Research Article

Joint Behaviour of Semirecursive Kernel Estimators of the Location and of the Size of the Mode of a Probability Density Function

Abdelkader Mokkadem,[1] Mariane Pelletier,[1] and Baba Thiam[2]

[1] *Département de Mathématiques, Université de Versailles-Saint-Quentin, 45 avenue des Etats-Unis, 78035 Versailles Cedex, France*

[2] *EQUIPPE, Université Charles-De-Gaulle, Lille 3, Maison de la Recherche, Domaine Universitaire du Pont de Bois, BP 60149, 59653 Villeneuve d'Ascq Cedex, France*

Correspondence should be addressed to Mariane Pelletier, pelletier@math.uvsq.fr

Let θ and μ denote the location and the size of the mode of a probability density. We study the joint convergence rates of semirecursive kernel estimators of θ and μ. We show how the estimation of the size of the mode allows measuring the relevance of the estimation of its location. We also enlighten that, beyond their computational advantage on nonrecursive estimators, the semirecursive estimators are preferable to use for the construction of confidence regions.

1. Introduction

Let $X_1, \ldots, X_n$ be independent and identically distributed $\mathbb{R}^d$-valued random variables with unknown probability density f. The aim of this paper is to study the joint kernel estimation of the location θ and of the size $\mu = f(\theta)$ of the mode of f. The mode is assumed to be unique, that is, $f(x) < f(\theta)$ for any $x \neq \theta$, and nondegenerated, that is, the second order differential $D^2 f(\theta)$ at the point θ is nonsingular (in the sequel, $D^m g$ will denote the differential of order m of a multivariate function g).

The problem of estimating the location of the mode of a probability density was widely studied. Kernel methods were considered, among many others, by Parzen [1], Nadaraya [2], Van Ryzin [3], Rüschendorf [4], Konakov [5], Samanta [6], Eddy ([7, 8]), Romano [9], Tsybakov [10], Vieu [11], Mokkadem and Pelletier [12], and Abraham et al. ([13, 14]). The problem of estimating the size of the mode was brought up by several authors (see, e.g., Romano [9] and Vieu [11]), but, at our knowledge, the behaviour of estimators of the size

of the mode has not been investigated in detail, whereas there are at least two statistical motivations for estimating this parameter. First, the use of an estimator of the size is necessary for the construction of confidence regions for the location of the mode (see, e.g., Romano [9]). As a more important motivation, the estimation of the high of the peak gives information on the shape of a density in a neighbourhood of its mode and, consequently, allows measuring the pertinence of the parameter location of the mode; this motivation must be related to the remark made by Vieu [11], who pointed out that the location of the mode is more related to the shape of the derivative of f, whereas the size of the mode is more related to the shape of the density itself.

Let us mention that, even though the problem of estimating the size of the mode was not investigated in the framework of density estimation, it was studied in the framework of regression estimation. Müller [15] proves in particular the joint asymptotic normality and independence of kernel estimators of the location and of the size of the mode in the framework of nonparametric regression models with fixed design. In the framework of nonparametric regression with random design, a similar result is obtained by Ziegler ([16, 17]) for kernel estimators and by Mokkadem and Pelletier [18] for estimators issued from stochastic approximation methods.

This paper is focused on semirecursive kernel estimators of θ and $f(\theta)$. To explain why we chose this option of semirecursive estimators, let us first recall that the (nonrecursive) wellknown kernel estimator of the location of the mode introduced by Parzen [1] is defined as a random variable θ_n^* satisfying

$$f_n^*(\theta_n^*) = \sup_{y \in \mathbb{R}^d} f_n^*(y), \tag{1.1}$$

where f_n^* is Rosenblatt's estimator of f; more precisely,

$$f_n^*(x) = \frac{1}{n h_n^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), \tag{1.2}$$

where the bandwidth (h_n) is a sequence of positive real numbers going to zero and the kernel K is a continuous function satisfying $\lim_{\|x\| \to +\infty} K(x) = 0$, $\int_{\mathbb{R}^d} K(x) dx = 1$. The asymptotic behaviour of θ_n^* was widely studied (see, among others, [1–9, 11, 12]), but, on a computational point of view, the estimator θ_n^* has a main drawback: its update, from a sample size n to a sample size $n + 1$, is far from being immediate. Applying the stochastic approximation method, Tsybakov [10] introduced the recursive kernel estimator of θ defined as

$$T_n = T_{n-1} + \gamma_n \frac{1}{h_n^{d+1}} \nabla K\left(\frac{T_{n-1} - X_n}{h_n}\right), \tag{1.3}$$

where $T_0 \in \mathbb{R}^d$ is arbitrarily chosen and the stepsize (γ_n) is a sequence of positive real numbers going to zero. The great property of this estimator is that its update is very rapid. Unfortunately, for reasons inherent to stochastic approximation algorithms properties, very strong assumptions on the density f must be required to ensure its consistency. A recursive version f_n of Rosenblatt's density estimator was introduced by Wolverton and Wagner [19] (and discussed, among others, by Yamato [20], Davies [21], Devroye [22], Menon et al. [23],

Wertz [24], Wegman and Davies [25], Roussas [26], and Mokkadem et al. [27]). Let us recall that f_n is defined as

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^d}K\left(\frac{x-X_i}{h_i}\right). \tag{1.4}$$

Its update from a sample of size n to one of size $n+1$ is immediate since f_n clearly satisfies the recursive relation

$$f_n(x) = \left(1 - \frac{1}{n}\right)f_{n-1}(x) + \frac{1}{nh_n^d}K\left(\frac{x-X_n}{h_n}\right). \tag{1.5}$$

This property of rapid update of the density estimator is particularly important in the framework of mode estimation, since the number of points where f must be estimated is very large. We thus define a semirecursive version of Parzen's estimator of the location of the mode by using Wolverton-Wagner's recursive density estimator, rather than Rosenblatt's density estimator. More precisely, our estimator θ_n of the location θ of the mode is a random variable satisfying

$$f_n(\theta_n) = \sup_{y \in \mathbb{R}^d} f_n(y). \tag{1.6}$$

Let us now come back to the problem of estimating the size $f(\theta)$ of the mode. The ordinarily used estimator is defined as $\mu_n^* = f_n^*(\theta_n^*)$ (f_n^* being Rosenblatt's density estimator and θ_n^* Parzen's mode estimator); the consistency of μ_n^* is sufficient to allow the construction of confidence regions for θ (see, e.g., Romano [9]). Adapting the construction of μ_n^* to the semirecursive framework would lead us to estimate $f(\theta)$ by

$$\mu_n = f_n(\theta_n). \tag{1.7}$$

However, this estimator has two main drawbacks (as well as μ_n^*). First, the use of a higher order kernel K is necessary for $(\mu_n - \mu)$ to satisfy a central limit theorem and thus for the construction of confidence intervals of μ (and of confidence regions for (θ, μ)). Moreover, in the case when a higher order kernel is used, it is not possible to choose a bandwidth for which both estimators θ_n and μ_n converge at the optimal rate. These observations lead us to use two different bandwidths, one for the estimation of θ, the other one for the estimation of μ. More precisely, let $\widetilde{f}_n$ be the recursive kernel density estimator defined as

$$\widetilde{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\widetilde{h}_i^d}K\left(\frac{x-X_i}{\widetilde{h}_i}\right), \tag{1.8}$$

where the bandwidth $(\widetilde{h}_n)$ may be different from (h_n) used in the definition of f_n (see (1.4)); we estimate the size of the mode by

$$\widetilde{\mu}_n = \widetilde{f}_n(\theta_n), \tag{1.9}$$

where θ_n is still defined by (1.6) and thus with the first bandwidth (h_n).

The purpose of this paper is the study of the joint asymptotic behaviour of θ_n and $\tilde{\mu}_n$. We first prove the strong consistency of both estimators. We then establish the joint weak convergence rate of θ_n and $\tilde{\mu}_n$. We prove in particular that adequate choices of the bandwidths lead to the asymptotic normality and independence of these estimators and that the use of different bandwidths allow obtaining simultaneously the optimal convergence rate of both estimators. We then apply our weak convergence rate result to the construction of confidence regions for (θ, μ) and illustrate this application with a simulations study. This application enlightens the advantage of using semirecursive estimators rather than nonrecursive estimators. It also shows how the estimation of the size of the mode gives information on the relevance of estimating its location. Finally, we establish the joint strong convergence rate of θ_n and $\tilde{\mu}_n$.

2. Assumptions and Main Results

Throughout this paper, (h_n) and $(\tilde{h}_n)$ are defined as $h_n = h(n)$ and $\tilde{h}_n = \tilde{h}(n)$ for all $n \geq 1$, where h and $\tilde{h}$ are two positive functions.

2.1. Strong Consistency

The conditions we require for the strong consistency of θ_n and $\tilde{\mu}_n$ are the following.

(A1)(i) K is an integrable, differentiable, and even function such that $\int_{\mathbb{R}^d} K(z)dz = 1$.

   (ii) There exists $\zeta > 0$ such that $\int_{\mathbb{R}^d} \|z\|^{\zeta} |K(z)| dz < \infty$.

   (iii) K is Hölder continuous.

   (iv) There exists $\gamma > 0$ such that $z \mapsto \|z\|^{\gamma} |K(z)|$ is a bounded function.

(A2)(i) f is uniformly continuous on $\mathbb{R}^d$.

   (ii) There exists $\xi > 0$ such that $\int_{\mathbb{R}^d} \|x\|^{\xi} f(x) dx < \infty$.

   (iii) There exists $\eta > 0$ such that $z \mapsto \|z\|^{\eta} f(z)$ is a bounded function.

   (iv) There exists $\theta \in \mathbb{R}^d$ such that $f(x) < f(\theta)$ for all $x \neq \theta$.

(A3)(i) The function h is locally bounded and varies regularly with exponent $(-a)$ with $a \in]0, 1/d[$.

   (ii) The function $\tilde{h}$ is locally bounded and varies regularly with exponent $(-\tilde{a})$ with $\tilde{a} \in]0, 1/d[$.

Remark 2.1. Note that (A1)(iv) implies that K is bounded.

Remark 2.2. Let us recall that a positive function (not necessarily monotone) $\mathcal{L}$ defined on $]0, \infty[$ is slowly varying if $\lim_{t \to \infty} \mathcal{L}(tx)/\mathcal{L}(t) = 1$ and that a function G varies regularly with exponent ρ, $\rho \in \mathbb{R}$, if and only if it is of the form $G(x) = x^{\rho} \mathcal{L}(x)$ with $\mathcal{L}$ slowly varying (see, e.g., Feller [28, page 275]). Typical examples of regularly varying functions are x^{ρ}, $x^{\rho} \log x$, $x^{\rho} \log \log x$, $x^{\rho} \log x / \log \log x$, and so on.

Proposition 2.3. *Let θ_n and $\tilde{\mu}_n$ be defined by (1.6) and (1.9), respectively.*

   (i) *Under (A1), (A2), and (A3)(i), $\lim_{n\to\infty}\theta_n = \theta$ a.s.*

   (ii) *Under (A1)–(A3), $\lim_{n\to\infty}\tilde{\mu}_n = \mu$ a.s.*

Let us mention that the assumptions required on the probability density to establish the strong consistency of the semirecursive estimator of the location of the mode are slightly stronger than those needed for the nonrecursive estimator θ_n^* (see, e.g., [9, 12]), but are much weaker than those needed for the recursive estimator (see [10]). Let us also note that the strong consistency of $\tilde{\mu}_n^*$ can be proved in the same way as that of $\tilde{\mu}_n$.

2.2. Weak Convergence Rate

In order to state the weak convergence rate of θ_n and $\tilde{\mu}_n$, we need the following additional assumptions on K, f, h, and $\tilde{h}$.

   (A4)(i) K is twice differentiable on $\mathbb{R}^d$.

   (ii) $z \mapsto z\nabla K(z)$ is integrable.

   (iii) For any $(i,j) \in \{1,\ldots,d\}^2$, $\partial^2 K/\partial x_i\partial x_j$ is bounded integrable and Hölder continuous.

   (iv) K is a kernel of order $q \geq 2$, that is, for all $s \in \{1,\ldots,q-1\}$, for all $j \in \{1,\ldots,d\}$, $\int_{\mathbb{R}^d} y_j^s K(y)dy_j = 0$ and $\int_{\mathbb{R}^d} |y_j^q K(y)|dy < \infty$.

   (A5)(i) $D^2 f(\theta)$ is nonsingular.

   (ii) $D^2 f$ is q-times differentiable; ∇f and $D^q f$ are bounded.

   (iii) For any $(i,j) \in \{1,\ldots,d\}^2$, $\sup_{x\in\mathbb{R}^d}\|D^q(\partial^2 f/\partial x_i\partial x_j)\| < \infty$, and for any $k \in \{1,\ldots,d\}$, $\sup_{x\in\mathbb{R}^d}\|D^q(\partial f/\partial x_k)\| < \infty$.

   (A6)(i) $a \in]0, 1/(d+4)[$.

   (ii) $\tilde{a} \in]0, 1/(d+2)[$.

Remark 2.4. Note that (A4)(ii) and (A4)(iii) imply that ∇K is Lipschitz-continuous and integrable; it is thus straightforward to see that $\lim_{\|x\|\to\infty}\|\nabla K(x)\| = 0$ (and in particular ∇K is bounded).

Let G be the $d \times d$ matrix defined by $G^{(i,j)} = \int_{\mathbb{R}^d}(\partial K/\partial x_i)(x)(\partial K/\partial x_j)(x)dx$ and set

$$\beta_j^q = \int_{\mathbb{R}^d} y_j^q K(y)dy,$$

$$B_{a,q}^{(\theta)} = \begin{cases} \dfrac{(-1)^q}{q!(1-aq)}\nabla\left(\displaystyle\sum_{j=1}^d \beta_j^q \dfrac{\partial^q f}{\partial x_j^q}(\theta)\right) & \text{if } aq < 1, \\ 0 & \text{otherwise,} \end{cases} \tag{2.1}$$

$$\Sigma_a^{(\theta)} = \frac{f(\theta)}{1 + a(d+2)}G.$$

The following theorem gives the weak convergence rate of the semirecursive kernel mode estimator θ_n.

Theorem 2.5 (Weak convergence rate of θ_n). *Let θ_n be defined by (1.6), and assume that (A1), (A2), (A3)(i), (A4), (A5), and (A6)(i) hold.*

(1) *If there exists $c \geq 0$ such that $\lim_{n \to \infty} nh_n^{d+2+2q} = c$, then*

$$\sqrt{nh_n^{d+2}}(\theta_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}\left(-\sqrt{c}\left[D^2 f(\theta)\right]^{-1} B_{a,q}^{(\theta)}, \left[D^2 f(\theta)\right]^{-1} \Sigma_a^{(\theta)} \left[D^2 f(\theta)\right]^{-1}\right). \qquad (2.2)$$

(2) *If $\lim_{n \to \infty} nh_n^{d+2+2q} = \infty$, then*

$$h_n^{-q}(\theta_n - \theta) \xrightarrow{\mathbb{P}} -\left[D^2 f(\theta)\right]^{-1} B_{a,q}^{(\theta)}. \qquad (2.3)$$

Remark 2.6. If $\lim_{n \to \infty} nh_n^{d+2+2q} = c \neq 0$ or if $\lim_{n \to \infty} nh_n^{d+2+2q} = \infty$, then $aq < 1$, and thus $B_{a,q}^{(\theta)}$ is usually nonzero.

In order to compare the semirecursive estimator θ_n with the well-known nonrecursive Parzen's estimator θ_n^*, let us recall that Theorem 2.5 holds when θ_n is replaced by θ_n^* and a by zero (see e.g., Parzen [1] in the case $d = 1$, and Mokkadem and Pelletier [12] in the case $d \geq 1$). The main advantage of θ_n on θ_n^* is that, due to the factor $[1 + a(d+2)]^{-1}$ standing in the definition of $\Sigma_a^{(\theta)}$, the asymptotic covariance of θ_n^* is smaller than that of θ_n; this property will be discussed again in Section 2.3.

In order to state the weak convergence rate of $\tilde{\mu}_n$, we need the following notation (where $\gamma > 0$ and where (b_n) is a positive sequence):

$$w(b_n, \gamma) = \frac{(\log n)^{1+\gamma}}{nb_n^{d+2}} + b_n^{2q},$$

$$P_n(\gamma) = \left[w(h_n, \gamma)\right]^2 + w(h_n, \gamma) w\left(\tilde{h}_n, \gamma\right),$$

$$B_{\tilde{a},q}^{(\mu)} = \begin{cases} \dfrac{(-1)^q}{q!(1 - \tilde{a}q)} \sum_{j=1}^{d} \beta_j^q \dfrac{\partial^q f}{\partial x_j^q}(\theta) & \text{if } \tilde{a}q < 1, \\ 0 & \text{otherwise,} \end{cases} \qquad (2.4)$$

$$\Sigma_{\tilde{a}}^{(\mu)} = \frac{f(\theta)}{1 + \tilde{a}d} \int_{\mathbb{R}^d} K^2(z)dz.$$

Theorem 2.7 (Weak convergence rate of $\tilde{\mu}_n$). *Let $\tilde{\mu}_n$ be defined by (1.9), and assume that (A1)–(A6) hold.*

(1) *If there exists $\tilde{c} \geq 0$ such that $\lim_{n \to \infty} n\tilde{h}_n^{d+2q} = \tilde{c}$ and if there exists $\gamma > 0$ such that $\lim_{n \to \infty} n\tilde{h}_n^d P_n(\gamma) = 0$, then*

$$\sqrt{n\tilde{h}_n^d}(\tilde{\mu}_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}\left(-\sqrt{\tilde{c}} B_{\tilde{a},q}^{(\mu)}, \Sigma_{\tilde{a}}^{(\mu)}\right). \qquad (2.5)$$

(2) If $\lim_{n\to\infty} n\widetilde{h}_n^{d+2q} = \infty$ and if there exists $\gamma > 0$ such that $\lim_{n\to\infty} \widetilde{h}_n^{-2q} P_n(\gamma) = 0$, then

$$\widetilde{h}_n^{-q}(\widetilde{\mu}_n - \mu) \xrightarrow{\mathbb{P}} B_{\widetilde{a},q}^{(\mu)}. \tag{2.6}$$

Remark 2.8. If $\lim_{n\to\infty} n\widetilde{h}_n^{d+2q} = \widetilde{c} \neq$ or if $\lim_{n\to\infty} n\widetilde{h}_n^{d+2q} = \infty$, then $\widetilde{a}q < 1$, so that $B_{\widetilde{a},q}^{(\mu)}$ is usually nonzero.

Remark 2.9. Following the proof of Theorem 2.7, it can be shown that the results of Theorem 2.7 also hold when $\widetilde{\mu}_n$ is replaced by $\widetilde{\mu}_n^*$ and $\widetilde{a}$ by zero.

Let us consider the case when $(h_n) = (\widetilde{h}_n)$, that is, the case when the same bandwidth is used to define θ_n and $\widetilde{\mu}_n$ (or θ_n^* and $\widetilde{\mu}_n^*$). If K is a two-order kernel (i.e., if $q = 2$), then the condition $\lim_{n\to\infty} n\widetilde{h}_n^{d+2q} = \widetilde{c} \geq 0$ implies that $\lim_{n\to\infty} n\widetilde{h}_n^d P_n(\gamma) = \infty$ for all $\gamma > 0$, so that the condition required in Part 1 of Theorem 2.7 is not satisfied: the limit of $(\widetilde{\mu}_n - \mu)$ (or of $(\widetilde{\mu}_n^* - \mu)$) suitably normalized is then necessary degenerated. This is the first main drawback of using the same bandwidth to estimate the location and the size of the mode: to construct confidence intervals for μ, the use of higher-order kernels is unavoidable.

The estimation of the size of the mode is of course not independent of the estimation of the location, since the estimator $\widetilde{\mu}_n$ is constructed with the help of the estimator θ_n. To get a good estimation of the size of the mode, it seems obvious that θ_n should be computed with a bandwidth (h_n) leading to its optimal convergence rate (or, at least, to a convergence rate close to the optimal one). The main information given by Theorem 2.7 is that, for $\widetilde{\mu}_n$ to converge at the optimal rate, the use of a second bandwidth $(\widetilde{h}_n)$ is then necessary.

Now, set

$$B_{a,\widetilde{a},q} = \begin{pmatrix} B_{a,q}^{(\theta)} \\ B_{\widetilde{a},q}^{(\mu)} \end{pmatrix}, \qquad \Sigma_{a,\widetilde{a}} = \begin{pmatrix} \Sigma_a^{(\theta)} & 0 \\ 0 & \Sigma_{\widetilde{a}}^{(\mu)} \end{pmatrix}, \qquad A = \begin{pmatrix} -[D^2 f(\theta)]^{-1} & 0 \\ 0 & 1 \end{pmatrix}, \tag{2.7}$$

and, for any $c, \widetilde{c} \geq 0$, $D(c,\widetilde{c}) = \begin{pmatrix} \sqrt{c}I_d & 0 \\ 0 & \sqrt{\widetilde{c}} \end{pmatrix}$ where I_d is the $d \times d$ identity matrix. The following corollary gives a central limit theorem for the couple $(\theta_n, \widetilde{\mu}_n)$.

Corollary 2.10 (Joint weak convergence rate of θ_n and $\widetilde{\mu}_n$). *Let θ_n and $\widetilde{\mu}_n$ be defined by (1.6) and (1.9), respectively, and let the assumptions of the first parts of Theorems 2.5 and 2.7 hold. Then,*

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n - \theta) \\ \sqrt{n\widetilde{h}_n^d}(\widetilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathfrak{D}} \mathcal{N}(D(c,\widetilde{c})AB_{a,\widetilde{a},q}, A\Sigma_{a,\widetilde{a}}A). \tag{2.8}$$

Remark 2.11. Following the proof of Corollary 2.10, it can be shown that the result of Corollary 2.10 also holds when θ_n is replaced by θ_n^*, $\widetilde{\mu}_n$ by $\widetilde{\mu}_n^*$, and a and $\widetilde{a}$ by zero.

Let us enlighten that, in view of Corollary 2.10, in the case when the couple $(\theta_n, \widetilde{\mu}_n)$ satisfies a central limit theorem, the estimators θ_n and $\widetilde{\mu}_n$ are asymptotically independent, although, in its definition, the estimator of the size of the mode is heavily connected to that

of the location of the mode. This property is not quite surprising since, as pointed out by Vieu [11], the location of the mode gives information on the shape of the density derivative, whereas the size of the mode gives information on the shape of the density itself. This constatation must be related to the fact that the weak (and strong) convergence rate of θ_n is given by that of the gradient of f_n, whereas the weak (and strong) convergence rate of $\tilde{\mu}_n$ is given by that of $\tilde{f}_n$ itself; the variance of the density estimators converging to zero faster than that of the estimators of the density derivatives, the asymptotic independence of θ_n and $\tilde{\mu}_n$ is completely explained.

Let us finally say one word on our assumptions on the bandwidths. In the framework of nonrecursive estimation, there is no need to assume that (h_n) and $(\tilde{h}_n)$ are regularly varying sequences. In the case of semirecursive estimation, this assumption cannot obviously be omitted, since the exponents a and $\tilde{a}$ stand in the expressions of the asymptotic bias $B_{a,\tilde{a},q}$ and variance $\Sigma_{a,\tilde{a}}$. This might be seen as a slight inconvenient of semirecursive estimation; however, as it is enlightened in the following section, it turns out to be an advantage, since the asymptotic variances of the semirecursive estimators are smaller than those of the nonrecursive estimators.

2.3. Construction of Confidence Regions and Simulations Studies

The application of Theorems 2.5, 2.7, and Corollary 2.10 allows the construction of confidence regions of the location and of the size of the mode, as well as confidence ellipsoids of the couple (θ, μ). Hall [29] shows that, in order to construct confidence regions, avoiding bias estimation by a slight undersmoothing is more efficient than explicit bias correction. In the framework of undersmoothing, the asymptotic bias of the estimator is negligible in front of its asymptotic variance; according to the estimation by confidence regions point of view, the parameter to minimize is thus the asymptotic variance. Now, set

$$\Sigma^* = \begin{pmatrix} [1 + a(d+2)]I_d & 0 \\ 0 & [1 + \tilde{a}d] \end{pmatrix} \Sigma_{a,\tilde{a}}, \tag{2.9}$$

and note that, in view of Corollary 2.10 and of Remark 2.11, $A\Sigma_{a,\tilde{a}}A$ (resp. $A\Sigma^*A$) is the asymptotic covariance matrix of the semirecursive estimators $(\theta_n, \tilde{\mu}_n)$ (resp., of the nonrecursive estimators $(\theta_n^*, \tilde{\mu}_n^*)$). In order to construct confidence regions for the location and/or size of the mode, it is thus much preferable to use semirecursive estimators rather than nonrecursive estimators. Simulations studies confirm this theoretical conclusion, whatever the parameter (θ, μ or (θ, μ)) for which confidence regions are constructed is. For sake of succinctness, we do not give all these simulations results here but focus on the construction of confidence ellipsoid for (θ, μ); the aim of this example is of course to enlighten the advantage of using semirecursive estimators rather than nonrecursive estimators but also to show how this confidence region gives information on the shape of the density and consequently allows measuring the pertinence of the parameter location of the mode.

To construct confidence regions for (θ, μ), we consider the case $d = 1$. The following corollary is a straightforward consequence of Corollary 2.10.

Corollary 2.12. *Let θ_n and $\tilde{\mu}_n$ be defined by (1.6) and (1.9), respectively, and let the assumptions of the first parts of Theorems 2.5 and 2.7 hold. We then have*

$$\frac{(1+3a)nh_n^3\left[f''(\theta)\right]^2}{f(\theta)\int_{\mathbb{R}}K'^2(x)dx}(\theta_n-\theta)^2 + \frac{(1+\tilde{a})n\tilde{h}_n}{f(\theta)\int_{\mathbb{R}}K^2(x)dx}(\tilde{\mu}_n-\mu)^2 \xrightarrow{\mathcal{D}} \chi^2(2). \qquad (2.10)$$

Moreover, (2.10) *still holds when the parameters $f(\theta)$ and $f''(\theta)$ are replaced by consistent estimators.*

Remark 2.13. In view of **Remark 2.11**, in the case when the nonrecursive estimators θ_n^* and $\tilde{\mu}_n^*$ are used, (2.10) becomes

$$\frac{nh_n^3\left[f''(\theta)\right]^2}{f(\theta)\int_{\mathbb{R}}K'^2(x)dx}(\theta_n^*-\theta)^2 + \frac{n\tilde{h}_n}{f(\theta)\int_{\mathbb{R}}K^2(x)dx}(\tilde{\mu}_n^*-\mu)^2 \xrightarrow{\mathcal{D}} \chi^2(2) \qquad (2.11)$$

(and, again, this convergence still holds when the parameters $f(\theta)$ and $f''(\theta)$ are replaced by consistent estimators).

Let $\check{f}_n''$ (resp., $\check{f}_n^{*''}$) be the recursive estimator (resp., the nonrecursive Rosenblatt's estimator) of f'' computed with the help of a bandwidth $\check{h}_n$, and set

$$P_n = \frac{(1+3a)nh_n^3\left[\check{f}_n''(\theta_n)\right]^2}{\tilde{f}_n(\theta_n)\int_{\mathbb{R}}K'^2(x)dx}, \qquad Q_n = \frac{(1+\tilde{a})n\tilde{h}_n}{\tilde{f}_n(\theta_n)\int_{\mathbb{R}}K^2(x)dx},$$

$$\qquad\qquad (2.12)$$

$$P_n^* = \frac{nh_n^3\left[\check{f}_n^{*''}(\theta_n^*)\right]^2}{\tilde{f}_n^*(\theta_n^*)\int_{\mathbb{R}}K'^2(x)dx}, \qquad Q_n^* = \frac{n\tilde{h}_n}{\tilde{f}_n^*(\theta_n^*)\int_{\mathbb{R}}K^2(x)dx}.$$

Moreover, let c_α be such that $\mathbb{P}(Z \leq c_\alpha) = 1 - \alpha$, where Z is $\chi^2(2)$-distributed; in view of **Corollary 2.12** and **Remark 2.13**, the sets

$$\mathcal{E}_\alpha = \left\{\frac{(\theta,\mu)}{P_n(\theta_n-\theta)^2 + Q_n(\tilde{\mu}_n-\mu)^2} \leq c_\alpha\right\}$$

$$\qquad\qquad (2.13)$$

$$\mathcal{E}_\alpha^* = \left\{\frac{(\theta,\mu)}{P_n^*(\theta_n^*-\theta)^2 + Q_n^*(\tilde{\mu}_n^*-\mu)^2} \leq c_\alpha\right\}$$

are confidence ellipsoids for (θ,μ) with asymptotic coverage level $1 - \alpha$. Let us dwell on the fact that both confidence regions have the same asymptotic level, but the lengths of the axes of the first one (constructed with the help of the semirecursive estimators θ_n and $\tilde{\mu}_n$) are smaller than those of the second one (constructed with the help of the nonrecursive estimators θ_n^* and $\tilde{\mu}_n^*$).

We now present simulations results. In order to see the relationship between the shape of the confidence ellipsoids and that of the density, the density f we consider is the density of the $\mathcal{N}(0,\sigma^2)$-distribution, the parameter σ taking the values 0.3, 0.4, 0.5, 0.7, 0.75, 1, 1.5, 2, and

2.5. We use the sample size $n = 100$ and the coverage level $1 - \alpha = 95\%$ (and thus $c_\alpha = 5.99$). In each case, the number of simulations is $N = 5000$. The kernel we use is the standard Gaussian density; the bandwidths are

$$h_n = \frac{n^{-1/7}}{(\log n)}, \qquad \widetilde{h}_n = \frac{n^{-1/5}}{(\log n)}, \qquad \check{h}_n = n^{-1/9}. \tag{2.14}$$

Table 1 gives, for each value of σ, the empirical values of θ_n, θ_n^*, μ_n, μ_n^* (with respect to the 5000 simulations) and

    b: the empirical length of the θ-axis of the confidence ellipsoid $\mathcal{E}_{5\%}$;
    b^*: the empirical length of the θ-axis of the confidence ellipsoid $\mathcal{E}_{5\%}^*$;
    a: the empirical length of the μ-axis of the confidence ellipsoid $\mathcal{E}_{5\%}$;
    a^*: the empirical length of the μ-axis of the confidence ellipsoid $\mathcal{E}_{5\%}^*$;
    p: the empirical coverage level of the confidence ellipsoid $\mathcal{E}_{5\%}$;
    p^*: the empirical coverage level of the confidence ellipsoid $\mathcal{E}_{5\%}^*$.

    Confirming our theoretical results, we see that the empirical coverage levels of both confidence ellipsoids $\mathcal{E}_{5\%}^*$ and $\mathcal{E}_{5\%}$ are similar but that the empirical areas of the ellipsoids $\mathcal{E}_{5\%}$ (constructed with the help of the semirecursive estimators) are always smaller than those of the ellipsoids $\mathcal{E}_{5\%}^*$ (constructed with the help of the nonrecursive estimators).

    Let us now discuss the interest of the estimation of the size of the mode and that of the joint estimation of the location and size of the mode. Both estimations give information on the shape of the probability density and, consequently, allow measuring the pertinence of the parameter location of the mode. Of course, the parameter θ is significant only in the case when the high of the peak is large enough; since we consider here the example of the $\mathcal{N}(0, \sigma^2)$-distribution, this corresponds to the case when σ is small enough. Estimating only the size of the mode gives a first idea of the shape of the density around the location of the mode (for instance, when the size is estimated around 0.16, it is clear that the density is very flat). Now, the shape of the confidence ellipsoids allows getting a more precise idea. As a matter of fact, for small values of σ, the length of the μ-axis is larger than that of the θ-axis; as σ increases, the length of the μ-axis decreases, and the one of the θ-axis increases (for $\sigma = 2.5$, the length of the θ-axis is larger than 20 times the one of the μ-axis). Let us underline that these variations of the lengths of the axes are not due to bad estimations results; Table 2 gives the values of the lengths b (resp., b^*) of the θ-axis, a (resp., a^*) of the μ-axis of the ellipsoids computed with the semirecursive estimators θ_n and $\widetilde{\mu}_n$ (resp., with the nonrecursive estimators θ_n^* and $\widetilde{\mu}_n^*$) in the case when the true values of the parameters $f(\theta)$ and $f''(\theta)$ are used (that is, by straightforwardly applying (2.10) and (2.11)).

2.4. Strong Convergence Rate

To establish the strong convergence rate of θ_n and $\widetilde{\mu}_n$, we need the following additional assumption.

  (A7)(i) h is differentiable, and its derivative varies regularly with exponent $(-a - 1)$.

      (ii) $\widetilde{h}$ is differentiable, and its derivative varies regularly with exponent $(-\widetilde{a} - 1)$.

    The following two theorems give the almost sure convergence rate of θ_n and of $\widetilde{\mu}_n$, respectively. Before stating them, let us enlighten that Proposition 2.3 in Mokkadem and Pelletier [12] ensures that the matrix G (and thus the matrix $\Sigma_a^{(\theta)}$) is nonsingular.

Table 1

σ	0.3	0.4	0.5	0.7	0.75	1	1.5	2	2.5
θ_n	−0.002	0.004	0.001	0.003	0.002	0.014	−0.005	−0.009	0.014
θ_n^*	0.003	0.005	0.001	0.005	−0.008	0.016	0.003	−0.020	−0.046
b	1.154	1.346	1.805	2.898	3.160	5.218	10.094	17.866	17.405
b^*	1.166	1.458	1.968	3.300	3.582	5.925	12.943	21.946	23.715
μ_n	1.335	0.989	0.782	0.564	0.522	0.401	0.263	0.196	0.155
μ_n^*	1.312	0.979	0.783	0.562	0.512	0.388	0.269	0.193	0.163
a	0.444	0.399	0.365	0.322	0.315	0.283	0.247	0.224	0.210
a^*	0.514	0.459	0.420	0.369	0.363	0.327	0.287	0.261	0.246
p	98.7%	97.8%	98.2%	98.4%	97.7%	97.8%	97.5%	97.2%	98.4%
p^*	98.6%	98.1%	98.4%	98.2%	96.8%	96.6%	96.9%	97.7%	98.2%

Table 2

σ	0.3	0.4	0.5	0.7	0.75	1	1.5	2	2.5
b	0.159	0.327	0.571	1.357	1.572	3.227	8.895	18.260	31.899
b^*	0.190	0.390	0.682	1.622	1.879	3.858	10.631	21.825	38.127
μ	1.333	0.998	0.798	0.570	0.532	0.399	0.266	0.199	0.159
a	0.465	0.403	0.360	0.303	0.294	0.255	0.208	0.180	0.161
a^*	0.509	0.441	0.395	0.332	0.322	0.279	0.228	0.197	0.176

Theorem 2.14 (Strong convergence rate of θ_n). *Let θ_n be defined by (1.6), and assume that (A1), (A2), (A3)(i), (A4), (A5), (A6)(i), and (A7)(i) hold.*

(1) *If there exists $c \geq 0$ such that $\lim_{n \to \infty} n h_n^{d+2+2q} / [2 \log \log n] = c$, then, with probability one, the sequence $((\sqrt{n h_n^{d+2}} / \sqrt{2 \log \log n})(\theta_n - \theta))$ is relatively compact and its limit set is the ellipsoid*

$$\mathcal{E} = \left\{ v \in \mathbb{R}^d \text{ such that } \left(-D^2 f(\theta) v - \sqrt{c} B_{a,q}^{(\theta)} \right)^T \left[\Sigma_a^{(\theta)} \right]^{-1} \left(-D^2 f(\theta) v - \sqrt{c} B_{a,q}^{(\theta)} \right) \leq 1 \right\}. \quad (2.15)$$

(2) *If $\lim_{n \to \infty} n h_n^{d+2+2q} / [2 \log \log n] = \infty$, then, with probability one, $\lim_{n \to \infty} h_n^{-q}(\theta_n - \theta) = -[D^2 f(\theta)]^{-1} B_{a,q}^{(\theta)}$.*

Theorem 2.15 (Strong convergence rate of $\tilde{\mu}_n$). *Let $\tilde{\mu}_n$ be defined by (1.9), and assume that (A1)– (A7) hold.*

(1) *If there exists $\tilde{c} \geq 0$ such that $\lim_{n \to \infty} n \tilde{h}_n^{d+2q} / [2 \log \log n] = \tilde{c}$ and if there exists $\gamma > 0$ such that $\lim_{n \to \infty} n \tilde{h}_n^d P_n(\gamma) / \log \log n = 0$, then, with probability one, the sequence*

$$\left(\frac{\sqrt{n \tilde{h}_n^d}}{\sqrt{2 \log \log n}} (\tilde{\mu}_n - \mu) \right) \quad (2.16)$$

is relatively compact, and its limit set is the interval $[\sqrt{\tilde{c}} B_{\tilde{a},q}^{(\mu)} - \sqrt{\Sigma_{\tilde{a}}^{(\mu)}}; \sqrt{\tilde{c}} B_{\tilde{a},q}^{(\mu)} + \sqrt{\Sigma_{\tilde{a}}^{(\mu)}}]$.

(2) *If* $\lim_{n\to\infty} n\widetilde{h}_n^{d+2q} / \log\log n = \infty$ *and if there exists* $\gamma > 0$ *such that* $\lim_{n\to\infty} \widetilde{h}_n^{-2q} P_n(\gamma) = 0$, *then, with probability one,* $\lim_{n\to\infty} \widetilde{h}_n^{-q}(\widetilde{\mu}_n - \mu) = B_{\widetilde{a},q}^{(\mu)}$.

To establish a law of the iterated logarithm for the couple $(\theta_n, \widetilde{\mu}_n)$, we need the following additional assumption.

(A8) There exists $n_0 \in \mathbb{N}$ such that

$$n \geq m \geq n_0 \Longrightarrow \max\left\{ \frac{mh_m^{-(d+2)}}{nh_n^{-(d+2)}} ; \frac{m\widetilde{h}_m^{-d}}{n\widetilde{h}_n^{-d}} \right\} = \frac{\min\left\{ mh_m^{-(d+2)} ; m\widetilde{h}_m^{-d} \right\}}{\min\left\{ nh_n^{-(d+2)} ; n\widetilde{h}_n^{-d} \right\}}. \tag{2.17}$$

Remark 2.16. Assumption (A8) holds when $a \neq \widetilde{a}$. In the case when $a = \widetilde{a}$, set $\mathcal{L}_\theta(n) = n^a h_n$ and $\mathcal{L}_\mu(n) = n^a \widetilde{h}_n$; (A8) is then satisfied when $\mathcal{L}_\theta(n) = (\mathcal{L}_\mu(n))^{d/(d+2)}$ for n large enough.

Corollary 2.17 (Joint strong convergence rate of θ_n and $\widetilde{\mu}_n$). *Let* θ_n *and* $\widetilde{\mu}_n$ *be defined by (1.6) and (1.9), respectively; let the assumptions of Parts 1 of Theorems 2.14 and 2.15 hold, as well as (A8). Then, with probability one, the sequence*

$$\frac{1}{\sqrt{2\log\log n}} \left(\begin{array}{c} \sqrt{nh_n^{d+2}}(\theta_n - \theta) \\ \sqrt{n\widetilde{h}_n^{d}}(\widetilde{\mu}_n - \mu) \end{array} \right) \tag{2.18}$$

is relatively compact, and its limit set is the ellipsoid

$$\mathcal{E} = \left\{ v \in \mathbb{R}^{d+1} \text{ such that } \left(A^{-1}v - D(c,\widetilde{c})B_{a,\widetilde{a},q} \right)^T \Sigma_{a,\widetilde{a}}^{-1} \left(A^{-1}v - D(c,\widetilde{c})B_{a,\widetilde{a},q} \right) \leq 1 \right\}. \tag{2.19}$$

Laws of the iterated logarithm for Parzen's nonrecursive kernel mode estimator in the multivariate framework were established by Mokkadem and Pelletier [12]. The technics of demonstration used in the framework of nonrecursive estimators are totally different from those employed to prove Theorem 2.14. This is due to the following fundamental difference between the nonrecursive estimator θ_n^* and the semirecursive estimator θ_n: the study of the asymptotic behaviour of θ_n^* comes down to the one of a *triangular sum* of independent variables, whereas the study of the asymptotic behaviour of θ_n reduces to the one of a *sum* of independent variables. Of course, this difference is not quite important for the study of the weak convergence rate. But, for the study of the strong convergence rate, it makes the case of the semirecursive estimation much easier than the case of the nonrecursive estimation. In particular, on the opposite to the weak convergence rate, the joint strong convergence rate of the nonrecursive estimators θ_n^* and $\widetilde{\mu}_n^*$ cannot be obtained by following the lines of the proof of Theorem 2.14 and remains an open question.

3. Proofs

Let us first note that an important consequence of (A3)(i) is that

$$\text{if } \beta a < 1, \quad \text{then } \lim_{n \to \infty} \frac{1}{nh_n^\beta} \sum_{i=1}^{n} h_i^\beta = \frac{1}{1 - a\beta}. \tag{3.1}$$

Moreover, for all $\varepsilon > 0$,

$$\frac{1}{n} \sum_{i=1}^{n} h_i^q = O\left(h_n^q + n^{\varepsilon - 1}\right). \tag{3.2}$$

As a matter of fact: (i) if $aq < 1$, (3.2) follows easily from (3.1); (ii) if $aq > 1$, since $\sum_i h_i^q$ is summable, (3.2) holds; (iii) if $aq = 1$, then $h_n^q = O(n^{\varepsilon - 1})$, so that $\sum_{i=1}^{n} h_i^q = O(n^\varepsilon)$, and thus (3.2) follows.

Of course, in view of (A3)(ii), (3.1) and (3.2) also hold when (h_n) and a are replaced by $(\widetilde{h}_n)$ and $\widetilde{a}$, respectively.

Our proofs are now organized as follows. Section 3.1 is devoted to the proof of Proposition 2.3. In Section 3.2, we state some preliminary lemmas, which are crucial in the proof of the convergence rates of θ_n and $\widetilde{\mu}_n$ and which are proved in Section 3.6. Section 3.3 is reserved to the proof of Theorems 2.5 and 2.14, Section 3.4 to that of Theorems 2.7 and 2.15, and Section 3.5 to that of Corollaries 2.10 and 2.17.

3.1. Consistency of θ_n and $\widetilde{\mu}_n$: Proof of Proposition 2.3

Since θ_n is the mode of f_n and θ the mode of f, we have

$$0 \le f(\theta) - f(\theta_n) = \left[f(\theta) - f_n(\theta_n)\right] + \left[f_n(\theta_n) - f(\theta_n)\right] \le \left[f(\theta) - f_n(\theta)\right] + \left[f_n(\theta_n) - f(\theta_n)\right]$$
$$\le \left|f(\theta) - f_n(\theta)\right| + \left|f_n(\theta_n) - f(\theta_n)\right| \le 2\|f_n - f\|_\infty. \tag{3.3}$$

The application of Theorem 5 in Mokkadem et al. [27] with $|\alpha| = 0$ and $v_n = \log n$ ensures that, for any $\delta > 0$, there exists $c(\delta) > 0$ such that $\mathbb{P}[(\log n)\|f_n - \mathbb{E}(f_n)\|_\infty \ge \delta] \le \exp(-c(\delta) \sum_{i=1}^{n} h_i^d / (\log n)^2)$. In view of (3.1), since $ad < 1$, we can write

$$n^2 \exp\left(-c(\delta) \frac{\sum_{i=1}^{n} h_i^d}{(\log n)^2}\right) = n^2 \exp\left(-c(\delta) \frac{nh_n^d}{(\log n)^2} \frac{\sum_{i=1}^{n} h_i^d}{nh_n^d}\right) = o(1). \tag{3.4}$$

Borell-Cantelli's Lemma then ensures that $\lim_{n \to \infty} \|f_n - \mathbb{E}(f_n)\|_\infty = 0$ a.s. Since $\lim_{n \to \infty} \|\mathbb{E}(f_n) - f\|_\infty = 0$, it follows from (3.3) that $\lim_{n \to \infty} f(\theta_n) = f(\theta)$ a.s. Since f is continuous, since $\lim_{\|z\| \to \infty} f(z) = 0$, and since θ is the unique mode of f, we deduce that $\lim_{n \to \infty} \theta_n = \theta$ a.s. Now, we have

$$\left|\widetilde{\mu}_n - \mu\right| \le \left|\widetilde{f}_n(\theta_n) - f(\theta_n)\right| + \left|f(\theta_n) - f(\theta)\right| \le \left\|\widetilde{f}_n - f\right\|_\infty + 2\|f_n - f\|_\infty, \tag{3.5}$$

where the last inequality follows from (3.3). Following the proof of the strong uniform convergence of f_n, we show that $\lim_{n \to \infty} \|\widetilde{f}_n - f\|_\infty = 0$ a.s. It follows that $\lim_{n \to \infty} \widetilde{\mu}_n = \mu$ a.s., which concludes the proof of Proposition 2.3.

3.2. Some Preliminary Lemmas

The aim of this section is to state some properties of the density estimators and of their derivatives, which are crucial in the proof of Theorems 2.5–2.15 and of Corollaries 2.10 and 2.17.

3.2.1. Strong Uniform Convergence Rate of the Derivatives of the Density Estimators

For any d-uplet $[\alpha] = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$, set $|\alpha| = \alpha_1 + \cdots + \alpha_d$ and, for any function g, let $\partial^{[\alpha]} g(x) = \partial^{|\alpha|} g / (\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d})(x)$ denote the $[\alpha]$-th partial derivative of g (if $|\alpha| = 0$, then $\partial^{[\alpha]} g = g$). In order to prove Theorems 2.5–2.15 and Corollaries 2.10 and 2.17, we need to know the behaviour of $\partial^{[\alpha]} f_n$ for $|\alpha| \in \{0, 1, 2\}$ and that of $\partial^{[\alpha]} \widetilde{f}_n$ for $|\alpha| \in \{0, 1\}$. For the sake of conciseness, we state the preliminary lemmas 3.1 and 3.2 for $\partial^{[\alpha]} g_n$ where either $g_n = f_n$ and $|\alpha| \in \{0, 1, 2\}$ or $g_n = \widetilde{f}_n$ and $|\alpha| \in \{0, 1\}$. Moreover, we set $(b_n) = (h_n)$ if $g_n = f_n$, and $(b_n) = (\widetilde{h}_n)$ if $g_n = \widetilde{f}_n$.

Lemma 3.1. *Let (A1), (A2), (A3)(i), (A4), (A5), and (A6)(i) hold. Moreover, if $g_n = \widetilde{f}_n$, then let (A3)(ii) and (A6)(ii) hold. We have*

$$\lim_{n \to \infty} \frac{n}{\sum_{i=1}^n b_i^q} \left[\mathbb{E}\left[\partial^{[\alpha]} g_n(x)\right] - \partial^{[\alpha]} f(x) \right] = \frac{(-1)^q}{q!} \partial^{[\alpha]} \left(\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q} \right)(x), \tag{3.6}$$

where β_j^q is defined in (2.1). Moreover, if we set $M_q = \sup_{x \in \mathbb{R}^d} \|D^q \partial^{[\alpha]} f(x)\|$, then

$$\lim_{n \to \infty} \frac{n}{\sum_{i=1}^n b_i^q} \sup_{x \in \mathbb{R}^d} \left| \mathbb{E}\left(\partial^{[\alpha]} g_n(x)\right) - \partial^{[\alpha]} f(x) \right| \le \frac{M_q}{q!} \int_{\mathbb{R}^d} \|z\|^q |K(z)| dz. \tag{3.7}$$

Lemma 3.2. *Let U be a compact set of $\mathbb{R}^d$, and assume that (A1), (A2), (A3)(i), (A4), (A5), and (A6)(i) hold. Moreover, if $g_n = \widetilde{f}_n$, then let (A3)(ii) and (A6)(ii) hold. Then, for all $\gamma > 0$, we have*

$$\sup_{x \in U} \left| \partial^{[\alpha]} g_n(x) - \mathbb{E}\left(\partial^{[\alpha]} g_n(x)\right) \right| = O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{\sum_{i=1}^n b_i^{d+2|\alpha|}}} \right) \quad a.s. \tag{3.8}$$

Lemma 3.1 is proved in Mokkadem et al. [27], Lemma 3.2 in Section 3.6.

3.2.2. Convergence Rate of $(\nabla f_n(\theta) - \mathbb{E}(\nabla f_n(\theta)), \widetilde{f}_n(\theta) - \mathbb{E}(\widetilde{f}_n(\theta)))$

Lemma 3.3.

(1) *Under Assumptions (A1), (A2), (A3)(i), (A4), (A5), and (A6)(i), we have*

$$\sqrt{nh_n^{d+2}}\left[\nabla f_n(\theta) - \mathbb{E}(\nabla f_n(\theta))\right] \xrightarrow{\mathfrak{D}} \mathcal{N}\left(0, \Sigma_a^{(\theta)}\right). \tag{3.9}$$

(2) *Under Assumptions (A1)–(A6), we have*

$$W_n = \begin{pmatrix} \sqrt{nh_n^{d+2}}\left[\nabla f_n(\theta) - \mathbb{E}(\nabla f_n(\theta))\right] \\ \sqrt{n\widetilde{h}_n^d}\left[\widetilde{f}_n(\theta) - \mathbb{E}\left(\widetilde{f}_n(\theta)\right)\right] \end{pmatrix} \xrightarrow{\mathfrak{D}} \mathcal{N}(0, \Sigma_{a,\widetilde{a}}). \tag{3.10}$$

Lemma 3.4.

(1) *Under Assumptions (A1), (A2), (A3)(i), (A4), (A5), (A6)(i), and (A7)(i), with probability one, the sequence $((\sqrt{nh_n^{d+2}}[\nabla f_n(\theta) - \mathbb{E}(\nabla f_n(\theta))])/\sqrt{2\log\log n})$ is relatively compact, and its limit set is $\{\nu \in \mathbb{R}^d \text{ such that } \nu^T [\Sigma_a^{(\theta)}]^{-1} \nu \leq 1\}$.*

(2) *Under Assumptions (A1)–(A7), with probability one, the sequence $((\sqrt{n\widetilde{h}_n^d}[\widetilde{f}_n(\theta) - \mathbb{E}(\widetilde{f}_n(\theta))])/\sqrt{2\log\log n})$ is relatively compact, and its limit set is the interval $[-\sqrt{\Sigma_{\widetilde{a}}^{(\mu)}}; \sqrt{\Sigma_{\widetilde{a}}^{(\mu)}}]$.*

(3) *Under Assumptions (A1)–(A8), with probability one, the sequence*

$$\frac{1}{\sqrt{2\log\log n}} \begin{pmatrix} \sqrt{nh_n^{d+2}}\left[\nabla f_n(\theta) - \mathbb{E}(\nabla f_n(\theta))\right] \\ \sqrt{n\widetilde{h}_n^d}\left[\widetilde{f}_n(\theta) - \mathbb{E}\left(\widetilde{f}_n(\theta)\right)\right] \end{pmatrix} \tag{3.11}$$

is relatively compact, and its limit set is $\{\nu \in \mathbb{R}^{d+1} \text{ such that } \nu^T \Sigma_{a,\widetilde{a}}^{-1} \nu \leq 1\}$.

3.3. Convergence Rate of θ_n: Proof of Theorems 2.5 and 2.14

In order to prove Theorems 2.5 and 2.14, we first show that the weak and strong asymptotic behaviours of $\theta_n - \theta$ are given by those of $-[D^2 f(\theta)]^{-1}\nabla f_n(\theta)$ (see Section 3.3.1) and then deduce the convergence rates of $\theta_n - \theta$ from those of $\nabla f_n(\theta)$ (see Section 3.3.2).

3.3.1. Relationship between $(\theta_n - \theta)$ and $\nabla f_n(\theta)$

By definition of θ_n, we have $\nabla f_n(\theta_n) = 0$, so that

$$\nabla f_n(\theta_n) - \nabla f_n(\theta) = -\nabla f_n(\theta). \tag{3.12}$$

For each $i \in \{1, \ldots, d\}$, a Taylor expansion applied to the real valued application $\partial f_n / \partial x_i$ implies the existence of $\varepsilon_n(i) = (\varepsilon_n^{(1)}(i), \ldots, \varepsilon_n^{(d)}(i))^t$ such that

$$\frac{\partial f_n}{\partial x_i}(\theta_n) - \frac{\partial f_n}{\partial x_i}(\theta) = \sum_{j=1}^{d} \frac{\partial^2 f_n}{\partial x_i \partial x_j}(\varepsilon_n(i))\left(\theta_n^{(j)} - \theta^{(j)}\right),$$

$$\left|\varepsilon_n^{(j)}(i) - \theta^{(j)}\right| \le \left|\theta_n^{(j)}(i) - \theta^{(j)}\right| \quad \forall j \in \{1, \ldots, d\}. \tag{3.13}$$

Define the $d \times d$ matrix $H_n = (H_n^{(i,j)})_{1 \le i,j \le d}$ by setting $H_n^{(i,j)} = (\partial^2 f_n / \partial x_i \partial x_j)(\varepsilon_n(i))$; (3.12) can then be rewritten as

$$H_n(\theta_n - \theta) = -\nabla f_n(\theta). \tag{3.14}$$

Now, let U be a compact set of $\mathbb{R}^d$ containing θ. The combination of Lemmas 3.1 and 3.2 with $|\alpha| = 2$, $g_n = f_n$, and $b_n = h_n$ ensures that, for any $\gamma > 0$ and any $\varepsilon \in]0, 1[$,

$$\sup_{x \in U}\left|\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right| = O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{\sum_{i=1}^{n} h_i^{d+4}} + \frac{\sum_{i=1}^{n} h_i^q}{n}}\right) \text{ a.s.}$$

$$= O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{nh_n^{d+4}}} + h_n^q + n^{\varepsilon-1}\right) = o(1) \text{ a.s.} \tag{3.15}$$

Since $D^2 f$ is continuous in a neighbourhood of θ and since $\lim_{n \to \infty} \theta_n = \theta$ a.s., (3.15) ensures that

$$\lim_{n \to \infty} H_n = D^2 f(\theta) \text{ a.s.} \tag{3.16}$$

In view of (3.14), we can thus state the following lemma.

Lemma 3.5. *Under Assumptions (A1), (A2), (A3)(i), (A4), (A5), and (A6)(i), the weak and strong asymptotic behaviour of $(\theta_n - \theta)^T$ is given by that of $-[D^2 f(\theta)]^{-1} \nabla f_n(\theta)$.*

3.3.2. Proof of Theorems 2.5 and 2.14

Theorem 2.5 (resp., Theorem 2.14) straightforwardly follows from the application of Lemma 3.5 of the first part of Lemma 3.3 (resp., of the first part of Lemma 3.4) and of the following lemma.

Lemma 3.6. *Let (A1), (A2), (A3)(i), (A4), (A5), and (A6)(i) hold.*

(1) *If $\lim_{n \to \infty} nh_n^{d+2+2q} \in]0, \infty]$, then $\lim_{n \to \infty} h_n^{-q} \mathbb{E}(\nabla f_n(\theta)) = B_{a,q}^{(\theta)}$.*

(2) *If $\lim_{n \to \infty} nh_n^{d+2+2q} = 0$, then $\lim_{n \to \infty} \sqrt{nh_n^{d+2}} \mathbb{E}(\nabla f_n(\theta)) = 0$.*

Let us now prove Lemma 3.6. The application of Lemma 3.1 (with $g_n = f_n$ and $|\alpha| = 1$) ensures that

$$\lim_{n \to \infty} \frac{n}{\sum_{i=1}^{n} h_i^q} \mathbb{E}(\nabla f_n(\theta)) = \frac{(-1)^q}{q!} \nabla \left(\sum_{j=1}^{d} \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta) \right). \tag{3.17}$$

Let us first consider the case when $aq < 1$. By application of (3.1), we have

$$\lim_{n \to \infty} h_n^{-q} \mathbb{E}(\nabla f_n(\theta)) = B_{a,q}^{(\theta)}. \tag{3.18}$$

(i) If $\lim_{n \to \infty} n h_n^{d+2+2q} \in]0, \infty]$, then $1 - a(d + 2 + 2q) \geq 0$, and thus $aq < 1/2$; Part 1 of Lemma 3.6 thus follows.

(ii) If $\lim_{n \to \infty} n h_n^{d+2+2q} = 0$, then $\lim_{n \to \infty} \sqrt{n h_n^{d+2}} h_n^q = 0$, and Part 2 of Lemma 3.6 also follows.

It remains to prove Part 2 of Lemma 3.6 in the case when $aq \geq 1$. The application of (3.2) then ensures that, for all $\varepsilon \in]0, 1/2[$,

$$\sqrt{n h_n^{d+2}} \frac{\sum_{i=1}^{n} h_i^q}{n} = O\left(\sqrt{n h_n^{d+2}} \left[h_n^q + n^{\varepsilon-1} \right] \right) = o(1). \tag{3.19}$$

In view of (3.17), Part 2 of Lemma 3.6 then follows.

3.4. Convergence Rate of $\tilde{\mu}_n$: Proof of Theorems 2.7 and 2.15

In order to prove Theorems 2.7 and 2.15, we show that the weak and strong convergence rates of $\tilde{\mu}_n - \mu$ are given by those of $\tilde{f}_n(\theta) - f(\theta)$. More precisely, set

$$R_n = (\tilde{\mu}_n - \mu) - \left(\tilde{f}_n(\theta) - f(\theta) \right), \tag{3.20}$$

and assume that the following two lemmas hold.

Lemma 3.7. *Let (A1)–(A6) hold.*

(1) *If $\lim_{n \to \infty} n \tilde{h}_n^{d+2q} \in]0, \infty]$, then $\lim_{n \to \infty} \tilde{h}_n^{-q} [\mathbb{E}(\tilde{f}_n(\theta)) - f(\theta)] = B_{\tilde{a},q}^{(\mu)}$.*

(2) *If $\lim_{n \to \infty} n \tilde{h}_n^{d+2q} = 0$, then $\lim_{n \to \infty} \sqrt{n \tilde{h}_n^d} [\mathbb{E}(\tilde{f}_n(\theta)) - f(\theta)] = 0$.*

Lemma 3.8. *Let (A1)–(A6) hold. For all $\gamma > 0$, $R_n^2 = O(P_n(\gamma))$ a.s.*

In order to prove Theorem 2.7, we first note that the application of the second part of Lemma 3.3 yields

$$\sqrt{n \tilde{h}_n^d} \left[\tilde{f}_n(\theta) - \mathbb{E}\left(\tilde{f}_n(\theta) \right) \right] \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma_{\tilde{a}}^{(\mu)} \right). \tag{3.21}$$

In view of Lemma 3.7, it follows that if $\lim_{n \to \infty} n\tilde{h}_n^{d+2q} = \tilde{c} \geq 0$, then

$$\sqrt{n\tilde{h}_n^d}\left[\tilde{f}_n(\theta) - f(\theta)\right] \xrightarrow{\mathfrak{D}} \mathcal{N}\left(\sqrt{\tilde{c}}B_{\tilde{a},q}^{(\mu)}, \Sigma_{\tilde{a}}^{(\mu)}\right), \tag{3.22}$$

and if $\lim_{n \to \infty} n\tilde{h}_n^{d+2q} = \infty$, then

$$\tilde{h}_n^{-q}\left[\tilde{f}_n(\theta) - f(\theta)\right] \xrightarrow{\mathbb{P}} B_{\tilde{a},q}^{(\mu)}. \tag{3.23}$$

Part 1 (resp., Part 2) of Theorem 2.7 straightforwardly follows from the combination of (3.22) (resp., of (3.23)) and of Lemma 3.8.

The proof of Theorem 2.15 follows that of Theorem 2.7 (except that the second part of Lemma 3.4 stands instead of (3.21)); this proof is thus omitted.

It remains to prove Lemmas 3.7 and 3.8. The proof of Lemma 3.7 follows that of Lemma 3.6 and is omitted. We now prove Lemma 3.8. We first note that $R_n = \tilde{f}_n(\theta_n) - \tilde{f}_n(\theta)$; a Taylor's expansion implies the existence of ζ_n such that $\|\zeta_n - \theta_n\| \leq \|\theta_n - \theta\|$ and

$$\begin{aligned} R_n &= (\theta_n - \theta)^T \nabla \tilde{f}_n(\zeta_n) \\ &= (\theta_n - \theta)^T\left[\nabla \tilde{f}_n(\zeta_n) - \nabla f(\zeta_n) + \nabla f(\zeta_n) - \nabla f(\theta)\right]. \end{aligned} \tag{3.24}$$

Let $\mathcal{U}$ be a compact set that contains θ; for n large enough, we get

$$\begin{aligned} |R_n| &= O\left(\|\theta_n - \theta\|\left[\sup_{x \in \mathcal{U}}\left\|\nabla \tilde{f}_n(x) - \nabla f(x)\right\| + \|\zeta_n - \theta\|\right]\right) \\ &= O\left(\|\theta_n - \theta\|\sup_{x \in \mathcal{U}}\left\|\nabla \tilde{f}_n(x) - \nabla f(x)\right\| + \|\theta_n - \theta\|^2\right). \end{aligned} \tag{3.25}$$

To get an upper bound of $|R_n|$, we thus need to establish upper bounds of $\|\theta_n - \theta\|$ and of $\sup_{x \in \mathcal{U}}\|\nabla \tilde{f}_n(x) - \nabla f(x)\|$.

(i) Let us recall that the a.s. convergence rate of $(\theta_n - \theta)$ is given by that of $[D^2 f(\theta)]^{-1} \nabla f_n(\theta)$ (see Lemma 3.5). Theorem 2.14 can be applied to obtain the exact a.s. convergence rate of $\theta_n - \theta$. However, to avoid assuming (A7)(i), we apply here Lemmas 3.1 and 3.2 (with $|\alpha| = 1$ and $(g_n, b_n) = (f_n, h_n)$) and get the following upper bound of the a.s. convergence rate of $\nabla f_n(\theta)$ and thus of $\theta_n - \theta$: for any $\gamma > 0$ and $\varepsilon \in]0, (1 + a(d + 2))/2[$,

$$\begin{aligned} \|\theta_n - \theta\|^2 &= O\left(\frac{(\log n)^{1+\gamma}}{\sum_{i=1}^n h_i^{d+2}} + \left[\frac{\sum_{i=1}^n h_i^q}{n}\right]^2\right) = O\left(\frac{(\log n)^{1+\gamma}}{nh_n^{d+2}} + \left[h_n^q + n^{\varepsilon-1}\right]^2\right) \text{ a.s.} \\ &= O(w(h_n, \gamma)) \text{ a.s.} \end{aligned} \tag{3.26}$$

(ii) Noting that

$$\sup_{x\in\mathcal{U}}\left\|\nabla\widetilde{f}_n(x)-\nabla f(x)\right\| \le \sup_{x\in\mathcal{U}}\left\|\nabla\widetilde{f}_n(x)-\mathbb{E}\left(\nabla\widetilde{f}_n(x)\right)\right\| + \sup_{x\in\mathcal{U}}\left\|\mathbb{E}\left(\nabla\widetilde{f}_n(x)\right)-\nabla f(x)\right\|, \qquad (3.27)$$

the application of Lemmas 3.1 and 3.2 with $|\alpha| = 1$, $(g_n, b_n) = (\widetilde{f}_n, \widetilde{h}_n)$ ensures that, for any $\gamma > 0$ and $\varepsilon \in]0, (1 + \widetilde{a}(d+2))/2[$,

$$\begin{aligned}
\sup_{x\in\mathcal{U}}\left\|\nabla\widetilde{f}_n(x)-\nabla f(x)\right\|^2 &= O\left(\frac{(\log n)^{1+\gamma}}{n\widetilde{h}_n^{d+2}} + \left[\frac{\sum_{i=1}^n \widetilde{h}_i^q}{n}\right]^2\right) \text{ a.s.} \\
&= O\left(\frac{(\log n)^{1+\gamma}}{n\widetilde{h}_n^{d+2}} + \left[\widetilde{h}_n^q + n^{\varepsilon-1}\right]^2\right) \text{ a.s.} \\
&= O\left(w\left(\widetilde{h}_n, \gamma\right)\right) \text{ a.s.}
\end{aligned} \qquad (3.28)$$

It follows from (3.25), (3.26), and (3.28) that

$$\begin{aligned}
|R_n|^2 &= O\left(w(h_n, \gamma)w\left(\widetilde{h}_n, \gamma\right) + [w(h_n, \gamma)]^2\right) \text{ a.s.} \\
&= O(P_n(\gamma)) \text{ a.s.,}
\end{aligned} \qquad (3.29)$$

which concludes the proof of Lemma 3.8.

3.5. Joint Convergence Rate of θ_n and $\widetilde{\mu}_n$: Proof of Corollaries 2.10 and 2.17

In view of (3.14) and (3.20), we have

$$\begin{pmatrix} H_n & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \theta_n - \theta \\ \widetilde{\mu}_n - \mu \end{pmatrix} = \begin{pmatrix} -\nabla f_n(\theta) \\ \widetilde{f}_n(\theta) - f(\theta) \end{pmatrix} + \begin{pmatrix} 0 \\ R_n \end{pmatrix}, \qquad (3.30)$$

and, in view of (3.16), the weak and strong asymptotic behaviour of $\begin{pmatrix} \theta_n - \theta \\ \widetilde{\mu}_n - \mu \end{pmatrix}$ is given by that of

$$\begin{aligned}
&\begin{pmatrix} [D^2 f(\theta)]^{-1} & 0 \\ 0 & 1 \end{pmatrix}\left[\begin{pmatrix} -\nabla f_n(\theta) \\ \widetilde{f}_n(\theta) - f(\theta) \end{pmatrix} + \begin{pmatrix} 0 \\ R_n \end{pmatrix}\right] \\
&= \begin{pmatrix} [-D^2 f(\theta)]^{-1}[\nabla f_n(\theta) - \mathbb{E}(\nabla f_n(\theta))] \\ \widetilde{f}_n(\theta) - \mathbb{E}\left(\widetilde{f}_n(\theta)\right) \end{pmatrix} + \begin{pmatrix} [-D^2 f(\theta)]^{-1}\mathbb{E}(\nabla f_n(\theta)) \\ \mathbb{E}\left(\widetilde{f}_n(\theta)\right) - f(\theta) + R_n \end{pmatrix}.
\end{aligned} \qquad (3.31)$$

Under the assumptions of Corollary 2.10, the application of Part 1 of Lemma 3.6, that of Part 1 of Lemma 3.7, and that of Lemma 3.8 ensure that

$$\lim_{n \to \infty} \begin{pmatrix} \sqrt{nh_n^{d+2}} \left[-D^2 f(\theta) \right]^{-1} \mathbb{E}(\nabla f_n(\theta)) \\ \sqrt{n\widetilde{h}_n^d} \left[\mathbb{E}\left(\widetilde{f}_n(\theta) \right) - f(\theta) + R_n \right] \end{pmatrix} = D(c, \widetilde{c}) A B_{a, \widetilde{a}, q}. \tag{3.32}$$

Corollary 2.10 then follows from the application of Part 2 of Lemma 3.3. The proof of Corollary 2.17 follows that of Corollary 2.10 (except that Part 3 of Lemma 3.4 is applied instead of Part 2 of Lemma 3.3) and is omitted.

3.6. Proof of Lemmas 3.2–3.4

3.6.1. Proof of Lemma 3.2

Set $v_n = \left[\sum_{i=1}^n b_i^{d+2|\alpha|} \right]^{1/2} \left[(\log n)^{1+\gamma} \right]^{-1/2}$. Applying Proposition 3 in Mokkadem et al. [27], it holds that for any $\delta > 0$, there exists $c(\delta) > 0$ such that

$$\mathbb{P}\left[\sup_{x \in U} v_n \left| \partial^{[\alpha]} g_n(x) - \mathbb{E}\left(\partial^{[\alpha]} g_n(x) \right) \right| \geq \delta \right] \leq \exp\left(-c(\delta) \frac{\sum_{i=1}^n b_i^{d+2|\alpha|}}{2v_n^2} \right). \tag{3.33}$$

Since $\lim_{n \to \infty} \sum_{i=1}^n b_i^{d+2|\alpha|} / (v_n^2 \log n) = \infty$, we have, for n large enough, $c(\delta) \sum_{i=1}^n b_i^{d+2|\alpha|} / 2v_n^2 \geq 2 \log n$, and Lemma 3.2 follows from the application of Borel-Cantelli's Lemma.

3.6.2. Proof of Lemma 3.3

We only prove the second part of Lemma 3.3: the proof of its first part can be easily deduced. We first establish that

$$\lim_{n \to \infty} \mathbb{E}\left(W_n W_n^T \right) = \Sigma_{a, \widetilde{a}} \tag{3.34}$$

and then check that (W_n) satisfies Lyapounov's condition. Set

$$Y_{k,n} = \frac{1}{\sqrt{nh_n^{-d-2}}} h_k^{-d-1} \left[\nabla K \left(\frac{\theta - X_k}{h_k} \right) - \mathbb{E}\left(\nabla K \left(\frac{\theta - X_k}{h_k} \right) \right) \right],$$

$$Z_{k,n} = \frac{1}{\sqrt{n\widetilde{h}_n^{-d}}} \widetilde{h}_k^{-d} \left[K \left(\frac{\theta - X_k}{\widetilde{h}_k} \right) - \mathbb{E}\left(K \left(\frac{\theta - X_k}{\widetilde{h}_k} \right) \right) \right], \tag{3.35}$$

and note that

$$\mathbb{E}\left(W_n W_n^T \right) = \sum_{k=1}^n \begin{pmatrix} \mathbb{E}\left(Y_{k,n} Y_{k,n}^T \right) & \mathbb{E}\left(Y_{k,n} Z_{k,n} \right) \\ \mathbb{E}\left(Y_{k,n}^T Z_{k,n} \right) & \mathbb{E}\left(Z_{k,n}^2 \right) \end{pmatrix}. \tag{3.36}$$

Now, for any $s, t \in \{1, \ldots, d\}$, we have

$$\mathbb{E}\left[\frac{\partial K}{\partial x_s}\left(\frac{\theta - X_k}{h_k}\right)\frac{\partial K}{\partial x_t}\left(\frac{\theta - X_k}{h_k}\right)\right] = \int_{\mathbb{R}^d} \frac{\partial K}{\partial x_s}\left(\frac{\theta - y}{h_k}\right)\frac{\partial K}{\partial x_t}\left(\frac{\theta - y}{h_k}\right)f(y)dy$$

$$= h_k^d f(\theta)G_{s,t} + o\left(h_k^d\right), \tag{3.37}$$

and since $\mathbb{E}[(\partial K/\partial x_s)(\theta - X_k/h_k)] = O(h_k^d)$, we deduce that

$$\mathbb{E}\left(\left[\nabla K\left(\frac{\theta - X_k}{h_k}\right) - \mathbb{E}\left(\nabla K\left(\frac{\theta - X_k}{h_k}\right)\right)\right]\left[\nabla K\left(\frac{\theta - X_k}{h_k}\right) - \mathbb{E}\left(\nabla K\left(\frac{\theta - X_k}{h_k}\right)\right)\right]^T\right)$$

$$= f(\theta)G h_k^d [1 + o(1)], \tag{3.38}$$

which implies that $\lim_{n \to \infty} \sum_{k=1}^n \mathbb{E}(Y_{k,n} Y_{k,n}^T) = f(\theta)[1 + a(d+2)]^{-1}G$. In the same way, we have

$$\mathbb{E}\left(\left[K\left(\frac{\theta - X_k}{\tilde{h}_k}\right) - \mathbb{E}\left(K\left(\frac{\theta - X_k}{\tilde{h}_k}\right)\right)\right]^2\right) = \tilde{h}_k^d f(\theta) \int_{\mathbb{R}^d} K^2(z)dz[1 + o(1)], \tag{3.39}$$

and thus $\lim_{n \to \infty} \sum_{k=1}^n \mathbb{E}(Z_{k,n}^2) = f(\theta)[1 + \tilde{a}d]^{-1} \int_{\mathbb{R}^d} K^2(z)dz$. Moreover, set $h_n^* = \min(h_n, \tilde{h}_n)$; we have

$$\mathbb{E}\left[\nabla K\left(\frac{\theta - X_k}{h_k}\right)K\left(\frac{\theta - X_k}{\tilde{h}_k}\right)\right] = h_k^{*d} \int_{\mathbb{R}^d} \nabla K\left(\frac{h_k^*}{h_k}z\right)K\left(\frac{h_k^*}{\tilde{h}_k}z\right)f(\theta - h_k^* z)dz. \tag{3.40}$$

Noting that $f(\theta - h_k^* z) = f(\theta) + h_k^* R_k(\theta, z)$ with $|R_k(\theta, z)| \leq \|\nabla f\|_\infty \|z\|$, we get

$$\mathbb{E}\left[\nabla K\left(\frac{\theta - X_k}{h_k}\right)K\left(\frac{\theta - X_k}{\tilde{h}_k}\right)\right]$$

$$= h_k^{*d}\left[f(\theta)\int_{\mathbb{R}^d} \nabla K\left(\frac{h_k^*}{h_k}z\right)K\left(\frac{h_k^*}{\tilde{h}_k}z\right)dz + h_k^* \int_{\mathbb{R}^d} \nabla K\left(\frac{h_k^*}{h_k}z\right)K\left(\frac{h_k^*}{\tilde{h}_k}z\right)R_k(\theta, z)dz\right]. \tag{3.41}$$

Since the function $z \mapsto [\nabla K(z)]K(z)$ is odd (in each coordinate), the first right-handed integral is zero, and since h_k^* equals either h_k or $\tilde{h}_k$, we get

$$\left\|\mathbb{E}\left[\nabla K\left(\frac{\theta - X_k}{h_k}\right)K\left(\frac{\theta - X_k}{\tilde{h}_k}\right)\right]\right\|$$

$$\leq h_k^{*(d+1)}\|\nabla f\|_\infty\left[\|K\|_\infty \int_{\mathbb{R}^d} \|z\|\|\nabla K(z)\|dz + \|\nabla K\|_\infty \int_{\mathbb{R}^d} \|z\||K(z)|dz\right] = O\left(h_k^{*(d+1)}\right). \tag{3.42}$$

We then deduce that

$$
\mathbb{E}\left(\left[\nabla K\left(\frac{\theta-X_k}{h_k}\right)-\mathbb{E}\left(\nabla K\left(\frac{\theta-X_k}{h_k}\right)\right)\right]\left[K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)-\mathbb{E}\left(K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right)\right]\right)
$$
$$
= O\left(\left[\min\left(h_k,\tilde{h}_k\right)\right]^{d+1}\right) + O\left(h_k^d \tilde{h}_k^d\right) = O\left(h_k^{(d+1)/2}\tilde{h}_k^{(d+1)/2}\right),
\tag{3.43}
$$

and thus, in view of (3.1),

$$
\sum_{k=1}^{n}\mathbb{E}(Y_{k,n}Z_{k,n}) = O\left(\frac{1}{\sqrt{\left(nh_n^{-d-2}\right)\left(n\tilde{h}_n^{-d}\right)}}\sum_{k=1}^{n}h_k^{-(d+1)/2}\tilde{h}_k^{(1-d)/2}\right) = o(1),
\tag{3.44}
$$

which concludes the proof of (3.34). Let us now check that (W_n) satisfies the Lyapounov's condition. Set $p > 2$. Since K and ∇K are bounded and integrable, we have $\int_{\mathbb{R}^d}\|\nabla K(z)\|^p dz < \infty$ and $\int_{\mathbb{R}^d}|K(z)|^p dz < \infty$. It follows that

$$
\sum_{k=1}^{n}\mathbb{E}(\|Y_{k,n}\|^p) = O\left(\frac{1}{\left(nh_n^{-d-2}\right)^{p/2}}\sum_{k=1}^{n}h_k^{(-d-1)p}\int_{\mathbb{R}^d}\left\|\nabla K\left(\frac{\theta-y}{h_k}\right)\right\|^p f(y)dy\right)
$$
$$
= O\left(\frac{1}{\left(nh_n^{-d-2}\right)^{p/2}}\sum_{k=1}^{n}h_k^{(-d-1)p}h_k^d\right) = o(1),
$$

$$
\sum_{k=1}^{n}\mathbb{E}(|Z_{k,n}|^p) = O\left(\frac{1}{\left(n\tilde{h}_n^{-d}\right)^{p/2}}\sum_{k=1}^{n}\tilde{h}_k^{-dp}\int_{\mathbb{R}^d}\left|K\left(\frac{\theta-y}{\tilde{h}_k}\right)\right|^p f(y)dy\right)
$$
$$
= O\left(\frac{1}{\left(n\tilde{h}_n^{-d}\right)^{p/2}}\sum_{k=1}^{n}\tilde{h}_k^{-dp}\tilde{h}_k^d\right) = o(1),
\tag{3.45}
$$

which concludes the proof of Lemma 3.3.

3.6.3. Proof of Lemma 3.4

We only prove the third part of **Lemma 3.4**: the proofs of its first two parts can be easily deduced. Set

$$
\Gamma = f(\theta) \begin{pmatrix} G & 0 \\ 0 & \int_{\mathbb{R}^d} K^2(z)dz \end{pmatrix}, \qquad \Delta_n = \begin{pmatrix} \dfrac{1}{\sqrt{nh_n^{-d-2}}} I_d & 0 \\ 0 & \dfrac{1}{\sqrt{n\tilde{h}_n^{-d}}} \end{pmatrix},
$$

$$
Q_n = \begin{pmatrix} \sqrt{h_n^{-d-2}} I_d & 0 \\ 0 & \sqrt{\tilde{h}_n^{-d}} \end{pmatrix},
$$

(3.46)

let (ε_n) be a sequence of $\mathbb{R}^{d+1}$-valued, independent, and $\mathcal{N}(0, \Gamma)$-distributed random vectors, and set $S_n = \sum_{k=1}^{n} Q_k \varepsilon_k$ and $B_n = \mathbb{E}(S_n S_n^T)$. We clearly have $\lim_{n \to +\infty} \Delta_n B_n \Delta_n = \Sigma_{a,\tilde{a}}$. Moreover, all the eigenvalues $\rho_n^{(i)} (1 \le i \le d+1)$ of B_n satisfy $\lim_{n \to +\infty} \log \log \rho_n^{(i)} / \log \log n = 1$. The application of Theorem 1 in Mokkadem and Pelletier [30] then ensures that, with probability one, the sequence

$$
\left(\frac{\Delta_n S_n}{\sqrt{2 \log \log n}} \right)
$$

(3.47)

is relatively compact, and its limit set is the ellipsoid $\{v \in \mathbb{R}^{d+1} \text{ such that } v^T \Sigma_{a,\tilde{a}}^{-1} v \le 1\}$. Now, set

$$
\tilde{V}_k = \begin{pmatrix} h_k^{-d/2} \left[\nabla K \left(\dfrac{\theta - X_k}{h_k} \right) - \mathbb{E} \left(\nabla K \left(\dfrac{\theta - X_k}{h_k} \right) \right) \right] \\ \tilde{h}_k^{-d/2} \left[K \left(\dfrac{\theta - X_k}{\tilde{h}_k} \right) - \mathbb{E} \left(K \left(\dfrac{\theta - X_k}{\tilde{h}_k} \right) \right) \right] \end{pmatrix}
$$

(3.48)

and $\Gamma_k = \mathbb{E}(\tilde{V}_k \tilde{V}_k^T)$. In view of (3.38), (3.39), and (3.43), we have $\lim_{k \to \infty} \Gamma_k = \Gamma$. It follows that $\exists k_0 \ge 1$ such that for all $k \ge k_0$, Γ_k is inversible; without loss of generality, we assume $k_0 = 1$, and set $\tilde{U}_k = \Gamma_k^{-1/2} \tilde{V}_k$. Set $p \in]2, 4[$ and let $\mathcal{L}$ be a slowly varying function; we have

$$\frac{\mathbb{E}\left(\left\|\tilde{U}_k\right\|^p\right)}{(k\log\log k)^{p/2}} = O\left(\frac{h_k^{-dp/2}\mathbb{E}\left[\|\nabla K((\theta-X_k)/h_k)\|^p\right] + \tilde{h}_k^{-dp/2}\mathbb{E}\left[|K((\theta-X_k)/h_k)|^p\right]}{(k\log\log k)^{p/2}}\right)$$

$$= O\left(\frac{h_k^{d-dp/2} + \tilde{h}_k^{d-dp/2}}{(k\log\log k)^{p/2}}\right)$$

$$= O\left(\mathcal{L}(k)\left[k^{-[1+((p/2)-1)(1-ad)]} + k^{-[1+((p/2)-1)(1-\tilde{a}d)]}\right]\right)$$

$$(3.49)$$

so that $\sum_k (k\log\log k)^{-p/2}\mathbb{E}(\|\tilde{U}_k\|^p) < \infty$. By application of Theorem 2 of Einmahl [31], we deduce that $\sum_{k=1}^n \tilde{U}_k - \sum_{k=1}^n \eta_k = o(\sqrt{n\log\log n})$ a.s., where η_k are independent and $\mathcal{N}(0, I_{d+1})$-distributed random vectors. It follows that

$$\sum_{k=1}^n \Gamma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k - \sum_{k=1}^n \varepsilon_k = o\left(\sqrt{n\log\log n}\right) \text{ a.s.} \qquad (3.50)$$

Now,

$$\Delta_n\left[\sum_{k=1}^n Q_k\Gamma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k - \sum_{k=1}^n Q_k\varepsilon_k\right]$$

$$= \Delta_n\sum_{k=1}^n Q_k\left[\Gamma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k - \varepsilon_k\right]$$

$$= \Delta_n\sum_{k=1}^n Q_k\left(\sum_{j=1}^k\left[\Gamma^{1/2}\Gamma_j^{-1/2}\tilde{V}_j - \varepsilon_j\right] - \sum_{j=1}^{k-1}\left[\Gamma^{1/2}\Gamma_j^{-1/2}\tilde{V}_j - \varepsilon_j\right]\right) \quad \left(\text{with }\sum_{j=1}^0 = 0\right) \quad (3.51)$$

$$= \Delta_n\sum_{k=1}^{n-1}(Q_k - Q_{k+1})\left(\sum_{j=1}^k\left(\Gamma^{1/2}\Gamma_j^{-1/2}\tilde{V}_j - \varepsilon_j\right)\right) + \Delta_n Q_n\sum_{j=1}^n\left(\Gamma^{1/2}\Gamma_j^{-1/2}\tilde{V}_j - \varepsilon_j\right)$$

$$= \Delta_n\sum_{k=1}^{n-1}(Q_k - Q_{k+1})\left[o\left(\sqrt{k\log\log k}\right)\right] + \Delta_n Q_n\left[o\left(\sqrt{n\log\log n}\right)\right] \text{ a.s.}$$

Moreover,

$$\Delta_n\sum_{k=1}^{n-1}(Q_k - Q_{k+1})\left[o\left(\sqrt{k\log\log k}\right)\right]$$

$$= \begin{pmatrix} \sqrt{\dfrac{h_n^{d+2}}{n}}\displaystyle\sum_{k=1}^{n-1}\left(h_k^{-(d+2)/2} - h_{k+1}^{-(d+2)/2}\right)o\left(\sqrt{k\log\log k}\right) & 0 \\ \\ 0 & \sqrt{\dfrac{\tilde{h}_n^d}{n}}\displaystyle\sum_{k=1}^{n-1}\left(\tilde{h}_k^{-d/2} - \tilde{h}_{k+1}^{-d/2}\right) \end{pmatrix}$$

$$= \begin{pmatrix} o\left(\sqrt{h_n^{d+2}\log\log n}\right)\sum_{k=1}^{n-1}\left(h_k^{-(d+2)/2} - h_{k+1}^{-(d+2)/2}\right) & 0 \\ 0 & o\left(\sqrt{\tilde{h}_n^d\log\log n}\right)\sum_{k=1}^{n-1}\left(\tilde{h}_k^{-d/2} - \tilde{h}_{k+1}^{-d/2}\right) \end{pmatrix}.$$

(3.52)

Set $\phi(s) = [h(s)]^{-(d+2)/2}$ and $\tilde{\phi}(s) = [\tilde{h}(s)]^{-d/2}$, and let $u_k \in [k, k+1]$; since ϕ' and $\tilde{\phi}'$ vary regularly with exponent $(a(d+2)/2 - 1)$ and $(\tilde{a}d/2 - 1)$, respectively, we have

$$\sum_{k=1}^{n-1}\left(h_k^{-(d+2)/2} - h_{k+1}^{-(d+2)/2}\right) = O\left(\sum_{k=1}^{n-1}\phi'(u_k)\right) = O\left(\int_1^n \phi'(s)ds\right) = O\left(h_n^{-(d+2)/2}\right),$$

$$\sum_{k=1}^{n-1}\left(\tilde{h}_k^{-d/2} - \tilde{h}_{k+1}^{-d/2}\right) = O\left(\sum_{k=1}^{n-1}\tilde{\phi}'(u_k)\right) = O\left(\int_1^n \tilde{\phi}'(s)ds\right) = O\left(\tilde{h}_n^{-d/2}\right)$$

(3.53)

so that

$$\Delta_n \sum_{k=1}^{n-1}(Q_k - Q_{k+1})\left[o\left(\sqrt{k\log\log k}\right)\right] = o\left(\sqrt{\log\log n}\right).$$

(3.54)

Since $\Delta_n Q_n[o(\sqrt{n\log\log n})] = o(\sqrt{\log\log n})$, we deduce that

$$\frac{\Delta_n \sum_{k=1}^n Q_k \Gamma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k}{\sqrt{2\log\log n}} - \frac{\Delta_n \sum_{k=1}^n Q_k \varepsilon_k}{\sqrt{2\log\log n}} = o(1) \text{ a.s.}$$

(3.55)

The application of (3.47) then ensures that, with probability one, the sequence

$$\left(\frac{\Delta_n \sum_{k=1}^n Q_k \Gamma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k}{\sqrt{2\log\log n}}\right)$$

(3.56)

is relatively compact, and its limit set is $\mathcal{E} = \{v \in \mathbb{R}^{d+1} \text{ such that } v^T\Sigma_{a,\tilde{a}}^{-1}v \leq 1\}$. Since

$$\frac{\Delta_n \sum_{k=1}^n Q_k \tilde{V}_k}{\sqrt{2\log\log n}} = \frac{\Delta_n \sum_{k=1}^n Q_k \Gamma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k}{\sqrt{2\log\log n}} + \frac{\Delta_n \sum_{k=1}^n Q_k\left(I_{d+1} - \Gamma^{1/2}\Gamma_k^{-1/2}\right)\tilde{V}_k}{\sqrt{2\log\log n}}$$

(3.57)

with $\lim_{k\to\infty}(I_{d+1} - \Gamma^{1/2}\Gamma_k^{-1/2}) = 0$, Lemma 3.4 follows.

References

[1] E. Parzen, "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.

[2] È. A. Nadaraya, "On non-parametric estimates of density functions and regression," *Theory of Probability and Its Applications*, vol. 10, pp. 186–190, 1965.

[3] J. van Ryzin, "On strong consistency of density estimates," *Annals of Mathematical Statistics*, vol. 40, pp. 1765–1772, 1969.

[4] L. Rüschendorf, "Consistency of estimators for multivariate density functions and for the mode," *Sankhyā Series A*, vol. 39, no. 3, pp. 243–250, 1977.

[5] V. D. Konakov, "The asymptotic normality of the mode of multivariate distributions," *Theory of Probability and Its Applications*, vol. 18, pp. 836–842, 1973.

[6] M. Samanta, "Nonparametric estimation of the mode of a multivariate density," *South African Statistical Journal*, vol. 7, pp. 109–117, 1973.

[7] W. F. Eddy, "Optimum kernel estimators of the mode," *Annals of Statistics*, vol. 8, no. 4, pp. 870–882, 1980.

[8] W. F. Eddy, "The asymptotic distributions of kernel estimators of the mode," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 59, no. 3, pp. 279–290, 1982.

[9] J. P. Romano, "On weak convergence and optimality of kernel density estimates of the mode," *Annals of Statistics*, vol. 16, no. 2, pp. 629–647, 1988.

[10] A. B. Tsybakov, "Recurrent estimation of the mode of a multidimensional distribution," *Problems of Information Transmission*, vol. 26, no. 1, pp. 31–37, 1990.

[11] P. Vieu, "A note on density mode estimation," *Statistics & Probability Letters*, vol. 26, no. 4, pp. 297–307, 1996.

[12] A. Mokkadem and M. Pelletier, "The law of the iterated logarithm for the multivariable kernel mode estimator," *ESAIM. Probability and Statistics*, vol. 7, pp. 1–21, 2003.

[13] C. Abraham, G. Biau, and B. Cadre, "Simple estimation of the mode of a multivariate density," *Canadian Journal of Statistics*, vol. 31, no. 1, pp. 23–34, 2003.

[14] C. Abraham, G. Biau, and B. Cadre, "On the asymptotic properties of a simple estimate of the mode," *ESAIM. Probability and Statistics*, vol. 8, pp. 1–11, 2004.

[15] H.-G. Müller, "Adaptive nonparametric peak estimation," *Annals of Statistics*, vol. 17, no. 3, pp. 1053–1069, 1989.

[16] K. Ziegler, "On the asymptotic normality of kernel regression estimators of the mode in the nonparametric random design model," *Journal of Statistical Planning and Inference*, vol. 115, no. 1, pp. 123–144, 2003.

[17] K. Ziegler, "Adaptive kernel estimation of the mode in a nonparametric random design regression model," *Probability and Mathematical Statistics*, vol. 24, no. 2, pp. 213–235, 2004.

[18] A. Mokkadem and M. Pelletier, "A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm," *Annals of Statistics*, vol. 35, no. 4, pp. 1749–1772, 2007.

[19] C. T. Wolverton and T. J. Wagner, "Asymptotically optimal discriminant functions for pattern classification," *IEEE Transactions on Information Theory*, vol. 15, pp. 258–265, 1969.

[20] H. Yamato, "Sequential estimation of a continuous probability density function and mode," *Bulletin of Mathematical Statistics*, vol. 14, pp. 1–12, 1971.

[21] H. I. Davies, "Strong consistency of a sequential estimator of a probability density function," *Bulletin of Mathematical Statistics*, vol. 15, no. 3-4, pp. 49–54, 1973.

[22] L. P. Devroye, "On the pointwise and the integral convergence of recursive kernel estimates of probability densities," *Utilitas Mathematica*, vol. 15, pp. 113–128, 1979.

[23] V. V. Menon, B. Prasad, and R. S. Singh, "Nonparametric recursive estimates of a probability density function and its derivatives," *Journal of Statistical Planning and Inference*, vol. 9, no. 1, pp. 73–82, 1984.

[24] W. Wertz, "Sequential and recursive estimators of the probability density," *Statistics*, vol. 16, no. 2, pp. 277–295, 1985.

[25] E. J. Wegman and H. I. Davies, "Remarks on some recursive estimators of a probability density," *Annals of Statistics*, vol. 7, no. 2, pp. 316–327, 1979.

[26] G. G. Roussas, "Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function: application to regression and hazard rate estimation," *Journal of Nonparametric Statistics*, vol. 1, no. 3, pp. 171–195, 1992.

[27] A. Mokkadem, M. Pelletier, and B. Thiam, "Large and moderate deviations principles for recursive kernel estimator of a multivariate density and its partial derivatives," *Serdica. Mathematical Journal*, vol. 32, no. 4, pp. 323–354, 2006.

[28] W. Feller, *An introduction to probability Theory and Its Applications*, vol. 2, John Wiley & Sons, New York, NY, USA, 2nd edition, 1970.

[29] P. Hall, "Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density," *Annals of Statistics*, vol. 20, no. 2, pp. 675–694, 1992.

[30] A. Mokkadem and M. Pelletier, "Compact law of the iterated logarithm for matrix-normalized sums of random vectors," *Theory of Probability and Its Applications*, vol. 52, no. 4, pp. 752–767, 2007.

[31] U. Einmahl, "A useful estimate in the multidimensional invariance principle," *Probability Theory and Related Fields*, vol. 76, no. 1, pp. 81–101, 1987.