Applied Mathematics, 2015, 6, 1592-1610
Published Online August 2015 in SciRes. http://www.scirp.org/journal/am
http://dx.doi.org/10.4236/am.2015.69142

From Nonparametric Density Estimation to Parametric Estimation of Multidimensional Diffusion Processes

Julien Apala N'drin[1], Ouagnina Hili[2]

[1]Laboratory of Applied Mathematics and Computer Science, University Felix Houphouët Boigny, Abidjan, Côte d'Ivoire
[2]Laboratory of Mathematics and New Technologies of Information, National Polytechnique Institute Houphouët-Boigny of Yamoussoukro, Yamoussoukro, Côte d'Ivoire
Email: lecorrige@yahoo.fr, o_hili@yahoo.fr

Received 17 June 2015; accepted 18 August 2015; published 21 August 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

The paper deals with the estimation of parameters of multidimensional diffusion processes that are discretely observed. We construct estimator of the parameters based on the minimum Hellinger distance method. This method is based on the minimization of the Hellinger distance between the density of the invariant distribution of the diffusion process and a nonparametric estimator of this density. We give conditions which ensure the existence of an invariant measure that admits density with respect to the Lebesgue measure and the strong mixing property with exponential rate for the Markov process. Under this condition, we define an estimator of the density based on kernel function and study his properties (almost sure convergence and asymptotic normality). After, using the estimator of the density, we construct the minimum Hellinger distance estimator of the parameters of the diffusion process and establish the almost sure convergence and the asymptotic normality of this estimator. To illustrate the properties of the estimator of the parameters, we apply the method to two examples of multidimensional diffusion processes.

Keywords

Hellinger Distance Estimation, Multidimensional Diffusion Processes, Strong Mixing Process, Consistence, Asymptotic Normality

1. Introduction

Diffusion processes are widely used for modeling purposes in various fields, especially in finance. Many papers

How to cite this paper: N'drin, J.A. and Hili, O. (2015) From Nonparametric Density Estimation to Parametric Estimation of Multidimensional Diffusion Processes. *Applied Mathematics*, **6**, 1592-1610. http://dx.doi.org/10.4236/am.2015.69142

are devoted to the parameter estimation of the drift and diffusion coefficients of diffusion processes by discrete observation. As a diffusion process is Markovian, the maximum likelihood estimation is the natural choice for parameter estimation to get consistent and asymptotical normally estimator when the transition probability density is known [1]. However, in the discrete case, for most diffusion processes, the transition probability density is difficult to calculate explicitly which prevents the use of this method. To solve this problem, several methods have been developed such as the approximation of the likelihood function [2] [3], the approximation of the transition density [4], schemes of approximation of the diffusion [5] or methods based on martingale estimating functions [6].

In this paper, we study the multidimensional diffusion model

$$\mathrm{d}X_t = a(X_t,\theta)\mathrm{d}t + b(X_t,\theta)\mathrm{d}W_t,\ t \ge 0$$

under the condition that X_t is positive recurrent and exponentially strong mixing. We assume that the diffusion process is observed at regular spaced times $t_k = k\Delta$ where Δ is a positive constant. Using the density of the invariant distribution of the diffusion, we construct an estimator of θ based on minimum Hellinger distance method.

Let f_θ denote the density of the invariant distribution of the diffusion. The estimator of θ is that value (or values) $\hat{\theta}_n$ in the parameter space Θ which minimizes the Hellinger distance between f_θ and $\hat{f}_n$, where $\hat{f}_n$ is a nonparametric density estimator of f_θ.

The interest for this method of parametric estimation is that the minimum Hellinger distance estimation method gives efficient and robust estimators [7]. The minimum Hellinger distance estimators have been used in parameter estimation for independent observations [7], for nonlinear time series models [8] and recently for univariate diffusion processes [9].

The paper is organized as follows. In Section 2, we present the statistical model and some conditions which imply that X_t is positive recurrent and exponentially strong mixing. Consistence and asymptotic normality of the kernel estimator of the density of the invariant distribution are studied in the same section. Section 3 defines the minimum Hellinger distance estimator of θ and studies its properties (consistence and asymptotic normality). Section 4 is devoted to some examples and simulations. Proofs of some results are presented in Appendix.

2. Nonparametric Density Estimation

We consider the *d*-dimensional diffusion process solution of the multivariate stochastic differential equation:

$$\mathrm{d}X_t = a(X_t,\theta)\mathrm{d}t + b(X_t,\theta)\mathrm{d}W_t,\ t \ge 0, \tag{1}$$

where $\{W_t\}_{t\ge 0}$ is a standard *l*-dimensional Wiener process, θ is an unknown parameter which varies in a compact subset Θ of $\mathbb{R}^s$, $a:\mathbb{R}^d\times\Theta\to\mathbb{R}^d$ is the drift coefficient and $b:\mathbb{R}^d\times\Theta\to\mathbb{R}^d\times\mathbb{R}^l$ is the diffusion coefficient.

We assume that the functions *a* and *b* are known up to the parameter θ and *b* is bounded.

We denote by θ_0 the unknown true value of the parameter.

For a matrix A, the notation A^t denote the transpose of the matrix *A*. We will use the notation $|.|$ to denote a vectorial norm or a matricial norm.

The process X_t is observed at discrete time $t_k = k\Delta$ where Δ is a positive constant.

We make the following assumptions on the model:

(A_1): there exists a constant *C* such that

$$|a(x,\theta)-a(y,\theta)| + |b(x,\theta)-b(y,\theta)| \le C|x-y|$$

(A_2): there exist constants $M_0 > 0$ and $r > 0$ such that

$$\langle a(x,\theta),x\rangle \le -r|x|,\ |x| \ge M_0 \quad \text{where } \langle .,.\rangle \text{ denotes the scalar product in } \mathbb{R}^d$$

(A_3): the matrix function $b(x,\theta)$ is non degenerate, that is

$$\inf_x \inf_{|\lambda|=1} \lambda^t b(x,\theta) b(x,\theta)^t \lambda > 0,\ \lambda \in \mathbb{R}^d.$$

Assumptions (A_1)-(A_3) ensure the existence of a unique strong solution for the Equation (1) and an invariant measure for the process $\{X_t\}$ that admits a density with respect to the Lebesgue measure and the strong mixing property for $\{X_t\}$ with exponential rate [10]-[12]. We denote by α the strong mixing coefficient.

In the sequel, we assume that the initial value X_0 follows the invariant law; which implies that the process $\{X_t\}$ is strictly stationary.

We consider the kernel estimator $\hat{f}_n(x)$ of $f_\theta(x)$ that is,

$$\hat{f}_n(x) = \frac{1}{nb_n^d}\sum_{k=1}^{n} K\left(\frac{x - X_k}{b_n}\right), \quad x \in \mathbb{R}^d$$

where (b_n) is a sequence of bandwidths such that $b_n \to 0$ and $nb_n^d \to +\infty$ as $n \to +\infty$ and $K : \mathbb{R}^d \to \mathbb{R}^+$ is a non negative kernel function which satisfies the following assumptions:

(A_4)

(1) There exists $N_1 > 0$ such that $K(.) \le N_1 < +\infty$,

(2) $\int K(x)\mathrm{d}x = 1$ and $|x|^d K(x) \to 0$ as $|x| \to \infty$,

(A_5) $\int u_i K(u)\mathrm{d}u = 0$ and $\int u_i^2 K(u)\mathrm{d}u < \infty$ for $i = 1, \cdots, d$.

We finish with assumptions concerning the density of the invariant distribution:

(A_6) $f_\theta(.)$ is twice continuously differentiable with respect to x.

(A_7) $\theta_1 \ne \theta_2$ implies that $f_{\theta_1}(x) \ne f_{\theta_2}(x)$ for all $x \in \mathbb{R}^d$.

Properties (consistence and asymptotic normality) of the kernel density estimator are examined in the following theorems. The proof of the two theorems can be found in the Appendix.

Theorem 1. *Under assumptions* (A_1)-(A_4), *if the function* $f_\theta(x)$ *is continuous with respect to x for all* $\theta \in \Theta$, *then for any positive sequence* (b_n) *such that* $b_n \to 0$ *and* $nb_n^d \to +\infty$ *as* $n \to +\infty$, $\hat{f}_n(x) \to f_\theta(x)$ *almost surely*.

Theorem 2. *Under assumptions* (A_1)-(A_6), *if* (b_n) *is such that* $nb_n^{d+4} \to 0$ *as* $n \to +\infty$ *then the limiting distribution of* $\sqrt{nb_n^d}\left(\hat{f}_n(x) - f_\theta(x)\right)$ *is* $N\left(0, \tau^2(x)\right)$ *where*

$$\tau^2(x) = f_\theta(x)\int_{\mathbb{R}^d} K^2(u)\mathrm{d}u.$$

3. Estimation of the Parameter

The minimum Hellinger distance estimator of θ is defined by:

$$\hat{\theta}_n = \operatorname{Arg}\min_{\theta\in\Theta} H_2\left(\hat{f}_n, f_\theta\right)$$

where

$$H_2\left(\hat{f}_n, f_\theta\right) = \left\{\int_{\mathbb{R}^d}\left|\hat{f}_n^{1/2}(x) - f_\theta^{1/2}(x)\right|^2 \mathrm{d}x\right\}^{1/2}.$$

Let $\mathcal{G}$ denote the set of squared integrable functions with respect to the Lebesgue measure on $\mathbb{R}^d$.

Define the functional $T : \mathcal{G} \to \Theta$ as follows: let $g \in \mathcal{G}$ and denote:

$$A(g) = \left\{\theta \in \Theta : H_2(f_\theta, g) = \min_{\gamma\in\Theta} H_2(f_\gamma, g)\right\}$$

where H_2 is the Hellinger distance.

If $A(g)$ is reduced to an unique element, then $T(g)$ is defined as the value of this element. Elsewhere, we choose an arbitrary but unique element of $A(g)$ and call it $T(g)$.

Theorem 3. (*almost sure consistency*)

Assume that assumptions (A_1)-(A_4) and (A_7) hold. If for all $x \in \mathbb{R}^d$, $f_\theta(x)$ is continuous at θ_0, then for any positive sequence (b_n) such that $b_n \to 0$ and $nb_n^d \to +\infty$, $\hat{\theta}_n$ converges almost surely to θ_0 as $n \to +\infty$.

Proof. By Theorem 1, $\hat{f}_n(x) \to f_{\theta_0}(x)$ almost surely.

Using the inequality $\left(a^{1/2}-b^{1/2}\right)^2 \le |a-b|$ for $a,b \ge 0$, we get

$$H_2^2\left(\hat{f}_n, f_\theta\right) = \int_{\mathbb{R}^d} \left|\hat{f}_n^{1/2}(x) - f_\theta^{1/2}(x)\right|^2 \mathrm{d}x \le \int_{\mathbb{R}^d} \left|\hat{f}_n(x) - f_\theta(x)\right| \mathrm{d}x.$$

Since

$$\int_{\mathbb{R}^d} \hat{f}_n(x)\mathrm{d}x = \int_{\mathbb{R}^d} f_\theta(x)\mathrm{d}x = 1,$$

$H_2^2\left(\hat{f}_n, f_{\theta_0}\right) \to 0$ almost surely [13] [14].

By theorem 1 [7], $T\left(f_{\theta_0}\right) = \theta_0$ uniquely on Θ; then the functional T is continuous at f_{θ_0} in the Hellinger topology. Therefore $\hat{\theta}_n = T\left(\hat{f}_n\right) \to T\left(f_{\theta_0}\right) = \theta_0$ almost surely.

This achieves the proof of the theorem.

Denote

$$g_\theta = f_\theta^{1/2},\ \dot{g}_\theta = \frac{\partial g_\theta}{\partial \theta},\ \ddot{g}_\theta = \frac{\partial^2 g_\theta}{\partial \theta^t \partial \theta}$$

when these quantities exist. Furthermore, let

$$V_\theta(x) = \left\{\int_{\mathbb{R}^d} \dot{g}_\theta(x)\dot{g}_\theta^t(x)\mathrm{d}x\right\}^{-1} \dot{g}_\theta(x) \text{ and } h_\theta(x) = \frac{\dot{g}_\theta(x)}{2 f_\theta^{1/2}(x)}.$$

To prove asymptotic normality of the estimator of the parameter, we begin with two lemmas.

Lemma 1. *Let* E_n *be a subset of* $\mathbb{R}^d$ *and denote* E_n^c *the complementary set of* E_n. *Assume that*

(1) *assumptions* (A_1)-(A_5) *are satisfied,*

(2) $h_{\theta_0}(.)$ *is twice continuously differentiable with respect to* x *and*

$$\sqrt{n}b_n^2 \int_{E_n} f_{\theta_0}(y)\mathrm{d}y \to 0 \text{ and } \sqrt{n}b_n^2 \int_{E_n} \left|\frac{\partial^2 h_{\theta_0}(y)}{\partial y_i^2}\right| f_{\theta_0}(y)\mathrm{d}y \to 0 \text{ for } i = 1,\cdots,d,$$

(3) $\sqrt{n}\int_{E_n^c}\left(\int_{\mathbb{R}^d} \left|h_{\theta_0}(y+ub_n)\right| K(u)\mathrm{d}u\right) f_{\theta_0}(y)\mathrm{d}y \to 0$ *and* $\sqrt{n}\int_{E_n^c} \left|\dot{g}_{\theta_0}(y)\right| f_{\theta_0}^{1/2}(y)\mathrm{d}y \to 0,$

(4) $\mathbb{E}\left|h_{\theta_0}^{2+\delta}(X_1)\right| < \infty$ *for some* $\delta > 0,$

then for any positive sequence (b_n) *such that* $b_n \to 0$, *the limiting distribution of*

$$\int_{\mathbb{R}^d} \sqrt{n} h_{\theta_0}(x)\hat{f}_n(x)\mathrm{d}x \text{ is } N(0,\Gamma) \text{ where } \Gamma = \frac{1}{4}\int \dot{g}_{\theta_0}(x)\dot{g}_{\theta_0}^t(x)\mathrm{d}x.$$

The proof can be found in the Appendix.

Remark 1. *The two dimensional stochastic process* (*see Section* 4) *with invariant density*

$$f_\theta(x,y) = \frac{\sqrt{\beta\sigma}}{\pi}\exp\left(-\beta x^2 - \sigma y^2\right),\ \beta > 0,\ \sigma > 0$$

where $\theta = (\beta,\sigma)$, *satisfies the conditions of Lemma* 1 *with for example* $E_n = [w_n;+\infty[\times[w_n;+\infty[$ *a subset of* $\mathbb{R}^2$ *where* $w_n = n$.

Lemma 2. *Let* G_n *be a compact set of* $\mathbb{R}^d$ *and denote by* G_n^c *the complementary set of* G_n. *Suppose that assumptions* (A_1)-(A_6) *are satisfied and*:

(1) $\frac{1}{n^{1/2}b_n^d}\int_{G_n} \left|h_{\theta_0}(x)\right| f_{\theta_0}^{-1}(x)\left(\int_{\mathbb{R}^d} K^2(t) f_{\theta_0}(x-tb_n)\mathrm{d}t\right)\mathrm{d}x \to 0$

(2) p, q *and* r *are such that* $\frac{1}{p}+\frac{1}{q}+\frac{1}{r} = 1$ *and*

$$\frac{1}{n^{3/2}b_n^{d+d/p}}\int_{G_n}\left|h_{\theta_0}(x)\right| f_{\theta_0}^{-1}(x)\left(\left[\int_{\mathbb{R}^d}K^2(t)\,f_{\theta_0}(x-tb_n)\,\mathrm{d}t\right]^{1/q+1/r}\right)\mathrm{d}x\to 0$$

(3) $n^{1/2}b_n^4\int_{G_n}\left|h_{\theta_0}(x)\right| f_{\theta_0}^{-1}(x)\left(1+\left(\frac{\partial^2 f_{\theta_0}(x)}{\partial x_i^2}\right)^2\right)\mathrm{d}x\to 0$ for $i=1,\cdots,d$

(4) $\sqrt{n}\int_{G_n^c}\left|\dot{g}_{\theta_0}(x)\right| f_{\theta_0}^{1/2}(x)\,\mathrm{d}x\to 0$

(5) $\sqrt{n}\int_{G_n^c}\left|h_{\theta_0}(x)\right|\left(\int_{\mathbb{R}^d}K(u)\,f_{\theta_0}(x+ub_n)\,\mathrm{d}u\right)\mathrm{d}x\to 0$

then

$$R_n=\int_{\mathbb{R}^d}\sqrt{n}h_{\theta_0}(x)\left(\hat{f}_n^{1/2}(x)-f_{\theta_0}^{1/2}(x)\right)^2\mathrm{d}x\to 0 \text{ in probability as } n\to+\infty.$$

The proof can be found in the Appendix.

Remark 2. *Let* $G_n=[-v_n;v_n]\times[-v_n;v_n]$ *a compact set of* $\mathbb{R}^2$ *where* $\{v_n,n\geq 1\}$ *is a sequence of positive numbers diverging to infinity. Let* $b_n=\frac{(\log(n))^{1/2}}{n^r}$, $\frac{1}{8}<r<\frac{1}{4}$ *and* $v_n=(\log(n))^q$, $\frac{1}{2}<q<1$, *then the two dimensional stochastic process with invariant density* $f_\theta(x,y)=\frac{\sqrt{\beta\sigma}}{\pi}\exp\left(-\beta x^2-\sigma y^2\right)$, $\beta>0$, $\sigma>0$ *where* $\theta=(\beta,\sigma)$, *satisfies the conditions of Lemma* 2.

Theorem 4. (*asymptotic normality*)

Under assumption (A_7) *and conditions of Lemma* 1 *and Lemma* 2, *if*

(1) *for all* $x\in\mathbb{R}^d$, g_θ *is twice continuously differentiable at* θ_0,

(2) *the components of* $\dot{g}_{\theta_0}$ *and* $\ddot{g}_{\theta_0}$ *belong to* L_2 *and if the norms of these components are continuous functions at* θ_0,

(3) θ_0 *is in the interior of* Θ *and* $\int\ddot{g}_{\theta_0}(x)g_{\theta_0}(x)\mathrm{d}x$ *is a non-singular matrix, then the limiting distribution of* $\sqrt{n}\left[\hat{\theta}_n-\theta_0\right]$ *is* $N(0,\lambda^2)$ *where*

$$\lambda^2=\frac{1}{4}\left\{\int_{\mathbb{R}^d}\dot{g}_{\theta_0}(x)\dot{g}_{\theta_0}^t(x)\mathrm{d}x\right\}^{-1}.$$

Proof. From Theorem 2 [7], we have:

$$\begin{aligned}\sqrt{n}\left[\hat{\theta}_n-\theta_0\right]&=\sqrt{n}\left\{\int_{\mathbb{R}^d}V_{\theta_0}(x)\left[\hat{f}_n^{1/2}(x)-f_{\theta_0}^{1/2}(x)\right]\mathrm{d}x\right\}\\&\quad+\sqrt{n}\left\{A_n\int_{\mathbb{R}^d}\dot{g}_{\theta_0}(x)\left[\hat{f}_n^{1/2}(x)-f_{\theta_0}^{1/2}(x)\right]\mathrm{d}x\right\}\\&=\left\{\int_{\mathbb{R}^d}\dot{g}_{\theta_0}(x)\dot{g}_{\theta_0}^t(x)\mathrm{d}x\right\}^{-1}\int_{\mathbb{R}^d}\sqrt{n}\dot{g}_{\theta_0}(x)\left[\hat{f}_n^{1/2}(x)-f_{\theta_0}^{1/2}(x)\right]\mathrm{d}x\\&\quad+A_n\int_{\mathbb{R}^d}\sqrt{n}\dot{g}_{\theta_0}(x)\left[\hat{f}_n^{1/2}(x)-f_{\theta_0}^{1/2}(x)\right]\mathrm{d}x\end{aligned}$$

where A_n is a $(m\times m)$ matrix which tends to 0 as $n\to+\infty$.

We have

$$\hat{f}_n^{1/2}(x)-f_{\theta_0}^{1/2}(x)=\frac{\hat{f}_n(x)-f_{\theta_0}(x)}{2f_{\theta_0}^{1/2}(x)}-\frac{\left(\hat{f}_n^{1/2}(x)-f_{\theta_0}^{1/2}(x)\right)^2}{2f_{\theta_0}^{1/2}(x)}.$$

Denote

$$D_n=\int_{\mathbb{R}^d}\sqrt{n}\dot{g}_{\theta_0}(x)\left[\hat{f}_n^{1/2}(x)-f_{\theta_0}^{1/2}(x)\right]\mathrm{d}x.$$

We have

$$D_n = \int_{\mathbb{R}^d} \sqrt{n}\dot{g}_{\theta_0}(x)\frac{\hat{f}_n(x) - f_{\theta_0}(x)}{2f_{\theta_0}^{1/2}(x)}\mathrm{d}x - \int_{\mathbb{R}^d}\sqrt{n}\dot{g}_{\theta_0}(x)\frac{\left(\hat{f}_n^{1/2}(x) - f_{\theta_0}^{1/2}(x)\right)^2}{2f_{\theta_0}^{1/2}(x)}\mathrm{d}x$$

$$= \int_{\mathbb{R}^d}\sqrt{n}\frac{\dot{g}_{\theta_0}(x)}{2f_{\theta_0}^{1/2}(x)}\hat{f}_n(x)\mathrm{d}x - \frac{1}{2}\int_{\mathbb{R}^d}\sqrt{n}\dot{g}_{\theta_0}(x)f_{\theta_0}^{1/2}(x)\mathrm{d}x - R_n$$

$$= \int_{\mathbb{R}^d}\sqrt{n}h_{\theta_0}(x)\hat{f}_n(x)\mathrm{d}x - R_n$$

where

$$R_n = \int_{\mathbb{R}^d}\sqrt{n}h_{\theta_0}(x)\left(\hat{f}_n^{1/2}(x) - f_{\theta_0}^{1/2}(x)\right)^2\mathrm{d}x.$$

By Lemma 2, $R_n \to 0$ in probability as $\to \infty$; then, the limiting distribution of $\sqrt{n}\left[\hat{\theta}_n - \theta_0\right]$ is reduced to that of

$$\left\{\int_{\mathbb{R}^d}\dot{g}_{\theta_0}(x)\dot{g}_{\theta_0}^t(x)\mathrm{d}x\right\}^{-1}\int_{\mathbb{R}^d}\sqrt{n}h_{\theta_0}(x)\hat{f}_n(x)\mathrm{d}x$$

since $A_n \to 0$. But

$$\int_{\mathbb{R}^d}\sqrt{n}h_{\theta_0}(x)\hat{f}_n(x)\mathrm{d}x \xrightarrow{\mathcal{L}} N(0,\Gamma) \text{ with } \Gamma = \frac{1}{4}\int_{\mathbb{R}^d}\dot{g}_{\theta_0}(x)\dot{g}_{\theta_0}^t(x)\mathrm{d}x \text{ from lemma 1.}$$

Therefore the limiting distribution of

$$\left\{\int_{\mathbb{R}^d}\dot{g}_{\theta_0}(x)\dot{g}_{\theta_0}^t(x)\mathrm{d}x\right\}^{-1}\int_{\mathbb{R}^d}\sqrt{n}h_{\theta_0}(x)\hat{f}_n(x)\mathrm{d}x \text{ is } N(0,\lambda^2)$$

where

$$\lambda^2 = \left\{\int_{\mathbb{R}^d}\dot{g}_{\theta_0}(x)\dot{g}_{\theta_0}^t(x)\mathrm{d}x\right\}^{-2}\frac{1}{4}\int_{\mathbb{R}^d}\dot{g}_{\theta_0}(x)\dot{g}_{\theta_0}^t(x)\mathrm{d}x = \frac{1}{4}\left\{\int_{\mathbb{R}^d}\dot{g}_{\theta_0}(x)\dot{g}_{\theta_0}^t(x)\mathrm{d}x\right\}^{-1}.$$

This completes the proof of the theorem.

4. Examples and Simulations

4.1. Example 1

We consider the two-dimensional Ornstein-Uhlenbeck process solution of the stochastic differential equation

$$\mathrm{d}Z_t = AZ_t\mathrm{d}t + \mathrm{d}W_t,\ Z_0 = z_0 \tag{2}$$

where

$$A = \begin{pmatrix} -\beta & 0 \\ 0 & -\sigma \end{pmatrix},\ \beta > 0,\ \sigma > 0$$

Let $Z = (X,Y)$ and $z = (x,y)$, we have:

$$a(z,\theta) = \begin{pmatrix} -\beta & 0 \\ 0 & -\sigma \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix},\ b(z,\theta) = b = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \theta = (\beta,\sigma)$$

- $a(z,\theta)$ and $b(z,\theta)$ satisfy assumptions (A_1)-(A_3). Therefore, Z_t is exponentially strong mixing and the invariant distribution μ_θ admits a density f_θ with respect to the Lebesgue measure.

Furthermore [15], $\mu_\theta = N(0,\Gamma)$, the Gaussian distribution on $\mathbb{R}^2$ with Γ the unique symmetric solution of the equation is

$$C + A\Gamma + \Gamma A^t = 0 \text{ where } C = bb^t. \tag{3}$$

The solution of the Equation (3) is $\Gamma = \begin{pmatrix} \frac{1}{2\beta} & 0 \\ 0 & \frac{1}{2\sigma} \end{pmatrix}$.

Therefore [16], the density of the invariant distribution is

$$f_\theta(x,y) = \frac{\sqrt{\beta\sigma}}{\pi}\exp\left(-\beta x^2 - \sigma y^2\right)$$

- The minimum Hellinger distance estimator of θ is defined by:

$$\hat{\theta}_n = \operatorname{Arg}\min_{\theta\in\Theta} H_2\left(\hat{f}_n, f_\theta\right)$$

where

$$H_2\left(\hat{f}_n, f_\theta\right) = \left\{\int_{\mathbb{R}^2}\left|\hat{f}_n^{1/2}(x,y) - f_\theta^{1/2}(x,y)\right|^2 \mathrm{d}x\mathrm{d}y\right\}^{1/2}$$

with

$$\hat{f}_n(x,y) = \frac{1}{nb_n^2}\sum_{k=1}^{n} K_1\left(\frac{x-X_k}{b_n}\right)K_1\left(\frac{y-Y_k}{b_n}\right) \quad \text{and} \quad f_\theta(x,y) = \frac{\sqrt{\beta\sigma}}{\pi}\exp\left(-\beta x^2 - \sigma y^2\right)$$

where K_1 is a kernel function which satisfies conditions (A_4) and (A_5) such that $K_1(x)K_1(y) = K(x,y)$.

Let $W = \left(W^{(1)}, W^{(2)}\right)$, we can write Equation (2) as follows:

$$\begin{pmatrix} \mathrm{d}X_t \\ \mathrm{d}Y_t \end{pmatrix} = \begin{pmatrix} -\beta & 0 \\ 0 & -\sigma \end{pmatrix}\begin{pmatrix} X_t \\ Y_t \end{pmatrix}\mathrm{d}t + \begin{pmatrix} \mathrm{d}W_t^{(1)} \\ \mathrm{d}W_t^{(2)} \end{pmatrix}$$

which gives the the following system

$$\begin{cases} \mathrm{d}X_t = -\beta X_t \mathrm{d}t + \mathrm{d}W_t^{(1)} \\ \mathrm{d}Y_t = -\sigma Y_t \mathrm{d}t + \mathrm{d}W_t^{(2)} \end{cases}$$

Thus, $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are two independent univariate Ornstein-Uhlenbeck processes of parameters β and σ respectively.

We now give simulations for different parameter values using the R language. For each process, we generate sample paths using the package "sde" [17] and to compute a value of the estimator, we use the function "nlm" [18] of the R language. The kernel function K_1 is the density of the standard normal distribution. We use the bandwidth $b_n = \frac{\sqrt{\log(n)}}{n^{0.24}}$ according to conditions on the bandwidth in the paper.

Simulations are based on 1000 observations of the Ornstein-Uhlenbeck process with 200 replications.

Simulation results are given in the **Table 1**.

Table 1. Means and standard errors of the minimum Hellinger distance estimator.

$\theta_0 = (\beta_0, \sigma_0)$	$\hat{\theta} = (\hat{\beta}, \hat{\sigma})$	
	Means	Standard errors
(0.3, 0.7)	(0.2985977, 0.6998527)	(0.01076013, 0.01032311)
(0.5, 2)	(0.4954066, 1.998997)	(0.0341282, 0.008429909)
(1, 2.4)	(0.9987882, 2.398991)	(0.009604874, 0.01262858)
(1, 3)	(0.998918, 2.999193)	(0.008726621, 0.01034987)
(0.223, 0.6)	(0.2223449, 0.6006928)	(0.01048311, 0.01224315)

In **Table 1**, θ_0 denotes the true value of the parameter and $\hat{\theta}$ denotes an estimation of θ_0 given by the minimum Hellinger distance estimator. Simulation results illustrate the good properties of the estimator. Indeed, the means of the estimator are quite close to the true values of the parameter in all cases and the standard errors are low.

4.2. Example 2

We consider the Homogeneous Gaussian diffusion process [19] solution of the stochastic differential equation

$$\mathrm{d}X_t = \left(A + BX_t\right)\mathrm{d}t + \sigma \mathrm{d}W_t, \quad X_0 = x_0 \tag{4}$$

where $\sigma > 0$ is known, W is a two-dimensional Brownian motion, B is a 2×2 matrix with eigenvalues with strictly negative parts and A is a 2×1 matrix. By condition on the matrix B, X has an invariant probability $\mu = N\left(m,\Gamma\right)$ where $m = -B^{-1}A$ and Γ is the unique symetric solution of the equation

$$C + B\Gamma + \Gamma B^t = 0 \text{ where } C = DD^t \text{ and } D = \sigma I \text{ with } I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{5}$$

Let

$$A = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \text{ and } B = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{12} & \beta_{22} \end{pmatrix} \text{ with } \beta_{11} < 0 \text{ and } \beta_{22} < 0.$$

As in [19], we suppose that $\sigma = \sqrt{2}$. In the following, we suppose that $\beta_{11}\beta_{22} - \beta_{12}^2 \neq 0$.

Then we have

$$B^{-1} = \frac{1}{\beta_{11}\beta_{22} - \beta_{12}^2}\begin{pmatrix} \beta_{22} & -\beta_{12} \\ -\beta_{12} & \beta_{11} \end{pmatrix} \text{ and } m = \frac{1}{\beta_{11}\beta_{22} - \beta_{12}^2}\begin{pmatrix} \alpha_2\beta_{12} - \alpha_1\beta_{22} \\ \alpha_1\beta_{12} - \alpha_2\beta_{11} \end{pmatrix}.$$

Let $\Gamma = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$, we have $C = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

$$C + B\Gamma + \Gamma B^t = 0 \Leftrightarrow \begin{cases} a\beta_{11} + b\beta_{12} = -1 \\ a\beta_{12} + b\left(\beta_{11} + \beta_{22}\right) + d\beta_{12} = 0 \\ b\beta_{12} + d\beta_{22} = -1 \end{cases}$$

$$\Leftrightarrow \begin{pmatrix} \beta_{11} & \beta_{12} & 0 \\ \beta_{12} & \beta_{11} + \beta_{22} & \beta_{12} \\ 0 & \beta_{12} & \beta_{22} \end{pmatrix}\begin{pmatrix} a \\ b \\ d \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix}$$

Let $G = \begin{pmatrix} \beta_{11} & \beta_{12} & 0 \\ \beta_{12} & \beta_{11} + \beta_{22} & \beta_{12} \\ 0 & \beta_{12} & \beta_{22} \end{pmatrix}$ and $H = \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix}$, we have $\det\left(G\right) = \left(\beta_{11} + \beta_{22}\right)\left(\beta_{11}\beta_{22} - \beta_{12}^2\right) \neq 0$; G is invertible and we have

$$\begin{pmatrix} a \\ b \\ d \end{pmatrix} = G^{-1}H = \frac{1}{\beta_{11}\beta_{22} - \beta_{12}^2}\begin{pmatrix} -\beta_{22} \\ \beta_{12} \\ -\beta_{11} \end{pmatrix} \text{ and } \Gamma = \frac{1}{\beta_{11}\beta_{22} - \beta_{12}^2}\begin{pmatrix} -\beta_{22} & \beta_{12} \\ \beta_{12} & -\beta_{11} \end{pmatrix}.$$

Γ is invertible and we have $\Gamma^{-1} = \begin{pmatrix} -\beta_{11} & -\beta_{12} \\ -\beta_{12} & -\beta_{22} \end{pmatrix}$. Hence, the invariant density of μ is

Table 2. Means and standard errors of the estimators.

True values of θ	$\hat{\theta}$ (MHD)		$\hat{\theta}$ (Estimating function)	
	Means	Standard errors	Means	Standard errors
$\alpha_1 = 4$	3.996942	0.0005203	4.0349	0.2904
$\alpha_2 = 1$	1.00776	0.001311968	1.0035	0.2891
$\beta_{11} = -2$	−2.007696	0.001315799	−2.0155	0.1248
$\beta_{22} = -3$	−2.982749	0.002923666	−3.0247	0.1978
$\beta_{12} = 1$	1.009081	0.001513984	1.0078	0.1177

$$f(x) = \frac{1}{\left(\sqrt{2\pi}\right)^2 \sqrt{\left|\det(\Gamma)\right|}} \exp\left(-\frac{1}{2}(x-m)^t \Gamma^{-1}(x-m)\right)$$
$$= \frac{\sqrt{\left|\beta_{11}\beta_{22} - \beta_{12}^2\right|}}{2\pi} \exp\left(\frac{1}{2}\beta_{11}\left(x_1 - \frac{\alpha_2\beta_{12} - \alpha_1\beta_{22}}{\beta_{11}\beta_{22} - \beta_{12}^2}\right)^2 + \frac{1}{2}\beta_{22}\left(x_2 - \frac{\alpha_1\beta_{12} - \alpha_2\beta_{11}}{\beta_{11}\beta_{22} - \beta_{12}^2}\right)^2\right)$$
$$\times \exp\left(\beta_{12}\left(x_1 - \frac{\alpha_2\beta_{12} - \alpha_1\beta_{22}}{\beta_{11}\beta_{22} - \beta_{12}^2}\right)\left(x_2 - \frac{\alpha_1\beta_{12} - \alpha_2\beta_{11}}{\beta_{11}\beta_{22} - \beta_{12}^2}\right)\right).$$

For simulation, we must write the stochastic differential Equation (4) in matrix form as follows:

$$\begin{pmatrix} \mathrm{d}X_t^{(1)} \\ \mathrm{d}X_t^{(2)} \end{pmatrix} = \left(\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{12} & \beta_{22} \end{pmatrix}\right)\begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \end{pmatrix}\mathrm{d}t + \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}\begin{pmatrix} \mathrm{d}W_t^{(1)} \\ \mathrm{d}W_t^{(2)} \end{pmatrix}$$
$$= \begin{pmatrix} \alpha_1 + \beta_{11}X_t^{(1)} + \beta_{12}X_t^{(2)} \\ \alpha_2 + \beta_{12}X_t^{(1)} + \beta_{22}X_t^{(2)} \end{pmatrix}\mathrm{d}t + \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}\begin{pmatrix} \mathrm{d}W_t^{(1)} \\ \mathrm{d}W_t^{(2)} \end{pmatrix}$$

As in [19], the true values of the parameter $\theta = (\alpha_1, \alpha_2, \beta_{11}, \beta_{22}, \beta_{12})$ are $\theta_0 = (4, 1, -2, -3, 1)$ and $\sigma = \sqrt{2}$. Then, we have

$$\begin{pmatrix} \mathrm{d}X_t^{(1)} \\ \mathrm{d}X_t^{(2)} \end{pmatrix} = \begin{pmatrix} 4 - 2X_t^{(1)} + X_t^{(2)} \\ 1 + X_t^{(1)} - 3X_t^{(2)} \end{pmatrix}\mathrm{d}t + \begin{pmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix}\begin{pmatrix} \mathrm{d}W_t^{(1)} \\ \mathrm{d}W_t^{(2)} \end{pmatrix}$$

Now, we can simulate a sample path of the Homogeneous Gaussian diffusion using the "yuima" package of R language [20]. We use the function "nlm" to compute a value of the estimator.

We generate 500 sample paths of the process, each of size 500. The kernel function and the bandwidth are those of the previous example.

We compare the estimator obtained by the minimum Hellinger distance method (MHD) of this paper and the estimator obtained in [19] by estimating function. **Table 2** summarizes results of simulation of means and standard errors of the different estimators.

Table 2 shows that the two estimators have good behavior. For the two methods, the means of the estimators are close to the true values of the parameter. But the standard errors of the MHD estimator are lower than those of the estimating function estimator.

References

[1] Dacunha-Castelle, D. and Florens-Zmirou, D. (1986) Estimation of the Coefficients of a Diffusion from Discrete Observations. *Stochastics*, **19**, 263-284. http://dx.doi.org/10.1080/17442508608833428

[2] Pedersen, A.R. (1995) A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics*, **22**, 55-71.

[3] Yoshida, N. (1992) Estimation for Diffusion Processes from Discrete Observation. *Journal of Multivariate Analysis*, **41**, 220-242. http://dx.doi.org/10.1016/0047-259X(92)90068-Q

[4] Aït-Sahalia, Y. (2002) Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica*, **70**, 223-262. http://dx.doi.org/10.1111/1468-0262.00274

[5] Florens-Zmirou, D. (1989) Approximate Discrete-Time Schemes for Statistics of Diffusion Processes. *Statistics*, **20**, 547-557. http://dx.doi.org/10.1080/02331888908802205

[6] Bibby, B.M. and Sørensen, M. (1995) Martingale Estimation Functions for Discretely Observed Diffusion Processes. *Bernoulli*, **1**, 17-39. http://dx.doi.org/10.2307/3318679

[7] Beran, R. (1977) Minimum Hellinger Distance Estimates for Parametric Models. *Annals of Statistics*, **5**, 445-463. http://dx.doi.org/10.1214/aos/1176343842

[8] Hili, O. (1995) On the Estimation of Nonlinear Time Series Models. *Stochastics*: *An International Journal of Probability and Stochastic Processes*, **52**, 207-226. http://dx.doi.org/10.1080/17442509508833972

[9] N'drin, J.A. and Hili, O. (2013) Parameter Estimation of One-Dimensional Diffusion Process by Minimum Hellinger Distance Method. *Random Operators and Stochastic Equations*, **21**, 403-424. http://dx.doi.org/10.1515/rose-2013-0019

[10] Bianchi, A. (2007) Nonparametric Trend Coefficient Estimation for Multidimensional Diffusions. *Comptes Rendus de l'Académie des Sciences*, **345**, 101-105.

[11] Pardoux, E. and Veretennikov, Y.A. (2001) On the Poisson Equation and Diffusion Approximation. I. *The Annals of Probability*, **29**, 1061-1085. http://dx.doi.org/10.1214/aop/1015345596

[12] Veretennikov, Y.A. (1997) On Polynomial Mixing Bounds for Stochastic Differential Equations. *Stochastic Process*, **70**, 115-127. http://dx.doi.org/10.1016/S0304-4149(97)00056-2

[13] Devroye, L. and Györfi, L. (1985) Nonparametric Density Estimation: The L_1 View. Wiley, New York.

[14] Glick, N. (1974) Consistency Conditions for Probability Estimators and Integrals of Density Estimators. *Utilitas Mathematica*, **6**, 61-74.

[15] Jacobsen, M. (2001) Examples of Multivariate Diffusions: The Time-Reversibility, a Cox-Ingersoll-Ross Type Process. Department of Statistics and Operations Research, University of Copenhagen, Copenhagen.

[16] Caumel, Y. (2011) Probabilités et processus stochastiques. Springer-Verlag, Paris. http://dx.doi.org/10.1007/978-2-8178-0163-6

[17] Iacus, S.M. (2008) Simulation and Inference for Stochastic Differential Equations. Springer Series in Statistics, Springer, New York. http://dx.doi.org/10.1007/978-0-387-75839-8

[18] Lafaye de Micheaux, P., Drouilhet, R. and Liquet, B. (2011) Le logiciel R: Maitriser le langage-Effectuer des analyses statistiques. Springer-Verlag, Paris. http://dx.doi.org/10.1007/978-2-8178-0115-5

[19] Sørensen, H. (2001) Discretely Observed Diffusions: Approximation of the Continuous-Time Score Function. *Scandinavian Journal of Statistics*, **28**, 113-121. http://dx.doi.org/10.1111/1467-9469.00227

[20] Iacus, S.M. (2011) Option Pricing and Estimation of Financial Models with R. John Wiley and Sons, Ltd., Chichester. http://dx.doi.org/10.1002/9781119990079

[21] Roussas, G.G. (1969) Nonparametric Estimation in Markov Processes. *Annals of the Institute of Statistical Mathematics*, **21**, 73-87. http://dx.doi.org/10.1007/BF02532233

[22] Bosq, D. (1998) Nonparametric Statistics for Stochastic Processes: Estimation and Prediction. Second Edition, Springer-Verlag, New York. http://dx.doi.org/10.1007/978-1-4612-1718-3

[23] Dharmenda, S. and Masry, E. (1996) Minimum Complexity Regression Estimation with Weakly Dependent Observations. *IEEE Transactions on Information Theory*, **42**, 2133-2145. http://dx.doi.org/10.1109/18.556602

[24] Dominique, F. and Aimé, F. (1998) Calcul des probabilities: Cours, exercices et problèmes corrigés. 2e edition, Dunod, Paris.

Appendix

A1. Proof of Theorem 1

Proof.

$$\left|\hat{f}_n(x)-f_\theta(x)\right|=\left|\left(\hat{f}_n(x)-\mathbb{E}\hat{f}_n(x)\right)+\left(\mathbb{E}\hat{f}_n(x)-f_\theta(x)\right)\right|$$
$$\leq\left|\left(\hat{f}_n(x)-\mathbb{E}\hat{f}_n(x)\right)\right|+\left|\left(\mathbb{E}\hat{f}_n(x)-f_\theta(x)\right)\right|$$

We have:

Step 1:

$$\mathbb{E}\hat{f}_n(x)=\mathbb{E}\left[\frac{1}{nb_n^d}\sum_{k=1}^{n}K\left(\frac{x-X_k}{b_n}\right)\right]=\frac{1}{b_n^d}\mathbb{E}K\left(\frac{x-X_1}{b_n}\right)$$
$$=\frac{1}{b_n^d}\int_{\mathbb{R}^d}K\left(\frac{x-u}{b_n}\right)f_\theta(u)\mathrm{d}u$$
$$=\frac{1}{b_n^d}\int_{\mathbb{R}^d}K\left(\frac{t}{b_n}\right)f_\theta(x-t)\mathrm{d}t\to f_\theta(x).$$

by Theorem 2.1 [21].

Hence

$$\mathbb{E}\hat{f}_n(x)-f_\theta(x)\to 0. \tag{6}$$

Step 2:

$$\left|\hat{f}_n(x)-\mathbb{E}\hat{f}_n(x)\right|=\frac{1}{nb_n^d}\left|\sum_{k=1}^{n}Y_k\right|$$

where

$$Y_k=K\left(\frac{x-X_k}{b_n}\right)-\mathbb{E}K\left(\frac{x-X_k}{b_n}\right)$$

- $\mathbb{E}(Y_k)=0$
- $|Y_k|\leq 2N_1$

Then by theorem 2.1 [9], we have for all $\epsilon>0$

$$\mathbb{P}\left\{\frac{1}{nb_n^d}\left|\sum_{k=1}^{n}Y_k\right|>\epsilon\right\}=\mathbb{P}\left\{\frac{1}{n}\left|\sum_{k=1}^{n}Y_k\right|>\epsilon b_n^d\right\}\leq 2C\exp\left(-\frac{\epsilon^2 nb_n^{2d}}{2\left(\mathbb{E}|Y_1|^2+\frac{2\epsilon N_1 b_n^d}{3}\right)}\right).$$

We have

$$\mathbb{E}|Y_1|^2=\mathbb{E}\left|K\left(\frac{x-X_1}{b_n}\right)-\mathbb{E}K\left(\frac{x-X_1}{b_n}\right)\right|^2=\mathbb{E}K^2\left(\frac{x-X_1}{b_n}\right)-\left(\mathbb{E}K\left(\frac{x-X_1}{b_n}\right)\right)^2$$
$$=b_n^d\left(\frac{1}{b_n^d}\mathbb{E}K^2\left(\frac{x-X_1}{b_n}\right)\right)-b_n^{2d}\left(\frac{1}{b_n^d}\mathbb{E}K\left(\frac{x-X_1}{b_n}\right)\right)^2$$
$$=b_n^d\left[\frac{1}{b_n^d}\mathbb{E}K^2\left(\frac{x-X_1}{b_n}\right)-b_n^d\left(\frac{1}{b_n^d}\mathbb{E}K\left(\frac{x-X_1}{b_n}\right)\right)^2\right]=b_n^d A_n$$

where

$$A_n \to f_\theta(x)\int_{\mathbb{R}^d} K^2(u)\,\mathrm{d}u.$$

Then

$$\mathbb{P}\left\{\frac{1}{nb_n^d}\left|\sum_{k=1}^n Y_k\right| > \epsilon\right\} \le 2C\exp\left(-\frac{\epsilon^2 nb_n^{2d}}{2\left(b_n^d A_n + \frac{2\epsilon N_1 b_n^d}{3}\right)}\right) \le 2C\exp\left(-\frac{\epsilon^2 nb_n^d}{2\left(A_n + \frac{2\epsilon N_1}{3}\right)}\right).$$

Therefore

$$\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) \to 0 \text{ almost surely,} \tag{7}$$

by the Borel-Cantelli's lemma.

(6) and (7) imply that

$$\hat{f}_n(x) \to f_\theta(x) \text{ almost surely.}$$

This achieves the proof of the theorem.

A2. Proof of Theorem 2

Proof.

$$\sqrt{nb_n^d}\left(\hat{f}_n(x) - f_\theta(x)\right) = \sqrt{nb_n^d}\left(\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)\right) + \sqrt{nb_n^d}\left(\mathbb{E}\hat{f}_n(x) - f_\theta(x)\right).$$

(1)

By making the change of variable $t = \frac{x-u}{b_n}$ and using assumptions (A_4) and (A_5), we get:

$$\begin{aligned}
&\sqrt{nb_n^d}\left(\mathbb{E}\hat{f}_n(x) - f_\theta(x)\right)\\
&= \sqrt{nb_n^d}\left\{\int_{\mathbb{R}^d}\frac{1}{b_n^d}K\left(\frac{x-u}{b_n}\right)f_\theta(u)\,\mathrm{d}u - f_\theta(x)\right\}\\
&= \sqrt{nb_n^d}\left\{\int_{\mathbb{R}^d}K(t)\left[f_\theta(x-tb_n) - f_\theta(x)\right]\mathrm{d}t\right\}\\
&= \sqrt{nb_n^d}\left\{\int_{\mathbb{R}^d}K(t)\left[\frac{1}{2!}\sum_{i,j}^d\frac{\partial^2 f_\theta(x)}{\partial x_i\partial x_j}t_i t_j(-b_n)^2 + o\left(b_n^2\right)\right]\mathrm{d}t\right\}\\
&= \sqrt{nb_n^{d+4}}\left\{\frac{1}{2}\int_{\mathbb{R}^d}\sum_{i=1}^d\frac{\partial^2 f_\theta(x)}{\partial x_i^2}t_i^2 K(t)\,\mathrm{d}t + o(1)\right\}\\
&= \sqrt{nb_n^{d+4}}\left\{\frac{1}{2}\sum_{i=1}^d\frac{\partial^2 f_\theta(x)}{\partial x_i^2}\int_{\mathbb{R}^d}t_i^2 K(t)\,\mathrm{d}t + o(1)\right\} \to 0 \quad \text{as } n \to +\infty.
\end{aligned}$$

(2)

$$\sqrt{nb_n^d}\left(\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)\right) = \left(nb_n^d\right)^{-1/2}\sum_{k=1}^n Y_k$$

where

$$Y_k = K\left(\frac{x-X_k}{b_n}\right) - \mathbb{E}K\left(\frac{x-X_k}{b_n}\right).$$

We have $\mathbb{E}(Y_k)=0$ and $|Y_k|\le 2N_1$.

Let $p=p(n)$, $q=q(n)$ and $r=r(n)$ be positive integers which tend to infinity as $n\to\infty$ such that $r(p+q)\le n<r(p+q+1)$.

Define U_m and V_m by

$$U_m=\sum_{k=(m-1)(p+q)+1}^{(m-1)(p+q)+p}Y_k,\quad V_m=\sum_{k=(m-1)(p+q)+p+1}^{m(p+q)}Y_k,\ m=1,\cdots,r$$

and

$$V_{r+1}=\sum_{k=r(p+q)+1}^{n}Y_k.$$

We have

$$\sum_{k=1}^{n}Y_k=\sum_{m=1}^{r}U_m+\sum_{m=1}^{r+1}V_m.$$

Step 1: We prove that $\left(nb_n^d\right)^{-1/2}\sum_{m=1}^{r+1}V_m\to 0$ in probability.

By Minkowski's inequality, we have

$$\begin{aligned}\left(\mathbb{E}\left|\left(nb_n^d\right)^{-1/2}\sum_{m=1}^{r+1}V_m\right|^2\right)^{1/2}&\le\left(nb_n^d\right)^{-1/2}\left[\sum_{m=1}^{r}\left(\mathbb{E}|V_m|^2\right)^{1/2}+\left(\mathbb{E}|V_{r+1}|^2\right)^{1/2}\right]\\&\le\left(nb_n^d\right)^{-1/2}\left[r\left(\mathbb{E}|V_1|^2\right)^{1/2}+\left(\mathbb{E}|V_{r+1}|^2\right)^{1/2}\right]\end{aligned}$$

(1) Using Billingsley's inequality [22],

$$\begin{aligned}\mathbb{E}|V_1|^2&=\mathbb{E}\left(\sum_{k=p+1}^{p+q}Y_k\right)^2=q\mathbb{E}\left(Y_1^2\right)+2\sum_{i<j}\mathbb{E}\left(Y_iY_j\right)\\&\le q\mathbb{E}\left(Y_1^2\right)+2q\sum_{j=p+1}^{p+q-1}\left|\mathbb{E}\left(Y_{p+1}Y_{j+1}\right)\right|\le q\left[\mathbb{E}\left(Y_1^2\right)+32N_1^2\sum_{j=p+1}^{p+q-1}\alpha(j-p)\right]\\&\le q\left[\mathbb{E}\left(Y_1^2\right)+32N_1^2\sum_{j=1}^{q-1}\alpha(j)\right]\le qC\quad\text{with } C=\mathbb{E}\left(Y_1^2\right)+32N_1^2\sum_{j=1}^{\infty}\alpha(j).\end{aligned}$$

(2)

$$\mathbb{E}|V_{r+1}|^2=\mathbb{E}\left(\sum_{k=r(p+q)+1}^{n}Y_k\right)^2=\left(n-r(p+q)\right)\mathbb{E}\left(Y_1^2\right)+2\sum_{i<j}\mathbb{E}\left(Y_iY_j\right)\le rC.$$

Hence,

$$\left(\mathbb{E}\left|\left(nb_n^d\right)^{-1/2}\sum_{m=1}^{r+1}V_m\right|^2\right)^{1/2}\le C\left(nb_n^d\right)^{-1/2}\left(r\sqrt{q}+\sqrt{r}\right)$$

Therefore, choosing $q=q(n), r=r(n)$ and b_n such that

$$\frac{r\sqrt{q}}{\sqrt{nb_n^d}}\to 0\ \text{ as } n\to\infty,\tag{8}$$

we get

$$\mathbb{E}\left|\left(nb_n^d\right)^{-1/2}\sum_{m=1}^{r+1}V_m\right|^2\to 0.$$

which implies that

$$\left(nb_n^d\right)^{-1/2}\sum_{m=1}^{r+1}V_m \to 0 \text{ in probability.}$$

Step 2: asymptotic normality of $\left(nb_n^d\right)^{-1/2}\sum_{m=1}^{r}U_m$.

U_m, $m=1,\cdots,r$ have the same distribution; so that

$$\prod_{m=1}^{r}\mathbb{E}\exp\left(itU_m\right)=\left(\mathbb{E}\exp\left(itU_1\right)\right)^r.$$

From Lemma 4.2 [23], we have

$$\begin{aligned}&\left|\mathbb{E}\left[\exp\left(it\sum_{m=1}^{r}U_m\right)\right]-\left[\mathbb{E}\exp\left(itU_1\right)\right]^r\right|=\left|\mathbb{E}\left[\exp\left(it\sum_{m=1}^{r}U_m\right)\right]-\prod_{m=1}^{r}\mathbb{E}\exp\left(itU_m\right)\right|\\&=\left|\mathbb{E}\left(\prod_{m=1}^{r}\exp\left(itU_m\right)\right)-\prod_{m=1}^{r}\mathbb{E}\exp\left(itU_m\right)\right|\le 4\left(r-1\right)\alpha\left(1+q\right)\le 4r\alpha\left(q\right).\end{aligned}$$

Setting $\phi_1\left(t\right)=\mathbb{E}\exp\left(itU_1\right)$. If $q=q\left(n\right)$ and $r=r\left(n\right)$ are chosen such that

$$r\alpha\left(q\right)\to 0 \text{ as } n\to\infty, \tag{9}$$

the charasteristic function of $\left(nb_n^d\right)^{-1/2}\sum_{m=1}^{r}U_m$ is $\phi_1^r\left(t\left(nb_n^d\right)^{-1/2}\right)$ which is the charasteristic function of $\sum_{m=1}^{r}Z_m$ where Z_m, $m=1,\cdots,r$ are independent random variables with distribution that of $\left(nb_n^d\right)^{-1/2}U_1$.

We have $\mathbb{E}\left(Z_m\right)=0$ and

$$\begin{aligned}\mathbb{E}\left(\sum_{m=1}^{r}Z_m\right)^2&=r\mathbb{E}\left(Z_1^2\right)=\left(nb_n^d\right)^{-1}r\mathbb{E}\left(U_1^2\right)=\left(nb_n^d\right)^{-1}r\left[p\mathbb{E}\left(Y_1^2\right)+2\sum_{i<j}^{p}\mathbb{E}\left(Y_iY_j\right)\right]\\&=\frac{rp}{nb_n^d}\mathbb{E}\left(Y_1^2\right)+\frac{2r}{nb_n^d}\sum_{i<j}^{p}\mathbb{E}\left(Y_iY_j\right).\end{aligned}$$

(1) $\frac{rp}{nb_n^d}\mathbb{E}\left(Y_1\right)^2=\frac{rp}{n}\left\{\frac{1}{b_n^d}\mathbb{E}K^2\left(\frac{x-X_1}{b_n}\right)-b_n^d\left(\frac{1}{b_n^d}\mathbb{E}K\left(\frac{x-X_1}{b_n}\right)\right)^2\right\}\to f_\theta\left(x\right)\int_{\mathbb{R}^d}K^2\left(u\right)\mathrm{d}u.$

(2) Note that $\alpha\left(k\right)\le\exp\left(-\lambda k\right)=\varphi\left(k\right)$ with $\lambda>0$.

$$\begin{aligned}\frac{2r}{nb_n^d}\sum_{i<j}\left|\mathbb{E}\left(Y_iY_j\right)\right|&\le\frac{2r}{nb_n^d}\sum_{j=1}^{p-1}j\left|\mathbb{E}\left(Y_1Y_{j+1}\right)\right|\le\frac{2r}{nb_n^d}16N_1^2\sum_{j=1}^{p-1}j\alpha\left(j\right)\\&\le\frac{32N_1^2r}{nb_n^d}\sum_{k=1}^{p-1}\sum_{l=k}^{p-1}\alpha\left(l\right)\le\frac{32N_1^2r}{nb_n^d}\sum_{k=1}^{p-1}\sum_{l=k}^{p-1}\alpha\left(l\right)\\&\le\frac{32N_1^2r}{nb_n^d}\sum_{k=1}^{p-1}\sum_{l=k}^{p-1}\left(\varphi\left(l\right)\right)^{1/2}\left(\varphi\left(k\right)\right)^{1/2}\\&\le\frac{32N_1^2r}{nb_n^d}\sum_{k=1}^{\infty}\left(\varphi\left(k\right)\right)^{1/2}\sum_{l=1}^{\infty}\left(\varphi\left(l\right)\right)^{1/2}\\&\le\frac{32N_1^2r}{nb_n^d}\left(\sum_{i=1}^{\infty}\left(\varphi\left(i\right)\right)^{1/2}\right)^2\to 0 \quad\text{if } \frac{r}{nb_n^d}\to 0 \text{ as } n\to\infty.\end{aligned}\tag{10}$$

Therefore

$$\mathbb{E}\left(\sum_{m=1}^{r} Z_m\right)^2 \to f_\theta(x)\int_{\mathbb{R}^d} K^2(u)\,\mathrm{d}u \quad \text{as } n\to\infty.$$

Since the random variables $U_m\left(m=1,\cdots,r\right)$ have the same distribution, then by Lyapunov's theorem [24], the limiting distribution of $\left(nb_n^d\right)^{-1/2}\sum_{m=1}^{r} U_m$ is $N\left(0,\tau^2(x)\right)$ where

$$\tau^2(x) = f_\theta(x)\int_{\mathbb{R}^d} K^2(u)\,\mathrm{d}u.$$

The condition (8), (9) and (10) are satisfied, for example, with

$$r(n)\sim\log(n),\ p(n)\sim\frac{n}{\log(n)}-n^{1/4},\ q(n)\sim n^{1/4} \text{ and } b_n^d=\frac{\log(n)}{n^\lambda} \text{ with } 0<\lambda<\frac{3}{4}.$$

This achieves the proof of the theorem.

A3. Proof of Lemma 1

Proof. The proof of the lemma is done in two steps.

Step 1: we prove that

$$Y_n=\sqrt{n}\left\{\int_{\mathbb{R}^d} h_{\theta_0}(x)\,\hat{f}_n(x)\,\mathrm{d}x-\frac{1}{n}\sum_{i=1}^{n} h_{\theta_0}(X_i)\right\}\to 0 \text{ in probability.}$$

$$\begin{aligned}
\mathbb{E}\left|Y_n\right| &= \sqrt{n}\int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d} h_{\theta_0}(x)\frac{1}{nb_n^d}\sum_{i=1}^{n} K\left(\frac{x-y}{b_n}\right)\mathrm{d}x-\frac{1}{n}\sum_{i=1}^{n} h_{\theta_0}(y)\right| f_{\theta_0}(y)\,\mathrm{d}y\\
&= \sqrt{n}\int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d} h_{\theta_0}(x)\frac{1}{b_n^d}K\left(\frac{x-y}{b_n}\right)\mathrm{d}x-h_{\theta_0}(y)\right| f_{\theta_0}(y)\,\mathrm{d}y\\
&= \sqrt{n}\int_{E_n}\left|\int_{\mathbb{R}^d} h_{\theta_0}(y+ub_n)K(u)\,\mathrm{d}u-h_{\theta_0}(y)\right| f_{\theta_0}(y)\,\mathrm{d}y\\
&\quad+\sqrt{n}\int_{E_n^c}\left|\int_{\mathbb{R}^d} h_{\theta_0}(y+ub_n)K(u)\,\mathrm{d}u-h_{\theta_0}(y)\right| f_{\theta_0}(y)\,\mathrm{d}y\\
&= I_{1n}+I_{2n}.
\end{aligned}$$

With assumptions (A_4) and (A_5), we have

$$\begin{aligned}
I_{1n} &= \sqrt{n}\int_{E_n}\left|\int_{\mathbb{R}^d}\left[h_{\theta_0}(y+ub_n)-h_{\theta_0}(y)\right]K(u)\,\mathrm{d}u\right| f_{\theta_0}(y)\,\mathrm{d}y\\
&= \sqrt{n}\int_{E_n}\left|\int_{\mathbb{R}^d}\left[\frac{1}{2!}\sum_{i,j}^{d}\frac{\partial^2 h_{\theta_0}(y)}{\partial y_i\partial y_j}u_iu_j\left(b_n^2\right)+o\left(b_n^2\right)\right]K(u)\,\mathrm{d}u\right| f_{\theta_0}(y)\,\mathrm{d}y\\
&= \sqrt{n}b_n^2\int_{E_n}\left|\int_{\mathbb{R}^d}\frac{1}{2}\sum_{i=1}^{d}\frac{\partial^2 h_{\theta_0}(y)}{\partial y_i^2}u_i^2K(u)\,\mathrm{d}u+o(1)\right| f_{\theta_0}(y)\,\mathrm{d}y\\
&\le \sqrt{n}b_n^2\int_{E_n}\left|\frac{1}{2}\sum_{i=1}^{d}\frac{\partial^2 h_{\theta_0}(y)}{\partial y_i^2}\int_{\mathbb{R}^d}u_i^2K(u)\,\mathrm{d}u+o(1)\right| f_{\theta_0}(y)\,\mathrm{d}y\\
&\le \sqrt{n}b_n^2\int_{E_n} f_{\theta_0}(y)\left\{\frac{1}{2}\sum_{i=1}^{d}\left|\frac{\partial^2 h_{\theta_0}(y)}{\partial y_i^2}\right|\left(\int_{\mathbb{R}^d}u_i^2K(u)\,\mathrm{d}u\right)+o(1)\right\}\mathrm{d}y\\
&\to 0 \quad \text{as } n\to+\infty.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
I_{2n} &= \sqrt{n}\int_{E_n^c}\left|\int_{\mathbb{R}^d} h_{\theta_0}(y+ub_n)K(u)\mathrm{d}u - h_{\theta_0}(y)\right| f_{\theta_0}(y)\mathrm{d}y \\
&\le \sqrt{n}\left\{\int_{E_n^c}\left(\int_{\mathbb{R}^d}\left|h_{\theta_0}(y+ub_n)K(u)\mathrm{d}u\right|\right) f_{\theta_0}(y)\mathrm{d}y + \int_{E_n^c}\left|h_{\theta_0}(y)\right| f_{\theta_0}(y)\mathrm{d}y\right\} \\
&\le \sqrt{n}\left\{\int_{E_n^c}\left(\int_{\mathbb{R}^d}\left|h_{\theta_0}(y+ub_n)\right|K(u)\mathrm{d}u\right) f_{\theta_0}(y)\mathrm{d}y + \int_{E_n^c}\left|\dot{g}_{\theta_0}(y)\right| f_{\theta_0}^{1/2}(y)\mathrm{d}y\right\} \\
&\to 0 \quad \text{as } n\to+\infty.
\end{aligned}$$

Therefore

$$Y_n = \sqrt{n}\left\{\int_{\mathbb{R}^d} h_{\theta_0}(x)\hat{f}_n(x)\mathrm{d}x - \frac{1}{n}\sum_{i=1}^n h_{\theta_0}(X_i)\right\} = I_{1n} + I_{2n} \to 0 \text{ in probability.}$$

Step 2: asymptotic normality of $\frac{1}{\sqrt{n}}\sum_{i=1}^n h_{\theta_0}(X_i)$, $\theta_0 \in \mathbb{R}^s$, $s \ge 1$

(1) $\theta_0 \in \mathbb{R}$

Proof is similar to that of theorem 2; we use the inequality of Davidov [22] instead of that of Billingsley. Note that:

$$\mathbb{E}\left(h_{\theta_0}(X_1)\right) = \int_{\mathbb{R}^d} h_{\theta_0}(x) f_{\theta_0}(x)\mathrm{d}x = \frac{1}{2}\int_{\mathbb{R}^d}\dot{g}_{\theta_0}(x) f_{\theta_0}^{1/2}(x)\mathrm{d}x = \frac{1}{4}\int_{\mathbb{R}^d} 2\dot{g}_{\theta_0}(x) g_{\theta_0}(x)\mathrm{d}x = 0.$$

and

$$\mathbb{E}\left(h_{\theta_0}^2(X_1)\right) = \int_{\mathbb{R}^d} h_{\theta_0}^2(x) f_{\theta_0}(x)\mathrm{d}x = \frac{1}{4}\int_{\mathbb{R}^d}\dot{g}_{\theta_0}^2(x)\mathrm{d}x = \frac{1}{4}\int_{\mathbb{R}^d}\dot{g}_{\theta_0}(x)\dot{g}_{\theta_0}^t(x)\mathrm{d}x.$$

(2) $\theta_0 \in \mathbb{R}^s$, $s > 1$

Recall that $X_n \xrightarrow{\mathcal{L}} X$ if and only if $u^t X_n \xrightarrow{\mathcal{L}} u^t X$ for all $u \in \mathbb{R}^s$.

Let $u \in \mathbb{R}^s$, $Y_i = h_{\theta_0}(X_i) = \frac{\dot{g}_{\theta_0}}{2 f_{\theta_0}^{1/2}}(X_i)$ and $T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$, the real random variables $\left(u^t Y_i, i \ge 1\right)$ are strongly mixing with mean zero and variance $u^t \Gamma u$ where Γ is the covariance matrix of Y_1; $\Gamma = \mathbb{E}\left(Y_1 Y_1^t\right)$.

From (1), $u^t T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n u^t Y_i \xrightarrow{\mathcal{L}} N\left(0, u^t \Gamma u\right)$.

Therefore,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n h_{\theta_0}(X_k) \xrightarrow{\mathcal{L}} N(0,\Gamma) \quad \text{where } \Gamma = \frac{1}{4}\int_{\mathbb{R}^d}\dot{g}_{\theta_0}(x)\dot{g}_{\theta_0}^t(x)\mathrm{d}x.$$

This completes the proof of the lemma.

A4. Proof of Lemma 2

Proof.

$$\begin{aligned}
R_n &= \int_{\mathbb{R}^d}\sqrt{n}h_{\theta_0}(x)\left(\hat{f}_n^{1/2}(x) - f_{\theta_0}^{1/2}(x)\right)^2 \mathrm{d}x \\
&= \int_{G_n}\sqrt{n}h_{\theta_0}(x)\left(\hat{f}_n^{1/2}(x) - f_{\theta_0}^{1/2}(x)\right)^2 \mathrm{d}x + \int_{G_n^c}\sqrt{n}h_{\theta_0}(x)\left(\hat{f}_n^{1/2}(x) - f_{\theta_0}^{1/2}(x)\right)^2 \mathrm{d}x \\
&= \int_{G_n}\sqrt{n}h_{\theta_0}(x)\frac{\left(\hat{f}_n(x) - f_{\theta_0}(x)\right)^2}{\left(\hat{f}_n^{1/2}(x) + f_{\theta_0}^{1/2}(x)\right)^2}\mathrm{d}x + \int_{G_n^c}\sqrt{n}h_{\theta_0}(x)\left(\hat{f}_n^{1/2}(x) - f_{\theta_0}^{1/2}(x)\right)^2 \mathrm{d}x \\
&= R_{n1} + R_{n2}.
\end{aligned}$$

We have,

$$|R_{n1}| \le \int_{G_n} \sqrt{n}\left|h_{\theta_0}(x)\right| \frac{\left(\hat{f}_n(x) - f_{\theta_0}(x)\right)^2}{\left(\hat{f}_n^{1/2}(x) + f_{\theta_0}^{1/2}(x)\right)^2} \mathrm{d}x \le \int_{G_n} \sqrt{n}\left|h_{\theta_0}(x)\right| \frac{\left(\hat{f}_n(x) - f_{\theta_0}(x)\right)^2}{f_{\theta_0}(x)} \mathrm{d}x.$$

Now,

$$\begin{aligned}\mathbb{E}\left(|R_{n1}|\right) &\le \int_{G_n} \left|h_{\theta_0}(x)\right| f_{\theta_0}^{-1}(x)\sqrt{n}\mathbb{E}\left(\hat{f}_n(x) - f_{\theta_0}(x)\right)^2 \mathrm{d}x \\ &\le \int_{G_n} \left|h_{\theta_0}(x)\right| f_{\theta_0}^{-1}(x)\sqrt{n}\left[\mathbb{E}\left(\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)\right)^2 + \left(\mathbb{E}\hat{f}_n(x) - f_{\theta_0}(x)\right)^2\right]\mathrm{d}x\end{aligned}$$

(1)

$$\begin{aligned}&\sqrt{n}\mathbb{E}\left(\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)\right)^2 \\ &= \frac{\sqrt{n}}{\left(nb_n^d\right)^2}\mathbb{E}\left(\sum_{i=1}^n Y_i\right)^2 \quad \text{with } Y_i = K\left(\frac{x - X_i}{b_n}\right) - \mathbb{E}K\left(\frac{x - X_i}{b_n}\right) \\ &= \frac{\sqrt{n}}{\left(nb_n^d\right)^2}\left[n\mathbb{E}\left(Y_1^2\right) + 2\sum_{i<j}^n \mathbb{E}\left(Y_iY_j\right)\right] \\ &\le \frac{1}{n^{3/2}b_n^{2d}}\left[n\mathbb{E}\left(Y_1^2\right) + 2\sum_{i<j}^n \left|\mathbb{E}\left(Y_iY_j\right)\right|\right] \\ &\le \frac{1}{n^{3/2}b_n^{2d}}\left[n\mathbb{E}K^2\left(\frac{x - X_1}{b_n}\right) + 2\sum_{j=1}^{n-1} j\left|\mathbb{E}\left(Y_1Y_{j+1}\right)\right|\right]\end{aligned}$$

Using Davidov's inequality for mixing processes, we get

$$\begin{aligned}\sum_{j=1}^{n-1} j\left|\mathbb{E}\left(Y_1Y_{j+1}\right)\right| &\le \sum_{j=1}^{n-1} j\left[2p\left(2\alpha(j)\right)^{1/p}\left(\mathbb{E}|Y_1|^q\right)^{1/q}\left(\mathbb{E}\left|Y_{j+1}\right|^r\right)^{1/r}\right] \\ &\le 2p\left(\mathbb{E}|Y_1|^q\right)^{1/q}\left(\mathbb{E}|Y_1|^r\right)^{1/r}\sum_{j=1}^{n-1} j\left(2\alpha(j)\right)^{1/p}\end{aligned}$$

Choose $q \ge 2$ and $r \ge 2$, we obtain

$$\begin{aligned}\sum_{j=1}^{n-1} j\left|\mathbb{E}\left(Y_1Y_{j+1}\right)\right| &\le 2p\left(\mathbb{E}|Y_1|^2|Y_1|^{q-2}\right)^{1/q}\left(\mathbb{E}|Y_1|^2|Y_1|^{r-2}\right)^{1/r}\sum_{j=1}^{n-1} j\left(2\alpha(j)\right)^{1/p} \\ &\le 2p\left((2N_1)^{q-2}\right)^{1/q}\left((2N_1)^{r-2}\right)^{1/r}\left(\mathbb{E}|Y_1|^2\right)^{1/q}\left(\mathbb{E}|Y_1|^2\right)^{1/r}\sum_{j=1}^{n-1} j\left(2\alpha(j)\right)^{1/p} \\ &\le 2p\left(2^{1/p}\right)\left((2N_1)^{q-2}\right)^{1/q}\left((2N_1)^{r-2}\right)^{1/r}\left(\mathbb{E}|Y_1|^2\right)^{1/q+1/r}\sum_{j=1}^{n-1} j\left(\alpha(j)\right)^{1/p} \\ &\le C\left[\mathbb{E}K^2\left(\frac{x - X_1}{b_n}\right)\right]^{1/q+1/r}\sum_{j=1}^{n-1} j\left(\varphi(j)\right) \quad \text{where } \varphi(j) = \left(\alpha(j)\right)^{1/p} \\ &\le C\left[\mathbb{E}K^2\left(\frac{x - X_1}{b_n}\right)\right]^{1/q+1/r}\left(\sum_{j=1}^{n-1}\varphi^{1/2}(i)\right)^2 \le \frac{C_1}{2}\left[\mathbb{E}K^2\left(\frac{x - X_1}{b_n}\right)\right]^{1/q+1/r}.\end{aligned}$$

Hence,

$$\begin{aligned}
&\sqrt{n}\mathbb{E}\left(\hat{f}_n(x)-\mathbb{E}\hat{f}_n(x)\right)^2\\
&\le \frac{1}{n^{3/2}b_n^{2d}}\left[n\mathbb{E}K^2\left(\frac{x-X_1}{b_n}\right)\right]+\frac{C_1}{n^{3/2}b_n^{2d}}\left[\mathbb{E}K^2\left(\frac{x-X_1}{b_n}\right)\right]^{1/q+1/r}\\
&\le \frac{1}{n^{1/2}b_n^{d}}\left[\frac{1}{b_n^d}\mathbb{E}K^2\left(\frac{x-X_1}{b_n}\right)\right]+\frac{C_1}{n^{3/2}b_n^{d+d/p}}\left[\frac{1}{b_n^d}\mathbb{E}K^2\left(\frac{x-X_1}{b_n}\right)\right]^{1/q+1/r}\\
&\le \frac{1}{n^{1/2}b_n^{d}}\int_{\mathbb{R}^d}K^2(t)f_{\theta_0}(x-tb_n)\mathrm{d}t+\frac{C_1}{n^{3/2}b_n^{d+d/p}}\left[\int_{\mathbb{R}^d}K^2(t)f_{\theta_0}(x-tb_n)\mathrm{d}t\right]^{1/q+1/r}.
\end{aligned}$$

(2)

$$\begin{aligned}
&\sqrt{n}\left(\mathbb{E}\hat{f}_n(x)-f_{\theta_0}(x)\right)^2\\
&=\sqrt{n}\left\{\int_{\mathbb{R}^d}K(t)\left[\frac{1}{2!}\sum_{i,j}^{d}\frac{\partial^2 f_{\theta_0}(x)}{\partial x_i\partial x_j}t_it_j(-b_n)^2+o\left(b_n^2\right)\right]\mathrm{d}t\right\}^2\\
&=n^{1/2}b_n^4\left\{\frac{1}{2}\int_{\mathbb{R}^d}\sum_{i=1}^{d}\frac{\partial^2 f_{\theta_0}(x)}{\partial x_i^2}t_i^2K(t)\mathrm{d}t+o(1)\right\}^2\\
&=n^{1/2}b_n^4\left\{\frac{1}{2}\sum_{i=1}^{d}\frac{\partial^2 f_{\theta_0}(x)}{\partial x_i^2}\int_{\mathbb{R}^d}t_i^2K(t)\mathrm{d}t+o(1)\right\}^2\\
&\le 2dn^{1/2}b_n^4\left\{\frac{1}{4}\sum_{i=1}^{d}\left(\frac{\partial^2 f_{\theta_0}(x)}{\partial x_i^2}\right)^2\left(\int_{\mathbb{R}^d}t_i^2K(t)\mathrm{d}t\right)^2+o(1)\right\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left(|R_{n1}|\right)\le{}&\frac{1}{n^{1/2}b_n^d}\int_{G_n}\left|h_{\theta_0}(x)\right|f_{\theta_0}^{-1}(x)\left(\int_{\mathbb{R}^d}K^2(t)f_{\theta_0}(x-tb_n)\mathrm{d}t\right)\mathrm{d}x\\
&+\frac{C_1}{n^{3/2}b_n^{d+d/p}}\int_{G_n}\left|h_{\theta_0}(x)\right|f_{\theta_0}^{-1}(x)\left(\left[\int_{G_n}K^2(t)f_{\theta_0}(x-tb_n)\mathrm{d}t\right]^{1/q+1/r}\right)\mathrm{d}x\\
&+2dn^{1/2}b_n^4\int_{G_n}\left|h_{\theta_0}(x)\right|f_{\theta_0}^{-1}(x)\left\{\frac{1}{4}\sum_{i=1}^{d}\left(\frac{\partial^2 f_{\theta_0}(x)}{\partial x_i^2}\right)^2\left(\int_{\mathbb{R}^d}t_i^2K(t)\mathrm{d}t\right)^2+o(1)\right\}\mathrm{d}x\\
\to{}&0\quad\text{as } n\to+\infty.
\end{aligned}$$

The last relation implies that

$$R_{n1}\to 0\ \text{ in probability as }\ n\to+\infty. \tag{11}$$

Furthermore,

$$\begin{aligned}
|R_{n2}|&\le\sqrt{n}\int_{G_n^c}\left|h_{\theta_0}(x)\right|\left(\hat{f}_n^{1/2}(x)-f_{\theta_0}^{1/2}(x)\right)^2\mathrm{d}x\\
&\le 2\sqrt{n}\int_{G_n^c}\left|h_{\theta_0}(x)\right|\left(\hat{f}_n(x)+f_{\theta_0}(x)\right)\mathrm{d}x\\
&\le 2\sqrt{n}\int_{G_n^c}\left|h_{\theta_0}(x)\right|f_{\theta_0}(x)\mathrm{d}x+2\sqrt{n}\int_{G_n^c}\left|h_{\theta_0}(x)\right|\hat{f}_n(x)\mathrm{d}x\\
&\le 2\sqrt{n}\int_{G_n^c}\left|\dot{g}_{\theta_0}(x)\right|f_{\theta_0}^{1/2}(x)\mathrm{d}x+2\sqrt{n}\int_{G_n^c}\left|h_{\theta_0}(x)\right|\hat{f}_n(x)\mathrm{d}x\\
&\le 2\sqrt{n}\int_{G_n^c}\left|\dot{g}_{\theta_0}(x)\right|f_{\theta_0}^{1/2}(x)\mathrm{d}x+R_{n22}.
\end{aligned}$$

We have,

$$\begin{aligned}\mathbb{E}\left(R_{n22}\right) &= 2\sqrt{n}\int_{G_n^c}\left|h_{\theta_0}(x)\right|\mathbb{E}\left(\hat{f}_n(x)\right)\mathrm{d}x \\ &= 2\sqrt{n}\int_{G_n^c}\left|h_{\theta_0}(x)\right|\left\{\int_{\mathbb{R}^d}\frac{1}{nb_n^d}\sum_{i=1}^{n}K\left(\frac{x-y}{b_n}\right)f_{\theta_0}(y)\mathrm{d}y\right\}\mathrm{d}x \\ &= 2\sqrt{n}\int_{G_n^c}\left|h_{\theta_0}(x)\right|\left\{\int_{\mathbb{R}^d}\frac{1}{b_n^d}K\left(\frac{x-y}{b_n}\right)f_{\theta_0}(y)\mathrm{d}y\right\}\mathrm{d}x \\ &= 2\sqrt{n}\int_{G_n^c}\left|h_{\theta_0}(x)\right|\left\{\int_{\mathbb{R}^d}K(u)\,f_{\theta_0}(x+ub_n)\mathrm{d}u\right\}\mathrm{d}x\end{aligned}$$

Therefore, if

$$\sqrt{n}\int_{G_n^c}\left|\dot{g}_{\theta_0}(x)\right| f_{\theta_0}^{1/2}(x)\mathrm{d}x \to 0 \quad \text{and} \quad \sqrt{n}\int_{G_n^c}\left|h_{\theta_0}(x)\right|\left\{\int_{\mathbb{R}^d}K(u)\,f_{\theta_0}(x+ub_n)\mathrm{d}u\right\}\mathrm{d}x \to 0$$

then

$$R_{n2} \to 0 \quad \text{in probability as} \quad n \to +\infty. \tag{12}$$

(11) and (12) imply that

$$R_n = \int_{\mathbb{R}^d}\sqrt{n}h_{\theta_0}(x)\left(\hat{f}_n^{1/2}(x) - f_{\theta_0}^{1/2}(x)\right)^2\mathrm{d}x \to 0 \quad \text{in probability as} \quad n \to +\infty.$$