BANDWIDTH SELECTION FOR THE WOLVERTON-WAGNER ESTIMATOR

FABIENNE COMTE* AND NICOLAS MARIE**

ABSTRACT. For n independent random variables having the same Hölder continuous density, this paper deals with controls of the Wolverton-Wagner's estimator MSE and MISE. Then, for a bandwidth $h_n(\beta)$, estimators of β are obtained by a Goldenshluger-Lepski type method and a Lacour-Massart-Rivoirard type method. Some numerical experiments are provided for this last method.

CONTENTS

1. INTRODUCTION

Consider $n \in \mathbb{N}^*$ independent random variables $X_1, \dots, X_n$ having the same probability distribution of density f with respect to Lebesgue's measure.

The usual Parzen [11] - Rosenblatt [12] kernel estimator of f is defined by

$$\widehat{f}_{n,h}(x) := \frac{1}{nh}\sum_{k=1}^{n} K\left(\frac{X_k - x}{h}\right) \ ; \ x \in \mathbb{R},$$

where $h > 0$ and $K : \mathbb{R} \to \mathbb{R}_+$ is a kernel. In 1969, Wolverton and Wagner introduced in [15] a variant of $\widehat{f}_{n,h}(x)$ defined by

$$\widehat{f}_{n,\mathbf{h}_n}(x) := \frac{1}{n}\sum_{k=1}^{n} \frac{1}{h_k} K\left(\frac{X_k - x}{h_k}\right), \tag{1}$$

where $\mathbf{h}_n = (h_1, \dots, h_n)$ and $0 < h_n < \dots < h_1$. Thanks to its recursive form, this type of estimator is well-suited to online treatment of data: by denoting $\mathbf{h}_{n+1} =$

$(h_1, \dots, h_n, h_{n+1})$,

$$\widehat{f}_{n+1,\mathbf{h}_{n+1}}(x) = \frac{n}{n+1}\widehat{f}_{n,\mathbf{h}_n}(x) + \frac{1}{(n+1)h_{n+1}}K\left(\frac{X_{n+1}-x}{h_{n+1}}\right).$$

Thus, up-dating the estimator when new observations are available is easy and fast.

We can mention here that several variants or generalizations of the Wolverton and Wagner (WW) estimator have been proposed: see Yamato [16], Wegman and Davies [14], Hall and Patil [5]. They were studied from almost sure convergence point of view, or asymptotic rates of convergence under fixed regularity assumptions. We choose to focus on Wolverton and Wagner estimator but our results and discussions may be applied to these.

Theoretical developments concerning either classical Parzen-Rosenblatt or WW recursive kernels estimators occurred recently following different and independent roads.
On the one hand, several recent works are dedicated to efficient and data-driven bandwidth selection, see Goldenshluger and Lespki [4] and several companion papers by these authors, or Lacour *et al.* [8] who proposed a modification of the method. The original Goldenshluger and Lepski (GL) method was difficult to implement because it turned out to be numerically consuming and with calibration difficulties, see Comte and Rebafka [3]. This is why the improvement proposed in Lacour *et al.* [8] has both theoretical and practical interest.
On the other hand, the increase of computer speed and of data sets sizes made fast up-dating of estimators mandatory. The theoretical developments in this context are in the field of stochastic algorithms (see e.g. Mokkadem *et al.* [10]) or in view of specific applications (see Bercu *et al.* [2]).

Bandwidths have to be chosen for WW estimators as for Parzen-Rosenblatt ones, and this choice is crucial to obtain good performances. This is why we propose to extend to this context general risk study as described in Tsybakov [13] and the GL method as improved by Lacour *et al.* [8]. More precisely, considering for instance $h_k = k^{-\gamma}$ for a parameter $\gamma > 0$ in formula (1), we study adaptive selection of γ. We prove risk bounds for the Mean Integrated Squares Error (MISE) of the resulting estimator $\widehat{f}_{n,\widehat{\mathbf{h}}_n}$ where $\widehat{\mathbf{h}}_n = (\widehat{h}_1, \dots, \widehat{h}_n)$ and $\widehat{h}_k = k^{-\widehat{\gamma}}$.
Amiri [1] proved that for f with regularity 2 and an adequate choice of the bandwidth, Parzen-Rosenblatt's estimator had asymptotical smaller risk than the WW estimator. We propose an empirical finite sample study of this question, together with an interesting insight on the gain brought by higher order kernels.
Now, clearly, plugging $\widehat{\gamma} = \widehat{\gamma}(X_1, \dots, X_n)$ in the estimator makes the recursivity fail. Therefore, a mixed strategy is required with initial estimation of γ on the first n-sample and recursive up-dating relying on this "freezed" value on the following N-sample, where N should have the same order as n. This is what is experimented in our final section, and empirically proved to be an appropriate strategy.

This paper provides in Section 2 controls of the MSE and of the MISE of the estimator $\widehat{f}_{n,\mathbf{h}_n}$ under general regularity conditions on f. Then, in Section 3, the well-known Goldenshluger-Lepski's bandwidth selection method for Parzen-Rosenblatt's estimator is extended to Wolverton-Wagner's estimator. Lastly, an estimator in the spirit of Lacour *et al.* [8] is studied from both theoretical and practical point of view in Section 4. Concluding remarks present a mixed strategy in Section 5. Proofs are relegated in Section 6.

Notations:

(1) Consider $\alpha \in]0,1[$. The space of α-Hölder continuous functions from $\mathbb{R}$ into itself is denoted by $C^{\alpha}(\mathbb{R})$ and equipped with the α-Hölder semi-norm $\|.\|_{\alpha}$ defined by

$$\|\varphi\|_{\alpha} := \sup_{x,y\in\mathbb{R}:x\neq y} \frac{|\varphi(y)-\varphi(x)|}{|y-x|^{\alpha}} \text{ ; } \forall\varphi \in C^{\alpha}(\mathbb{R}).$$

(2) For $\beta > 0$ and $l := \lfloor\beta\rfloor$,

$$\Sigma(\beta) := \{\varphi \in C^{l}(\mathbb{R}) : \|\varphi^{(l)}\|_{\beta-l} < \infty\}.$$

(3) For every square integrable functions $f, g : \mathbb{R} \to \mathbb{R}$,

$$(f * g)(x) := \int_{-\infty}^{\infty} f(x-y)g(y)dy \text{ ; } x \in \mathbb{R}.$$

(4) $K_{\varepsilon} := (1/\varepsilon)K(\cdot/\varepsilon)$ for every $\varepsilon > 0$.

2. Bounds on the MSE and the MISE of Wolverton-Wagner's estimator

Consider $\beta > 0$ and $l := \lfloor\beta\rfloor$. Throughout this section, the map K fulfills the following assumption.

Assumption 2.1. *The map $y \in \mathbb{R} \mapsto y^{i}K(y)$ is integrable for every $i \in [\![0, l]\!]$,*

$$\int_{-\infty}^{\infty} K(y)dy = 1, \quad \int_{-\infty}^{\infty} K^{2}(y)dy < +\infty \quad \text{and} \quad \int_{-\infty}^{\infty} y^{i}K(y)dy = 0,\ \forall i \in [\![1, l]\!].$$

Let us establish a control of the MSE of Wolverton-Wagner's estimator under the following condition on f.

Assumption 2.2. *The map f belongs to $\Sigma(\beta)$.*

Proposition 2.3. *Under Assumptions 2.1 and 2.2, there exists a constant $c > 0$, not depending on n and $h_1, \dots, h_n$, such that*

$$\mathbb{E}(|\widehat{f}_{n,\mathbf{h}_n}(x) - f(x)|^{2}) \leqslant \frac{c}{n^{2}}\left(\left|\sum_{k=1}^{n}\frac{h_k^{\beta}}{l!}\right|^{2} + \sum_{k=1}^{n}\frac{1}{h_k}\right).$$

Now, let us establish a control of the MISE of Wolverton-Wagner's estimator under Nikolski's condition on f.

Assumption 2.4. *The map f belongs to $C^{l}(\mathbb{R})$ and there exists a constant $N(f) > 0$ such that*

$$\left(\int_{-\infty}^{\infty} |f^{(l)}(y+\varepsilon) - f^{(l)}(y)|^{2}dy\right)^{1/2} \leqslant N(f)|\varepsilon|^{\beta-l} \text{ ; } \forall\varepsilon \in \mathbb{R}.$$

Proposition 2.5. *Under Assumptions 2.1 and 2.4, there exists a constant $c > 0$, not depending on n and $h_1, \dots, h_n$, such that*

$$\int_{-\infty}^{\infty} \mathbb{E}(|\widehat{f}_{n,\mathbf{h}_n}(x) - f(x)|^{2})dx \leqslant \frac{c}{n^{2}}\left(\left|\sum_{k=1}^{n}\frac{h_k^{\beta}}{(l-1)!}\right|^{2} + \sum_{k=1}^{n}\frac{1}{h_k}\right).$$

Remark. Assumptions 2.1, 2.2 and 2.4 are standard for density estimation, see Tsybakov (2009). Moreover, if we set $h_k = h$, we recover the results stated in Section 1.2.1 for Proposition 2.3 and in Theorem 1.3 for Proposition 2.5 in Tsybakov (2009) (that is a squared bias term of order $h^{2\beta}$ and a variance term of order

$1/(nh)$).

The estimator is consistent if the risk tends to zero when n grows to infinity, that is if

$$\mathbb{B}_n := \frac{1}{n^2}\left|\sum_{k=1}^{n} h_k^{\beta}\right|^2 \quad \text{and} \quad \mathbb{V}_n := \frac{1}{n^2}\sum_{k=1}^{n}\frac{1}{h_k} \tag{2}$$

satisfy

$$\mathbb{B}_n \xrightarrow[n\to\infty]{} 0 \quad \text{and} \quad \mathbb{V}_n \xrightarrow[n\to\infty]{} 0.$$

Let us consider

$$h_k = k^{-\gamma} \ ;\ k \in [\![1,n]\!]$$

with $\gamma \in]0,1[$ (otherwise $\mathbb{B}_n$ or $\mathbb{V}_n$ cannot tend to zero). Then

$$\mathbb{B}_n = O\left(\frac{1}{n^2}\right) \text{ if } \gamma\beta > 1 \quad \text{and} \quad \mathbb{B}_n = O\left(\frac{1}{n^{2\gamma\beta}}\right) \text{ if } \gamma\beta < 1 \tag{3}$$

with the intermediate case

$$\mathbb{B}_n = O\left(\frac{\log(n)}{n^2}\right) \text{ if } \gamma\beta = 1.$$

Indeed, if $\gamma\beta < 1$, then

$$\mathbb{B}_n = \left|\frac{1}{n}\sum_{k=1}^{n} k^{-\gamma\beta}\right|^2 = n^{-2\gamma\beta}\left|\frac{1}{n}\sum_{k=1}^{n}\left(\frac{k}{n}\right)^{-\gamma\beta}\right|^2 \sim \frac{n^{-2\gamma\beta}}{(1-\gamma\beta)^2}.$$

On the other hand,

$$\mathbb{V}_n = O(n^{\gamma-1}). \tag{4}$$

As a consequence, we have the following result:

Corollary 2.6. *Under Assumptions 2.1 and 2.4, choosing*

$$h_k = k^{-\gamma} \ ;\ k \in [\![1,n]\!],$$

with

$$\gamma = \frac{1}{2\beta+1}$$

yields the rate

$$\int_{-\infty}^{\infty} \mathbb{E}(|\widehat{f}_{n,\mathbf{h}_n}(x) - f(x)|^2)dx \leqslant cn^{-\frac{2\beta}{2\beta+1}},$$

where c is a positive constant which does not depend on n.

Clearly, this is the optimal rate in the minimax sense, see Goldenshluger and Lepski [4] and the references therein.

Proof. Consider

$$\varphi_n(\gamma) := n^{-2\gamma\beta} + n^{\gamma-1}.$$

Then,

$$\frac{\partial\varphi_n(\gamma)}{\partial\gamma} = \log(n)(-2\beta e^{-2\gamma\beta\log(n)} + e^{(\gamma-1)\log(n)}).$$

Moreover, $\partial_\gamma\varphi_n(\gamma) = 0$ if and only if,

$$\gamma = \frac{1}{2\beta+1} + \frac{\log(2\beta)}{\log(n)(1+2\beta)} \sim \frac{1}{2\beta+1}.$$

Therefore, $\gamma = 1/(2\beta+1)$ makes the upper bound on the risk minimal. $\square$

3. Goldenshluger-Lepski's method for Wolverton-Wagner's estimator

This section provides an extension of the well-known Goldenshluger-Lepski's bandwidth selection method for Parzen-Rosenblatt's estimator to Wolverton-Wagner's estimator.

Throughout this section, assume that

$$h_k = h_k(\gamma) \ ; \ \forall k \in [\![1,n]\!],$$

where $\gamma \in [0,1]$ and the maps $h_1(.),\dots,h_n(.)$ from $[0,1]$ into $]0,\infty[$ fulfill the following assumption.

Assumption 3.1. *For every $\gamma' \in [0,1]$,*

$$0 < h_n(\gamma') < \dots < h_1(\gamma').$$

Moreover, $h_n(.)$ is decreasing and one to one from $[0,1]$ into $]0,1]$.

For instance, one can take as above $h_k(\gamma') := k^{-\gamma'}$ for every $k \in [\![1,n]\!]$ and $\gamma' \in [0,1]$.

Consider

$$\mathbf{h}_n(\gamma) := (h_1(\gamma),\dots,h_n(\gamma))$$

and the set $\Gamma_n := \{\gamma_1,\dots,\gamma_{N(n)}\} \subset [0,1]$, where $N(n) \in [\![1,n]\!]$ and

$$0 < \gamma_1 < \dots < \gamma_{N(n)} \leqslant h_n^{-1}(1/n).$$

Consider also

$$\widehat{f}_{n,\gamma,\gamma'}(x) := \frac{1}{n}\sum_{k=1}^{n}(K_{h_k(\gamma')} * K_{h_k(\gamma)})(X_k - x),$$

where $\gamma' \in [0,1]$.

A way to extend the Goldenshluger-Lepski bandwidth selection method to Wolverton-Wagner's estimator is to solve the minimization problem

$$\min_{\gamma\in\Gamma_n}(A_n(\gamma) + V_n(\gamma)), \tag{5}$$

where

$$A_n(\gamma) := \sup_{\gamma'\in\Gamma_n}(\|\widehat{f}_{n,\mathbf{h}_n(\gamma')} - \widehat{f}_{n,\gamma,\gamma'}\|_2^2 - V_n(\gamma'))_+, \quad V_n(\gamma') := \frac{\upsilon}{n\,\mathfrak{h}_n(\gamma')}$$

with $\upsilon > 0$ not depending on n and

$$\frac{1}{\mathfrak{h}_n(\gamma')} := \frac{1}{n}\sum_{k=1}^{n}\frac{1}{h_k(\gamma')}.$$

In the sequel, the map $\mathfrak{h}_n(.)$ fulfills the following assumption.

Assumption 3.2. *For every $c > 0$ and $r \in \{1/2, 1\}$,*

$$\sup_{n\in\mathbb{N}^*}\sum_{\gamma'\in\Gamma_n}\exp(-c/\mathfrak{h}_n(\gamma')^r) < \infty.$$

Example. Consider

$$h_k(\gamma') = k^{-\gamma'} \ ; \ \forall k \in [\![1,n]\!], \forall \gamma' \in [0,1]$$

and

$$\Gamma_n = \left\{\left(\frac{i}{\log(n)}\right)^{1/2} \ ; \ i \in [\![1,[\log(n)]]\!]\right\}. \tag{6}$$

For every $\gamma' \in \Gamma_n$,

$$\frac{1}{\mathfrak{h}_n(\gamma')} = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{h_k(\gamma')} = n^{\gamma'-1}\sum_{k=1}^{n}\left(\frac{k}{n}\right)^{\gamma'}$$
$$\geqslant n^{\gamma'-2}\sum_{k=1}^{n}k \geqslant \frac{n^{\gamma'}}{2} \geqslant \frac{1}{2}\exp(\log(n)^{1/2}).$$

Then, for any $c > 0$ and $r \in \{1/2, 1\}$,

$$\sup_{n\in\mathbb{N}^*}\sum_{\gamma'\in\Gamma_n}\exp(-c/\mathfrak{h}_n(\gamma')^r) \leqslant \sup_{n\in\mathbb{N}^*}\log(n)\exp\left(-\frac{c}{2^r}\exp(r\log(n)^{1/2})\right) < \infty.$$

Proposition 3.3. *Under Assumptions 3.1 and 3.2, if f is bounded and $\widehat{\gamma}_n$ is a solution of the minimization problem (5), then there exists a constant $c > 0$, not depending on n, such that*

$$\mathbb{E}(\|\widehat{f}_{n,\mathbf{h}_n(\widehat{\gamma}_n)} - f\|_2^2) \leqslant c\left\{\inf_{\gamma\in\Gamma_n}\left(V_n(\gamma) + \frac{1}{n^2}\left|\sum_{k=1}^{n}\|f - K_{h_k(\gamma)} * f\|_2\right|^2\right) + \frac{1}{n}\right\}.$$

If in addition Assumptions 2.1 and 2.4 hold, then

$$\mathbb{E}(\|\widehat{f}_{n,\mathbf{h}_n(\widehat{\gamma}_n)} - f\|_2^2) \leqslant c\left\{\inf_{\gamma\in\Gamma_n}(\mathbb{B}_n(\gamma) + \mathbb{V}_n(\gamma)) + \frac{1}{n}\right\} \qquad (7)$$

where $\mathbb{B}_n(\gamma)$ and $\mathbb{V}_n(\gamma)$ are defined in (2), (3) and (4).

Remark. By Corollary 2.6, the infimum in bound (7) has the order of the optimal rate, and is reached automatically by the data driven estimator. This result is more precise than the heuristics associated with cross-validation.
We mentioned previously that the optimal theoretical choice for γ under Assumptions 2.1 and 2.4 is $\gamma = 1/(2\beta+1)$. Here, the selected γ should be at nearest of this value, e.g. if Γ_n is as in (6), distant from less than $1/\sqrt{\log(n)}$ of the good choice. We may therefore consider that $\widehat{\gamma}_n$ provides an estimate of $1/(2\beta+1)$ and thus an estimate of the regularity β of f (at least for huge values of n).

4. The Lacour-Massart-Rivoirard (LMR) estimator

4.1. **Estimator and main result.** The Goldenshluger-Lepski method has been acknowledged as being difficult to implement, due to the square grid in γ, γ' required to compute intermediate versions of the criterion and to the lack of intuition in the choice of the constant υ which should be calibrated from preliminary simulation experiments. This is the reason why Lacour *et al.* [8] investigated and proposed a simplified criterion relying on deviation inequalities for U-statistics due to Houdré and Reynaud-Bouret [6]. This inequality applies in our more complicated context and Lacour-Massart-Rivoirard's result can be extended here as follows.

Let us recall that $K_\varepsilon := (1/\varepsilon)K(\cdot/\varepsilon)$ for every $\varepsilon > 0$ and set

$$f_{n,\gamma}(x) := \mathbb{E}(\widehat{f}_{\mathbf{h}_n(\gamma)}(x)) = \frac{1}{n}\sum_{k=1}^{n}(K_{h_k(\gamma)} * f)(x).$$

Let $\gamma_{\max}$ be the maximal proposal in Γ_n and consider

$$\mathrm{Crit}(\gamma) := \|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})}\|_2^2 + \mathrm{pen}(\gamma)$$

with

$$\mathrm{pen}(\gamma) := \frac{2}{n^2}\sum_{k=1}^{n}\langle K_{h_k(\gamma_{\max})}, K_{h_k(\gamma)}\rangle_2.$$

Then, we define

$$\widetilde{\gamma}_n \in \arg\min_{\gamma \in \Gamma_n} \mathrm{Crit}(\gamma).$$

In the sequel, K, f and $\mathbf{h}_n$ fulfill the following assumption.

Assumption 4.1. *The kernel K is symmetric, $K(0) > 0$,*

$$\int_{-\infty}^{\infty} K(y)dy = 1, \qquad \frac{\|K\|_\infty \|K\|_1}{nh_n(\gamma_{\max})} \leqslant 1$$

and $\|f\|_\infty < \infty$.

Proposition 4.2. *Consider $\lambda \in [1,\infty[$ and $\varepsilon \in]0,1[$. Under Assumption 4.1, there exists three deterministic constants $c_1, c_2, c_3 > 0$, not depending on n, λ and γ, such that with probability larger than $1 - c_1|\Gamma_n|e^{-\lambda}$,*

$$\|\widehat{f}_{n,\mathbf{h}_n(\widetilde{\gamma}_n)} - f\|_2^2 \leqslant (1+\varepsilon)\min_{\gamma\in\Gamma_n} \|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - f\|_2^2 + \frac{c_2}{\varepsilon}\|f_{n,\gamma_{\max}} - f\|_2^2 + \frac{c_3}{\varepsilon}\left(\frac{\lambda^2}{n} + \frac{\lambda^3}{n^2 h_n(\gamma_{\max})}\right).$$

Remark. The term $\|f_{n,\gamma_{\max}} - f\|_2^2$ is negligible because it is a pure bias term for smallest bandwidth (e.g., under Assumption 2.4, it has order $n^{-2\beta\gamma_{\max}}$, see (3), and thus $o(1/n)$ if $\gamma_{\max}$ is near of 1 and $\beta > 1/2$). The terms following are of order $O(1/n)$ and are always negligible compared to nonparametric rates in our setting. Therefore, the bound given in Proposition 4.2 says that the MISE of the adaptive estimator has the order of the best estimator of the collection, up to a multplicative factor larger than 1. This is the method we implement in the next section: it is faster than GL method and with no constant to calibrate in the penalty.

4.2. **Simulation experiments.** We consider basic densities with different types and orders of regularity:

- $X \rightsquigarrow \mathcal{N}(0,1)$, density f_1,
- a mixed gaussian $X \rightsquigarrow 0.5\mathcal{N}(-2,1) + 0.5\mathcal{N}(2,1)$, density $f_{m,1}$,
- $X \rightsquigarrow \beta(3,3)$, density f_2,
- a mixed beta $X \rightsquigarrow 0.5(\beta(3,3)-1) + 0.5\beta(3,3)$, density $f_{m,2}$,
- $X \rightsquigarrow \gamma(5,5)/10$, density f_3,
- a mixed gamma $X \rightsquigarrow 0.4.\gamma(2,1/3) + 0.6\gamma(7,6)/10$, density $f_{m,3}$,
- $X \rightsquigarrow f_4$ with $f_4(x) = e^{-|x|}$, a Laplace density.

The densities f_1 and $f_{m,1}$ have infinite regularity, f_2 and $f_{m,2}$ should rather have regularity of order less than 2, f_3 and $f_{m,3}$ less than 4, and f_4 less than 1. This choice should allow to study the influence of the order of the kernel.

Denoting by $n_j(x)$ the density of a centered Gaussian random variable with variance equal to j, we consider the following kernels:

- a Gaussian kernel, $K_1(x) = e^{-x^2/2}/\sqrt{2\pi}$ which is of order 1,
- a Gaussian-type kernel of order 3, $K_3(x) = 2n_1(x) - n_2(x)$,
- a Gaussian-type kernel of order 5, $K_5(x) = 3n_1(x) - 3n_2(x) + n_3(x)$,
- a Gaussian-type kernel of order 7, $K_7(x) = 4n_1(x) - 6n_2(x) + 4n_3(x) - n_4(x)$.

With all these kernels, the penalty terms are computed analytically and without approximation. Indeed, for $n_{i,h}(x) = (1/h)n_i(x/h)$, it holds that

$$\langle n_{i,h_1}, n_{j,h_2}\rangle_2 = \int_{-\infty}^{\infty} n_{i,h_1}(x)n_{j,h_2}(x)dx = \frac{1}{\sqrt{2\pi}} \times \frac{1}{\sqrt{ih_1^2 + jh_2^2}}.$$

		LMR for WW				Original LMR				
	$n =$	K_1	K_3	K_5	K_7	K_1	K_3	K_5	K_7	`ks`
f_1	250	0.442	0.318	0.285	0.256	0.412	0.315	0.290	0.268	0.285
		(0.252)	(0.213)	(0.193)	(0.162)	(0.241)	(0.214)	(0.205)	(0.193)	(0.174)
	1000	0.144	0.091	0.080	0.075	0.133	0.088	0.079	0.076	0.101
		(0.079)	(0.065)	(0.061)	(0.059)	(0.076)	(0.064)	(0.062)	(0.061)	(0.059)
$f_{m,1}$	250	0.400	0.316	0.287	0.255	0.387	0.327	0.291	0.256	1.115
		(0.204)	(0.189)	(0.176)	(0.162)	(0.208)	(0.202)	(0.179)	(0.170)	(0.150)
	1000	0.141	0.101	0.090	0.084	0.135	0.101	0.094	0.091	0.585
		(0.0623)	(0.051)	(0.049)	(0.046)	(0.062)	(0.053)	(0.051)	(0.050)	(0.076)
f_2	250	3.586	2.141	1.840	1.709	1.865	1.343	1.221	1.178	1.272
		(1.403)	(1.230)	(1.155)	(1.116)	(1.108)	(0.930)	(0.884)	(0.885)	(0.789)
	1000	1.056	0.646	0.555	0.515	0.602	0.429	0.382	0.372	0.506
		(0.394)	(0.306)	(0.283)	(0.270)	(0.312)	(0.270)	(0.250)	(0.235)	(0.282)
$f_{m,2}$	250	3.071	2.040	1.778	1.654	1.825	1.362	1.217	1.157	8.912
		(0.851)	(0.743)	(0.706)	(0.681)	(0.655)	(0.584)	(0.605)	(0.565)	(0.909)
	1000	0.905	0.593	0.508	0.476	0.657	0.438	0.389	0.358	4.876
		(0.246)	(0.201)	(0.187)	(0.182)	(0.257)	(0.188)	(0.159)	(0.163)	(0.367)
f_3	250	0.449	0.358	0.340	0.326	0.419	0.356	0.343	0.327	0.298
		(0.263)	(0.236)	(0.221)	(0.198)	(0.259)	(0.241)	(0.224)	(0.201)	(0.202)
	1000	0.174	0.132	0.124	0.121	0.162	0.130	0.126	0.126	0.125
		(0.085)	(0.071)	(0.067)	(0.065)	(0.081)	(0.071)	(0.071)	(0.076)	(0.065)
$f_{m,3}$	250	1.257	1.129	1.106	1.103	1.140	1.117	1.138	1.162	4.089
		(0.597)	(0.555)	(0.537)	(0.532)	(0.564)	(0.562)	(0.568)	(0.576)	(0.355)
	1000	0.491	0.448	0.444	0.446	0.449	0.441	0.454	0.466	3.172
		(0.171)	(0.158)	(0.158)	(0.160)	(0.168)	(0.174)	(0.189)	(0.204)	(0.201)
f_4	250	0.683	0.642	0.642	0.649	0.663	0.680	0.706	0.708	0.519
		(0.353)	(0.318)	(0.301)	(0.294)	(0.347)	(0.343)	(0.339)	(0.322)	(0.260)
	1000	0.281	0.254	0.254	0.258	0.273	0.268	0.278	0.284	0.242
		(0.135)	(0.122)	(0.120)	(0.122)	(0.141)	(0.147)	(0.163)	(0.172)	(0.105)

TABLE 1. 100 × MISE with 100 × std in parenthesis, computed over 200 simulations.

We compute the variable bandwidth estimator as described in Section 4 and select $\widetilde{\gamma}_n$ in a collection of $M = 40$ equispaced values between 0 and 0.5 while the bandwidth associated with observation i is $h_i(\gamma) = i^{-\gamma}$. We also compute the original estimator of Lacour *et al.* [8] with bandwidth h which does not depend on the observation and is selected among $M = 40$ values in the set $\{k/M\ ;\ k = 1, \dots, M\}$.

For comparison, we give the performance of the Matlab density estimator obtained from `ksdensity` function (denoted by `ks` in Table 1), which entails a different bandwidth selection method and relies on a gaussian kernel.

We compute the integrated $\mathbb{L}^2$-risk associated with all the final estimators, evaluated at $P = 100$ equispaced points in the range $[a, b]$ of the observations, averaged over $K = 200$ repetitions:

$$\frac{1}{K}\sum_{j=1}^{K}\frac{b-a}{P}\sum_{\ell=1}^{P}(\widehat{f}^{(j)}_{\widehat{h}^{(j)}}(x_\ell) - f(x_\ell))^2, \quad x_\ell = a + \ell\frac{b-a}{P},$$

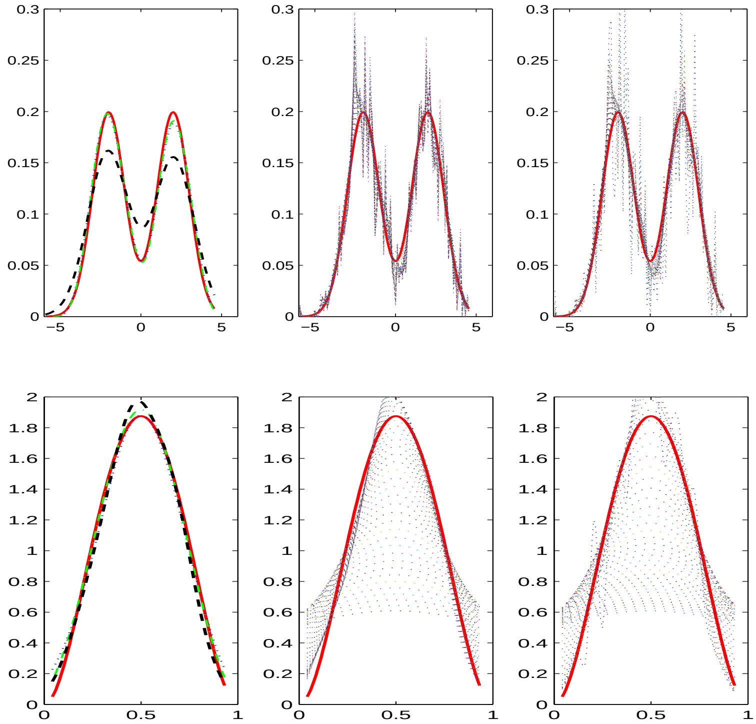

FIGURE 1. Left: The three estimators (dotted blue LMR-WW, green dash-dotted LMR, black dashed ks, the true in bold red. Middle: the 40 proposals for LMR-WW. Right: the 40 proposals for LMR. First line $n = 1000$, density $f_{1,m}$, second line $n = 250$, density f_2. In all cases, kernel K_7.

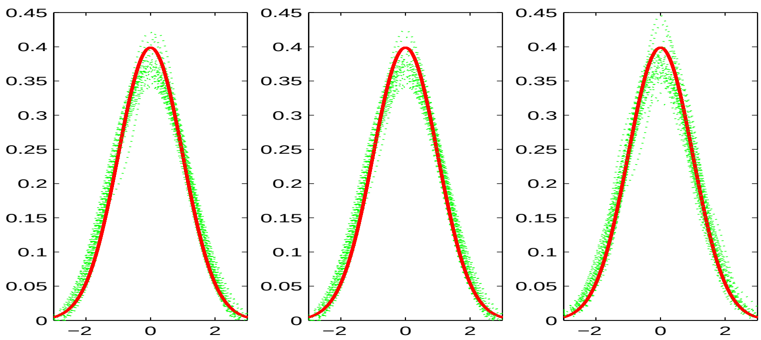

FIGURE 2. Beams of 30 estimators in dotted green of density f_1 for $n = 250$ and kernel K_7, and the true in bold red. Left: LMR-WW estimator. Middle: LMR estimator. Right: ks estimator.

where $\widehat{f}^{(j)}_{\widehat{h}^{(j)}}$ is the estimator computed for path j. Results are gathered in Table 1 and deserve some comments. As expected, when increasing n from 250 to 1000, the resulting MSEs decrease and seem to be more improved in LMR methods of both types than for ks estimator. Increasing the order of the kernel systematically improves the results, except for the lowest regularity density f_4, which is at best with K_3, but it is interesting to note that taking higher order kernel is always a good strategy: if a loss occurs, it is negligible while the improvement, when it happens, is in all cases significant. Estimator `ks` fails for all mixed densities $f_{m,1}$, $f_{m,2}$ and $f_{m,3}$ and provides rather bad results in these cases, for both sample sizes. For the other densities (f_1, f_2, f_3, f_4), the results obtained with kernel K_7 and LMR method are better than with `ks` for f_1 (Gaussian case), and of comparable order in all other cases. Now if we compare the LMR and WW-LMR results both with kernel K_7, we conclude that the WW-LMR method wins in 10 cases out of 14, but not significantly.

The first line of Figure 1 illustrates in the left picture the way Matlab estimator fails for mixed densities (here the mixed Gaussian $f_{m,1}$) by probably selecting a too large bandwidth, here $n = 1000$. The two LMR estimators are almost confounded. The middle and right pictures present the $M = 40$ estimators among which the LMR procedure makes the selection, for the same path: we observe that the collection of proposals are rather different. The second line of Figure 1 presents the same type of results for density f_2, and sample size $n = 250$. Figure 2 shows beams of 30 final estimators for sample size $n = 250$, for the three estimators LMR-WW with K_7, LMR with K_7 and `ks`, showing very similar behaviours.

A last remark corresponding to numerical results we do not report in detail is the following. For most densities, the value of γ selected by the LMR strategy decreases, and the value of h increases, when the order of the kernel increases. Exceptions are densities with lower regularity (the beta f_2, mixed beta $f_{2,m}$ and Laplace f_4 densities) for which the last value of selected h with K_7 is less than the one selected with K_5. This illustrates the fact that, asymptotically, if β is the regularity index of the density and ℓ the order of the kernel, the optimal choice is for h of order $n^{-1/(2\min(\beta,\ell)+1}$ and for γ, $1/(2\min(\beta,\ell)+1)$.

5. Concluding remarks

Our study illustrates that bandwidth selection is an important step for kernel functional estimation, and recent methods are really powerful whatever the type of density to recover.
Our simulations show also that, even if it implies non necessarily nonnegative kernels and thus density estimators, increasing the order of the kernel improves the estimation both in the theory and in practice. Also, we proved that variable bandwidth for WW-type estimators can reach excellent rates, again both in theory and in practice, provided that adaptive choice of this variable bandwidth is performed. The orders of practical MISEs show that this WW-strategy provides results of the same order as the more classical bandwidth methods.

However, one may wonder how to keep these ideas compatible with recursive procedures and online updating of the kernel estimator. We believe that the adaptive bandwidth, whatever its type, can be selected on a preliminary sample and then, "freezed" to this selected value and plugged in the estimator. We tested this strategy on simulations: for each sample, another independent sample is generated, and a second estimator is computed based on the new sample, with the value of h or γ selected for the first data set. Table 2 provides the MISEs obtained for the estimator with adaptive bandwidth with kernel K_7 for WW-estimator (column $\widehat{\gamma}$) or for NW-estimator (column $\widehat{h}$), for comparison with MISEs of the estimator computed on an independent sample with the values based on the previous selections (columns

"Fixed γ" and "Fixed h"). Surprisingly, the results have exactly the same orders (we expected a slight deterioration): this proves that the proposal of preliminary selection is really fine.

	$n = 250$				$n = 1000$			
	$\widehat{\gamma}$	Fixed γ	$\widehat{h}$	Fixed h	$\widehat{\gamma}$	Fixed γ	$\widehat{h}$	Fixed h
f_1	0.262	0.245	0.263	0.245	0.077	0.074	0.075	0.072
	(0.182)	(0.157)	(0.187)	(0.156)	(0.057)	(0.049)	(0.056)	(0.050)
$f_{1,m}$	0.257	0.248	0.260	0.247	0.085	0.076	0.091	0.084
	(0.144)	(0.134)	(0.149)	(0.136)	(0.049)	(0.038)	(0.053)	(0.046)
f_2	1.741	1.642	1.210	1.110	0.512	0.485	0.367	0.338
	(1.120)	(0.883)	(0.877)	(0.740)	(0.270)	(0.271)	(0.247)	(0.207)
$f_{2,m}$	1.749	1.603	1.246	1.085	0.480	0.493	0.364	0.367
	(0.720)	(0.683)	(0.623)	(0.567)	(0.190)	(0.208)	(0.166)	(0.188)
f_3	0.359	0.360	0.362	0.364	0.103	0.105	0.108	0.107
	(0.188)	(0.191)	(0.184)	(0.193)	(0.055)	(0.056)	(0.058)	(0.058)
$f_{3,m}$	1.035	1.041	1.092	1.092	0.435	0.426	0.457	0.448
	(0.447)	(0.358)	(0.479)	(0.447)	(0.153)	(0.131)	(0.210)	(0.182)
f_4	0.619	0.615	0.677	0.674	0.241	0.237	0.277	0.276
	(0.270)	(0.252)	(0.290)	(0.283)	(0.113)	(0.097)	(0.172)	(0.161)

TABLE 2. Comparison of $100 \times$ MISE (with $100\times$ std in parenthesis) for adaptive estimators and estimators based on a new sample with the previous $\widehat{\gamma}$ or $\widehat{h}$ plugged in.

6. PROOFS

6.1. **Proof of Proposition 2.3.** First, by the bias-variance decomposition,

$$\mathbb{E}(|\widehat{f}_{n,\mathbf{h}_n}(x) - f(x)|^2) = b_n(f,x)^2 + \mathrm{var}(\widehat{f}_{n,\mathbf{h}_n}(x)) \tag{8}$$

where

$$b_n(f,x) := \mathbb{E}(\widehat{f}_{n,\mathbf{h}_n}(x)) - f(x).$$

Let us find controls for $b_n(f,x)$ and $\mathrm{var}(\widehat{f}_{n,\mathbf{h}_n}(x))$.

On the one hand,

$$\begin{aligned}
\mathrm{var}(\widehat{f}_{n,\mathbf{h}_n}(x)) &= \frac{1}{n^2}\sum_{k=1}^{n}\frac{1}{h_k^2}\mathrm{var}\left(K\left(\frac{X_k - x}{h_k}\right)\right) \\
&\leqslant \frac{1}{n^2}\sum_{k=1}^{n}\frac{1}{h_k^2}\int_{-\infty}^{\infty} K\left(\frac{y-x}{h_k}\right)^2 f(y)dy \\
&= \frac{1}{n^2}\sum_{k=1}^{n}\frac{1}{h_k}\int_{-\infty}^{\infty} K(z)^2 f(h_k z + x)dz \\
&\leqslant \frac{c_1}{n^2}\sum_{k=1}^{n}\frac{1}{h_k},
\end{aligned}$$

where $c_1 := \|f\|_{\infty}\int_{-\infty}^{\infty}K(z)^2dz$. On the other hand,

$$\begin{aligned}
b_n(f,x) &= -f(x)+\frac{1}{n}\sum_{k=1}^{n}\frac{1}{h_k}\mathbb{E}\left(K\left(\frac{X_k-x}{h_k}\right)\right)\\
&= -f(x)+\frac{1}{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty}K(z)f(h_kz+x)dz\\
&= \frac{1}{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty}K(z)(f(h_kz+x)-f(x))dz.
\end{aligned}$$

For every $k\in[\![1,n]\!]$ and $z\in\mathbb{R}$, by Taylor-Lagrange's formula there exists $\tau\in[0,1]$ such that

$$f(h_kz+x)-f(x)=\sum_{i=1}^{l-1}\frac{(h_kz)^i}{i!}f^{(i)}(x)+\frac{(h_kz)^l}{l!}f^{(l)}(\tau h_kz+x).$$

Then, by Assumption 2.1,

$$\begin{aligned}
b_n(f,x) &= \frac{1}{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty}K(z)(f(h_kz+x)-f(x))dz\\
&= \frac{1}{n}\sum_{k=1}^{n}\left(\sum_{i=1}^{l-1}\frac{h_k^i}{i!}f^{(i)}(x)\int_{-\infty}^{\infty}z^iK(z)dz+\frac{h_k^l}{l!}\int_{-\infty}^{\infty}z^lK(z)f^{(l)}(\tau h_kz+x)dz\right)\\
&= \frac{1}{n}\sum_{k=1}^{n}\frac{h_k^l}{l!}\int_{-\infty}^{\infty}z^lK(z)f^{(l)}(\tau h_kz+x)dz\\
&= \frac{1}{n}\sum_{k=1}^{n}\frac{h_k^l}{l!}\int_{-\infty}^{\infty}z^lK(z)(f^{(l)}(\tau h_kz+x)-f^{(l)}(x))dz.
\end{aligned}$$

Therefore, by Assumption 2.2,

$$\begin{aligned}
|b_n(f,x)| &\leqslant \frac{1}{n}\sum_{k=1}^{n}\frac{h_k^l}{l!}\int_{-\infty}^{\infty}z^l\cdot|K(z)|\cdot|f^{(l)}(\tau h_kz+x)-f^{(l)}(x)|dz\\
&\leqslant \frac{c_2}{n}\sum_{k=1}^{n}\frac{h_k^{\beta}}{l!}
\end{aligned}$$

where $c_2 := \|f^{(l)}\|_{\beta-l}\int_{-\infty}^{\infty}z^{\beta}|K(z)|dz$. In conclusion, by Equation (8), setting $c := c_1\vee c_2^2$, we get

$$\mathbb{E}(|\widehat{f}_{n,\mathbf{h}_n}(x)-f(x)|^2)\leqslant\frac{c}{n^2}\left(\left|\sum_{k=1}^{n}\frac{h_k^{\beta}}{l!}\right|^2+\sum_{k=1}^{n}\frac{1}{h_k}\right).\quad\square$$

6.2. **Proof of Proposition 2.5.** In order to prove Proposition 2.5, the two following generalizations of Minkowski's inequality are required.

Lemma 6.1. *For any Borel function $\varphi:\mathbb{R}^2\to\mathbb{R}$, if $y\mapsto\varphi(y,z)$ is integrable and*

$$y\longmapsto\int_{-\infty}^{\infty}\varphi(y,z)dz$$

is a Borel function, then

$$(1)\quad \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\varphi(y,z)dy\right)^2dz\leqslant\left(\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\varphi(y,z)^2dz\right)^{1/2}dy\right)^2.$$

$$(2)\quad \int_{-\infty}^{\infty}\left(\sum_{k=1}^{n}\varphi(k,z)\right)^2 dz \leqslant \left(\sum_{k=1}^{n}\left(\int_{-\infty}^{\infty}\varphi(k,z)^2dz\right)^{1/2}\right)^2.$$

Proof. Result (1) is proved in Tsybakov [13], see Lemma A.1 for a proof.

The proof of Lemma 6.1.(2) is nearly the same, but we provide it for the sake of completeness. Assume that

$$M_{\varphi} := \sum_{k=1}^{n}\left(\int_{-\infty}^{\infty}\varphi(k,z)^2dz\right)^{1/2} < \infty$$

and consider

$$S(z) = \sum_{k=1}^{n}\varphi(k,z)\ ;\ \forall z\in\mathbb{R}.$$

For every $f\in\mathbb{L}^2(\mathbb{R},dz)$,

$$\begin{aligned}\left|\int_{-\infty}^{\infty}f(z)S(z)dz\right| &\leqslant \int_{-\infty}^{\infty}|f(z)|\sum_{k=1}^{n}|\varphi(k,z)|dz = \sum_{k=1}^{n}\int_{-\infty}^{\infty}|f(z)|\cdot|\varphi(k,z)|dz\\ &\leqslant \|f\|_2\sum_{k=1}^{n}\left(\int_{-\infty}^{\infty}\varphi(k,z)^2dz\right)^{1/2} = M_{\varphi}\|f\|_2.\end{aligned}$$

Then, the linear map

$$L : f\in\mathbb{L}^2(\mathbb{R},dz)\longmapsto L(f) := \int_{-\infty}^{\infty}f(z)S(z)dz$$

is continuous. Therefore, by the equality case of Cauchy-Schwarz's inequality,

$$\begin{aligned}\|S\|_2^2 = \int_{-\infty}^{\infty}\left(\sum_{k=1}^{n}\varphi(k,z)\right)^2 dz &= \left(\sup_{f\in\mathbb{L}^2(\mathbb{R},dz)\setminus\{0\}}\frac{L(f)}{\|f\|_2}\right)^2\\ &\leqslant M_{\varphi}^2 = \left(\sum_{k=1}^{n}\left(\int_{-\infty}^{\infty}\varphi(k,z)^2dz\right)^{1/2}\right)^2.\end{aligned}$$

□

It has been established in the proof of Proposition 2.3 that

$$(9)\qquad \text{var}(\widehat{f}_{n,\mathbf{h}_n}(x)) \leqslant \frac{1}{n^2}\sum_{k=1}^{n}\frac{1}{h_k^2}\int_{-\infty}^{\infty}K\left(\frac{y-x}{h_k}\right)^2 f(y)dy$$

and

$$(10)\qquad b_n(f,x) = \frac{1}{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty}K(z)(f(h_kz+x)-f(x))dz.$$

On the one hand, by Inequality (9),

$$\begin{aligned}\int_{-\infty}^{\infty}\text{var}(\widehat{f}_{n,\mathbf{h}_n}(x))dx &\leqslant \frac{1}{n^2}\sum_{k=1}^{n}\frac{1}{h_k^2}\int_{-\infty}^{\infty}f(y)\int_{-\infty}^{\infty}K\left(\frac{y-x}{h_k}\right)^2dxdy\\ &= \frac{1}{n^2}\sum_{k=1}^{n}\frac{1}{h_k}\left(\int_{-\infty}^{\infty}f(y)dy\right)\left(\int_{-\infty}^{\infty}K(z)^2dz\right)\\ &\leqslant \frac{c_1}{n^2}\sum_{k=1}^{n}\frac{1}{h_k}\end{aligned}$$

where $c_1 := \|K\|^2$. On the other hand, by Taylor's formula with integral remainder,

$$f(h_kz+x)-f(x)=\sum_{i=1}^{l-1}\frac{(h_kz)^i}{i!}f^{(i)}(x)+\frac{(h_kz)^l}{(l-1)!}\int_0^1(1-\tau)^{l-1}f^{(l)}(\tau h_kz+x)d\tau.$$

Then, by Assumption 2.1,

$$\begin{aligned}
b_n(f,x) &= \frac{1}{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty}K(z)(f(h_kz+x)-f(x))dz\\
&= \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{l-1}\frac{h_k^i}{i!}f^{(i)}(x)\int_{-\infty}^{\infty}z^iK(z)dz\\
&\quad+\frac{1}{n}\sum_{k=1}^{n}\frac{h_k^l}{(l-1)!}\int_{-\infty}^{\infty}z^lK(z)\int_0^1(1-\tau)^{l-1}f^{(l)}(\tau h_kz+x)d\tau dz\\
&= \frac{1}{n}\sum_{k=1}^{n}\frac{h_k^l}{(l-1)!}\int_{-\infty}^{\infty}z^lK(z)\int_0^1(1-\tau)^{l-1}(f^{(l)}(\tau h_kz+x)-f^{(l)}(x))d\tau dz.
\end{aligned}$$

By Lemma 6.1.(2),

$$\int_{-\infty}^{\infty}b_n(f,x)^2dx\leqslant\frac{1}{n^2}\left(\sum_{k=1}^{n}\frac{h_k^l}{(l-1)!}u_k^{1/2}\right)^2$$

where

$$u_k:=\int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty}z^lK(z)\int_0^1(1-\tau)^{l-1}(f^{(l)}(\tau h_kz+x)-f^{(l)}(x))d\tau dz\right|^2dx.$$

By Lemma 6.1.(1), for every $k\in[\![1,l]\!]$,

$$u_k\leqslant\left(\int_{-\infty}^{\infty}z^l|K(z)|\int_0^1(1-\tau)^{l-1}\left(\int_{-\infty}^{\infty}|f^{(l)}(\tau h_kz+x)-f^{(l)}(x)|^2dx\right)^{1/2}d\tau dz\right)^2.$$

Therefore, by Assumption 2.4,

$$\int_{-\infty}^{\infty}b_n(f,x)^2dx\leqslant\frac{c_2}{n^2}\left(\sum_{k=1}^{n}\frac{h_k^\beta}{(l-1)!}\right)^2$$

where

$$c_2:=N(f)^2\left(\int_{-\infty}^{\infty}z^\beta|K(z)|dz\right)^2\left(\int_0^1(1-\tau)^{l-1}\tau^{\beta-l}d\tau\right)^2.$$

In conclusion, by Equation (8), setting $c:=c_1\vee c_2$, we get

$$\int_{-\infty}^{\infty}\mathbb{E}(|\widehat{f}_{n,\mathbf{h}_n}(x)-f(x)|^2)dx\leqslant\frac{c}{n^2}\left(\left|\sum_{k=1}^{n}\frac{h_k^\beta}{(l-1)!}\right|^2+\sum_{k=1}^{n}\frac{1}{h_k}\right).\quad\square$$

6.3. **Proof of Proposition 3.3.** There exists a universal constant $c_1>0$ such that

$$\begin{aligned}
\|\widehat{f}_{n,\mathbf{h}_n(\widehat{\gamma}_n)}-f\|_2^2\leqslant c_1(&\|\widehat{f}_{n,\mathbf{h}_n(\widehat{\gamma}_n)}-\widehat{f}_{n,\widehat{\gamma}_n,\gamma'}\|_2^2\\
&+\|\widehat{f}_{n,\mathbf{h}_n(\gamma')}-\widehat{f}_{n,\widehat{\gamma}_n,\gamma'}\|_2^2\\
&+\|\widehat{f}_{n,\mathbf{h}_n(\gamma')}-f\|_2^2).
\end{aligned}$$

By the definition of A_n,

$$\begin{aligned}
\|\widehat{f}_{n,\mathbf{h}_n(\widehat{\gamma}_n)}-\widehat{f}_{n,\widehat{\gamma}_n,\gamma'}\|_2^2 &= \|\widehat{f}_{n,\mathbf{h}_n(\widehat{\gamma}_n)}-\widehat{f}_{n,\gamma',\widehat{\gamma}_n}\|_2^2\\
&\leqslant A_n(\gamma')+V_n(\widehat{\gamma}_n)
\end{aligned}$$

and

$$\|\widehat{f}_{n,\mathbf{h}_n(\gamma')}-\widehat{f}_{n,\widehat{\gamma}_n,\gamma'}\|_2^2\leqslant A_n(\widehat{\gamma}_n)+V_n(\gamma').$$

Since

$$\widehat{\gamma}_n\in\arg\min_{\gamma\in\Gamma_n}(A_n(\gamma)+V_n(\gamma)),$$

there exists a universal constant $c_2>0$ such that for any $\gamma\in\Gamma_n$,

$$\begin{aligned}\mathbb{E}(\|\widehat{f}_{n,\mathbf{h}_n(\widehat{\gamma}_n)}-f\|_2^2) &\leqslant c_2(\mathbb{E}(\|\widehat{f}_{n,\mathbf{h}_n(\gamma)}-f\|_2^2)+\mathbb{E}(A_n(\gamma))+V_n(\gamma))\\ &\leqslant c_2(2\mathbb{E}(\|\widehat{f}_{n,\mathbf{h}_n(\gamma)}-f_{n,\gamma}\|_2^2)\\ &\quad+2\mathbb{E}(\|f_{n,\gamma}-f\|_2^2)+\mathbb{E}(A_n(\gamma))+V_n(\gamma)),\end{aligned}\tag{11}$$

where

$$f_{n,\gamma}:=\frac{1}{n}\sum_{k=1}^{n}K_{h_k(\gamma)}*f.$$

Now, let us find a suitable control for $\mathbb{E}(A_n(\gamma))$. For that, consider

$$f_{n,\gamma,\gamma'}:=\frac{1}{n}\sum_{k=1}^{n}K_{h_k(\gamma')}*K_{h_k(\gamma)}*f.$$

There exists a universal constant $c_3>0$ such that

$$\begin{aligned}\|\widehat{f}_{n,\mathbf{h}_n(\gamma')}-\widehat{f}_{n,\gamma,\gamma'}\|_2^2 &\leqslant c_3(\|\widehat{f}_{n,\mathbf{h}_n(\gamma')}-f_{n,\gamma'}\|_2^2\\ &\quad+\|f_{n,\gamma'}-f_{n,\gamma,\gamma'}\|_2^2\\ &\quad+\|\widehat{f}_{n,\gamma,\gamma'}-f_{n,\gamma,\gamma'}\|_2^2).\end{aligned}$$

Then,

$$\begin{aligned}A_n(\gamma)\leqslant c_3\Bigg(&\sup_{\gamma'\in\Gamma_n}\left(\|\widehat{f}_{n,\mathbf{h}_n(\gamma')}-f_{n,\gamma'}\|_2^2-\frac{V_n(\gamma')}{2c_3}\right)_+\\ &+\sup_{\gamma'\in\Gamma_n}\left(\|\widehat{f}_{n,\gamma,\gamma'}-f_{n,\gamma,\gamma'}\|_2^2-\frac{V_n(\gamma')}{2c_3}\right)_++\|f_{n,\gamma'}-f_{n,\gamma,\gamma'}\|_2^2\Bigg).\end{aligned}\tag{12}$$

Let us control each terms of the right-hand side of Inequality (12). On the one hand, by Lemma 6.1.(2),

$$\begin{aligned}\|f_{n,\gamma'}-f_{n,\gamma,\gamma'}\|_2^2 &= \left\|\frac{1}{n}\sum_{k=1}^{n}K_{h_k(\gamma')}*(f-K_{h_k(\gamma)}*f)\right\|_2^2\\ &\leqslant \frac{1}{n^2}\left|\sum_{k=1}^{n}\|K_{h_k(\gamma')}*(f-K_{h_k(\gamma)}*f)\|_2\right|^2\\ &\leqslant \frac{\|K\|_1^2}{n^2}\left|\sum_{k=1}^{n}\|f-K_{h_k(\gamma)}*f\|_2\right|^2.\end{aligned}$$

On the other hand, let $\mathcal{C}$ be a countable and dense subset of the unit sphere of $\mathbb{L}^2(\mathbb{R},dx)$. Then,

$$\mathbb{E}\left(\sup_{\gamma'\in\Gamma_n}\left(\|\widehat{f}_{n,\mathbf{h}_n(\gamma')}-f_{n,\gamma'}\|_2^2-\frac{V_n(\gamma')}{2c_3}\right)_+\right)\leqslant\sum_{\gamma'\in\Gamma_n}\mathbb{E}\left(\left(\sup_{\psi\in\mathcal{C}}\mathfrak{v}_{n,\gamma'}(\psi)^2-\frac{V_n(\gamma')}{2c_3}\right)_+\right)$$

where, for every $\psi\in\mathcal{C}$,

$$\begin{aligned}\mathfrak{v}_{n,\gamma'}(\psi) &:= \langle\psi,\widehat{f}_{n,\mathbf{h}_n(\gamma')}-f_{n,\gamma'}\rangle_2\\ &= \frac{1}{n}\sum_{k=1}^{n}(v_\psi(h_k(\gamma'),X_k)-\mathbb{E}(v_\psi(h_k(\gamma'),X_k)))\end{aligned}$$

and

$$v_\psi(h,y) := \int_{-\infty}^{\infty}\psi(x)K_h(y-x)dx\ ;\ \forall(h,y)\in]1/n,1[\times\mathbb{R}.$$

In order to apply Talagrand's inequality (see Klein and Rio [7]):

(1) For every $\psi\in\mathcal{C}$, $h\in]1/n,1[$ and $y\in\mathbb{R}$,

$$|v_\psi(h,y)| \leqslant \int_{-\infty}^{\infty}|\psi(x)|K_h(y-x)dx$$
$$\leqslant \|K_h(y-\cdot)\|_2 = \frac{1}{\sqrt{h}}\|K\|_2 \leqslant \|K\|_2\sqrt{n}.$$

Then,

$$\sup_{\psi\in\mathcal{C}}\|v_\psi\|_\infty \leqslant M_1(n) := \|K\|_2\sqrt{n}.$$

(2) For every $\psi\in\mathcal{C}$,

$$\mathfrak{v}_{n,\gamma'}(\psi)^2 = \langle\psi,\widehat{f}_{n,\mathbf{h}_n(\gamma')}-f_{n,\gamma'}\rangle_2^2$$
$$\leqslant \|\widehat{f}_{n,\mathbf{h}_n(\gamma')}-f_{n,\gamma'}\|_2^2 = \int_{-\infty}^{\infty}|\widehat{f}_{n,\mathbf{h}_n(\gamma')}(x)-\mathbb{E}(\widehat{f}_{n,\mathbf{h}_n(\gamma')}(x))|^2dx.$$

Then, as established in the proof of Proposition 2.5,

$$\mathbb{E}\left(\sup_{\psi\in\mathcal{C}}|\mathfrak{v}_{n,\gamma'}(\psi)|\right) \leqslant \left|\int_{-\infty}^{\infty}\mathrm{var}(\widehat{f}_{n,\mathbf{h}_n(\gamma')}(x))dx\right|^{1/2}$$
$$\leqslant \frac{\|K\|_2}{n}\left|\sum_{k=1}^{n}\frac{1}{h_k(\gamma')}\right|^{1/2} = M_2(n,\gamma') := \frac{\|K\|_2}{\upsilon}V_n(\gamma')^{1/2}.$$

(3) For every $\psi\in\mathcal{C}$ and $k\in[\![1,n]\!]$,

$$\mathrm{var}(v_\psi(h_k(\gamma'),X_k)) \leqslant \mathbb{E}\left(\left|\int_{-\infty}^{\infty}K_{h_k(\gamma')}(X_k-x)\psi(x)dx\right|^2\right)$$
$$= \int_{-\infty}^{\infty}(K_{h_k(\gamma')}*\psi)(y)^2f(y)dy \leqslant \|f\|_\infty\|K\|_1^2.$$

Then,

$$\sup_{\psi\in\mathcal{C}}\frac{1}{n}\sum_{k=1}^{n}\mathrm{var}(v_\psi(h_k(\gamma'),X_k)) \leqslant M_3 := \|f\|_\infty\|K\|_1^2.$$

By applying Talagrand's inequality to $(v_\psi)_{\psi\in\mathcal{C}}$ and to the independent random variables $(h_1(\gamma'),X_1),\dots,(h_n(\gamma'),X_n)$, there exist four constants $c_5,c_6,c_7,c_8>0$, depending only on f, K and υ, such that

$$\mathbb{E}\left(\left(\sup_{\psi\in\mathcal{C}}\mathfrak{v}_{n,\gamma'}(\psi)^2-4M_2(n,\gamma')\right)_+\right) \leqslant c_5\left(\frac{M_3}{n}\exp\left(-\frac{c_6}{M_3}nM_2(n,\gamma')^2\right)\right.$$
$$\left.+\frac{M_1(n)^2}{n^2}\exp\left(-c_6\frac{nM_2(n,\gamma')}{M_1(n)}\right)\right)$$
$$\leqslant \frac{c_7}{n}(\exp(-c_8/\mathfrak{h}_n(\gamma'))+\exp(-c_8/\mathfrak{h}_n(\gamma')^{1/2})).$$

Then, by Assumption 3.2, with $\upsilon = 8\|K\|_2c_3$, there exists a constant $c_9>0$, not depending on n, such that

$$\mathbb{E}\left(\sup_{\gamma'\in\Gamma_n}\left(\|\widehat{f}_{n,\mathbf{h}_n(\gamma')}-f_{n,\gamma'}\|_2^2-\frac{V_n(\gamma')}{2c_3}\right)_+\right) \leqslant \frac{c_9}{n}. \tag{13}$$

The same ideas give that there exists a constant $c_{10} > 0$, not depending on n, such that

$$\sup_{\gamma\in\Gamma_n}\mathbb{E}\left(\sup_{\gamma'\in\Gamma_n}\left(\|\widehat{f}_{n,\gamma,\gamma'} - f_{n,\gamma,\gamma'}\|_2^2 - \frac{V_n(\gamma')}{2c_3}\right)_+\right) \leqslant \frac{c_{10}}{n}. \tag{14}$$

Therefore, by Inequalities (11)-(14), there exists a constant $c_{11} > 0$, not depending on n, such that

$$\mathbb{E}(\|\widehat{f}_{n,\mathbf{h}_n(\widehat{\gamma}_n)} - f\|_2^2) \leqslant c_{11}\left\{\inf_{\gamma\in\Gamma_n}\left(V_n(\gamma) + \frac{1}{n^2}\left|\sum_{k=1}^n \|f - K_{h_k(\gamma)} * f\|_2\right|^2\right) + \frac{1}{n}\right\}.$$

Moreover,

$$V_n(\gamma) = \frac{\upsilon}{n\,\mathfrak{h}_n(\gamma)} = \upsilon\mathbb{V}_n(\gamma)$$

and, if Assumptions 2.1 and 2.4 hold, as established in the proof of Proposition 2.5, there exists a constant $c_{12} > 0$ which does not depend on n such that

$$\begin{aligned}\sum_{k=1}^n \|f - K_{h_k(\gamma)} * f\|_2 &= \sum_{k=1}^n\left(\int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty} K(z)(f(h_k(\gamma)z + x) - f(x))dz\right|^2 dx\right)^{1/2}\\ &\leqslant c_{12}\sum_{k=1}^n h_k(\gamma)^\beta = c_{12}n\mathbb{B}_n(\gamma)^{1/2}.\end{aligned}$$

6.4. **Proof of Proposition 4.2.** For this proof, we use the tools and follow the lines given in the proof of Theorem 2 in Lacour *et al.* [8].

Throughout this section, for every $h > 0$, we consider $f_h := f * K_h$, where $*$ is the convolution product and we recall that $K_h = 1/hK(\cdot/h)$. Note that for every $h > 0$ and $k \in [\![1, n]\!]$,

$$\mathbb{E}(K_h(X_k - x)) = \int_{-\infty}^{\infty} K_h(y - x)f(y)dy = f_h(x).$$

We also consider $\lambda \in [1, \infty[$, $\varepsilon, \theta \in]0, 1[$ and $\gamma \in \Gamma_n$.

In order to prove Proposition 4.2, let us first establish the three following lemmas providing suitable bounds for key quantities involved in the decomposition of

$$\|\widehat{f}_{n,\mathbf{h}_n(\widetilde{\gamma}_n)} - f\|_2^2.$$

Throughout this section, the positive constants κ_i ; $i \in \mathbb{N}^*$ are not depending on n, λ, θ, ε and γ.

6.4.1. *Steps of the proof.* The proof relies on three Lemmas, which are stated first. Lemma 6.2 follows from an exponential inequality for U-statistics, which is applied here in a non indentically distributed context.

Lemma 6.2. *Consider the U-statistic*

$$U_n(\gamma, \gamma_{\max}) := \sum_{k\neq l}\langle K_{h_k(\gamma)}(X_k - \cdot) - f_{h_k(\gamma)}, K_{h_l(\gamma_{\max})}(X_l - \cdot) - f_{h_l(\gamma_{\max})}\rangle_2.$$

There exists a universal constant $\mathfrak{c} > 0$ such that with probability larger than $1 - 5.54|\Gamma_n|e^{-\lambda}$,

$$\frac{|U_n(\gamma, \gamma_{\max})|}{n^2} \leqslant \frac{\theta\|K\|_2^2}{n\mathfrak{h}_n(\gamma)} + \frac{\mathfrak{c}}{\theta}\left(\frac{\|K\|_1^2\|f\|_\infty}{n}\lambda^2 + \frac{\|K\|_1\|K\|_\infty}{n^2 h_n(\gamma_{\max})}\lambda^3\right).$$

Lemmas 6.3 and 6.4 rely on Bernstein's inequality for non identically distributed variables.

Lemma 6.3. *There exists a deterministic constant $c > 0$, not depending on n, λ, θ and γ, such that for every $\gamma' \in \Gamma_n$, with probability larger than $1 - 2e^{-\lambda}$,*

$$V_n(\gamma, \gamma') := \langle \widehat{f}_{n,\mathbf{h}_n(\gamma)} - f_{n,\gamma}, f_{n,\gamma'} - f\rangle$$

satisfies

$$|V_n(\gamma, \gamma')| \leqslant \theta \|f_{n,\gamma'} - f\|_2^2 + \frac{c\lambda}{\theta n}.$$

Lemma 6.4. *Under Assumption 4.1, there exists two deterministic constants $c_1, c_2 > 0$, not depending on n, λ, ε and γ, such that with probability larger than $1 - c_1|\Gamma_n|e^{-\lambda}$,*

$$\|f_{n,\gamma} - f\|_2^2 + \frac{\|K\|_2^2}{n\,\mathfrak{h}_n(\gamma)} \leqslant (1+\varepsilon)\|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - f\|_2^2 + c_2\frac{(1+\varepsilon)^2}{\varepsilon}\left(\frac{\lambda^2}{n} + \frac{\lambda^3}{n^2 h_n(\gamma_{\max})}\right).$$

The proof of Proposition 4.2 is dissected in three steps.

Step 1. In this step, a suitable decomposition of

$$\|\widehat{f}_{n,\mathbf{h}_n(\widetilde{\gamma}_n)} - f\|_2^2$$

is provided. On the one hand,

$$\begin{aligned}
\|\widehat{f}_{n,\mathbf{h}_n(\widetilde{\gamma}_n)} - f\|_2^2 + \mathrm{pen}(\widetilde{\gamma}_n) = & \|\widehat{f}_{n,\mathbf{h}_n(\widetilde{\gamma}_n)} - \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})}\|_2^2 + \mathrm{pen}(\widetilde{\gamma}_n)\\
& + \|\widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - f\|_2^2\\
& - 2\langle \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - f, \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - \widehat{f}_{n,\mathbf{h}_n(\widetilde{\gamma}_n)}\rangle_2.
\end{aligned}$$

Since $\widetilde{\gamma}_n \in \arg\min_{\gamma\in\Gamma_n} \mathrm{Crit}(\gamma)$,

$$\begin{aligned}
\|\widehat{f}_{n,\mathbf{h}_n(\widetilde{\gamma}_n)} - f\|_2^2 \leqslant & \|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})}\|_2^2 + \mathrm{pen}(\gamma)\\
& - \mathrm{pen}(\widetilde{\gamma}_n) + \|\widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - f\|_2^2\\
& - 2\langle \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - f, \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - \widehat{f}_{n,\mathbf{h}_n(\widetilde{\gamma}_n)}\rangle_2\\
= & \|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - f\|_2^2\\
& - [\mathrm{pen}(\widetilde{\gamma}_n) - 2\|\widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - f\|_2^2\\
& + 2\langle \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - f, \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - \widehat{f}_{n,\mathbf{h}_n(\widetilde{\gamma}_n)}\rangle_2]\\
& + \mathrm{pen}(\gamma) - 2\langle \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - f, \widehat{f}_{n,\mathbf{h}_n(\gamma)} - f\rangle_2\\
= & \|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - f\|_2^2 + \mathrm{pen}(\gamma) - 2\psi_n(\gamma) - (\mathrm{pen}(\widetilde{\gamma}_n) - 2\psi_n(\widetilde{\gamma}_n))
\end{aligned} \tag{15}$$

with

$$\psi_n := \langle \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - f, \widehat{f}_{n,\mathbf{h}_n(.)} - f\rangle_2.$$

On the other hand,

$$\begin{aligned}
\psi_n(\gamma) = & \langle \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - f_{n,\gamma_{\max}}, \widehat{f}_{n,\mathbf{h}_n(\gamma)} - f_{n,\gamma}\rangle_2 + \langle \widehat{f}_{n,\mathbf{h}_n(\gamma_{\max})} - f_{n,\gamma_{\max}}, f_{n,\gamma} - f\rangle_2\\
& + \langle f_{n,\gamma_{\max}} - f, \widehat{f}_{n,\mathbf{h}_n(\gamma)} - f_{n,\gamma}\rangle_2 + \langle f_{n,\gamma_{\max}} - f, f_{n,\gamma} - f\rangle_2\\
= & \psi_{1,n}(\gamma) + \psi_{2,n}(\gamma) + \psi_{3,n}(\gamma),
\end{aligned}$$

where

$$\psi_{1,n}(\gamma) := \frac{1}{n^2}\sum_{k=1}^{n}\langle K_{h_k(\gamma)}, K_{h_k(\gamma_{\max})}\rangle_2 + \frac{U_n(\gamma,\gamma_{\max})}{n^2},$$

$$\psi_{2,n}(\gamma) := \frac{1}{n^2}\left(-\sum_{k=1}^{n}\langle K_{h_k(\gamma)}(X_k-\cdot), f_{h_k(\gamma_{\max})}\rangle_2 - \sum_{k=1}^{n}\langle K_{h_k(\gamma_{\max})}(X_k-\cdot), f_{h_k(\gamma)}\rangle_2 + \sum_{k=1}^{n}\langle f_{h_k(\gamma)}, f_{h_k(\gamma_{\max})}\rangle_2\right) \text{ and}$$

$$\psi_{3,n}(\gamma) := V_n(\gamma,\gamma_{\max}) + V_n(\gamma_{\max},\gamma) + \langle f_{n,\gamma}-f, f_{n,\gamma_{\max}}-f\rangle_2$$

Step 2. Some bounds for $\psi_{n,1}(\gamma)$, $\psi_{n,2}(\gamma)$ and $\psi_{n,3}(\gamma)$ are provided in this step.

(1) Consider

$$\widetilde{\psi}_{1,n}(\gamma) := \psi_{1,n}(\gamma) - \frac{1}{n^2}\sum_{k=1}^{n}\langle K_{h_k(\gamma)}, K_{h_k(\gamma_{\max})}\rangle_2.$$

By Lemma 6.2, with probability larger than $1-5.54|\Gamma_n|e^{-\lambda}$,

$$|\widetilde{\psi}_{1,n}(\gamma)| = \frac{|U_n(\gamma,\gamma_{\max})|}{n^2} \leqslant \frac{\theta\|K\|_2^2}{n\,\mathfrak{h}_n(\gamma)} + \frac{\mathfrak{c}}{\theta}\left(\frac{\|K\|_1^2\|f\|_\infty}{n}\lambda^2 + \frac{\|K\|_\infty\|K\|_1}{n^2h_n(\gamma_{\max})}\lambda^3\right).$$

(2) On the one hand, for any $\gamma'\in\Gamma_n$,

$$\frac{1}{n}\left|\sum_{k=1}^{n}\langle K_{h_k(\gamma)}(X_k-\cdot), f_{h_k(\gamma')}\rangle_2\right| \leqslant \max_{k\in[\![1,n]\!]}\int_{-\infty}^{\infty}|K_{h_k(\gamma)}(X_k-x)f_{h_k(\gamma')}(x)|dx \leqslant \|K\|_1\max_{k\in[\![1,n]\!]}\|K_{h_k(\gamma')}*f\|_\infty \leqslant \|K\|_1^2\|f\|_\infty.$$

On the other hand,

$$\frac{1}{n}\left|\sum_{k=1}^{n}\langle f_{h_k(\gamma)}, f_{h_k(\gamma')}\rangle_2\right| \leqslant \max_{k\in[\![1,n]\!]}\int_{-\infty}^{\infty}|f_{h_k(\gamma)}(x)f_{h_k(\gamma')}(x)|dx \leqslant \max_{k\in[\![1,n]\!]}\|K_{h_k(\gamma)}*f\|_1\|K_{h_k(\gamma')}*f\|_\infty \leqslant \|K\|_1^2\|f\|_\infty.$$

Therefore,

$$\|\psi_{2,n}\|_\infty \leqslant \frac{3\|K\|_1^2\|f\|_\infty}{n}.$$

(3) By applying Lemma 6.3 to $V_n(\gamma,\gamma_{\max})$ and $V_n(\gamma_{\max},\gamma)$, with probability larger than $1-2e^{-\lambda}$,

$$|\psi_{n,3}(\gamma)| \leqslant \frac{\theta}{2}(\|f_{n,\gamma}-f\|_2^2 + \|f_{n,\gamma_{\max}}-f\|_2^2) + \frac{\kappa_1\lambda}{\theta n} + \frac{\theta}{2}\|f_{n,\gamma}-f\|_2^2 + \frac{1}{2\theta}\|f_{n,\gamma_{\max}}-f\|_2^2 \leqslant \theta\|f_{n,\gamma}-f\|_2^2 + \left(\frac{\theta}{2}+\frac{1}{2\theta}\right)\|f_{n,\gamma_{\max}}-f\|_2^2 + \frac{\kappa_1\lambda}{\theta n}.$$

Step 3. Consider

$$\widetilde{\psi}_n(\gamma) := \psi_n(\gamma) - \frac{1}{n^2}\sum_{k=1}^{n}\langle K_{h_k(\gamma)}, K_{h_k(\gamma_{\max})}\rangle_2.$$

By Step 2 and Lemma 6.4, with probability larger than $1 - \kappa_2|\Gamma_n|e^{-\lambda}$,

$$\begin{aligned}
|\widetilde{\psi}_n(\gamma)| &\leqslant \theta\|f_{n,\gamma} - f\|_2^2 + \frac{\theta\|K\|_2^2}{n\,\mathfrak{h}_n(\gamma)}\\
&\quad + \left(\frac{\theta}{2} + \frac{1}{2\theta}\right)\|f_{n,\gamma_{\max}} - f\|_2^2 + \frac{\kappa_3}{\theta}\left(\frac{\lambda^2}{n} + \frac{\lambda^3}{n^2h_n(\gamma_{\max})}\right)\\
&\leqslant 2\theta\|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - f\|_2^2\\
&\quad + \left(\frac{\theta}{2} + \frac{1}{2\theta}\right)\|f_{n,\gamma_{\max}} - f\|_2^2 + \frac{\kappa_4}{\theta}\left(\frac{\lambda^2}{n} + \frac{\lambda^3}{n^2h_n(\gamma_{\max})}\right).
\end{aligned}$$

Therefore, by Inequality (15), with probability larger than $1 - \kappa_5|\Gamma_n|e^{-\lambda}$,

$$\begin{aligned}
\|\widehat{f}_{n,\mathbf{h}_n(\widetilde{\gamma}_n)} - f\|_2^2 &\leqslant (1+\varepsilon)\|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - f\|_2^2 + \frac{\kappa_6}{\varepsilon}\|f_{n,\gamma_{\max}} - f\|_2^2\\
&\quad +\text{pen}(\gamma) - \frac{2}{n^2}\sum_{k=1}^{n}\langle K_{h_k(\gamma)}, K_{h_k(\gamma_{\max})}\rangle_2\\
&\quad - \left(\text{pen}(\widetilde{\gamma}_n) - \frac{2}{n^2}\sum_{k=1}^{n}\langle K_{h_k(\widetilde{\gamma}_n)}, K_{h_k(\gamma_{\max})}\rangle_2\right)\\
&\quad +\frac{\kappa_7}{\varepsilon}\left(\frac{\lambda^2}{n} + \frac{\lambda^3}{n^2h_n(\gamma_{\max})}\right)\\
&= (1+\varepsilon)\|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - f\|_2^2 + \frac{\kappa_6}{\varepsilon}\|f_{n,\gamma_{\max}} - f\|_2^2\\
&\quad +\frac{\kappa_7}{\varepsilon}\left(\frac{\lambda^2}{n} + \frac{\lambda^3}{n^2h_n(\gamma_{\max})}\right).
\end{aligned}$$

This concludes the proof. □

6.4.2. *Proof of Lemma 6.2.* Consider

$$\Delta_n := \{(k,l) \in \mathbb{N}^2 : 2 \leqslant k \leqslant n \text{ and } 1 \leqslant l \leqslant k-1\}.$$

The U-statistic satisfies

$$U_n(\gamma,\gamma_{\max}) = \sum_{k=2}^{n}\sum_{l<k}(G_{\gamma,\gamma_{\max}}^{k,l}(X_k,X_l) + G_{\gamma_{\max},\gamma}^{k,l}(X_k,X_l)),$$

where

$$G_{a,b}^{k,l}(\alpha,\beta) := \langle K_{h_k(a)}(\alpha - \cdot) - f_{h_k(a)}, K_{h_l(b)}(\beta - \cdot) - f_{h_l(b)}\rangle_2$$

for every $(k,l) \in \Delta_n$, $a,b \in \{\gamma,\gamma_{\max}\}$ and $(\alpha,\beta) \in \mathbb{R}^2$.

By Houdré and Reynaud-Bourret [6], Theorem 3.4, there exists a universal constant $\mathfrak{c} > 0$ such that

$$\mathbb{P}(|U_n(\gamma,\gamma_{\max})| \geqslant \mathfrak{c}(C\sqrt{\lambda} + D\lambda + B\lambda^{3/2} + A\lambda^2)) \leqslant 5.54e^{-\lambda} \tag{16}$$

where the constants A, B, C and D will be defined and controlled in the sequel.

- **The constant** A**.** Consider

$$A := \max_{(k,l)\in\Delta_n}\sup_{(\alpha,\beta)\in\mathbb{R}^2} A_{k,l}(\alpha,\beta)$$

with

$$A_{k,l}(\alpha,\beta) := |G_{\gamma,\gamma_{\max}}^{k,l}(\alpha,\beta) + G_{\gamma_{\max},\gamma}^{k,l}(\alpha,\beta)|\ ;\ \forall(k,l) \in \Delta_n,\ \forall(\alpha,\beta) \in \mathbb{R}^2.$$

For any $(k,l)\in\Delta_n$ and $(\alpha,\beta)\in\mathbb{R}^2$,

$$\begin{aligned}
A_{k,l}(\alpha,\beta) &\leqslant |\langle K_{h_k(\gamma)}(\alpha-\cdot)-f_{h_k(\gamma)},K_{h_l(\gamma_{\max})}(\beta-\cdot)-f_{h_l(\gamma_{\max})}\rangle_2|\\
&\quad +|\langle K_{h_k(\gamma_{\max})}(\alpha-\cdot)-f_{h_k(\gamma_{\max})},K_{h_l(\gamma)}(\beta-\cdot)-f_{h_l(\gamma)}\rangle_2|\\
&\leqslant 2(\|K_{h_k(\gamma_{\max})}\|_\infty+\|f_{h_k(\gamma_{\max})}\|_\infty)(\|K\|_1+\|f_{h_l(\gamma)}\|_1)\\
&\leqslant 8\frac{\|K\|_1\|K\|_\infty}{h_n(\gamma_{\max})}.
\end{aligned}$$

Therefore,

$$\frac{A\lambda^2}{n^2}\leqslant 8\frac{\|K\|_1\|K\|_\infty}{n^2h_n(\gamma_{\max})}\lambda^2.$$

- **The constant B.** Consider

$$B^2:=\max\left\{\sup_{\alpha,l}\sum_{k=1}^{l-1}\mathbb{E}(|G^{k,l}_{\gamma,\gamma_{\max}}(\alpha,X_l)|^2)\ ;\ \sup_{\alpha,l}\sum_{k=l+1}^{n}\mathbb{E}(|G^{k,l}_{\gamma_{\max},\gamma}(\alpha,X_l)|^2)\right\}.$$

For any $(k,l)\in\Delta_n$, $a,b\in\{\gamma,\gamma_{\max}\}$ and $(\alpha,\beta)\in\mathbb{R}^2$,

$$\begin{aligned}
\mathbb{E}(G^{k,l}_{a,b}(\alpha,X_l)^2) &= \mathbb{E}(\langle K_{h_k(a)}(\alpha-\cdot)-f_{h_k(a)},K_{h_l(b)}(X_l-\cdot)-f_{h_l(b)}\rangle_2^2)\\
&\leqslant \|K_{h_k(a)}(\alpha-\cdot)-f_{h_k(a)}\|_2^2\mathbb{E}(\|K_{h_l(b)}(X_l-\cdot)-f_{h_l(b)}\|_2^2)\\
&\leqslant 4\frac{\|K\|_2^2}{h_k(a)}\int_{-\infty}^{\infty}\mathbb{E}(|K_{h_l(b)}(X_l-y)-f_{h_l(b)}(y)|^2)dy\\
&\leqslant 4\frac{\|K\|_2^2}{h_n(a)}\|K_{h_l(b)}\|_2^2\leqslant 4\frac{\|K\|_2^4}{h_k(a)h_l(b)}\leqslant 4\frac{\|K\|_2^4}{h_k(a)h_n(b)}.
\end{aligned}$$

Then,

$$B^2\leqslant 4\frac{\|K\|_2^4}{h_n(\gamma_{\max})}\sum_{k=1}^{n}\frac{1}{h_k(\gamma)}.$$

Therefore,

$$\begin{aligned}
\frac{B\lambda^{3/2}}{n^2} &\leqslant 2\left(\frac{\theta}{3}\right)^{1/2}\|K\|_2\sqrt{\frac{1}{n^2}\sum_{k=1}^{n}\frac{1}{h_k(\gamma)}}\times\left(\frac{3}{\theta}\right)^{1/2}\frac{\|K\|_2}{(n^2h_n(\gamma_{\max}))^{1/2}}\lambda^{3/2}\\
&\leqslant \frac{\theta\|K\|_2^2}{3n\,\mathfrak{h}_n(\gamma)}+\frac{3}{\theta}\times\frac{\|K\|_2^2}{n^2h_n(\gamma_{\max})}\lambda^3.
\end{aligned}$$

- **The constant C.** Consider

$$C^2:=\sum_{(k,l)\in\Delta_n}\mathbb{E}((G^{k,l}_{\gamma,\gamma_{\max}}(X_k,X_l)+G^{k,l}_{\gamma_{\max},\gamma}(X_k,X_l))^2).$$

For any $(k,l)\in\Delta_n$ and $a,b\in\{\gamma,\gamma_{\max}\}$,

$$\begin{aligned}
\mathbb{E}(G^{k,l}_{a,b}(X_k,X_l)^2) &= \mathbb{E}(\langle K_{h_k(a)}(X_k-\cdot)-f_{h_k(a)},K_{h_l(b)}(X_l-\cdot)-f_{h_l(b)}\rangle_2^2)\\
&\leqslant \kappa_1(\mathbb{E}(\langle K_{h_k(a)}(X_k-\cdot),K_{h_l(b)}(X_l-\cdot)\rangle_2^2)\\
&\quad +\|f_{h_l(b)}\|_\infty^2\|K\|_1^2+\|f_{h_k(a)}\|_\infty^2\|K\|_1^2+\|f_{h_k(a)}\|_\infty^2\|f_{h_l(b)}\|_1^2)\\
&\leqslant \kappa_2\left(\mathbb{E}\left(\left|\int_{-\infty}^{\infty}K_{h_k(a)}(X_k-x)K_{h_l(b)}(X_l-x)dx\right|^2\right)+\|f\|_\infty^2\|K\|_1^4\right).
\end{aligned}$$

Moreover,

$$\mathbb{E}\left(\left|\int_{-\infty}^{\infty}K_{h_k(a)}(X_k-x)K_{h_l(b)}(X_l-x)dx\right|^2\right)\leqslant\frac{\|K\|_1^2\|K\|_2^2\|f\|_\infty}{h_k(a)}.$$

Then,

$$C \leqslant \kappa_3\sqrt{n}\|K\|_1\|f\|_{\infty}^{1/2}\left(\|K\|_2\sqrt{\sum_{k=1}^{n}\frac{1}{h_k(\gamma)}}+\|K\|_1\|f\|_{\infty}^{1/2}\right).$$

Therefore, since $\lambda\in[1,\infty[$,

$$\frac{C\lambda^{1/2}}{n^2}\leqslant\frac{\theta\|K\|_2^2}{3n\,\mathfrak{h}_n(\gamma)}+\kappa_4\frac{\|K\|_1^2\|f\|_{\infty}}{\theta n}\lambda.$$

- **The constant D.** Consider

$$D:=\sup_{(a,b)\in\mathcal{S}}\sum_{k=2}^{n}\sum_{l=1}^{k-1}\mathbb{E}((G_{\gamma,\gamma_{\max}}^{k,l}(X_k,X_l)+G_{\gamma_{\max},\gamma}^{k,l}(X_k,X_l))a_k(X_k)b_l(X_l)),$$

where

$$\mathcal{S}:=\left\{(a,b):\sum_{k=2}^{n}\mathbb{E}(a_k(X_k)^2)\leqslant 1\text{ and }\sum_{l=1}^{n-1}\mathbb{E}(b_l(X_l)^2)\leqslant 1\right\}.$$

For any $(a,b)\in\mathcal{S}$,

$$\sum_{k=2}^{n}\sum_{l=1}^{k-1}\mathbb{E}(G_{\gamma,\gamma_{\max}}^{k,l}(X_k,X_l)a_k(X_k)b_l(X_l))\leqslant D_2(a,b)\sup_{x\in\mathbb{R}}D_1(a,b,x)$$

with

$$\begin{aligned}
D_1(a,b,x) &:= \sum_{k=2}^{n}\mathbb{E}(|a_k(X_k)(K_{h_k(\gamma)}(X_k-x)-f_{h_k(\gamma)}(x))|)\\
&\leqslant \mathbb{E}\left[\left|\sum_{k=2}^{n}a_k(X_k)^2\right|^{1/2}\left|\sum_{k=1}^{n}(K_{h_k(\gamma)}(X_k-x)-f_{h_k(\gamma)}(x))^2\right|^{1/2}\right]\\
&\leqslant \mathbb{E}\left(\sum_{k=2}^{n}a_k(X_k)^2\right)^{1/2}\left|\sum_{k=1}^{n}\mathbb{E}(|K_{h_k(\gamma)}(X_k-x)-f_{h_k(\gamma)}(x)|^2)\right|^{1/2}\\
&\leqslant \left|\sum_{k=1}^{n}\mathbb{E}(K_{h_k(\gamma)}(X_k-x)^2)\right|^{1/2}\leqslant\left|\sum_{k=1}^{n}\frac{\|K\|_2^2\|f\|_{\infty}}{h_k(\gamma)}\right|^{1/2}
\end{aligned}$$

and

$$\begin{aligned}
D_2(a,b) &:= \sum_{l=1}^{n-1}\mathbb{E}\left(|b_l(X_l)|\int_{-\infty}^{\infty}|K_{h_l(\gamma_{\max})}(X_l-x)-f_{h_l(\gamma_{\max})}(x)|dx\right)\\
&\leqslant 2\|K\|_1\sum_{l=1}^{n-1}\mathbb{E}(|b_l(X_l)|)\leqslant 2\|K\|_1\sqrt{n}\left|\sum_{l=1}^{n-1}\mathbb{E}(b_l(X_l)^2)\right|^{1/2}\leqslant 2\sqrt{n}\|K\|_1.
\end{aligned}$$

Then,

$$D\leqslant 2\sqrt{n}\|f\|_{\infty}^{1/2}\|K\|_1\|K\|_2\left|\sum_{k=1}^{n}\frac{1}{h_k(\gamma)}\right|^{1/2}.$$

Therefore,

$$\frac{D\lambda}{n^2}\leqslant\frac{\theta\|K\|_2^2}{3n\,\mathfrak{h}_n(\gamma)}+\frac{12}{\theta}\times\frac{\|K\|_1^2\|f\|_{\infty}}{n}\lambda^2.$$

Plugging the bounds obtained for A,B,C,D in Inequality (16) gives the announced result and ends the proof. □

6.4.3. *Proof of Lemma 6.3.* For any $\gamma' \in \Gamma_n$,

$$V_n(\gamma,\gamma') = \frac{1}{n}\sum_{k=1}^{n}(g_{\gamma'}(h_k(\gamma),X_k) - \mathbb{E}(g_{\gamma'}(h_k(\gamma),X_k)))$$

where, for any $k \in [\![1,n]\!]$,

$$g_{\gamma'}(h_k(\gamma),X_k) := \langle K_{h_k(\gamma)}(X_k - \cdot), f_{n,\gamma'} - f\rangle_2.$$

Indeed,

$$\mathbb{E}(g_{\gamma'}(h_k(\gamma),X_k)) = \langle \mathbb{E}(K_{h_k(\gamma)}(X_k - \cdot)), f_{n,\gamma'} - f\rangle_2 = \langle f_{h_k(\gamma)}, f_{n,\gamma'} - f\rangle_2.$$

In order to apply Bernstein's inequality to $g_{\gamma'}(h_1(\gamma),X_1),\dots,g_{\gamma'}(h_n(\gamma),X_n)$, let us find suitable controls of

$$c_{\gamma'} := \frac{\|g_{\gamma'}\|_\infty}{3} \text{ and } \upsilon_n(\gamma,\gamma') := \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(g_{\gamma'}(h_k(\gamma),X_k)^2).$$

On the one hand,

$$\begin{aligned}
c_{\gamma'} &= \frac{1}{3}\sup_{h>0,x\in\mathbb{R}}|\langle K_h(x-\cdot), f_{n,\gamma'} - f\rangle_2| \\
&\leqslant \frac{1}{3}\sup_{h>0,x\in\mathbb{R}}\|K_h(x-\cdot)\|_1\|f_{n,\gamma'} - f\|_\infty \\
&\leqslant \frac{1}{3}\|K\|_1\max_{k\in[\![1,n]\!]}\|K_{h_k(\gamma')} * f - f\|_\infty \\
&\leqslant \frac{1}{3}\|K\|_1(1+\|K\|_1)\|f\|_\infty \leqslant \frac{2}{3}\|K\|_1^2\|f\|_\infty,
\end{aligned}$$

as $\|K\|_1 \geq 1$. On the other hand,

$$\begin{aligned}
\upsilon_n(\gamma,\gamma') &= \frac{1}{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K_{h_k(\gamma)}(y-x)(f_{n,\gamma'}(x) - f(x))dx\right)^2 f(y)dy \\
&\leqslant \|f\|_\infty\max_{k\in[\![1,n]\!]}\|K_{h_k(\gamma)} * (f_{n,\gamma'} - f)\|_2^2 \leqslant \|f\|_\infty\|K\|_1^2\|f_{n,\gamma'} - f\|_2^2.
\end{aligned}$$

Then, by Bernstein's inequality, with probability larger than $1 - 2e^{-\lambda}$,

$$\begin{aligned}
|V_n(\gamma,\gamma')| &\leqslant \sqrt{\frac{2\lambda}{n}\upsilon_n(\gamma,\gamma')} + \frac{c_{\gamma'}\lambda}{n} \\
&\leqslant \sqrt{\frac{2\lambda}{n}\|f\|_\infty\|K\|_1^2\|f_{n,\gamma'} - f\|_2^2} + \frac{\lambda\|K\|_1(1+\|K\|_1)\|f\|_\infty}{3n} \\
&\leqslant \theta\|f_{n,\gamma'} - f\|_2^2 + \frac{c\lambda}{\theta n},
\end{aligned}$$

with $c = 7/6\|f\|_\infty\|K\|_1^2$. This is the announced inequality. □

6.4.4. *Proof of Lemma 6.4.* First of all,

$$\|f_{n,\gamma} - f\|_2^2 = \|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - f\|_2^2 - \|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - f_{n,\gamma}\|_2^2 - 2V_n(\gamma,\gamma).$$

Then, by Lemma 6.3, with probability larger than $1 - 2e^{-\lambda}$,

$$(1-\theta)\|f_{n,\gamma} - f\|_2^2 + \frac{\|K\|_2^2}{n\,\mathfrak{h}_n(\gamma)} \leqslant \|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - f\|_2^2 + \Lambda_n(\gamma) + \frac{\kappa_1\lambda}{\theta n} \tag{17}$$

where

$$\begin{aligned}\Lambda_n(\gamma) &:= \left|\frac{\|K\|_2^2}{n\,\mathfrak{h}_n(\gamma)} - \|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - f_{n,\gamma}\|_2^2\right|\\ &= \left|\frac{U_n(\gamma,\gamma)}{n^2} + \frac{W_n(\gamma)}{n} - \frac{1}{n^2}\sum_{k=1}^{n}\|f_{h_k(\gamma)}\|_2^2\right|\end{aligned}$$

and

$$W_n(\gamma) := \frac{1}{n}\sum_{k=1}^{n}(Y_k(\gamma) - \mathbb{E}(Y_k(\gamma)))$$

with, for any $k \in [\![1,n]\!]$,

$$Y_k(\gamma) := \|K_{h_k(\gamma)}(X_k - \cdot) - f_{h_k(\gamma)}\|_2^2$$

and

$$\begin{aligned}\mathbb{E}(Y_k(\gamma)) &= \mathbb{E}(\|K_{h_k(\gamma)}(X_k - \cdot)\|_2^2) + \|f_{h_k(\gamma)}\|_2^2 - 2\langle\mathbb{E}(K_{h_k(\gamma)}(X_k - \cdot)), f_{h_k(\gamma)}\rangle_2\\ &= \frac{\|K\|_2^2}{h_k(\gamma)} - \|f_{h_k(\gamma)}\|_2^2.\end{aligned}$$

Since

$$|Y_k(\gamma)| \leqslant 4\|K_{h_k(\gamma)}\|_2^2 \leqslant M_n(\gamma) := 4\frac{\|K\|_2^2}{h_n(\gamma_{\max})}$$

and

$$\mathbb{E}(Y_k(\gamma)^2) \leqslant M_n(\gamma)\mathbb{E}(Y_k(\gamma)) \leqslant 16\frac{\|K\|_2^4}{h_n(\gamma_{\max})h_k(\gamma)},$$

by Bernstein's inequality, with probability larger than $1 - 2e^{-\lambda}$,

$$\begin{aligned}|W_n(\gamma)| &\leqslant 2\sqrt{\frac{16\|K\|_2^2\lambda}{\theta n h_n(\gamma_{\max})} \times \frac{\theta}{2}\frac{\|K\|_2^2}{\mathfrak{h}_n(\gamma)}} + \frac{4\|K\|_2^2\lambda}{3nh_n(\gamma_{\max})}\\ &\leqslant \frac{\theta}{2}\frac{\|K\|_2^2}{\mathfrak{h}_n(\gamma)} + \frac{\kappa_2\lambda}{\theta n h_n(\gamma_{\max})}.\end{aligned}$$

Moreover, by Jensen's inequality,

$$\begin{aligned}\|f_{h_k(\gamma)}\|_2^2 &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x + h_k(\gamma)y)K(y)dy\right)^2 dx\\ &\leqslant \|K\|_1^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x + h_k(\gamma)y)^2\frac{K(y)}{\|K\|_1}dydx \leqslant \|f\|_\infty\|K\|_1^2.\end{aligned}$$

Then, by Lemma 6.2, with probability larger than $1 - \kappa_3 e^{-\lambda}$,

$$\Lambda_n(\gamma) \leqslant \theta\frac{\|K\|_2^2}{n\,\mathfrak{h}_n(\gamma)} + \kappa_4\left(\frac{\lambda^2}{\theta n} + \frac{\lambda^3}{\theta n^2 h_n(\gamma_{\max})}\right).$$

Therefore, by Inequality (17), with probability larger than $1 - \kappa_5 e^{-\lambda}$,

$$\|f_{n,\gamma} - f\|_2^2 + \frac{\|K\|_2^2}{n\,\mathfrak{h}_n(\gamma)} \leqslant \frac{1}{1-\theta}\|\widehat{f}_{n,\mathbf{h}_n(\gamma)} - f\|_2^2 + \frac{\kappa_6}{\theta(1-\theta)}\left(\frac{\lambda^2}{n} + \frac{\lambda^3}{n^2 h_n(\gamma_{\max})}\right).$$

This is the announced result if we set $1 + \varepsilon = 1/(1-\theta)$, which gives $1/[\theta(1-\theta)] = (1+\varepsilon)^2/\varepsilon$. $\square$

References

[1] A. Amiri. On a parametric family of sequential estimators of density for a strong mixing process. *C. R. Math. Acad. Sci. Paris* **347**, no. 5-6, 309-314, 2009.

[2] B. Bercu, S. Capderou and G. Durrieu. Nonparametric recursive estimation of the derivative of the regression function with application to sea shores water quality. *Stat Inference Stoch Process*, to appear, 2019.

[3] F. Comte and T. Rebafka. Nonparametric weighted estimators for biased data. *J. Statist. Plann. Inference* **174**, 104-128, 2016.

[4] A. Goldenshluger, and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics* **39**, 1608-1632, 2011.

[5] P. Hall and P. Patil. On the efficiency of on-line density estimators. *IEEE Transactions on Information Theory*, **40**, 5 ,1504-1512, 1994 .

[6] C. Houdré and P. Reynaud-Bouret. *Exponential Inequalities, with Constants, for U-Statistics of Order Two.* Stochastic Inequalities and Applications, Progr. Probab., 56 Birkhäuser, Basel, 55-69, 2003.

[7] T. Klein and E. Rio. Concentration Around the Mean for Maxima of Empirical Processes. *The Annals of Probability* **33**, 1060-1077, 2005.

[8] C. Lacour, P. Massart and V. Rivoirard. Estimator Selection: a New Method with Applications to Kernel Density Estimation. *Sankhya* **79**, 298-335, 2017.

[9] M. Lerasle, N. Magalhaes and P. Reynaud-Bouret. Optimal Kernel Selection for Density Estimation. *High-Dimensional Probability VII : The Cargese Volume*, Prog. Probab. 71, Birkhaüser, 425-460, 2016.

[10] A. Mokkadem, M. Pelletier and Y. Slaoui. The stochastic approximation method for the estimation of a multivariate probability density. *J. Statist. Plann. Inference* **139**, 2459-2478, 2009.

[11] E. Parzen. On the estimation of a probability density function and the mode. *The Annals of Mathematical Statistics*, **33**, 1065-1076.

[12] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832-837, 1956.

[13] A. Tsybakov. *Introduction to Nonparametric Estimation.* Springer, 2009.

[14] E.J. Wegman and H.I. Davies. Remarks on Some Recursive Estimators of a Probability Density. *The Annals of Statistics* **7**, 316-327, 1979.

[15] C. Wolverton and T. Wagner. Asymptotically Optimal Discriminant Functions for Pattern Classification. *IEEE Trans. Information Theory* IT-15, 258-265, 1969.

[16] H. Yamato. Sequential estimation of a continuous probability density function and mode. *Bull. Math. Statist.* **14**, 1-12, 1970/71; correction, ibid. **15**, 133, 1972.

*Laboratoire MAP5, Université Paris Descartes, Paris, France
E-mail address: fabienne.comte@parisdescartes.fr

**Laboratoire Modal'X, Université Paris Nanterre, Nanterre, France
E-mail address: nmarie@parisnanterre.fr

**ESME Sudria, Paris, France
E-mail address: nicolas.marie@esme.fr