(X_s)\, ds}}{\sqrt{|x-y|}}.
\end{aligned} \tag{4.6}
$$

It follows from [6], theorem 3.4.2, that the random maps (4.5) are asymptotically tight if $\|M\|_t$ is asymptotically tight for $t \to \infty$. To prove that this is indeed the case, observe that for $x \le y$

and $x, y \in I$ we have

$$\begin{aligned}
\frac{1}{t}\int_0^t (\lambda_x - \lambda_y)^2(X_s)\,ds &= \int_I (\lambda_x - \lambda_y)^2(z) f_t(z)\,dz \\
&= \int_I (\lambda_x - \lambda_y)^2(z) \frac{f_t(z)}{f(z)} f(z)\,dz \\
&\leq \sup_{z\in I}\left|\frac{f_t(z)}{f(z)}\right| \int_{\mathbb{R}} (\lambda_x - \lambda_y)^2(z) f(z)\,dz \\
&= \sup_{z\in I}\left|\frac{f_t(z)}{f(z)}\right| 4\int_x^y \frac{1}{\sigma^2(z)f(z)}\,dz \\
&\leq 4\sup_{z\in I}\frac{1}{\sigma^2(z)f(z)} \sup_{z\in I}\left|\frac{f_t(z)}{f(z)}\right| |x-y|.
\end{aligned}$$

The function $1/(\sigma^2 f)$ is equal to a constant times the function s given by (2.1) (see definition (2.3)). In particular, it is continuous and therefore bounded on the compact interval I. We may thus conclude that there exists a constant $C > 0$ such that

$$\|M\|_t \leq C\left(\sup_{z\in I}\left|\frac{f_t(z)}{f(z)}\right|\right)^{1/2}. \tag{4.7}$$

By theorem 3.1, the quadratic modulus $\|M\|_t$ is therefore asymptotically tight and the proof is finished.

Remark 4.1. It follows from inequality (4.7) and theorem 2.4.4 of [6] that there exists a version of the random maps (4.5) that is continuous in x. So in fact, the weak convergence of these random maps to the limit G takes place in the space $C(I)$ of continuous functions on the interval I. In particular, we see that the random map G admits a continuous version. Since the invariant density f is continuous, the same holds for the limit H of corollary 3.3. We will use this little refinement in the proof of lemma 4.3 below.

4.2 Proof of theorems 3.4 and 3.5

The asymptotic bias of the kernel estimators can be treated in the same manner as the bias of a kernel estimator for the density of i.i.d. observations. We find that if the bandwidth $h = h_t$ tends to 0, then the bias also tends to 0, uniformly on compact intervals.

Lemma 4.2. *Let $m \geq 0$ be given. Suppose that f is $m+2$ times continuously differentiable and that K is m times continuously differentiable. Then for every compact interval $I \subseteq (l, r)$ we have*

$$\sup_{x\in I}\left|E\,\hat{f}_{t,h}^{(m)}(x) - f^{(m)}(x)\right| = O(h^2)$$

for $h \to 0$.

Proof. By stationarity of the process X and Fubini's theorem we have

$$E\,\hat{f}_{t,h}^{(m)}(x) = \int_l^r \frac{1}{h^{m+1}} K^{(m)}\left(\frac{x-y}{h}\right) f(y)\,dy.$$

Say that the support of K is contained in the compact interval J. Since both K and f are m times continuously differentiable, repeated partial integration and a change of variables yield

$$E\,\hat{f}_{t,h}^{(m)}(x) = \int_J K(z) f^{(m)}(x+hz)\,dz$$

for every $x \in I$ and $h > 0$ small enough. The invariant density f is assumed to be $m+2$ times continuously differentiable. So by Taylor's formula we have

$$f^{(m)}(x+hz) = f^{(m)}(x) + hzf^{(m+1)}(x) + h^2z^2\int_0^1 (1-t)f^{(m+2)}(x+thz)\,dt.$$

If we plug this in the preceding display and use the fact that K is symmetric and integrates to 1 we get

$$E\,\hat{f}_{t,h}^{(m)}(x) - f^{(m)}(x) = h^2 \int_J \int_0^1 z^2K(z)(1-t)f^{(m+2)}(x+thz)\,dtdz.$$

Since we are looking for an expansion for $h \to 0$ we can assume that the quantity $x+thz$ appearing as the argument of $f^{(m+2)}$ in the preceding display is contained in a compact subinterval of (l,r). Therefore, the quantity $|f^{(m+2)}(x+thz)|$ is bounded and we can find a constant $C > 0$ such that

$$\sup_{x\in I} |E\,\hat{f}_{t,h}^{(m)}(x) - f^{(m)}(x)| \leq Ch^2 \left|\int_{\mathbb{R}} z^2K(z)\,dz\right|$$

for all h small enough. This completes the proof of the lemma. $\square$

Now that we know that the kernel estimators are asymptotically unbiased, we consider the difference $\hat{f}_{t,h}^{(m)}(x) - E\,\hat{f}_{t,h}^{(m)}(x)$. Using theorem 3.2 we can prove the following lemma.

Lemma 4.3. *Suppose that C_1 and C_2 hold and let $I \subseteq (l,r)$ be a compact interval. Suppose that the kernel K is m times continuously differentiable, and that the bandwidths $h_t \downarrow 0$. Then as $t \to \infty$, the random maps*

$$x \mapsto h_t^m\sqrt{t}(\hat{f}_{t,h_t}^{(m)}(x) - E\,\hat{f}_{t,h_t}^{(m)}(x))$$

converge weakly to H_m in $\ell^\infty(I)$, where $H_0 = H$ is the zero-mean Gaussian random map with covariance function (3.3) and $H_m = 0$ for $m \geq 1$.

Proof. Say that the support of K is contained in the compact interval J. Then we have the relation

$$\begin{aligned} h_t^m \sqrt{t}(\hat{f}^{(m)}_{t,h_t}(x) - E\,\hat{f}^{(m)}_{t,h_t}(x)) \\ &= \int_{\mathbb{R}} \frac{1}{h_t} K^{(m)}\left(\frac{x-y}{h_t}\right) \sqrt{t}(f_t(y) - f(y))\,dy \\ &= \int_J K^{(m)}(z) P_t(x - h_t z)\,dz \end{aligned} \tag{4.8}$$

for every $x \in J$ and h_t small enough, where $P_t = \sqrt{t}(f_t - f)$. For $t > 0$ we define the function $Q_t : I \times J \to \mathbb{R}$ by $Q_t(x, z) = x - h_t z$. Since $h_t \downarrow 0$ there is a $t_0 \geq 0$ such that the function Q_t takes values in $I_1 = \{x \in (l, r) : d(x, I) \leq 1\}$ for every $t \geq t_0$. For $t \geq t_0$ we can therefore rewrite (4.8) as

$$h_t^m \sqrt{t}(\hat{f}^{(m)}_{t,h_t} - E\,\hat{f}^{(m)}_{t,h_t}) = \phi(P_t, Q_t),$$

where $\phi : \ell^\infty(I_1) \times \ell^\infty(I \times J) \to \ell^\infty(I)$ is given by

$$\phi(P, Q)(x) = \int_J K^{(m)}(z) P(Q(x, z))\,dz.$$

It is not hard to see that the map ϕ is continuous on the domain $\mathcal{D} \subseteq \ell^\infty(I_1) \times \ell^\infty(I \times J)$ of all pairs (P, Q) for which P is continuous. By corollary 3.3 we have $P_t \rightsquigarrow H$ in $\ell^\infty(I_1)$, where the covariance function of H is given by (3.3). Moreover, by remark 4.1, the random map H is continuous. Obviously the maps Q_t converge uniformly on $I \times J$ to the map $Q(x, z) = x$. By Slutsky's lemma we thus have the weak convergence $(P_t, Q_t) \rightsquigarrow (H, Q)$ in $\ell^\infty(I_1) \times \ell^\infty(I \times J)$, and we remarked that the random element (H, Q) takes values in the domain $\mathcal{D}$. Hence, it follows from the continuous mapping theorem that

$$h_t^m \sqrt{t}(\hat{f}^{(m)}_{t,h_t} - E\,\hat{f}^{(m)}_{t,h_t}) = \phi(P_t, Q_t) \rightsquigarrow \phi(H, Q) = \left(\int_J K^{(m)}(z)\,dz\right) H.$$

Since K is a probability density, the integral on the right hand side is equal to 1 for $m = 0$. For $m \geq 1$ it is equal to 0, by partial integration. $\square$

Combination of lemmas 4.2 and 4.3 yields theorems 3.4 and 3.5.

4.3 Proof of theorem 3.6

Let $\mathcal{D} \subseteq \ell^\infty(I) \times \ell^\infty(I)$ be the collection of all pairs (p, q) for which q is continuous and positive. Then the map $\phi : \mathcal{D} \to \ell^\infty(I)$, with

$$\phi(p, q) = \tfrac{1}{2}\sigma^2 \frac{p}{p} + \sigma\sigma'$$

is well-defined and easily seen to be continuous on $\mathcal{D}$. By theorem 3.5 we have

$$\left(\hat{f}^{(1)}_{t,h_t}, \hat{f}_{t,h_t}\right) \xrightarrow{\mathrm{P}} (f', f)$$

in $\ell^\infty(I) \times \ell^\infty(I)$ and the pair (f', f) is clearly and element of $\mathcal{D}$. So by the continuous mapping theorem,

$$\hat{b}_t = \phi\left(\hat{f}^{(1)}_{t,h_t}, \hat{f}_{t,h_t}\right) \xrightarrow{\mathrm{P}} \phi(f', f) = b$$

in $\ell^\infty(I)$, which proves the first part of the theorem. If $h_t^3\sqrt{t} \to 0$, then by theorems 3.4 and 3.5 and Slutsky's lemma

$$h_t\sqrt{t}\left(\hat{f}^{(1)}_{t,h_t} - f', \hat{f}_{t,h_t} - f\right) \rightsquigarrow 0$$

in $\ell^\infty(I) \times \ell^\infty(I)$. Since

$$h_t\sqrt{t}(\hat{b}_t - b) = h_t\sqrt{t}\left(\phi\left(\hat{f}^{(1)}_{t,h_t}, \hat{f}_{t,h_t}\right) - \phi(f', f)\right),$$

we can finish the proof by using the infinite dimensional delta-method (see [10], chapter 3.9). The only thing that remains to be shown is that the map ϕ is Hadamard-differentiable at the point (f, f'). To that end, write ϕ as a composition of maps

$$(p, q) \mapsto (p, 1/q) \mapsto p/q \mapsto \tfrac{1}{2}\sigma^2 p/q + \sigma\sigma'.$$

By lemma 3.9.25 of [10], the map $q \mapsto 1/q$ is differentiable on the domain of all functions that are bounded away from zero, which implies that the first map in the chain above is differentiable on $\mathcal{D}$. That the second map, multiplication of two functions, is Hadamard-differentiable is not hard to check. The last map is affine and therefore differentiable. So by the chain rule, the map ϕ is Hadamard-differentiable and the second assertion of the theorem follows by the delta-method.

4.4 Proof of theorem 3.7

We use the usual notations $D(\varepsilon, \mathcal{G}, d)$, $N(\varepsilon, \mathcal{G}, d)$ and $N_{[]}(\varepsilon, \mathcal{G}, d)$ for the packing, covering and bracketing numbers of a semimetric space $(\mathcal{G}, d)$ (see [10]).

Lemma 4.4. *Let $\mathcal{G}$ be a countable collection of functions on (l, r) that is bounded in $L^2(\mu)$. Then there exists a constant $C > 0$ such that*

$$E \sup_{g\in\mathcal{G}} \left|\int_a^b g(X_s)\, dW_s\right| \le C\, K(\mathcal{G})\, \sqrt{b-a},$$

for all $0 \le a \le b$, where

$$K(\mathcal{G}) = \sup_{g\in\mathcal{G}} \|g\|_{L^2(\mu)} + \int_0^{diam\,\mathcal{G}} D\left(\varepsilon, \mathcal{G}, \|\cdot\|_{L^2(\mu)}\right)^{1/2} d\varepsilon. \tag{4.9}$$

Proof. Let the random map Z on $\mathcal{G}$ be defined by

$$Z(g) = \frac{1}{\sqrt{b-a}} \int_a^b g(X_s)\, dW_s, \quad g \in \mathcal{G}.$$

If we fix some $g_0 \in \mathcal{G}$, then clearly

$$E \sup_{g\in\mathcal{G}} \left| \int_a^b g(X_s)\, dW_s \right| \leq \left(E\,|Z(g_0)| + E \sup_{g,h\in\mathcal{G}} |Z(g) - Z(h)| \right) \sqrt{b-a}. \tag{4.10}$$

For the first expectation we have $E\,|Z(g_0)| \leq (E\,|Z(g_0)|)^{1/2} = \|g_0\|_{L^2(\mu)} \leq \sup_g \|g\|_{L^2(\mu)}$. To bound the second expectation, we apply corollary 2.2.5 of [10]. Clearly, we have

$$\|Z(g) - Z(h)\|_{L^2(P)} = \|g - h\|_{L^2(\mu)}$$

for every $g, h \in L^2(\mu)$. Hence, by the cited result of [10], we have

$$\left\| \sup_{g,h\in\mathcal{G}} |Z(g) - Z(h)| \right\|_{L^2(P)} \leq B \int_0^{\operatorname{diam}\mathcal{G}} D\left(\varepsilon, \mathcal{G}, \|\cdot\|_{L^2(\mu)}\right)^{1/2} d\varepsilon$$

for some constant $B > 0$. $\square$

We can now give a bound for the uniform difference between Lebesgue integrals and their approximations. The infinitesimal operator of the diffusion X is denoted by A, i.e. for every twice continuously differentiable function g on (l, r) we put $Ag = bg' + (1/2)\sigma^2 g''$.

Lemma 4.5. *Let $\mathcal{G}$ be a countable collection of twice continuously differentiable functions on (l, r) such that $\sup_{g\in\mathcal{G}} |Ag| \leq G$ for some function $G \in L^1(\mu)$. Then there exists a constant $C > 0$ such that*

$$E \sup_{g\in\mathcal{G}} \left| \frac{1}{n\Delta} \int_0^{n\Delta} g(X_s)\, ds - \frac{1}{n} \sum_{i=1}^n g(X_{(i-1)\Delta}) \right| \leq C \left(\|G\|_{L^1(\mu)}\Delta + K(\sigma\mathcal{G}')\sqrt{\Delta} \right)$$

for every $\Delta > 0$ and $n \in \mathbb{N}$, where $\sigma\mathcal{G}' = \{\sigma g' : g \in \mathcal{G}\}$ and the constant $K(\sigma\mathcal{G}')$ is defined by (4.9).

Proof. By Itô's formula we have $dg(X_t) = (Ag)(X_t)\, dt + (\sigma g')(X_t)\, dW_t$. It follows that for every $a \leq b$

$$E \sup_g |g(X_b) - g(X_a)| \leq E \sup_g \left| \int_a^b (Ag)(X_s)\, ds \right| + E \sup_g \left| \int_a^b (\sigma g')(X_s)\, dW_s \right|.$$

The first expectation on the right hand side is clearly bounded by the quantity $(b - a)\|G\|_{L^1(\mu)}$. By the preceding lemma, the second one can be bounded by a constant times $K(\sigma\mathcal{G}')\sqrt{b-a}$.

Hence,

$$
\begin{aligned}
E \sup_{g\in\mathcal{G}} &\left| \frac{1}{n\Delta} \int_0^{n\Delta} g(X_s)\, ds - \frac{1}{n}\sum_{i=1}^n g(X_{(i-1)\Delta}) \right| \\
&= E \sup_g \left| \frac{1}{\Delta n} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \big(g(X_s) - g(X_{(i-1)\Delta})\big)\, ds \right| \\
&\le \frac{1}{\Delta n} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} E \sup_g \big|g(X_s) - g(X_{(i-1)\Delta})\big|\, ds \\
&\le c\,\|G\|_{L^1(\mu)}\Delta + d\,K(\sigma\mathcal{G}', p)\sqrt{\Delta}
\end{aligned}
$$

for some constants $c, d > 0$, which proves the assertion of the lemma. □

To proof theorem 3.7 we apply lemma 4.5. For every $x \in I$, define the function g_x by

$$
g_x(y) = \frac{1}{h^{m+1}} K^{(m)}\left(\frac{x-y}{h}\right).
$$

Let $I^* \subseteq I$ be a countable, dense subset and put $\mathcal{G} = \{g_x : x \in I^*\}$. The quantity that we have to bound is then equal to

$$
E \sup_{g\in\mathcal{G}} \left| \frac{1}{n\Delta} \int_0^{n\Delta} g(X_s)\, ds - \frac{1}{n}\sum_{i=1}^n g(X_{(i-1)\Delta}) \right|.
$$

It is easy to see that for every $x \in I$

$$
|Ag_x| \le \|K^{(m+1)}\|_\infty \frac{1}{h^{m+2}} b + \tfrac{1}{2}\|K^{(m+2)}\|_\infty \frac{1}{h^{m+3}}\sigma^2 =: G
$$

and that

$$
\|G\|_{L^1(\mu)} \le B\,\frac{1}{h^{m+3}} \tag{4.11}
$$

for some constant $B > 0$ and every h small enough. Another calculation shows that there exists a constant $C > 0$ such that

$$
\|\sigma g_x'\|_{L^2(\mu)} \le C\,\frac{1}{h^{m+3/2}}\|K^{(m+1)}\|_{L^2(\mathbb{R})}
$$

for all $x \in I$, so $\sigma\mathcal{G}'$ is bounded in $L^2(\mu)$. To bound the packing integral in $K(\sigma\mathcal{G}')$ we note that the class $\sigma\mathcal{G}'$ is pointwise Lipschitz in the following sense:

$$
\begin{aligned}
|\sigma g_x'(z) - \sigma g_y'(z)| &= \sigma(z)\frac{1}{h^{m+3}} \left| \int_x^y K^{(m+2)}\left(\frac{t-z}{h}\right) dt \right| \\
&\le \|K^{(m+2)}\|_\infty \frac{1}{h^{m+3}} |x-y|\sigma(z).
\end{aligned}
$$

It follows from theorem 2.7.4 of [10] that with $D = 2\|K^{(m+2)}\|_\infty \|\sigma\|_{L^2(\mu)}$ and $l = \mathrm{diam}(I)$

$$D(\varepsilon, \sigma\mathcal{G}', \|\cdot\|_{L^2(\mu)}) \leq N_{[\,]}(\varepsilon, \sigma\mathcal{G}', \|\cdot\|_{L^2(\mu)}) \leq N(h^{m+3}\varepsilon/D, I, |\cdot|) \leq \frac{D\,l}{h^{m+3}\varepsilon}.$$

Hence, we find that there exists a constant $E > 0$ such that

$$K(\sigma\mathcal{G}') \leq E\,\frac{1}{h^{m+3/2}} \tag{4.12}$$

for all h small enough. By inequalities (4.11) and (4.12) and lemma 4.5 we thus have for Δ and h small enough

$$\begin{aligned} E \sup_{x\in I}\left|\hat{f}^{(m)}_{n\Delta,h}(x) - \tilde{f}^{(m)}_{n,h,\Delta}(x)\right| &= E \sup_{g\in\mathcal{G}}\left|\frac{1}{n\Delta}\int_0^{n\Delta} g(X_s)\,ds - \frac{1}{n}\sum_{i=1}^n g(X_{(i-1)\Delta})\right| \\ &\leq C\,\left(\|G\|_{L^1(\mu)}\Delta + K(\sigma\mathcal{G}')\sqrt{\Delta}\right) \\ &\leq D\,\frac{\sqrt{\Delta}}{h^{m+3}}. \end{aligned}$$

This concludes the proof of theorem 3.7.

Acknowledgement

Thanks to Rob van der Horst for generating the pictures.

References

[1] Banon, G. (1978). Nonparametric identification for diffusion processes. *SIAM J. Control Optim.* **16**(3), 380–395.

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

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

[4] Kutoyants, Yu.A. (1998). Efficient density estimation for ergodic diffusion processes. *Stat. Inference Stoch. Process.* **1**(2), 131–155.

[5] Nguyen, H.T. (1979). Density estimation in a continuous-time stationary Markov process. *Ann. Statist.* **7**(2), 341–348.

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

[7] Revuz, D. and Yor, M. (1991). *Continuous martingales and Brownian motion.* Springer.

[8] Rosenblatt, M. (1970). Density estimates and Markov sequences, In *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)*, 199–213. Cambridge Univ. Press.

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

[10] Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes with applications to statistics.* Springer.

[11] Van Zanten, J.H. (2000). On the uniform convergence of the empirical density of an ergodic diffusion. *Stat. Inference Stoch. Process.*, to appear.

[12] Van Zanten, J.H. (2000). Rates of convergence and asymptotic normality of curve estimators for ergodic diffusion processes. In preparation.