ESAIM: Probability and Statistics January 2003, Vol. 7, 1–21
DOI: 10.1051/ps:2003004

THE LAW OF THE ITERATED LOGARITHM FOR THE MULTIVARIATE KERNEL MODE ESTIMATOR

Abdelkader Mokkadem[1] and Mariane Pelletier[1]

Abstract. Let θ be the mode of a probability density and θ_n its kernel estimator. In the case θ is nondegenerate, we first specify the weak convergence rate of the multivariate kernel mode estimator by stating the central limit theorem for $\theta_n - \theta$. Then, we obtain a multivariate law of the iterated logarithm for the kernel mode estimator by proving that, with probability one, the limit set of the sequence $\theta_n - \theta$ suitably normalized is an ellipsoid. We also give a law of the iterated logarithm for the l^p norms, $p \in [1, \infty]$, of $\theta_n - \theta$. Finally, we consider the case θ is degenerate and give the exact weak and strong convergence rate of $\theta_n - \theta$ in the univariate framework.

Mathematics Subject Classification. 62G05, 62G20, 60F05, 60F15.

Received March 1, 2002. Revised May 2, 2002.

1. Introduction

Let $X_1, \dots, X_n$ be a sequence of independent and identically distributed $\mathbb{R}^d$-valued random variables with unknown probability density f. We assume f has a unique mode θ, that is, we suppose that there exists $\theta \in \mathbb{R}^d$ such that $f(x) < f(\theta)$ for any $x \neq \theta$.

A natural approach to mode estimation is to locate the absolute maximum of a corresponding empirical density function; Parzen [19] gives consistency and asymptotic normality for such estimates based on kernel method. His results have been extended mainly by Chernoff [3], Nadaraya [17], Van Ryzin [28], Rüschendorf [23], Eddy [5, 6], Romano [22] and Grund and Hall [10]. An approach based on order statistics is introduced by Grenander [9] and Venter [29], and developed by Sager [24] and Hall [12]. A recursive method is proposed by Tsybakov [27]. A question closely related to the mode estimation is the estimation of the conditional mode developed by Collomb *et al* [4], and extended by Samanta and Thavaneswaran [26], Ould–Said [18], Quintela–Del–Rio and Vieu [21], Berlinet *et al.* [2] and Louani and Ould–Said [15].

The present paper deals with the classical Parzen mode estimator. Let $(h_n)_{n\geq 1}$ be a sequence of positive real numbers such that $\lim_{n\to\infty} h_n = 0$, $\lim_{n\to\infty} nh_n^d = \infty$, and let K be a continuous function satisfying $\int_{\mathbb{R}^d} K(x)\mathrm{d}x = 1$, $\lim_{\|x\|\to\infty} K(x) = 0$ (where $\|.\|$ is an arbitrary norm on $\mathbb{R}^d$). The well-known Parzen–Rosenblatt kernel estimator of the density f is defined by

$$f_n(x) = \frac{1}{nh_n^d} \sum_{k=1}^{n} K\left(\frac{x - X_k}{h_n}\right) \quad \text{for any} \quad x \in \mathbb{R}^d$$

Keywords and phrases: Density, mode, kernel estimator, central limit theorem, law of the iterated logarithm.

[1] Département de Mathématiques, Université de Versailles-Saint-Quentin, 45 avenue des États-Unis, 78035 Versailles Cedex, France; e-mail: `(mokkadem,pelletier)@math.uvsq.fr`

and the kernel mode estimator is any random variable θ_n satisfying

$$f_n(\theta_n) = \sup_{x\in\mathbb{R}^d} f_n(x). \tag{1}$$

Since K is continuous and vanishing at infinity, the choice of θ_n as a random variable satisfying (1) can be made with the help of an order on $\mathbb{R}^d$. For example, one can consider the following lexicographic order: $x \leq y$ if $x = y$ or if the first nonzero coordinate of $x - y$ is negative. The definition

$$\theta_n = \inf\left\{ y \in \mathbb{R}^d \ \text{ such that } \ f_n(y) = \sup_{x\in\mathbb{R}^d} f_n(x)\right\}$$

where the infimum is taken with respect to the lexicographic order on $\mathbb{R}^d$, ensures the measurability of the kernel mode estimator.

The weak consistency of θ_n was established by Parzen [19] and Yamato [31], its strong consistency by Nadaraya [17], Van Ryzin [28], Rüschendorf [23] and Romano [22]. In the univariate framework, that is when $d = 1$, Romano [22] proves the almost sure convergence of θ_n to θ under the optimal assumption on the bandwidth $\lim_{n\to\infty} nh_n[\ln n]^{-1} = \infty$. We first give a straightforward extension of his Theorem 1.1 to the multivariate case and obtain the strong consistency of θ_n under the condition $\lim_{n\to\infty} nh_n^d[\ln n]^{-1} = \infty$, which weakens the assumptions on the bandwidth made in Van Ryzin [28] and Rüschendorf [23].

Let us now assume θ is nondegenerate, that is, $D^2f(\theta)$ (the second order differential at the point θ) is nonsingular.

The weak convergence rate of θ_n to θ was first studied in the univariate framework by Parzen [19] who proved that, if h_n is chosen such that $\lim_{n\to\infty} nh_n^6 = \infty$ and $\lim_{n\to\infty} nh_n^7 = 0$, then, under appropriate smoothness assumptions,

$$\sqrt{nh_n^3}\,(\theta_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0\ ,\ \frac{f(\theta)}{[f''(\theta)]^2}\int_{\mathbb{R}} K'^2(x)\mathrm{d}x\right)$$

where $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution and $\mathcal{N}$ the Gaussian-distribution. Eddy [5,6] and Romano [22] then proved that this central limit theorem still holds when the condition $\lim_{n\to\infty} nh_n^6 = \infty$ is weakened to $\lim_{n\to\infty} nh_n^5[\ln n]^{-1} = \infty$.

The study of the weak convergence rate of θ_n to θ was extended by Konakov [13] and Samanta [25] to the multivariate framework. The key idea to establish the convergence rate of θ_n to θ is to note that, as soon as D^2f_n converges almost surely uniformly to D^2f in a neighborhood of θ, the asymptotic behaviour of $\theta_n - \theta$ is given by that of $-\left[D^2f(\theta)\right]^{-1}\nabla f_n(\theta)$, where ∇f_n denotes the gradient of f_n. The condition on the bandwidth required by Konakov [13] and Samanta [25] to ensure the strong uniform convergence of D^2f_n is $\lim_{n\to\infty} nh_n^{2d+4} = \infty$. Although this condition is equivalent to the one of Parzen [19] when $d = 1$, it is too strong to establish a central limit theorem as soon as $d \geq 2$. The reason is the following.

The weak convergence rate of $\nabla f_n(\theta)$ to zero is governed by the weak convergence rate of the variance term $\nabla f_n(\theta) - \mathrm{E}(\nabla f_n(\theta))$ on one hand and by the deterministic convergence rate of the bias term $\mathrm{E}(\nabla f_n(\theta))$ on the other hand. Since the variance term converges at the rate $\sqrt{nh_n^{d+2}}$ and the bias term at the rate h_n^{-2}, the condition $\lim_{n\to\infty} nh_n^{d+6} = 0$ is necessary to make the bias term negligible in front of the variance term, and thus to establish a central limit theorem for $\nabla f_n(\theta)$. The incompatibility of this last condition with the one required by Konakov [13] and Samanta [25] for the strong uniform convergence of D^2f_n prevents the transfer of a central limit theorem established for $\nabla f_n(\theta)$ to one which would hold for $\theta_n - \theta$. By weakening the condition $\lim_{n\to\infty} nh_n^{2d+4} = \infty$ of Konakov [13] and Samanta [25] to $\lim_{n\to\infty} nh_n^{d+4}[\ln n]^{-1} = \infty$, we make possible the choice of a bandwidth for which D^2f_n converges a.s. uniformly to D^2f and for which the bias of $\nabla f_n(\theta)$ is negligible in front of its variance term. This allows us to state the following central limit theorem:

$$\sqrt{nh_n^{d+2}}\,(\theta_n-\theta)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0\ ,\ f(\theta)\big[D^2f(\theta)\big]^{-1}G\big[D^2f(\theta)\big]^{-1}\right)$$

where G is the $d\times d$ matrix defined by

$$G_{i,j}=\int_{\mathbb{R}^d}\frac{\partial K}{\partial x_i}(x)\frac{\partial K}{\partial x_j}(x)\mathrm{d}x.$$

We now come to our main objective, which is to prove the law of the iterated logarithm for the multivariate kernel mode estimator.

In the univariate framework, upper bounds of the almost sure convergence rate of θ_n are given in Eddy [5], Vieu [30] and Leclerc and Pierre–Loti–Viaud [14]. The exact strong convergence rate of the univariate kernel mode estimator is given in Mokkadem and Pelletier [16] who proved the following law of the iterated logarithm:

$$\limsup_{n\to\infty}\sqrt{\frac{nh_n^3}{2\ln\ln n}}\,(\theta_n-\theta)=-\liminf_{n\to\infty}\sqrt{\frac{nh_n^3}{2\ln\ln n}}\,(\theta_n-\theta)=\frac{\sqrt{f(\theta)\int_{\mathbb{R}}K'^2(x)\mathrm{d}x}}{|f''(\theta)|}\quad\text{a.s.}\tag{2}$$

Our main result in the present paper is the following multivariate law of the iterated logarithm: with probability one, the sequence

$$\sqrt{\frac{nh_n^{d+2}}{2\ln\ln n}}\,(\theta_n-\theta)$$

is relatively compact and its limit set is the ellipsoid

$$\left\{\nu\in\mathbb{R}^d\ \text{ such that }\ \frac{1}{f(\theta)}\ \nu^t\left[D^2f(\theta)\right]G^{-1}\left[D^2f(\theta)\right]\nu\leq 1\right\}.$$

Note that the unidimensional version of this result can be written as: with probability one, the sequence

$$\sqrt{\frac{nh_n^3}{2\ln\ln n}}\,(\theta_n-\theta)$$

is relatively compact and its limit set is the interval

$$\left[\frac{-\sqrt{f(\theta)\int_{\mathbb{R}}K'^2(x)\mathrm{d}x}}{|f''(\theta)|}\ ;\ \frac{+\sqrt{f(\theta)\int_{\mathbb{R}}K'^2(x)\mathrm{d}x}}{|f''(\theta)|}\right]$$

and thus extends the univariate result (2).

We also establish a law of the iterated logarithm for the l^p norms, $p\in[1,\infty]$, of the vector $(\theta_n-\theta)$. For sake of simplicity, we state here our results in the two striking cases, that is, for $p=2$ and $p=\infty$. For any vector $x=(x_1,\dots,x_d)^t\in\mathbb{R}^d$, set $\|x\|_2=\left[\sum_{i=1}^d x_i^2\right]^{1/2}$ and $\|x\|_\infty=\max_{1\leq i\leq d}|x_i|$. We prove that, with probability one, the sequence

$$\sqrt{\frac{nh_n^{d+2}}{2\ln\ln n}}\ \|\theta_n-\theta\|_2$$

is relatively compact and its limit set is the interval $[0, \delta_2 \sqrt{f(\theta)}]$ where δ_2 is the spectral radius, *i.e.* the largest eigenvalue, of the matrix $-G^{1/2}\left[D^2 f(\theta)\right]^{-1}$. We also establish that, with probability one, the sequence

$$\sqrt{\frac{nh_n^{d+2}}{2\ln\ln n}}\ \|\theta_n - \theta\|_\infty$$

is relatively compact and its limit set is the interval $\left[0\ , \delta_\infty \sqrt{f(\theta)}\right]$ where δ_∞ is the square root of the largest diagonal term of the matrix $\left[D^2 f(\theta)\right]^{-1} G \left[D^2 f(\theta)\right]^{-1}$.

These different versions of the law of the iterated logarithm for the multivariate kernel mode estimator are proved by relating the strong behaviour of $(\theta_n - \theta)$ to the one of $\left[D^2 f(\theta)\right]^{-1} \nabla f_n(\theta)$ and by applying a result of Arcones [1].

Let us finally consider the case θ is degenerate. To our knowledge, the only result in this framework is an upper bound of the complete (and thus almost sure) convergence rate of $\theta_n - \theta$ stated in Vieu [30] in the univariate case. We specify this by establishing the exact weak and strong convergence rate of $\theta_n - \theta$ in the case $d = 1$. The degenerate multivariate case seems very intricate and the convergence rate of the kernel mode estimator in this framework remains an open question.

Our assumptions and results are stated in Section 2, whereas Section 3 is devoted to the proofs.

2. Assumptions and results

Before stating our assumptions, let us first define the *covering number condition.* Let Q be a probability on $\mathbb{R}^d$ and let $\mathcal{F} \subset \mathcal{L}_s(Q)$, $s \in \{1, 2\}$, be a class of Q-integrable functions. The $\mathcal{L}_s$-covering number (see Pollard [20]) is the smallest value $N_s(\varepsilon, Q, \mathcal{F})$ of m for which there exist m functions $g_1, \dots, g_m \in \mathcal{L}_s(Q)$ such that

$$\min_{i\in\{1,\dots,m\}} \|f - g_i\|_{\mathcal{L}_s(Q)} \le \varepsilon \quad \forall f \in \mathcal{F} \tag{3}$$

(if no such m exists, $N_s(\varepsilon, Q, \mathcal{F}) = \infty$). Now, let Λ be a $\mathbb{R}$-valued function defined on $\mathbb{R}^d$, and let $\mathcal{F}(\Lambda)$ be the class of functions defined by

$$\mathcal{F}(\Lambda) = \left\{ z \mapsto \Lambda\left(\frac{x - z}{h}\right),\quad h > 0,\quad x \in \mathbb{R}^d \right\}.$$

Λ is said to satisfy the $\mathcal{L}_s$-covering number condition if Λ is bounded and integrable on $\mathbb{R}^d$, and if there exist $A > 0$ and $w > 0$ such that, for any probability Q on $\mathbb{R}^d$ and any $\varepsilon \in]0, 1[$,

$$N_s(\varepsilon, Q, \mathcal{F}(\Lambda)) \le A\varepsilon^{-w}. \tag{4}$$

In the case $\mathcal{F}$ is uniformly bounded by a constant M, one can consider in (3) only the approximating functions g_i such that $\|g_i\|_\infty \le M$ (see Pollard [20]). In this case, simple inequalities show that the $\mathcal{L}_1$-covering number condition is equivalent to the $\mathcal{L}_2$-covering number condition. Since the only classes we consider are $\mathcal{F}(\Lambda)$ classes with Λ bounded, we shall only refer to the "covering number condition" without distinction.

The classes which satisfy (4) are often called VC classes. When $d = 1$, the real-valued kernels with bounded variations satisfy the covering number condition (see Pollard [20]). Some examples of multivariate kernels satisfying the covering number condition are the following:

– the kernels defined as $K(x) = \psi(\|x\|)$, where ψ is a real-valued function with bounded variations (see Pollard [20]);

– the kernels defined as $K(x) = \prod_{i=1}^{d} K_i(x_i)$ where the K_i, $1 \leq i \leq d$, are real-valued functions with bounded variations (this follows from Lem. A1 in Einmahl and Mason [7]);
– the kernels satisfying the assumption (K1) of Giné and Guillou [8].

The assumptions we require for the strong consistency of θ_n are the following:
(A1) i) $\lim_{\|x\|\to\infty} K(x) = 0$ and $\int_{\mathbb{R}^d} K(x)\mathrm{d}x = 1$;
ii) K is continuous on $\mathbb{R}^d$;
iii) K satisfies the covering number condition.

(A2) i) $\lim_{n\to\infty} h_n = 0$;
ii) $\lim_{n\to\infty} nh_n^d [\ln n]^{-1} = \infty$.

(A3) i) there exists $\theta \in \mathbb{R}^d$ such that $f(x) < f(\theta)$ for all $x \neq \theta$;
ii) f is continuous in a neighbourhood $\mathcal{V}$ of θ;
iii) $\sup_{x\notin\mathcal{V}} f(x) < f(\theta)$.

Theorem 2.1 (Strong consistency)**.** *Assume (A1–A3) hold. Moreover, assume either that K is nonnegative or that f is uniformly continuous on $\mathbb{R}^d$. Then,*

$$\lim_{n\to\infty} \theta_n = \theta \quad a.s.$$

Remarks.
1) In the case f is uniformly continuous on $\mathbb{R}^d$, (A3) iii) is a straightforward consequence of the unicity of the mode of f.
2) Theorem 2.1 is a straightforward extension of Theorem 1.1 of Romano [22] to the multivariate framework; it weakens the assumptions on the bandwidth made by Van Ryzin [28] and Rüschendorf [23] in the case $d \geq 2$.

In order to state the central limit theorem, we need the following additional assumptions:

(A4) $\lim_{n\to\infty} nh_n^{d+4} [\ln n]^{-1} = \infty$.

(A5) i) K is twice differentiable on $\mathbb{R}^d$ and, for any $(i,j) \in \{1,\ldots,d\}^2$, $\partial^2 K/\partial x_i \partial x_j$ satisfies the covering number condition;
ii) for any $i \in \{1,\ldots,d\}$, $\int_{\mathbb{R}^d} \left(\frac{\partial K}{\partial x_i}(x)\right)^2 \mathrm{d}x < \infty$ and $\int_{\mathbb{R}^d} \left|\frac{\partial K}{\partial x_i}(x)\right|^{2+\delta} \mathrm{d}x < \infty$ for some $\delta > 0$;
iii) there exists $q \geq 2$ such that for any $s \in \{1,\ldots,q-1\}$ and any $j \in \{1,\ldots,d\}$,

$$\int_{\mathbb{R}} y_j^s K(y)\mathrm{d}y_j = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} \left|y_j^q K(y)\right| \mathrm{d}y < \infty.$$

(A6) i) f is twice differentiable on $\mathbb{R}^d$, $D^2 f$ is continuous in a neighborhood of θ, and $D^2 f(\theta)$ is nonsingular;
ii) $D^2 f$ is bounded on $\mathbb{R}^d$;
iii) for any $i \in \{1,\ldots,d\}$, $\partial f/\partial x_i$ is bounded on $\mathbb{R}^d$;
iv) f is $q+1$ times differentiable at the point θ (q is defined in (A5)).

Remarks.
1) Some conditions in (A4–A6) clearly imply some other conditions already required in (A1–A3). For instance, since $\lim_{n\to\infty} h_n = 0$, (A2) ii) is included in (A4).
2) Since θ is assumed to be the unique mode of f, (A6) i) implies that $D^2 f(\theta)$ is negative definite.
3) Note that (A6) iii) implies the uniform continuity of f on $\mathbb{R}^d$.
4) Let us finally mention that the condition (A6) ii) is useless as soon as the support of K is bounded.

Let us set $B_q(\theta)$ the vector

$$B_q(\theta)=\frac{(-1)^{q+1}}{q!}\left[D^2f(\theta)\right]^{-1}\nabla\left(\sum_{j=1}^{d}\mu_j^q\frac{\partial^q f}{\partial x_j^q}(\theta)\right)\quad\text{with}\quad \mu_j^q=\int_{\mathbb{R}^d}y_j^qK(y)\mathrm{d}y$$

where ∇ denotes the gradient and recall that

$$G=(G_{i,j})_{1\le i,j\le d}\quad\text{with}\quad G_{i,j}=\int_{\mathbb{R}^d}\frac{\partial K}{\partial x_i}(x)\frac{\partial K}{\partial x_j}(x)\mathrm{d}x.$$

Theorem 2.2 (Central limit theorem)**.** *Let assumptions (A1–A6) hold.*

i) If $\lim_{n\to\infty}nh_n^{d+2q+2}=0$*, then*

$$\sqrt{nh_n^{d+2}}\,(\theta_n-\theta)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0\ ,\ f(\theta)\left[D^2f(\theta)\right]^{-1}G\left[D^2f(\theta)\right]^{-1}\right).$$

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

$$\sqrt{nh_n^{d+2}}\,(\theta_n-\theta)\xrightarrow{\mathcal{D}}\mathcal{N}\left(\sqrt{c}B_q(\theta)\ ,\ f(\theta)\left[D^2f(\theta)\right]^{-1}G\left[D^2f(\theta)\right]^{-1}\right).$$

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

$$\frac{1}{h_n^q}\,(\theta_n-\theta)\xrightarrow{P}B_q(\theta).$$

As mentioned in the introduction, the weak convergence rate of the kernel mode estimator is given by that of $\left[D^2f(\theta)\right]^{-1}\nabla f_n(\theta)$, which depends itself on the convergence rate of the variance term $\nabla f_n(\theta)-\mathrm{E}\left(\nabla f_n(\theta)\right)$ and of the bias term $\mathrm{E}\left(\nabla f_n(\theta)\right)$. Part i) of Theorem 2.2 corresponds to the case the bias term does not interfer, Part ii) holds when (h_n) is chosen such that the bias and the variance terms are balanced and Part iii) describes the case when the variance term does not interfer. Let us note that, in their study of the weak convergence rate of $(\theta_n-\theta)$, Konakov [13] and Samanta [25] consider kernels of order $q=2$. Their condition on the bandwidth $\lim_{n\to\infty}nh_n^{2d+4}=\infty$ required to ensure the strong uniform convergence of D^2f_n implies $\lim_{n\to\infty}nh_n^{d+6}=\infty$ and leads to the weak convergence of $h_n^{-2}(\theta_n-\theta)$ to a degenerate distribution. Let us finally mention that the higher order of weak convergence is attained for $h_n\sim n^{-1/(d+2q+2)}$.

The limit laws in Parts i) and ii) of Theorem 2.2 are nondegenerate. As a matter of fact, we shall prove the following proposition:

Proposition 2.3. *If* $K\not\equiv 0$ *is continuously differentiable and vanishing at infinity, then the matrix* G *is positive definite.*

In order to prove the law of the iterated logarithm for the kernel mode estimator, we require the following additional assumption:

(A7) There exists $M>0$ such that $K(x)=0$ if $\|x\|>M$.

Theorem 2.4 (Law of the iterated logarithm for the full vector)**.** *Let assumptions (A1–A7) hold.*
i) If $\lim_{n\to\infty} nh_n^{d+2q+2}/\ln\ln n = 0$, *then, with probability one, the sequence*

$$\sqrt{\frac{nh_n^{d+2}}{2\ln\ln n}}\;(\theta_n-\theta)$$

is relatively compact and its limit set is the ellipsoid

$$C=\left\{\nu\in\mathbb{R}^d \quad \text{such that} \quad \frac{1}{f(\theta)}\nu^t\left[D^2f(\theta)\right]G^{-1}\left[D^2f(\theta)\right]\nu\le 1\right\}.$$

ii) If there exists $c>0$ *such that* $\lim_{n\to\infty} nh_n^{d+2q+2}/\ln\ln n = c$, *then, with probability one, the sequence*

$$\sqrt{\frac{nh_n^{d+2}}{2\ln\ln n}}\;(\theta_n-\theta)$$

is relatively compact and its limit set is the ellipsoid

$$C=\left\{\nu\in\mathbb{R}^d \quad \text{such that} \quad \frac{1}{f(\theta)}\left(\nu-\sqrt{\frac{c}{2}}\,B_q(\theta)\right)^t\left[D^2f(\theta)\right]G^{-1}\left[D^2f(\theta)\right]\left(\nu-\sqrt{\frac{c}{2}}\,B_q(\theta)\right)\le 1\right\}.$$

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

$$\lim_{n\to\infty}\frac{1}{h_n^q}(\theta_n-\theta)=B_q(\theta)\quad a.s.$$

The strong convergence rate of $\theta_n-\theta$ is deduced from that of $\left[D^2f(\theta)\right]^{-1}\nabla f_n(\theta)$. Similarly to the study of its weak convergence rate, three cases have to be considered according to the choice of the bandwidth. Part i) of Theorem 2.4 corresponds to the case the bias term does not interfer, Part ii) to the case the bias and the variance terms are balanced and Part iii) to the case the variance term does not interfer. Let us underline that the conditions on the bandwidth which differentiate the three possible a.s. behaviours of the sequence $(\theta_n-\theta)$ are slightly different from those which determine the weak convergence rate of the kernel mode estimator. So, the choice of h_n which gives the optimal a.s. rate of convergence of θ_n, that is, $h_n\sim[\ln\ln n]^{1/(d+2q+2)}n^{-1/(d+2q+2)}$, is not the choice of the bandwidth which ensures the optimal weak convergence rate of θ_n.

For sake of simplicity, we shall state the next versions of the law of the iterated logarithm for the multivariate kernel mode estimator only in the case the bias term is negligible; the two other cases can be easily deduced.

Theorem 2.5 (Law of the iterated logarithm for the linear forms)**.** *Let assumptions (A1–A7) hold, assume that* $\lim_{n\to\infty} nh_n^{d+2q+2}/\ln\ln n = 0$, *and set* $u\in\mathbb{R}^d$. *Then, with probability one, the sequence*

$$\sqrt{\frac{nh_n^{d+2}}{2\ln\ln n}}\;u^t(\theta_n-\theta)$$

is relatively compact and its limit set is

$$I(u)=\left[\;-\sqrt{f(\theta)u^t[D^2f(\theta)]^{-1}G[D^2f(\theta)]^{-1}u}\;;\;+\sqrt{f(\theta)u^t[D^2f(\theta)]^{-1}G[D^2f(\theta)]^{-1}u}\;\right].$$

Note that the application of Theorem 2.5 to the i-th vector of the canonical basis of $\mathbb{R}^d$ gives the limit set of the sequence

$$\sqrt{\frac{nh_n^{d+2}}{2\ln\ln n}}\ (\theta_{n,i}-\theta_i)$$

that is, the law of the iterated logarithm for the i-th coordinate of θ_n.

To conclude our study on the multivariate kernel mode estimator in the case θ is nondegenerate, we finally state the law of the iterated logarithm for the l^p norms of $(\theta_n-\theta)$. To this end, for any matrix A and any $p\in[1,\infty]$, we denote by $\|A\|_{2,p}$ the matrix norm defined by

$$\|A\|_{2,p}=\sup_{\|x\|_2\leq 1}\|Ax\|_p$$

where $\|x\|_p$ is the l^p vector norm: $\|x\|_p=\left[\sum_{i=1}^d|x_i|^p\right]^{1/p}$ for $p\in[1,\infty[$ and $\|x\|_\infty=\max_{1\leq i\leq d}|x_i|$.

Theorem 2.6 (Law of the iterated logarithm for the l^p norms)**.** *Let assumptions (A1–A7) hold and assume that* $\lim_{n\to\infty}nh_n^{d+2q+2}/\ln\ln n=0$*. Set* $p\in[1,\infty]$*; with probability one, the sequence*

$$\sqrt{\frac{nh_n^{d+2}}{2\ln\ln n}}\ \|\theta_n-\theta\|_p$$

is relatively compact and its limit set is the interval $[0,\delta_p\sqrt{f(\theta)}]$ *with*

$$\delta_p=\left\|G^{1/2}\left[D^2f(\theta)\right]^{-1}\right\|_{2,p}.$$

In particular, for $p=2$*,* δ_2 *is the spectral radius of the matrix* $-G^{1/2}\left[D^2f(\theta)\right]^{-1}$ *and, for* $p=\infty$*,* δ_∞ *is the square root of the largest diagonal term of the matrix* $\left[D^2f(\theta)\right]^{-1}G\left[D^2f(\theta)\right]^{-1}$*.*

Let us finally consider the case θ may be degenerate. For that purpose, we set $d=1$ and require the following assumptions:

(A'4) i) There exists $p\geq 2$ such that f is p-times differentiable on $\mathbb{R}$, $f^{(j)}(\theta)=0$ for any $j\in\{1,\dots,p-1\}$, and $f^{(p)}(\theta)\neq 0$;
ii) for any $j\in\{1,\dots,p\}$, $f^{(j)}$ is bounded on $\mathbb{R}$.

(A'5) i) K is p-times differentiable on $\mathbb{R}$;
ii) for any $j\in\{1,\dots,p\}$, $K^{(j)}$ satisfies the covering number condition;
iii) there exists $q\geq p$ such that $\int_{\mathbb{R}}y^jK(y)\mathrm{d}y=0$ for any $j\in\{p-1,\dots,q-1\}$ and $\int_{\mathbb{R}}|y^qK(y)|\,\mathrm{d}y<\infty$;
iv) f is $q+1$ times differentiable on $\mathbb{R}$ and $f^{(q+1)}$ is bounded on $\mathbb{R}$.

(A'6) i) $\lim_{n\to\infty}nh_n^{2p+1}[\ln(1/h_n)]^{-p}=\infty$;
ii) (h_n) is a decreasing sequence and $\lim_{n\to\infty}\ln(1/h_n)[\ln\ln n]^{-1}=\infty$;
iii) either (nh_n) is an increasing sequence or there exists c such that $h_n\leq ch_{2n}$.

Remarks.
1) Since $f(\theta)$ is a maximum of f, the integer p defined in (A'4) i) is even.
2) Note that the case $f^{(p)}(\theta)=0$ for all p is not covered by (A'4).
3) Any even kernel with a finite moment of order p satisfies the assumption (A'5) iii) with $q=p$.

4) Note that (A'5) iv) implies the continuity of $f^{(p)}$.
Set

$$B_{p,q}(\theta) = (-1)^{q+1} \frac{(p-1)!\ f^{(q+1)}(\theta)}{q!\ f^{(p)}(\theta)} \int_{\mathbb{R}} y^q K(y)\mathrm{d}y.$$

Theorem 2.7 (Weak convergence rate of the univariate kernel mode estimator in the degenerate case)**.** *Let assumptions (A1–A3) and (A'4–A'6) hold.*

i) If there exists $c \geq 0$ such that $\lim_{n\to\infty} nh_n^{2q+3} = c$, then

$$\left(nh_n^3\right)^{1/2(p-1)}(\theta_n - \theta) \xrightarrow{\mathcal{D}} Z$$

where the random variable Z^{p-1} is $\mathcal{N}\left(\sqrt{c}B_{p,q}(\theta)\ ,\ \frac{[(p-1)!]^2 f(\theta)}{\left[f^{(p)}(\theta)\right]^2}\int_{\mathbb{R}} K'^2(x)\mathrm{d}x\right)$ – distributed.
ii) If $\lim_{n\to\infty} nh_n^{2q+3} = \infty$, then

$$\frac{1}{h_n^{q/(p-1)}}(\theta_n - \theta) \xrightarrow{P} \left[B_{p,q}(\theta)\right]^{1/(p-1)}.$$

Theorem 2.8 (Strong convergence rate of the univariate kernel mode estimator in the degenerate case)**.** *Let assumptions (A1–A3, A'4–A'6) and (A7) hold.*
i) If there exists $c \geq 0$ such that $\lim_{n\to\infty} nh_n^{2q+3}/\ln\ln n = c$, then, with probability one, the sequence

$$\left(\frac{nh_n^3}{2\ln\ln n}\right)^{1/2(p-1)} (\theta_n - \theta)$$

is relatively compact and its limit set is the interval

$$I = \left[\left(\sqrt{\frac{c}{2}}B_{p,q} - \frac{(p-1)!\sqrt{f(\theta)\int_{\mathbb{R}} K'^2(x)\mathrm{d}x}}{|f^{(p)}(\theta)|}\right)^{1/(p-1)} ;\right.$$
$$\left.\left(\sqrt{\frac{c}{2}}B_{p,q} + \frac{(p-1)!\sqrt{f(\theta)\int_{\mathbb{R}} K'^2(x)\mathrm{d}x}}{|f^{(p)}(\theta)|}\right)^{1/(p-1)}\right].$$

ii) If $\lim_{n\to\infty} nh_n^{2q+3}/\ln\ln n = \infty$, then

$$\lim_{n\to\infty} \frac{1}{h_n^{q/(p-1)}}(\theta_n - \theta) = \left[B_{p,q}(\theta)\right]^{1/(p-1)} \quad \text{a.s.}$$

Remark. Part i) of Theorem 2.8 implies that

$$\begin{cases} \liminf_{n\to\infty} \left(\dfrac{nh_n^3}{2\ln\ln n}\right)^{1/2(p-1)} (\theta_n - \theta) = \left(\sqrt{\dfrac{c}{2}}B_{p,q} - \dfrac{(p-1)!\sqrt{f(\theta)\int_{\mathbb{R}} K'^2(x)\mathrm{d}x}}{|f^{(p)}(\theta)|}\right)^{1/(p-1)} \\ \limsup_{n\to\infty} \left(\dfrac{nh_n^3}{2\ln\ln n}\right)^{1/2(p-1)} (\theta_n - \theta) = \left(\sqrt{\dfrac{c}{2}}B_{p,q} + \dfrac{(p-1)!\sqrt{f(\theta)\int_{\mathbb{R}} K'^2(x)\mathrm{d}x}}{|f^{(p)}(\theta)|}\right)^{1/(p-1)}. \end{cases}$$

This particular result has been obtained in Mokkadem and Pelletier [16] in the nondegenerate case (*i.e.* when $p = 2$) by applying a result of Hall [**?**] and without the assumption (A7) on the boundedness of the support of K.

3. Proofs

3.1. Consistency of the mode estimator

We first note that, following the proof of Theorem 1.1 in Romano [22], the application of Theorem 37 (p. 34) in Pollard [20] gives the following lemma:

Lemma 3.1. *Let Λ be a function on $\mathbb{R}^d$ satisfying the covering number condition. If there exists $j \geq 0$ such that $\lim_{n\to\infty} nh_n^{d+2j}[\ln n]^{-1} = \infty$, then*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}^d} \frac{1}{nh_n^{d+j}} \left|\sum_{i=1}^n \Lambda\left(\frac{x-X_i}{h_n}\right) - E\left[\Lambda\left(\frac{x-X_i}{h_n}\right)\right]\right| = 0 \quad a.s.$$

The application of Lemma 3.1 with $\Lambda = K$ ensures that, under the assumption $\lim_{n\to\infty} nh_n^d[\ln n]^{-1} = \infty$,

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}^d} |f_n(x) - \mathrm{E}(f_n(x))| = 0 \quad a.s. \tag{5}$$

Now, set $\delta > 0$ such that $\mathcal{B}(\theta,\delta) = \left\{x \in \mathbb{R}^d \ / \ \|\theta - x\| < \delta\right\} \subset \mathcal{V}$ and set $m_\delta = \sup_{x\notin\mathcal{B}(\theta,\delta)} f(x)$; by assumption, we have $m_\delta < f(\theta)$. Then, set $\varepsilon \in]0, f(\theta) - m_\delta[$; for n large enough, we have

$$\sup_{x\notin\mathcal{B}(\theta,\delta)} \mathrm{E}(f_n(x)) < f(\theta) - \varepsilon \tag{6}$$

this last inequality being either proved by following Romano [22] in the case K is nonnegative or being deduced from the uniform convergence of $E(f_n)$ to f in the case f is uniformly continuous on $\mathbb{R}^d$. The combination of (5) and (6) then ensures that almost surely, for n large enough,

$$\sup_{x\notin\mathcal{B}(\theta,\delta)} f_n(x) < f(\theta) - \frac{\varepsilon}{2}\cdot \tag{7}$$

Since $\lim_{n\to\infty} f_n(\theta) = f(\theta)$ a.s., we have a.s., for n large enough, $f_n(\theta) > f(\theta) - \varepsilon/2$ and thus $f_n(\theta_n) > f(\theta) - \varepsilon/2$. In view of (7), it follows that $\theta_n \in \mathcal{B}(\theta,\delta)$ a.s. for n large enough, which proves Theorem 2.1.

3.2. Connection between the convergence rate of the mode estimator and that of the variance term of the derivative density estimator

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

$$\nabla f_n(\theta_n) - \nabla f_n(\theta) = -\nabla f_n(\theta). \tag{8}$$

For each $i \in \{1, \dots, d\}$, Taylor's expansion applied to the real-valued application $\frac{\partial f_n}{\partial x_i}$ implies the existence of $\xi_n(i) = (\xi_{n,1}(i), \dots, \xi_{n,d}(i))^t$ such that

$$\begin{cases} \dfrac{\partial f_n}{\partial x_i}(\theta_n) - \dfrac{\partial f_n}{\partial x_i}(\theta) = \displaystyle\sum_{j=1}^{d} \dfrac{\partial^2 f_n}{\partial x_i \partial x_j}(\xi_n(i))\,(\theta_{n,j} - \theta_j) \\ |\xi_{n,j}(i) - \theta_j| \leq |\theta_{n,j}(i) - \theta_j| \quad \forall j \in \{1, \dots, d\}. \end{cases}$$

Define the $d \times d$ matrix $H_n = (H_{n,i,j})_{1 \leq i,j \leq d}$ by setting

$$H_{n,i,j} = \frac{\partial^2 f_n}{\partial x_i \partial x_j}(\xi_n(i)).$$

Equation (8) can then be rewritten as

$$H_n(\theta_n - \theta) = -\nabla f_n(\theta). \tag{9}$$

Now, under the assumption $\lim_{n\to\infty} n h_n^{d+4} [\ln n]^{-1} = \infty$, the application of Lemma 3.1 with $\Lambda = \partial^2 K / \partial x_i \partial x_j$ ensures that

$$\lim_{n\to\infty} \sup_{x \in \mathbb{R}^d} \left| \frac{\partial^2 f_n}{\partial x_i \partial x_j}(x) - E\left(\frac{\partial^2 f_n}{\partial x_i \partial x_j}(x)\right) \right| = 0 \quad \text{a.s.}$$

Moreover, classical computations give the uniform convergence of $E\left(\partial^2 f_n / \partial x_i \partial x_j\right)$ to $\partial^2 f / \partial x_i \partial x_j$ in a neighborhood of θ. Since $\lim_{n\to\infty} \theta_n = \theta$ a.s., we thus obtain

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

In view of (9), it follows that the convergence rate of $\theta_n - \theta$ is given by that of $-\left[D^2 f(\theta)\right]^{-1} \nabla f_n(\theta)$; Sections 3.3 and 3.4 are devoted to the study of the weak and almost sure asymptotic behaviour of the variance term of $\nabla f_n(\theta)$. The asymptotic behaviour of the bias term is given by the following lemma:

Lemma 3.2.

$$\lim_{n\to\infty} \frac{1}{h_n^q} E(\nabla f_n(\theta)) = \frac{(-1)^q}{q!} \nabla \left(\sum_{j=1}^{d} \mu_j^q \frac{\partial^q f}{\partial x_j^q}(\theta) \right).$$

Proof of Lemma 3.2. Let us set $f_i = \frac{\partial f}{\partial x_i}$ $(i \in \{1, \dots, d\})$, and $D^j f_i$ $(j \in \{1, \dots, p\})$ the j-th differential of f_i. For $x = (x_1, \dots, x_d)^t \in \mathbb{R}^d$, set $\underline{x}^{(j)} = (x, \dots, x) \in \left(\mathbb{R}^d\right)^j$; with these notations, the j-linear application $D^j f_i(\theta)$ satisfies

$$D^j f_i(\theta)\left(\underline{x}^{(j)}\right) = \sum_{\substack{1 \leq i_1 \leq \dots \leq i_k \leq d \\ \alpha_{i_1} + \dots + \alpha_{i_k} = j}} \frac{\partial^j f_i}{\partial x_{i_1}^{\alpha_1} \dots \partial x_{i_k}^{\alpha_k}}(\theta)\, x_{i_1}^{\alpha_1} \dots x_{i_k}^{\alpha_k}. \tag{10}$$

For $i \in \{1, \ldots, d\}$, we have

$$\mathrm{E}\left(\frac{\partial f_n}{\partial x_i}(\theta)\right) = \frac{1}{h_n^{d+1}} \int_{\mathbb{R}^d} \frac{\partial K}{\partial x_i}\left(\frac{\theta - x}{h_n}\right) f(x)\mathrm{d}x = \frac{1}{h_n^d} \int_{\mathbb{R}^d} K\left(\frac{\theta - x}{h_n}\right) \frac{\partial f}{\partial x_i}(x)\mathrm{d}x = \int_{\mathbb{R}^d} K(y) f_i\left(\theta - h_n y\right) \mathrm{d}y$$

so that

$$\mathrm{E}\left(\frac{\partial f_n}{\partial x_i}(\theta)\right) - \frac{\partial f}{\partial x_i}(\theta) = \int_{\mathbb{R}^d} K(y)\left[f_i\left(\theta - h_n y\right) - f_i(\theta)\right] \mathrm{d}y.$$

It follows from (10) and (A5) iii) that

$$\mathrm{E}\left(\frac{\partial f_n}{\partial x_i}(\theta)\right) - \frac{\partial f}{\partial x_i}(\theta) = h_n^q \int_{\mathbb{R}^d} \|y\|^q K(y) \left[\frac{f_i\left(\theta - h_n y\right) - f_i(\theta) - \sum_{j=1}^{q-1} \frac{(-1)^j h_n^j}{j!} D^j f_i(\theta)\left(\underline{y}^{(j)}\right)}{h_n^q \|y\|^q}\right] \mathrm{d}y. \quad (11)$$

The bracketed term in the last integral is bounded for all values of h_n and y, $\int_{\mathbb{R}^d} \|y\|^q |K(y)| \mathrm{d}y < \infty$ and, for any $y \neq 0$,

$$\lim_{n\to\infty} \left[\frac{f_i\left(\theta - h_n y\right) - f_i(\theta) - \sum_{j=1}^{q-1} \frac{(-1)^j h_n^j}{j!} D^j f_i(\theta)\left(\underline{y}^{(j)}\right)}{h_n^q \|y\|^q}\right] = \frac{(-1)^q}{q!\|y\|^q} D^q f_i(\theta)\left(\underline{y}^{(q)}\right).$$

Thus, we have

$$\lim_{n\to\infty} \frac{1}{h_n^q}\left[E\left(\frac{\partial f_n}{\partial x_i}(\theta)\right) - \frac{\partial f}{\partial x_i}(\theta)\right] = \frac{(-1)^q}{q!} \int_{\mathbb{R}^d} D^q f_i(\theta)\left(\underline{y}^{(q)}\right) K(y)\mathrm{d}y.$$

In view of (10) and (A5) iii), it comes:

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{h_n^q}\left[E\left(\frac{\partial f_n}{\partial x_i}(\theta)\right) - \frac{\partial f}{\partial x_i}(\theta)\right] &= \frac{(-1)^q}{q!} \sum_{j=1}^d \int_{\mathbb{R}^d} \frac{\partial^q f_i}{\partial x_j^q}(\theta) y_j^q K(y)\mathrm{d}y = \frac{(-1)^q}{q!} \sum_{j=1}^d \mu_j^q \frac{\partial^{q+1} f}{\partial x_i \partial x_j^q}(\theta) \\ &= \frac{(-1)^q}{q!} \frac{\partial}{\partial x_i}\left(\sum_{j=1}^d \mu_j^q \frac{\partial^q f}{\partial x_j^q}(\theta)\right)\end{aligned}$$

so that

$$\lim_{n\to\infty} \frac{1}{h_n^q}\left[E\left(\nabla f_n(\theta)\right) - \nabla f(\theta)\right] = \frac{(-1)^q}{q!} \nabla\left(\sum_{j=1}^d \mu_j^q \frac{\partial^q f}{\partial x_j^q}(\theta)\right)$$

which concludes the proof of Lemma 3.2. □

3.3. **Central limit theorem for the kernel mode estimator**

A straightforward application of Lyapounov's theorem gives the following central limit theorem fulfilled by the variance term of the kernel estimator of the density derivative:

$$\sqrt{n h_n^{d+2}}\left(\nabla f_n(\theta) - \mathrm{E}\left(\nabla f_n(\theta)\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0\ ,\ f(\theta) G\right). \quad (12)$$

In view of the previous section, Theorem 2.2 follows.

In order to justify that the limit laws in Parts i) and ii) of Theorem 2.2 are nondegenerate, we now prove Proposition 2.3 (which is straightforward in the case $d = 1$).

Proof of Proposition 2.3. Let $u = (u_1, \dots, u_d)^t \in \mathbb{R}^d$; we have

$$u^t G u \;=\; \int_{\mathbb{R}^d} \left(\sum_{i=1}^{d} u_i \frac{\partial K}{\partial x_i}(x) \right)^2 \mathrm{d}x \;\geq 0. \tag{13}$$

Assume there exists $u \neq 0$ such that $u^t G u = 0$. In view of (13), we then have $\sum_{i=1}^d u_i \partial K / \partial x_i \equiv 0$. Without loss of generality, assume that, for some $h \in \{1, \dots, d\}$, $u_1 \neq 0, \dots, u_h \neq 0$ and $u_{h+1} = \dots = u_d = 0$. Then,

$$u_1 \frac{\partial K}{\partial x_1} + \dots + u_h \frac{\partial K}{\partial x_h} \equiv 0.$$

The general solution of this first order linear partial differential equation is

$$\phi \left(u_2 x_1 - u_1 x_2, \dots, u_h x_1 - u_1 x_h, x_{h+1}, \dots, x_d \right)$$

where ϕ is an arbitrary real-valued differentiable function on $\mathbb{R}^{d-1}$. We thus have $K = \phi \circ M$ where M: $\mathbb{R}^d \to \mathbb{R}^{d-1}$ is a linear application. The rank of M is $d-1$ and $\ker M = \operatorname{span}\{v\}$ where $v = (u_1, u_2, \dots, u_h, 0, \dots, 0)^t$. Set $z \in \mathbb{R}^{d-1}$ and $x_0 \in M^{-1}(z)$; we have

$$K\left(x_0 + \lambda v\right) = \phi(z) \quad \forall \lambda \in \mathbb{R}.$$

Since $\lim_{\lambda \to \infty} K\left(x_0 + \lambda v\right) = 0$, it follows that $\phi(z) = 0$ for all $z \in \mathbb{R}^{d-1}$ and thus $K \equiv 0$, which is impossible. Thus, $u^t G u > 0$ for any $u \neq 0$, which proves Proposition 2.3.

3.4. **Law of the iterated logarithm for the kernel mode estimator**

Let $A = (A_{i,j})_{1 \leq i \leq d', 1 \leq j \leq d}$ be a given $d' \times d$ matrix, and set

$$\begin{aligned} R_n &= -\left[D^2 f(\theta)\right]^{-1} \left(\nabla f_n\left(\theta\right) - \mathrm{E}\left(\nabla f_n\left(\theta\right)\right)\right) \\ V_n &= \sqrt{\frac{n h_n^{d+2}}{2 \ln \ln n}}\, A R_n. \end{aligned}$$

From now on, we set $\Delta = -\left[D^2 f(\theta)\right]^{-1}$. For $i \in \{1, \dots, d\}$, we have

$$\begin{aligned} R_{n,i} &= \sum_{j=1}^{d} \Delta_{i,j} \frac{\partial f_n}{\partial x_j}(\theta) - E\left(\sum_{j=1}^{d} \Delta_{i,j} \frac{\partial f_n}{\partial x_j}(\theta) \right) \\ V_{n,i} &= \sqrt{\frac{n h_n^{d+2}}{2 \ln \ln n}} \left[\sum_{j=1}^{d} \sum_{l=1}^{d} A_{i,j} \Delta_{j,l} \frac{\partial f_n}{\partial x_l}(\theta) - E\left(\sum_{j=1}^{d} \sum_{l=1}^{d} A_{i,j} \Delta_{j,l} \frac{\partial f_n}{\partial x_l}(\theta) \right) \right] \\ &= \frac{1}{\sqrt{2 n h_n^d \ln \ln n}} \sum_{k=1}^{n} \left[\left(\sum_{j=1}^{d} \sum_{l=1}^{d} A_{i,j} \Delta_{j,l} \frac{\partial K}{\partial x_l} \right) \left(\frac{\theta - X_k}{h_n} \right) - E\left(\left(\sum_{j=1}^{d} \sum_{l=1}^{d} A_{i,j} \Delta_{j,l} \frac{\partial K}{\partial x_l} \right) \left(\frac{\theta - X_k}{h_n} \right) \right) \right]. \end{aligned}$$

Theorem 4.1 in Arcones [1] applies and gives the following result:

Lemma 3.3. *With probability one, the sequence (V_n) is relatively compact and its limit set is*

$$C=\left\{\left(\sqrt{f(\theta)}\int_{\mathbb{R}^d}\left[\sum_{j=1}^{d}\sum_{l=1}^{d}A_{i,j}\Delta_{j,l}\frac{\partial K}{\partial x_l}(x)\right]\ \alpha(x)\ \mathrm{d}x\right)_{i\in\{1,\dots,d'\}}\ :\ \int_{\mathbb{R}^d}\alpha^2(x)\ \mathrm{d}x\leq 1\right\}.$$

Note that C is the image of the closed unit ball of $L^2(\mathbb{R}^d)$ by the linear continuous application $\mathcal{H}:L^2(\mathbb{R}^d)\to\mathbb{R}^{d'}$ defined by

$$\mathcal{H}(\alpha)=\left(\sqrt{f(\theta)}\int_{\mathbb{R}^d}\left[\sum_{j=1}^{d}\sum_{l=1}^{d}A_{i,j}\Delta_{j,l}\frac{\partial K}{\partial x_l}(x)\right]\ \alpha(x)\ \mathrm{d}x\right)_{i\in\{1,\dots,d'\}}.$$

We now show how Theorems 2.4, 2.5 and 2.6 are deduced from Lemma 3.3.

3.4.1. *Law of the iterated logarithm for the full vector*

Let A be the identity matrix; in view of Lemma 3.3, the limit set of the sequence

$$\left(\sqrt{\frac{nh_n^{d+2}}{2\ln\ln n}}\ \Delta\ \left[\nabla f_n\left(\theta\right)-\mathrm{E}\left(\nabla f_n\left(\theta\right)\right)\right]\right)$$

is, with probability one,

$$C=\left\{\left(\sqrt{f(\theta)}\int_{\mathbb{R}^d}\left[\sum_{j=1}^{d}\Delta_{i,j}\frac{\partial K}{\partial x_j}(x)\right]\ \alpha(x)\ \mathrm{d}x\right)_{i\in\{1,\dots,d\}}\ :\ \int_{\mathbb{R}^d}\alpha^2(x)\ \mathrm{d}x\leq 1\right\}.$$

Now, let $\mathcal{U}$ be the vector subspace of $L^2(\mathbb{R}^d)$ spanned by $\left(\frac{\partial K}{\partial x_i}\right)_{i\in\{1,\dots,d\}}$. For any $\alpha\in L^2\left(\mathbb{R}^d\right)$, there exists $\gamma=(\gamma_1,\dots,\gamma_d)^t$ and $\overline{\alpha}\in\mathcal{U}^\perp$ such that

$$\alpha(x)=\sum_{k=1}^{d}\gamma_k\frac{\partial K}{\partial x_k}(x)+\overline{\alpha}(x)$$

and C can be rewritten as

$$C=\left\{\left(\sqrt{f(\theta)}\int_{\mathbb{R}^d}\left[\sum_{j=1}^{d}\Delta_{i,j}\frac{\partial K}{\partial x_j}(x)\right]\ \alpha(x)\ \mathrm{d}x\right)_{i\in\{1,\dots,d\}}\ :\ \alpha\in\mathcal{U}\text{ and }\int_{\mathbb{R}^d}\alpha^2(x)\ \mathrm{d}x\leq 1\right\}.$$

However, for any $\alpha\in\mathcal{U}$, $\alpha=\sum_{k=1}^{d}\gamma_k\frac{\partial K}{\partial x_k}$, we have

$$\int_{\mathbb{R}^d}\alpha^2(x)\ \mathrm{d}x=\sum_{k=1}^{d}\sum_{j=1}^{d}\gamma_k\gamma_j\int_{\mathbb{R}^d}\frac{\partial K}{\partial x_k}(x)\frac{\partial K}{\partial x_j}(x)\mathrm{d}x=\gamma^tG\gamma$$

and

$$\int_{\mathbb{R}^d}\left[\sum_{j=1}^{d}\Delta_{i,j}\frac{\partial K}{\partial x_j}(x)\right]\ \alpha(x)\ \mathrm{d}x=\int_{\mathbb{R}^d}\sum_{j=1}^{d}\sum_{k=1}^{d}\Delta_{i,j}\gamma_k\frac{\partial K}{\partial x_j}(x)\frac{\partial K}{\partial x_k}(x)\mathrm{d}x$$
$$=\sum_{j=1}^{d}\Delta_{i,j}\left(\sum_{k=1}^{d}\gamma_k\int_{\mathbb{R}^d}\frac{\partial K}{\partial x_j}(x)\frac{\partial K}{\partial x_k}(x)\mathrm{d}x\right)=\sum_{j=1}^{d}\Delta_{i,j}(G\gamma)_j=(\Delta G\gamma)_i.$$

Thus, we obtain

$$C=\left\{\left(\sqrt{f(\theta)}\ (\Delta G\gamma)_i\right)_{i\in\{1,\dots,d\}}\quad:\quad\gamma\in\mathbb{R}^d\ ,\ \gamma^tG\gamma\leq 1\right\}.$$

Finally, set $Z(\gamma)=\sqrt{f(\theta)}\ \Delta G\gamma$. We then have $\gamma=[f(\theta)]^{-1/2}G^{-1}\Delta^{-1}Z(\gamma)$ and

$$Z(\gamma)\in C\Leftrightarrow\gamma^tG\gamma\leq 1$$
$$\Leftrightarrow[f(\theta)]^{-1}Z(\gamma)^t\Delta^{-1}G^{-1}\Delta^{-1}Z(\gamma)\leq 1.$$

Thus, C is the ellipsoid

$$C=\left\{\nu\in\mathbb{R}^d\ \text{ such that }\ \frac{1}{f(\theta)}\nu^t\left[D^2f(\theta)\right]G^{-1}\left[D^2f(\theta)\right]\nu\leq 1\right\}.$$

Theorem 2.4 follows then from the considerations made in Section 3.2.

3.4.2. *Law of the iterated logarithm for the linear forms*

Set $u\in\mathbb{R}^d$ and $A=u^t$. In view of Lemma 3.3, the limit set of the sequence

$$\left(\sqrt{\frac{nh_n^{d+2}}{2\ln\ln n}}\ u^t\Delta\left(\nabla f_n\left(\theta\right)-\mathrm{E}\left(\nabla f_n\left(\theta\right)\right)\right)\right)$$

is, with probability one,

$$C=\left\{\left(\sqrt{f(\theta)}\int_{\mathbb{R}^d}J(x)\ \alpha(x)\ \mathrm{d}x\right)\quad:\quad\int_{\mathbb{R}^d}\alpha^2(x)\ \mathrm{d}x\leq 1\right\}$$

with

$$J(x)=\sum_{i=1}^{d}\sum_{j=1}^{d}u_i\Delta_{i,j}\frac{\partial K}{\partial x_j}(x).$$

Note that C is the image of the closed unit ball of $L^2(\mathbb{R}^d)$ by the linear continuous form

$$\mathcal{H}:\alpha\mapsto\sqrt{f(\theta)}\int_{\mathbb{R}^d}J(x)\ \alpha(x)\ \mathrm{d}x.$$

Since the unit ball of $L^2(\mathbb{R}^d)$ is connected and symmetric with respect to 0, it follows that C is a symmetric interval. Now, for any α such that $\int_{\mathbb{R}^d}\alpha^2(x)\ \mathrm{d}x\leq 1$, we have

$$\left|\int_{\mathbb{R}^d}J(x)\ \alpha(x)\ \mathrm{d}x\right|\leq\left(\int_{\mathbb{R}^d}J^2(x)\ \mathrm{d}x\right)^{1/2}$$

so that C is bounded by $\left(f(\theta)\int_{\mathbb{R}^d} J^2(x)\ \mathrm{d}x\right)^{1/2}$. But, setting $\alpha(x) = \left(\int_{\mathbb{R}^d} J^2(x)\ \mathrm{d}x\right)^{-1/2} J(x)$, we have $\int_{\mathbb{R}^d}\alpha^2(x)$ $\mathrm{d}x = 1$ and

$$\int_{\mathbb{R}^d} J(x)\ \alpha(x)\ \mathrm{d}x = \left(\int_{\mathbb{R}^d} J^2(x)\mathrm{d}x\right)^{1/2}$$

so that

$$C = \left[\ -\sqrt{f(\theta)\int_{\mathbb{R}^d} J^2(x)\mathrm{d}x}\ ;\ +\sqrt{f(\theta)\int_{\mathbb{R}^d} J^2(x)\mathrm{d}x}\ \right].$$

Finally, noting that

$$\int_{\mathbb{R}^d} J^2(x)\mathrm{d}x = \sum_{i,j,k,l} u_i u_l \Delta_{i,j}\Delta_{l,k}\int_{\mathbb{R}^d}\frac{\partial K}{\partial x_j}(x)\frac{\partial K}{\partial x_k}(x)\mathrm{d}x = \sum_{i,j,k,l} u_i u_l \Delta_{i,j}\Delta_{l,k} G_{j,k} = u^t\Delta G\Delta u$$

we obtain

$$C = \left[\ -\sqrt{f(\theta)u^t\Delta G\Delta u}\ ;\ +\sqrt{f(\theta)u^t\Delta G\Delta u}\ \right].$$

3.4.3. *Law of the iterated logarithm for the l^p norms*

Let N_p: $z \mapsto \|z\|_p$ be the l^p norm application on $\mathbb{R}^d$, and set

$$C = \left\{\nu\ /\ \nu^t\Sigma\nu \le 1\right\}\quad \text{with}\quad \Sigma = \left[D^2 f(\theta)\right] G^{-1}\left[D^2 f(\theta)\right]/f(\theta).$$

Since N_p is continuous, Theorem 2.4 implies that, with probability one, the sequence

$$\sqrt{\frac{nh_n^{d+2}}{2\ln\ln n}}\ \|\theta_n - \theta\|_p$$

is relatively compact and its limit set is $N_p(C)$. Moreover, since C is a compact and connected subset of $\mathbb{R}^d$, $N_p(C)$ is also compact and connected. Noting that $0 \in N_p(C)$, it follows that $N_p(C) = [0, \delta'_p]$ with $\delta'_p = \sup_{\nu\in C}\|\nu\|_p$. In view of the definition of C, we then have

$$\delta'_p = \sup_{\|\mu\|_2\le 1}\|\Sigma^{-1/2}\mu\|_p = \|\Sigma^{-1/2}\|_{2,p} = \|\sqrt{f(\theta)}G^{1/2}\left[D^2 f(\theta)\right]^{-1}\|_{2,p} = \sqrt{f(\theta)}\ \delta_p.$$

In particular, $\delta_2 = \|G^{1/2}\left[D^2 f(\theta)\right]^{-1}\|_{2,2}$; since $G^{1/2}\left[D^2 f(\theta)\right]^{-1}$ is definite negative, δ_2 is thus the spectral radius of the matrix $-G^{1/2}\left[D^2 f(\theta)\right]^{-1}$.

Now, for $i \in \{1,\dots,d\}$, set $k_i(x) = \sum_{j=1}^d \Delta_{i,j}\frac{\partial K}{\partial x_j}(x)$; we have

$$\int_{\mathbb{R}^d} k_i^2(x)\ \mathrm{d}x = \sum_{j,k}\Delta_{i,j}\Delta_{i,k}\int_{\mathbb{R}^d}\frac{\partial K}{\partial x_j}(x)\frac{\partial K}{\partial x_k}(x)\ \mathrm{d}x = \sum_{j,k}\Delta_{i,j}\Delta_{i,k}G_{j,k} = (\Delta G\Delta)_{i,i}.$$

The application of Lemma 3.3 when A is the identity matrix ensures that

$$C = \left\{\left(\sqrt{f(\theta)}\int_{\mathbb{R}^d} k_i(x)\ \alpha(x)\ \mathrm{d}x\right)_{i\in\{1,\dots,d\}}\ :\ \int_{\mathbb{R}^d}\alpha^2(x)\ \mathrm{d}x \le 1\right\}.$$

It thus follows that

$$\delta'^2_\infty \leq f(\theta) \max_{1\leq i\leq d} \int_{\mathbb{R}^d} k_i^2(x)\ \mathrm{d}x \leq f(\theta) \max_{1\leq i\leq d} (\Delta G\Delta)_{i,i}.$$

Finally, let $i_0 \in \{1, \ldots, d\}$ be the integer satisfying $(\Delta G\Delta)_{i_0,i_0} = \max_{1\leq i\leq d}(\Delta G\Delta)_{i,i}$. For

$$\alpha(x) = \left[\int_{\mathbb{R}^d} k_{i_0}^2(x)\ \mathrm{d}x\right]^{-1/2} k_{i_0}(x)$$

we have

$$\int_{\mathbb{R}^d} k_{i_0}(x)\ \alpha(x)\ \mathrm{d}x = \left[\int_{\mathbb{R}^d} k_{i_0}^2(x)\ \mathrm{d}x\right]^{1/2} = \left[(\Delta G\Delta)_{i_0,i_0}\right]^{1/2}$$

which ensures that $\delta_\infty = [\max_{1\leq i\leq d}(\Delta G\Delta)_{i,i}]^{1/2}$.

3.5. Convergence rate of the univariate kernel mode estimator in the case θ is degenerate

The key idea to prove Theorems 2.7 and 2.8 is to relate the weak and strong convergence rate of $(\theta_n - \theta)^{p-1}$ to that of $-(p-1)! f_n'(\theta) / f^{(p)}(\theta)$. For that purpose, we first note that a Taylor expansion of f_n' at the point θ ensures the existence of ξ_n such that

$$\begin{cases} f_n'(\theta_n) - f_n'(\theta) = \displaystyle\sum_{j=2}^{p-1} \frac{f_n^{(j)}(\theta)}{(j-1)!}\ (\theta_n - \theta)^{j-1} + \frac{f_n^{(p)}(\xi_n)}{(p-1)!}\ (\theta_n - \theta)^{p-1} \\ |\xi_n - \theta| \leq |\theta_n - \theta|. \end{cases}$$

(If $p = 2$, that is, in the case θ is nondegenerate, $\sum_{j=2}^{p-1} f_n^{(j)}(\theta)(\theta_n - \theta)^{j-1}/(j-1)! = 0$.) Since $f_n'(\theta_n) = 0$ it follows that

$$\frac{f_n^{(p)}(\xi_n)}{(p-1)!}\ (\theta_n - \theta)^{p-1} = -f_n'(\theta) - \sum_{j=2}^{p-1} \frac{f_n^{(j)}(\theta)}{(j-1)!}\ (\theta_n - \theta)^{j-1}.$$

Now, the a.s. uniform convergence of $f_n^{(p)}$ to $f^{(p)}$ in a neighborhood of θ and the strong consistency of θ_n ensure that $\lim_{n\to\infty} f_n^{(p)}(\xi_n) = f^{(p)}(\theta)$ a.s.; the weak and strong convergence rate of $(\theta_n - \theta)^{p-1}$ is thus given by that of

$$-\frac{(p-1)!}{f^{(p)}(\theta)} \left[f_n'(\theta) + \sum_{j=2}^{p-1} \frac{f_n^{(j)}(\theta)}{(j-1)!}\ (\theta_n - \theta)^{j-1}\right].$$

To prove Theorems 2.7 and 2.8, it remains to establish the convergence rate of $f_n'(\theta)$ on one hand, and, on the other hand, to prove that the term $\sum_{j=2}^{p-1} f_n^{(j)}(\theta)(\theta_n - \theta)^{j-1}/(j-1)!$ is negligible in front of $f_n'(\theta)$.

We first note that

$$\lim_{n\to\infty} \frac{1}{h_n^q}\ E\left(f_n'(\theta)\right) = \frac{(-1)^q}{q!}\ f^{(q+1)}(\theta) \int_{\mathbb{R}} y^q K(y)\mathrm{d}y. \tag{14}$$

As a matter of fact, we have $f^{(j+1)}(\theta) = 0\ \forall j \in \{1, \ldots, p-2\}$ and $\int_{\mathbb{R}} y^j K(y)\mathrm{d}y = 0\ \forall j \in \{p-1, \ldots, q-1\}$; it follows that equation (11) still holds and (14) is obtained by following the proof of Lemma 3.2. Moreover, when

$\|f^{(q+1)}\|_\infty < \infty$, the bracketed term in (11) in bounded by $\|f^{(q+1)}\|_\infty/q!$ uniformly with respect to θ. In this case, we have:

$$\|E\left(f_n'\right)\|_\infty = O\left(h_n^q\right). \tag{15}$$

Now, the weak convergence rate of $f_n'(\theta)$ is given by the application of the univariate version of the central limit theorem (12):

$$\sqrt{nh_n^3}\left(f_n'(\theta) - E\left(f_n'(\theta)\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0\ ,\ f(\theta)\int_{\mathbb{R}} K'^2(x)\ \mathrm{d}x\right).$$

For the strong convergence rate of $f_n'(\theta)$, we note that the application of Theorem 4.1 in Arcones [1] ensures that the sequence

$$\left(\sqrt{\frac{nh_n^3}{2\ln\ln n}}\left(f_n'(\theta) - E\left[f_n'(\theta)\right]\right)\right)$$

is relatively compact and its limit set is

$$\begin{aligned} C &= \left\{\sqrt{f(\theta)}\int_{\mathbb{R}} K'(x)\alpha(x)\ \mathrm{d}x \quad : \quad \int_{\mathbb{R}}\alpha^2(x)\ \mathrm{d}x \leq 1\right\} \\ &= \left[\ -\sqrt{f(\theta)\int_{\mathbb{R}} K'^2(x)\ \mathrm{d}x}\ ;\ +\sqrt{f(\theta)\int_{\mathbb{R}} K'^2(x)\ \mathrm{d}x}\ \right]. \end{aligned}$$

where the last equality is obtained by following the proof of Section 3.4.2.

It remains to prove that the term $\sum_{j=2}^{p-1} f_n^{(j)}(\theta)(\theta_n-\theta)^{j-1}/(j-1)!$ is negligible. This is given by the following lemma:

Lemma 3.4. *i) If there exists $c \in [0, \infty[$ such that $\lim_{n\to\infty} nh_n^{2q+3} = c$, then*

$$\lim_{n\to\infty}\sqrt{nh_n^3}\sum_{j=2}^{p-1}\frac{f_n^{(j)}(\theta)}{(j-1)!}\left(\theta_n-\theta\right)^{j-1} = 0 \quad a.s.$$

ii) If $\lim_{n\to\infty} nh_n^{2q+3} = \infty$, then

$$\lim_{n\to\infty}\frac{1}{h_n^q}\sum_{j=2}^{p-1}\frac{f_n^{(j)}(\theta)}{(j-1)!}\left(\theta_n-\theta\right)^{j-1} = 0 \quad a.s.$$

Proof of Lemma 3.4. To prove Lemma 3.4, we need an a.s. convergence rate upper bound of the sequences $\left(\left|f_n^{(j)}(\theta)\right|\right)$, $j \in \{2, \ldots, p-1\}$, as well as a preliminary upper bound of the a.s. convergence rate of $(\theta_n - \theta)$. At first, note that the application of Theorem 2.3 in Giné and Guillou [8] to the kernel $K^{(j)}$ ensures that

$$\left\|f_n^{(j)} - E\left(f_n^{(j)}\right)\right\|_\infty = O\left(\sqrt{\frac{\ln[1/h_n]}{nh_n^{2j+1}}}\right) \quad \text{a.s.} \tag{16}$$

Now, arguing as for the proof of (14), we have $\left|E\left(f_n^{(j)}(\theta)\right)\right| = O\left(h_n^{q+1-j}\right)$ and thus

$$\left|f_n^{(j)}(\theta)\right| = O\left(\sqrt{\frac{\ln[1/h_n]}{nh_n^{2j+1}}} + h_n^{q+1-j}\right) \quad \text{a.s.} \tag{17}$$

To obtain the preliminary upper bound of the strong convergence rate of $(\theta_n - \theta)$, we first note that a Taylor expansion of f' at the point θ ensures the existence of ζ_n such that

$$\begin{cases} f'(\theta_n) = \dfrac{f^{(p)}(\zeta_n)}{(p-1)!}\,(\theta_n - \theta)^{p-1} \\ |\zeta_n - \theta| \leq |\theta_n - \theta|. \end{cases}$$

Since $f_n'(\theta_n) = 0$, it follows that

$$f_n'(\theta_n) - f'(\theta_n) = \frac{-f^{(p)}(\zeta_n)}{(p-1)!}\,(\theta_n - \theta)^{p-1}.$$

The strong consistency of θ_n, the continuity of $f^{(p)}$ and the fact that $f^{(p)}(\theta) \neq 0$ imply that

$$|\theta_n - \theta|^{p-1} = O\left(\|f_n' - f'\|_\infty\right) \quad \text{a.s.}$$

In view of (15) and (16), we obtain:

$$|\theta_n - \theta|^{p-1} = O\left(\sqrt{\frac{\ln[1/h_n]}{nh_n^3}} + h_n^q\right) \quad \text{a.s.} \tag{18}$$

For any $j \in \{2, \dots, p-1\}$, the combination of (17) and (18) gives

$$\left[\left|f^{(j)}(\theta)\right| |\theta_n - \theta|^{j-1}\right]^2 = O\left(\left[\frac{\ln[1/h_n]}{nh_n^{2j+1}} + h_n^{2(q+1-j)}\right]\left[\left(\frac{\ln[1/h_n]}{nh_n^3}\right)^{\frac{j-1}{p-1}} + h_n^{\frac{2q(j-1)}{p-1}}\right]\right) \quad \text{a.s.} \tag{19}$$

• Let us at first assume that there exists $c \in [0, \infty[$ such that $\lim_{n\to\infty} nh_n^{2q+3} = c$. In this case, we note that, for any $j \geq 2$,

$$h_n^{2(q+1-j)} = o\left(\frac{\ln[1/h_n]}{nh_n^{2j+1}}\right) \quad \text{and} \quad h_n^{2q} = o\left(\frac{\ln[1/h_n]}{nh_n^3}\right).$$

In view of (19), it follows that

$$\begin{aligned} nh_n^3\,\left[\left|f^{(j)}(\theta)\right| |\theta_n - \theta|^{j-1}\right]^2 &= O\left(nh_n^3\left[\frac{\ln[1/h_n]}{nh_n^{2j+1}}\right]\left(\frac{\ln[1/h_n]}{nh_n^3}\right)^{\frac{j-1}{p-1}}\right) \quad \text{a.s.} \\ &= O\left(\left[\frac{[\ln[1/h_n]]^{1+\frac{p-1}{j-1}}}{nh_n^{2p+1}}\right]^{\frac{j-1}{p-1}}\right) \quad \text{a.s.} \\ &= o(1) \quad \text{a.s.} \end{aligned}$$

• Let us now assume that $\lim_{n\to\infty} nh_n^{2q+3} = \infty$. In view of (19), we have

$$
\begin{aligned}
&\frac{1}{h_n^{2q}}\left[\left|f^{(j)}(\theta)\right| |\theta_n-\theta|^{j-1}\right]^2\\
&= O\left(\left[\frac{\ln[1/h_n]}{nh_n^{2q+2j+1}} + h_n^{-2(j-1)}\right]\left[\left(\frac{\ln[1/h_n]}{nh_n^3}\right)^{\frac{j-1}{p-1}} + h_n^{\frac{2q(j-1)}{p-1}}\right]\right) \quad \text{a.s.}\\
&= O\left(\left[\frac{(\ln[1/h_n])^{p+j-2}}{\left(nh_n^{2q+2j+1}\right)^{p-1}(nh_n^3)^{j-1}}\right]^{\frac{1}{p-1}} + \left[\frac{(\ln[1/h_n])^{p-1}h_n^{2q(j-1)}}{\left(nh_n^{2q+2j+1}\right)^{p-1}}\right]^{\frac{1}{p-1}} + \left[\frac{\ln[1/h_n]}{nh_n^{3+2(p-1)}}\right]^{\frac{j-1}{p-1}}\right.\\
&\qquad \left. + \left[h_n^{-(p-1)+q}\right]^{\frac{2(j-1)}{p-1}}\right) \quad \text{a.s.}\\
&= O\left(\frac{1}{nh_n^{2q+3}}\left[\frac{(\ln[1/h_n])^{1+\frac{p-1}{j-1}}}{nh_n^{2p+1}}\right]^{\frac{j-1}{p-1}} + \left[\frac{(\ln[1/h_n])^{p-1}}{\left(nh_n^{2q+3}\right)^{p-1}h_n^{2(j-1)(p-q-1)}}\right]^{\frac{1}{p-1}} + \left[\frac{\ln[1/h_n]}{nh_n^{2p+1}}\right]^{\frac{j-1}{p-1}}\right.\\
&\qquad \left. + \left[h_n^{q-p+1}\right]^{\frac{2(j-1)}{p-1}}\right) \quad \text{a.s.}\\
&= o(1) \quad \text{a.s.}
\end{aligned}
$$

which concludes the proof of Lemma 3.4. □

We would like to thank the referees and the Associate Editor for their carefully reading and their helpful comments and suggestions.

References

[1] M.A. Arcones, The law of the iterated logarithm for a triangular array of empirical processes. *Electron. J. Probab.* **2** (1997) 1-39.

[2] A. Berlinet, A. Gannoun and E. Matzner–Loeber, Normalité asymptotique d'estimateurs convergents du mode conditionnel. *Can. J. Statist.* **26** (1998) 365-380.

[3] H. Chernoff, Estimation of the mode. *Ann. Inst. Stat. Math.* **16** (1964) 31-41.

[4] G. Collomb, W. Härdle and S. Hassani, A note on prediction *via* estimation of the conditional mode function. *J. Statist. Planning Inference* **15** (1987) 227-236.

[5] W.F. Eddy, Optimum kernel estimates of the mode. *Ann. Statist.* **8** (1980) 870-882.

[6] W.F. Eddy, The asymptotic distributions of kernel estimators of the mode. *Z. Warsch. Verw. Geb.* **59** (1982) 279-290.

[7] U. Einmahl and D.M. Mason, An empirical process approach to the uniform consistency of kernel-type functions estimators. *J. Theoret. Probab.* **13** (2000) 1-37.

[8] E. Giné and A. Guillou, *Rates of strong uniform consistency for multivariate kernel density estimators*, Preprint. Paris VI (2000).

[9] U. Grenander, Some direct estimates of the mode. *Ann. Math. Statist.* **36** (1965) 131-138.

[10] B. Grund and P. Hall, On the minimisation of L^p error in mode estimation. *Ann. Statist.* **23** (1995) 2264-2284.

[11] P. Hall, Laws of the iterated logarithm for nonparametric density estimators. *Z. Warsch. Verw. Geb.* **56** (1981) 47-61.

[12] P. Hall, Asymptotic theory of Grenander's mode estimator. *Z. Warsch. Verw. Geb.* **60** (1982) 315-334.

[13] V.D. Konakov, On asymptotic normality of the sample mode of multivariate distributions. *Theory Probab. Appl.* **18** (1973) 836-842.

[14] J. Leclerc and D. Pierre–Loti–Viaud, Vitesse de convergence presque sûre de l'estimateur à noyau du mode. *C. R. Acad. Sci. Paris Sér. I Math.* **331** (2000) 637-640.

[15] D. Louani and E. Ould–Said, Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. *J. Nonparametr. Statist.* **11** (1999) 413-442.

[16] A. Mokkadem and M. Pelletier, A law of the iterated logarithm for the kernel mode estimator. *Statist. Probab. Lett.* (submitted).
[17] E.A. Nadaraya, On non-parametric estimates of density functions and regression curves. *Theory Probab. Appl.* **10** (1965) 186-190.
[18] E. Ould–Said, A note on ergodic processes prediction *via* estimation of the conditional mode function. *Scand. J. Stat.* **24** (1997) 231-239.
[19] E. Parzen, On estimating probability density function and mode. *Ann. Math. Statist.* **33** (1962) 1065-1076.
[20] D. Pollard, *Convergence of Stochastic Processes.* Springer, New York (1984).
[21] A. Quintela–Del–Rio and P. Vieu, A nonparametric conditional mode estimate. *J. Nonparametr. Statist.* **8** (1997) 253-266.
[22] J. Romano, On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.* **16** (1988) 629-647.
[23] L. Rüschendorf, Consistency of estimators for multivariate density functions and for the mode. *Sankhya Ser. A* **39** (1977) 243-250.
[24] T.W. Sager, Consistency in nonparametric estimation of the mode. *Ann. Statist.* **3** (1975) 698-706.
[25] M. Samanta, Nonparametric estimation of the mode of a multivariate density. *South African Statist. J.* **7** (1973) 109-117.
[26] M. Samanta and A. Thavaneswaran, Nonparametric estimation of the conditional mode. *Commun Stat., Theory Methods* **19** (1990) 4515-4524.
[27] A.B. Tsybakov, Recurrent estimation of the mode of a multidimensional distribution. *Problems Inform. Transmission* **26** (1990) 31-37.
[28] J. Van Ryzin, On strong consistency of density estimates. *Ann. Math. Statist.* **40** (1969) 1765-1772.
[29] J.H. Venter, On estimation of the mode. *Ann. Math. Statist.* **38** (1967) 1446-1455.
[30] P. Vieu, A note on density mode estimation. *Statist. Probab. Lett.* **26** (1996) 297-307.
[31] H. Yamato, Sequential estimation of a continuous probability density function and the mode. *Bull. Math. Statist.* **14** (1971) 1-12.