



Hermite density deconvolution

Ousmane Sacko

MAP5 UMR 8145, Université Paris Descartes
45 Rue des Saints-Pères,
75006 Paris, France.
E-mail address: ousmane.sacko@parisdescartes.fr
URL: <http://helios.mi.parisdescartes.fr/~sousmane/>

Abstract. We consider the additive model: $Z = X + \varepsilon$, where X and ε are independent. We construct a new estimator of the density of X from n observations of Z . We propose a projection method which exploits the specific properties of the Hermite basis. We study the quality of the resulting estimator by proving a bound on the integrated quadratic risk. We show also that the results can be easily extended to dependent variables. We then propose an adaptive estimation procedure, that is a method of selecting a relevant model. We check that our estimator reaches the classical convergence speeds of deconvolution. Numerical simulations are proposed and a comparison with the results of the method proposed in [Comte and Lacour \(2011\)](#) is performed.

1. Introduction

Consider the additive noise model:

$$Z_k = X_k + \varepsilon_k, \quad k = 1, \dots, n \quad (1.1)$$

where

(**H**₁) $(X_k)_{k \geq 1}$ are independent and identically distributed (i.i.d.) with unknown density f , with respect to the Lebesgue measure,

(**H**₂) $(\varepsilon_k)_{k \geq 1}$ are i.i.d. with known common density f_ε , with respect to the Lebesgue measure,

(**H**₃) $(X_k)_{k \geq 1}$ and $(\varepsilon_k)_{k \geq 1}$ are independent.

We observe n copies Z_1, \dots, Z_n . We want to estimate f , the distribution of X_1 , using Z_1, \dots, Z_n only. Under (**H**₃), if we denote by f_Z the density of Z_1 , we can write

$$f_Z = f * f_\varepsilon, \quad (1.2)$$

Received by the editors March 2nd, 2019; accepted March 23th, 2020.

2010 Mathematics Subject Classification. 62G05, 62G07.

Key words and phrases. Deconvolution, Hermite basis, model selection, projection estimator.

where $g * h(x) = \int_{\mathbb{R}} g(u)h(x-u)du$ is the convolution product of the functions g and h under adequate assumptions. Formula (1.2) explains the term of "deconvolution" for density estimation in model (1.1). The deconvolution problem has been widely studied in the literature. It appears that two factors influence the rate of convergence: the regularity of f and the asymptotic decay of the Fourier transform of the errors f_{ε} , with slower rate of convergence if this decay is faster. Two types of errors are usually considered: "ordinary smooth" errors, when the Fourier transform of f_{ε} is polynomially decaying near infinity, and "super smooth" errors, when it is exponentially decaying near infinity. The first works proposed kernel nonadaptive estimators assuming that f is ordinary smooth and that f_{ε} is ordinary or super smooth. We can cite [Carroll and Hall \(1988\)](#), [Fan \(1991\)](#), [Fan \(1993\)](#), among others, see also the monograph of [Meister \(2009\)](#) on the topic. Adaptive estimation, based on a wavelet method, was first considered by [Pensky and Vidakovic \(1999\)](#). [Butucea \(2004\)](#) establishes the minimax rate in the case where f is super smooth and f_{ε} is ordinary smooth while [Butucea and Tsybakov \(2007\)](#) study optimality in the very difficult case where both functions are super smooth. Some more recent works were dedicated to this problem: [Comte and Lacour \(2011\)](#) consider the case where the noise density is unknown, and propose an adaptive estimator in this setting, later improved by [Kappus and Mabon \(2014\)](#). [Mabon \(2017\)](#) builds a projection estimator in Laguerre basis in the case where the variable of interest is positive. Recently [Comte and Genon-Catalot \(2018\)](#) and [Belomestny et al. \(2019\)](#) described nice properties of Hermite basis. Projection methods allow to summarize the information available on the unknown function through a small number of coefficients. This is why we go further in this direction, and we define a new estimator taking advantage of these convenient properties of Hermite basis. We propose also an adaptive model selection procedure. We obtain a simple, fast and powerful procedure, which preserves standard deconvolution rates. Moreover, its numerical performances are very good. The paper is organized as follows: we define our estimator in Section 2.2. We prove a bound on the risk in both the independent and β -dependent cases in Section 3, and discuss rates of convergence in Section 3.2. In Section 4, an adaptive estimation procedure is proposed in the independent case and a risk control of the resulting estimator is provided. We then illustrate the performance and stability of the adaptive estimation procedure in Section 5, and we compare our results with [Comte and Lacour \(2011\)](#). Proofs of most theoretical results are gathered in Section 7.

2. Estimation procedure and Hermite basis

2.1. Useful tools.

2.1.1. *Notations.* For $a, b \in \mathbb{R}$, let $a \vee b = \max(a, b)$, and $a_+ = \max(0, a)$. For f, g in $\mathbb{L}^2(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$, we denote by $\langle f, g \rangle = \int_{\mathbb{R}} f(u)\overline{g(u)}du$, $\|f\|^2 = \int_{\mathbb{R}} |f(u)|^2 du$, $f^*(x) = \int_{\mathbb{R}} e^{itu}f(u)du$ and $f * g(x) = \int_{\mathbb{R}} f(x-u)g(u)du \forall x \in \mathbb{R}$. Lastly, we recall Plancherel-Parseval formula $\langle f, g \rangle = (2\pi)^{-1}\langle f^*, g^* \rangle$.

Before proposing an estimator, we start by recalling the definition of the Hermite basis.

2.1.2. *Hermite basis.* The Hermite basis $(\varphi_j)_{j \geq 0}$ is a basis on $\mathbb{L}^2(\mathbb{R})$ defined from Hermite polynomials $(H_j)_{j \geq 0}$: $H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j} (e^{-x^2})$. The Hermite polynomials are orthogonal with respect to the weight function e^{-x^2} : $\int_{\mathbb{R}} H_j(x) H_k(x) e^{-x^2} dx = 2^j j! \sqrt{\pi} \delta_{j,k}$ (see [Abramowitz and Stegun, 1964](#), chap 22.2.14), where $\delta_{j,k}$ is the Kronecher symbol. Thus, we deduce that the basis:

$$\varphi_j(x) = c_j H_j(x) e^{-x^2/2}, \quad c_j = (2^j j! \sqrt{\pi})^{-1/2},$$

is orthonormal in $\mathbb{L}^2(\mathbb{R})$. The Hermite basis $(\varphi_j)_{j \geq 0}$ is a bounded basis verifying

$$\|\varphi_j\|_{\infty} = \sup_{x \in \mathbb{R}} |\varphi_j(x)| \leq \phi_0, \quad \text{with } \phi_0 = 1/\pi^{1/4}$$

(see [Abramowitz and Stegun, 1964](#), chap 22.14.17 and [Indritz, 1961](#)). The Fourier transform of $(\varphi_j)_{j \geq 0}$ verifies:

$$\varphi_j^* = \sqrt{2\pi} (i)^j \varphi_j. \tag{2.1}$$

Moreover, according to [Askey and Wainger \(1965\)](#), we have

$$|\varphi_j(x)| < C e^{-\xi x^2}, \quad |x| \geq \sqrt{2j+1}, \quad C > 0, \tag{2.2}$$

where ξ is a positive constant independent of x , $0 < \xi < \frac{1}{2}$.

2.1.3. *Assumptions on the noise.* For the definition of our estimator, we assume the following:

(H₄) the noise density f_{ε} is such that $f_{\varepsilon}^* \neq 0$.

We also assume that f_{ε} satisfies:

There exist $c_1 \geq c'_1 > 0$, and $\gamma \geq 0, \mu \geq 0, \delta \geq 0$ (with $\gamma > 0$ if $\delta = 0$) such that

$$c'_1 (1+t^2)^{\gamma} e^{\mu|t|^{\delta}} \leq \frac{1}{|f_{\varepsilon}^*(t)|^2} \leq c_1 (1+t^2)^{\gamma} e^{\mu|t|^{\delta}}, \quad \text{for all } t \in \mathbb{R}. \tag{2.3}$$

It is standard to assume a condition like (2.3) in the deconvolution setting. When $\delta = 0$ in (2.3), the function f_{ε} and the errors are called "ordinary smooth". When $\delta > 0$ (with the convention that $\delta > 0$ if and only if $\mu > 0$), they are called "super smooth".

Remark 2.1. According to [Lukacs \(1970\)](#), Theorem 4.1.1, the only characteristic function ϕ with $\phi(t) = 1 + o(t^2)$, as $t \rightarrow 0$, is the function $\phi(t) = 1$ for all t . That rules out characteristic functions of the form $e^{-\mu|t|^{\delta}}$ with $\delta > 2$. This implies that in definition (2.3), when $\gamma = 0$, if $|f_{\varepsilon}^*(t)|^2 = ce^{-\mu|t|^{\delta}}$ then necessarily $\delta \leq 2$. Indeed, $|f_{\varepsilon}^*(t)|^2$ is also the characteristic function of a probability density function (it is a characteristic function of $\varepsilon_1 - \varepsilon'_1$ where ε_1 and ε'_1 are i.i.d.).

2.2. *Estimation procedure.* We denote by $S_m = \text{span}\{\varphi_0, \dots, \varphi_{m-1}\}$, the linear space generated by $(\varphi_0, \dots, \varphi_{m-1})$ in $\mathbb{L}^2(\mathbb{R})$. Now, we construct an estimator of f relying on the data Z_1, \dots, Z_n , from model (1.1). We suppose that f belongs to $\mathbb{L}^2(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$, thus we can write $f = \sum_{j=0}^{+\infty} a_j \varphi_j$ with $a_j = \langle f, \varphi_j \rangle$ and the orthogonal projection of f on S_m is given by: $f_m = \sum_{j=0}^{m-1} a_j \varphi_j$. In fact, we estimate f_m and therefore, we build m estimators \hat{a}_j of $a_j, j = 0, \dots, m-1$. Under **(H₄)** and using (1.2), we have $f^* = \frac{f_Z^*}{f_{\varepsilon}^*}$. Therefore, using Parseval's Theorem and (2.1), we have:

$$a_j = \langle f, \varphi_j \rangle = \frac{1}{2\pi} \langle f^*, \varphi_j^* \rangle = \frac{(-i)^j}{\sqrt{2\pi}} \langle f^*, \varphi_j \rangle = \frac{(-i)^j}{\sqrt{2\pi}} \int \frac{f_Z^*(u)}{f_{\varepsilon}^*(u)} \varphi_j(u) du. \tag{2.4}$$

Thus, to estimate a_j , we replace f_Z^* by an estimate. As $f_Z^*(t) = \int e^{itu} f_Z(u) du = \mathbb{E}[e^{itZ_1}]$, we set:

$$\hat{f}_Z^*(t) = \frac{1}{n} \sum_{k=1}^n e^{itZ_k}. \tag{2.5}$$

Plugging (2.5) into (2.4), we can propose an estimator of f_m , provided that $\varphi_j/f_\varepsilon^*$ is integrable on \mathbb{R} , for $j = 0, \dots, m - 1$:

$$\hat{f}_m = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j, \quad \hat{a}_j = \frac{(-i)^j}{\sqrt{2\pi}} \int \frac{\hat{f}_Z^*(u)}{f_\varepsilon^*(u)} \varphi_j(u) du. \tag{2.6}$$

Note that the coefficients \hat{a}_j are real. Indeed, using that $\varphi_j(-x) = (-1)^j \varphi_j(x)$, it holds:

$$\overline{\hat{a}_j} = \frac{(i)^j}{\sqrt{2\pi}} \int \frac{\hat{f}_Z^*(-u)}{f_\varepsilon^*(-u)} \varphi_j(u) du = \frac{(i)^j}{\sqrt{2\pi}} \int \frac{\hat{f}_Z^*(u)}{f_\varepsilon^*(u)} \varphi_j(-u) du = \hat{a}_j,$$

where \bar{z} denotes the complex conjugate of the complex number z . The Hermite basis has the specificity of leading to integrable $\varphi_j/f_\varepsilon^*$ in a large number of cases. This estimator is different from the one proposed by Comte and Genon-Catalot (2018), who propose to take instead of \hat{a}_j , the estimator

$$\tilde{a}_{j,\sqrt{m}} = ((-i)^j / \sqrt{2\pi}) \int_{|u| \leq \pi\sqrt{m}} \hat{f}_Z^*(u) \varphi_j(u) / f_\varepsilon^*(u) du.$$

The drawback of the latter estimator is that it is biased and the coefficients depend on m , making the choice of m untractable in the sequel. Our estimator is an unbiased estimator of f_m and is easy to handle.

3. Risk study of the estimator

3.1. *Risk of the estimator for fixed m .* Under the additional assumption:

(H₅) f_Z is bounded,

we can study the risk of \hat{f}_m and the following proposition states our result.

Proposition 3.1. (i) Under (H₁), ..., (H₅) and for \hat{f}_m given by (2.6), we have for any $l > 0$

$$\mathbb{E}[\|\hat{f}_m - f\|^2] \leq \|f - f_m\|^2 + \frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} + \frac{2}{n} \|f_Z\|_\infty \sum_{j=0}^{m-1} \int_{|u| > \sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du, \tag{3.1}$$

(ii) If in addition we choose $l \geq 2$ and if f_ε satisfies (2.3) with $0 \leq \delta < 2$ or ($\delta = 2$, with $\mu < \xi$), where ξ is defined in (2.2), then

$$\frac{2}{n} \|f_Z\|_\infty \sum_{j=0}^{m-1} \int_{|u| > \sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du = \mathcal{O}\left(\frac{1}{n}\right). \tag{3.2}$$

Note that the constant l does not depend on m or n . The first right-hand side term of (3.1) is the bias term, it is decreasing with m as $\|f - f_m\|^2 = \sum_{j \geq m} a_j^2$. The second term is the main variance term, it is clearly increasing with m . The last term also comes from the variance computation, but we give in Proposition 3.1, part (ii) conditions ensuring that it is negligible. Thus, choosing m that minimizes the risk requires a bias-variance compromise.

So under the assumptions of Proposition 3.1, part (ii), (3.1) becomes:

$$\mathbb{E}[\|\hat{f}_m - f\|^2] \leq \|f - f_m\|^2 + \frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} + \frac{c}{n}, \quad c > 0, \quad l \geq 2.$$

Comment about (H₅): We both have, $\forall x \in \mathbb{R}, |f_Z(x)| = |f * f_\varepsilon(x)| \leq \min(\|f\|_\infty, \|f_\varepsilon\|_\infty)$ and $|f_Z(x)| \leq \|f\| \cdot \|f_\varepsilon\|$. Therefore, the density f_Z is bounded if f or f_ε is bounded, or if both functions are square integrable. Condition (H₅) is not very strong.

3.2. *Rate of convergence.* To obtain rates of convergence, we have to evaluate the order of bias and variance terms. In general, each basis is associated with a regularity space: here, we consider Sobolev-Hermite spaces.

3.2.1. *Rate on a Sobolev-Hermite space.* For $s > 0$, the Sobolev-Hermite space of regularity s (see Bongioanni and Torrea, 2006) is given by:

$$W_H^s = \{\theta : \mathbb{R} \rightarrow \mathbb{R}, \theta \in \mathbb{L}^2(\mathbb{R}), \sum_{k \geq 0} k^s a_k^2(\theta) < +\infty\}, \quad a_k(\theta) = \int \theta(u) \varphi_k(u) du$$

and the Sobolev-Hermite ball by:

$$W_H^s(D) = \{\theta \in \mathbb{L}^2(\mathbb{R}), \sum_{k \geq 0} k^s a_k^2(\theta) \leq D\}, \quad D > 0.$$

For s integer, θ belongs to W_H^s if and only if θ admits derivatives up to order s and the functions $\theta, \theta', \dots, \theta^{(s)}, x^{(s-k)}\theta^{(k)}$ belong to $\mathbb{L}^2(\mathbb{R})$, with $k = 0, \dots, s - 1$. We can compare this space with the classical Sobolev space with regularity s , defined by:

$$W^s = \{\theta \in \mathbb{L}^2(\mathbb{R}), \int (1 + u^{2s}) |\theta^*(u)|^2 du < +\infty\}.$$

Actually, Bongioanni and Torrea (2006) prove that, for $s > 0$, $W_H^s \subsetneq W^s$. It is also proved therein and in Belomestny et al. (2019) that, for s integer,

$$W^s = \left\{ \begin{array}{l} \theta \in \mathbb{L}^2(\mathbb{R}), \theta \text{ admits derivatives up to order } s, \text{ such that} \\ \|\theta\|_{s, sob} := \sum_{j=0}^s |\theta^{(j)}|^2 < +\infty \end{array} \right\}.$$

Consequently, for s integer, it follows that $W_H^s \subset W^s$. For more details on these regularity spaces, the reader is referred to Section 4.1 in Belomestny et al. (2019).

Thus, for f in $W_H^s(D)$, we have $\|f - f_m\|^2 = \sum_{j \geq m} j^s a_j^2 j^{-s} \leq Dm^{-s}$. Under the assumptions of Proposition 3.1 and for $f \in W_H^s(D)$, we get:

$$\mathbb{E}[\|\hat{f}_m - f\|^2] \lesssim Dm^{-s} + \frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2}, \quad (3.3)$$

where, for two functions u, v , we denote $u(x) \lesssim v(x)$ if $u(x) \leq cv(x)$, with c is a constant independent of x . This inequality is similar to the one in Comte and Lacour (2011), with m therein replaced now by \sqrt{m} . It is worth underlining that the role of the dimension m in projection methods is played here by \sqrt{m} : this is a specificity of the Hermite basis. The result is the similar in density estimation when X_k are directly observed, (see Comte and Genon-Catalot, 2018, Belomestny et al., 2019). Let us denote by m_{opt} the value of m for which the bias-variance compromise is obtained, relying on the same calculations as in Comte and Lacour (2011), the rates and the dimension m_{opt} are given in Table 3.1.

	$\delta = 0$	$0 < \delta < 2$ or $\delta = 2, \mu < \xi$
m_{opt}	$\lceil n^{\frac{2}{2s+2\gamma+1}} \rceil$	$\left\lceil \frac{1}{l} \left(\frac{\log n}{2\mu} \right)^{\frac{2}{\delta}} \right\rceil$
Rate	$n^{-\frac{2s}{2s+2\gamma+1}}$	$(\log n)^{-\frac{2s}{\delta}}$

TABLE 3.1. Rate of convergence for the MISE if $f \in W_H^s(D)$.

These rates coincide with the ones obtained by [Fan \(1993\)](#), [Pensky and Vidakovic \(1999\)](#). They are known to be optimal: lower bounds corresponding to these rates for f_ε verifying (2.3) are proved by [Fan \(1993\)](#) when f belongs to a Hölder class, and by [Pensky and Vidakovic \(1999\)](#) for f in a Sobolev class.

3.2.2. *Rates of convergence for specific function classes.* We can obtain for some specific classes of functions a bias term with much smaller order, for instance Gaussian densities or mixtures of Gaussian. Indeed, then, we can explicitly compute the coefficients a_j and obtain smaller bias than previously on $W_H^s(D)$. Let

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad g_{p,\sigma}(x) = \frac{x^{2p}}{\sigma^{2p}C_{2p}} f_{0,\sigma}(x), \quad C_{2p} = \mathbb{E}[X^{2p}],$$

for X a standard Gaussian variable. We also define the class of mean mixtures, respectively of variance mixtures of the Gaussian distribution by:

$$\mathcal{F}(C) = \left\{ f : f(x) = \phi \star \Pi(x) = \int \phi(x-u)d\Pi(u), \quad \Pi \in \mathcal{P}(C) \right\},$$

where $\mathcal{P}(C) := \{ \Pi \in \mathcal{P}(\mathbb{R}), \Pi(|u| > t) \leq C \exp(-t^2/C), \quad \forall t \in \mathbb{R}^+ \}$, respectively

$$\mathcal{G}(v) = \left\{ f : f(x) = \int_0^{+\infty} \frac{\phi(x/u)}{u} d\Pi(u), \quad \Pi([1/\sqrt{v}, \sqrt{v}]) = 1 \right\}, \quad v > 1,$$

with ϕ the density of standard Gaussian and $\mathcal{P}(\mathbb{R})$ the set of probability measures on \mathbb{R} . The following results are based on bias evaluation obtained in [Belomestny et al. \(2019\)](#). The rate is given by the order of variance term, since in all these cases, the bias term is exponentially small. We can prove the following proposition.

Proposition 3.2. *Assume the assumptions $(\mathbf{H}_1), \dots, (\mathbf{H}_5)$ hold and f_ε is ordinary smooth. For the choice $m_{opt} = \lceil \log(n)/C_1 \rceil$, with $C_1 = \log(2) + e\mu^2$ if $f = f_{\mu,1}$, $C_1 = \log\left(\frac{\sigma^2+1}{\sigma^2-1}\right)^2$ if $f = f_{0,\sigma}$, $C_1 = \frac{1}{(eC+1/\log(2))}$ if $f \in \mathcal{F}(C)$, $C_1 = \left(\frac{v^2-1}{v^2+1}\right)$ if $f \in \mathcal{G}(v)$, we have*

$$\mathbb{E} \left[\|\hat{f}_{m_{opt}} - f\|^2 \right] \lesssim \frac{(\log n)^{\gamma+\frac{1}{2}}}{n},$$

where γ is given in (2.3).

The same result holds for $f = g_{p,\sigma}$. This rate is similar to the one obtained in [Butucea \(2004\)](#) for super-smooth functions f .

However in all previous cases the choice $m = m_{opt}$ depends on the regularity of f and associated parameters, which are unknown. This is why we have to look for another method to make the bias-variance compromise, in a data-driven way (see Section 4).

3.3. *Comparison with the classical estimator in deconvolution.* The "standard" deconvolution estimator (see [Fan, 1991](#), and choose sinus cardinal kernel) is given by:

$$\check{f}_\ell(x) = \frac{1}{2\pi} \int_{-\pi\ell}^{\pi\ell} e^{-ixu} \frac{\hat{f}_Z^*(u)}{f_\varepsilon^*(-u)} du, \text{ where } \hat{f}_Z^* \text{ is defined by (2.5).}$$

We mention that this estimator can be decomposed in an orthonormal basis namely $\psi_{\ell,j}(x) = \sqrt{\ell}\psi(\ell x - j)$, $\psi(x) = \frac{\sin \pi x}{\pi x}$ (see [Comte et al., 2008](#), Section 3.2), but the development is infinite:

$$\check{f}_\ell(x) = \sum_{j \in \mathbb{Z}} \hat{a}_{\ell,j} \psi_{\ell,j}, \quad \hat{a}_{\ell,j} = \frac{1}{n} \sum_{k=1}^n \frac{1}{2\pi} \int \frac{\psi_{\ell,j}^*(-u)}{f_\varepsilon^*(u)} e^{iuZ_k} du$$

A finite (computable) development would require an additional approximation (truncation of the sum as in [Comte et al., 2008](#)) to $k_n \geq n$ coefficients. From computation point of view, the low complexity of \hat{f}_m in the Hermite basis is an advantage (see [Belomestny et al., 2019](#), Section 4.5). The risk of \check{f}_ℓ verifies

$$\mathbb{E}[\|\check{f}_\ell - f\|^2] \leq \frac{1}{2\pi} \int_{|t|>\pi\ell} |f^*(u)|^2 du + \frac{1}{2\pi n} \int_{|u|\leq\pi\ell} \frac{du}{|f_\varepsilon^*(u)|^2}.$$

In this context, the regularity spaces which are considered are Sobolev balls defined by

$$W^s(D') = \left\{ f \in \mathbb{L}^2(\mathbb{R}), \int (1 + u^{2s}) |f^*(u)|^2 du < D' \right\}, \quad D' > 0.$$

Note that it is proved in [Belomestny et al. \(2019\)](#) that $W_H^s(D) \subset W^s(D')$, for D and D' related constants. For $f \in W^s(D)$ the bias term is such that $\frac{1}{2\pi} \int_{|t|>\pi\ell} |f^*(u)|^2 du \leq \frac{D}{2\pi} (\pi\ell)^{-2s} = C\ell^{-2s}$, where $C = \frac{D}{2\pi} \pi^{-2s}$. Therefore, for $\ell = \sqrt{m}$, the risks of the two estimators have the same order on $W_H^s(D)$. This implies that they have the same rates of convergence.

3.4. *Extension to the dependent case.* Proposition 3.1 (and its consequences) may be extended to the context of dependent X_i 's. We first define the mixing coefficients.

Definition 3.3. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and \mathcal{U}, \mathcal{V} two σ -algebras of \mathcal{A} . The β -mixing coefficient is defined by

$$\beta(\mathcal{U}, \mathcal{V}) = \frac{1}{2} \sup \left\{ \sum_{i=1}^{\mathcal{I}} \sum_{j=1}^{\mathcal{J}} |\mathbb{P}(U_i \cap V_j) - \mathbb{P}(U_i)\mathbb{P}(V_j)| \right\}, \quad (3.4)$$

where the supremum is taken over all pairs finite partitions $\{U_1, \dots, U_{\mathcal{I}}\}$ and $\{V_1, \dots, V_{\mathcal{J}}\}$ of Ω , such that $U_i \in \mathcal{U}$ and $V_j \in \mathcal{V}$.

Let $(X_k)_{k \in \mathbb{Z}}$ a strictly stationary process. Let $\mathcal{F}_0 = \sigma(X_i, i \leq 0)$ and $\mathcal{F}_k = \sigma(X_i, i \geq k)$ for all $k \in \mathbb{Z}$, where \mathcal{F}_0 is the σ -algebra generated by the X_i for $i \leq 0$ and \mathcal{F}_k generated by X_i for $i \geq k$. The mixing coefficient β_k is defined by $\beta_k = \beta(\mathcal{F}_0, \mathcal{F}_k)$, where β is defined by (3.4).

The process $(X_k)_{k \in \mathbb{Z}}$ is β -mixing if the sequence β_k tends to zero at infinity.

In this section, we still consider model (1.1), but we replace (\mathbf{H}_1) by:

(\mathbf{H}'_1) $(X_k)_{k \geq 1}$ is strictly stationary and β -mixing.

The estimator is the same as in the independent case and we can prove a bound on the risk.

Proposition 3.4. *Let assumptions $(\mathbf{H}'_1), (\mathbf{H}_2), \dots, (\mathbf{H}_5)$ hold. Let $1 \leq p, q < +\infty$ two real numbers such that $\frac{1}{p} + \frac{1}{q} = 1$. Then, if $\mathbb{E}[|X_1|^{2q/3}] < +\infty$ and the mixing coefficient are such that $\sum_{k=0}^{+\infty} (k+1)^{p-1} \beta_k < +\infty$, we have*

$$\begin{aligned} \mathbb{E}[\|\hat{f}_m - f\|^2] &\leq \|f - f_m\|^2 + \frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} \\ &\quad + \frac{2}{n} \|f_Z\|_\infty \sum_{j=0}^{m-1} \int_{|u| > \sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du + c' \frac{\sqrt{m}}{n}, \end{aligned} \tag{3.5}$$

where $l \geq 2$ is a positive constant, and c' is a constant depending on $\mathbb{E}[|X_1|^{2q/3}]$ and $\sum_{k=0}^{+\infty} (k+1)^{p-1} \beta_k$.

Now, we comment this bound of risk. We remark that we have the same bias and variance terms as in the i.i.d. case with an additional term $c' \sqrt{m}/n$ which is clearly specific to the β -mixing case. As $|f_\varepsilon^*(u)| \leq 1$, we have, $\frac{1}{\pi} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} \geq \frac{2\sqrt{l}}{\pi} \sqrt{m}$. Consequently, \sqrt{m}/n has smaller order than $\frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2}$ and Inequality (3.5) implies that the risk of \hat{f}_m here has the same order as in the i.i.d. case. We have therefore the same rates of convergence.

We compare the result given in Proposition 3.4 to Proposition 4.1 in Comte et al. (2008). The first two right-hand side terms of (3.5) ($\|f - f_m\|^2 + \frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2}$) are the same as in Comte et al. (2008) with \sqrt{lm} replaced by πm (see Section 3.3). Under the assumptions of Proposition 3.1 (ii) the other terms (residual terms) are order $\mathcal{O}(n^{-1}) + \mathcal{O}(\sqrt{mn}^{-1})$. This order is smaller than the order of the residual term stated in (4.4) of Comte et al. (2008), which is $n^{-1}m^2$. Note that all estimators of their collection require to compute $k_n \geq n$ coefficients, which can make the procedure slow when n is large.

4. Adaptive estimation and model selection

For sake of brevity and simplicity, we only study the independent case (i.i.d case) hereafter.

From now on, l given in Proposition 3.1, part (ii) is assumed to be fixed. In this section we propose an automatic selection of m which performs the bias-variance compromise. The procedure does not depend on the regularity of the density f , but only on data Z_1, \dots, Z_n . Consider the contrast function defined by

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n} \sum_{k=1}^n \phi_t(Z_k), \quad \phi_t(x) = \frac{1}{2\pi} \int \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-ixu} du. \tag{4.1}$$

It is easy to check that $\hat{f}_m = \underset{t \in S_m}{\operatorname{argmin}} \gamma_n(t)$. Let

$$\Delta(m) = \frac{1}{\pi} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2}.$$

We consider \mathcal{M}_n , the collection of models,

$$\mathcal{M}_n = \{m \in \mathbb{N} \setminus \{0\}, \Delta(m) \leq n\}.$$

This collection is finite and contains models with bounded variance. More precisely, as already noticed, $|f_\varepsilon^*(u)| \leq 1$, implies $\Delta(m) \geq \frac{1}{\pi} \int_{|u| \leq \sqrt{lm}} du = \frac{2\sqrt{lm}}{\pi}$. Therefore,

the elements m of \mathcal{M}_n satisfy $m \lesssim n^2$. The cardinal of \mathcal{M}_n is therefore at most of order $\mathcal{O}(n^2)$. Our aim is to find the best model \hat{m} in \mathcal{M}_n , that is, to select \hat{m} such that, the risk of $\hat{f}_{\hat{m}}$ approximately performs the bias-variance trade-off, without any information on f . We set:

$$\hat{m} = \operatorname{argmin}_{m \in \mathcal{M}_n} \{ \gamma_n(\hat{f}_m) + \operatorname{pen}(m) \}, \tag{4.2}$$

where $\operatorname{pen}(m)$ is an increasing function defined by:

$$\operatorname{pen}(m) = \begin{cases} \kappa \frac{\Delta(m)}{n}, & \text{if } f_\varepsilon \text{ is ordinary smooth or super smooth with } \delta < \frac{1}{2}, \\ 2\kappa \left(1 + 24\mu l^{\delta/2} m^{\delta - \frac{1}{2}} \right) \frac{\Delta(m)}{n} & \text{if } f_\varepsilon \text{ is super smooth with } \frac{1}{2} \leq \delta \leq 2, \end{cases} \tag{4.3}$$

where $\kappa > 0$ is a numerical constant, μ is the constant given in (2.3) and $l \geq 2$ given in Proposition 3.1, fixed. As $\gamma_n(\hat{f}_m) = -\|\hat{f}_m\|^2 = -\sum_{j=0}^{m-1} \hat{a}_j^2$, it is worth emphasizing that computing \hat{m} is numerically fast. Clearly the choice of m given by (4.2) is entirely determined by the data. The constant κ is independent of the data. The theoretical results show that $\kappa > 17$ is suitable (see the proof of Lemma 7.2). In practice this value is too large and is calibrated by preliminary simulation experiments. They confirm that (see Section 5) smaller practical values must be chosen.

We can prove the following theorem.

Theorem 4.1. *Assume $(\mathbf{H}_1), \dots, (\mathbf{H}_5)$ hold and f_ε is square integrable. Let $\operatorname{pen}(m)$ defined by (4.3), $\hat{f}_m = \operatorname{argmin}_{t \in \mathcal{S}_m} \gamma_n(t)$ and \hat{m} selected by (4.2). Then, there exists a constant κ_0 such that, for all $\kappa > \kappa_0 = 17$, the estimator $\hat{f}_{\hat{m}}$ satisfies*

$$\mathbb{E} \left[\|\hat{f}_{\hat{m}} - f\|^2 \right] \leq C \inf_{m \in \mathcal{M}_n} (\|f - f_m\|^2 + \operatorname{pen}(m)) + \frac{C'}{n}, \tag{4.4}$$

where C is a numerical constants ($C=4$ suits) and C' a constant depending on f_ε .

Remark 4.2. Assume that the assumptions of Theorem 4.1 are satisfied. Then if $f \in W_H^s(D)$ the estimator $\hat{f}_{\hat{m}}$ converges to f with the rates obtained in Table 3.1. Indeed, the term C'/n in (4.4) does not change the order of the rate, and is negligible compared to the term $\|f - f_m\|^2 + \operatorname{pen}(m)$. Moreover, (4.3) induces a loss in the order of $\operatorname{pen}(m)$ compared to the variance term when $\delta > 1/2$, but this does not change the rate which is governed by the bias term in this case (see Table 3.1 and choice of m_{opt} of order $(\log n)$).

5. Simulation and numerical results

5.1. *Implementation of the adaptive estimator.* In this section, we propose some illustrations of the theoretical results. More precisely, we implement the projection estimator given by (2.6). To do this, we consider data simulated according to (1.1). For the density f , we choose the distributions (following Comte and Lacour, 2011):

- (i) Gaussian standard $\mathcal{N}(0, 1)$, $I = [-4, 4]$
- (ii) Cauchy standard: $f(x) = (\pi(1+x^2))^{-1}$, $I = [-10, 10]$
- (iii) Laplace density: $f(x) = e^{-\sqrt{2}|x|}/\sqrt{2}$, $I = [-5, 5]$
- (iv) Gamma density $\Gamma(4, 1/\sqrt{3})/\sqrt{12}$, $I = [0, 6]$
- (v) Mixed-Gaussian density $(0.5\mathcal{N}(-2, 1) + 0.5\mathcal{N}(2, 1))/\sqrt{5}$, $I = [-3, 3]$

where I is the interval on which we compute the risks. Except the Cauchy density, all the densities are normalized to have variance equal to 1. Note also densities (i) and (v) belong to W_H^s with $s = +\infty$, (iv) has regularity $s = 3 - \eta$, $\eta > 0$, (ii) and (iii) admit a regularity $s = 3/2 - \eta$ and $s = 1 - \eta$, $\eta > 0$, respectively (but (ii) is infinitely differentiable).

For noise distributions, we consider two cases with the same variance 1/10 and thus, except for the Cauchy density the signal to noise ratio is equal to 10.

• **Case 1 : Laplace noise ("ordinary smooth")**

We consider the density f_ε :

$$f_\varepsilon(x) = \frac{\lambda}{2} e^{-\lambda|x|}; \quad f_\varepsilon^*(x) = \frac{\lambda^2}{\lambda^2 + x^2}; \quad \lambda = 2\sqrt{5}.$$

The penalty term is given by:

$$\begin{aligned} \text{pen}(m) &= \frac{\kappa}{n} \Delta(m) = \frac{\kappa}{\pi n} \int_{|u| \leq \sqrt{lm}} \left(1 + \frac{u^2}{\lambda^2}\right)^2 du \\ &= \frac{2\kappa}{\pi n} \left(\sqrt{lm} + \frac{2}{3\lambda^2} (\sqrt{lm})^3 + \frac{(\sqrt{lm})^5}{5\lambda^4} \right), \end{aligned}$$

where $l = 6$.

• **Case 2 : Gaussian noise ("super smooth")**

We have:

$$f_\varepsilon(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-x^2/2\sigma^2}; \quad f_\varepsilon^*(x) = e^{-\sigma_\varepsilon^2 x^2/2}, \quad \sigma_\varepsilon^2 = 1/10.$$

The penalty proposed is:

$$\text{pen}(m) = 4\kappa \left(1 + 24\sigma_\varepsilon^2 lm^{3/2}\right) \frac{\sqrt{lm}}{\pi n} \left(\int_0^1 e^{u^2 \sigma_\varepsilon^2 lm} du \right),$$

where $l = 4$ here and the integral is computed by a Riemann sum discretized in 300 points. Then, we have to calibrate the penalty constant κ . This constant is fixed through preliminary simulations, by testing set of values on different densities f with a large number of repetitions. The comparison of the risks for these different values of κ makes it possible to make a reasonable choice. Finally, we choose $\kappa = 0.4$ for a Laplace noise, $\kappa = 10^{-3}$ for a Gaussian noise.

The estimation procedure is described as follows:

- For m in \mathcal{M}_n , compute $-\sum_{j=0}^{m-1} \hat{a}_j^2 + \text{pen}(m) = \text{Cr}(m)$, with \hat{a}_j given by (2.6),
- Select \hat{m} such that $\hat{m} = \underset{m \in \mathcal{M}_n}{\text{argmin}} \text{Cr}(m)$,
- Compute $\hat{f}_{\hat{m}} = \sum_{j=0}^{\hat{m}-1} \hat{a}_j \varphi_j$, and $\int_I (\hat{f}_{\hat{m}}(u) - f(u))^2 du$ by discretization.

5.2. *Simulations results.* Simulation results are given in Tables 5.2 and 5.3. The columns of Table 5.2 indicate the values of the MISE (Mean Integrated Squared Error) multiplied by 100 for a Laplace noise or a Gaussian noise, Table 5.3 gives the ratio of the risk values obtained in Comte and Lacour (2011) divided by the risk values obtained by our method: the larger it is, the better our method is. The errors obtained by our method are computed by a discretization of the integral as Riemann sums and averaged over 100 independent simulations. We remark that increasing the sample size makes the error smaller and thus improves the estimation. Globally the results of our simulations are satisfactory and our method is often better than

Comte and Lacour (2011) for both noise densities. The main exception concerns the Gamma density (iv). Some failures for Cauchy density (ii) and super smooth noise are also observed, especially when n increases.

		$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
f	Noise	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.
	Gaussian		0.44	0.37	0.12	0.06	9.5910^{-2}	4.310^{-2}	7.10^{-2}
Cauchy		0.28	0.89	0.20	0.56	0.14	0.37	0.10	0.29
Laplace		1.65	2.18	1.06	1.34	0.75	1.16	0.57	0.87
Gamma		1.70	1.27	0.98	0.97	0.50	0.90	0.28	0.83
Mixed-Gaussian		2.82	1.91	1.09	0.87	0.66	0.69	0.41	0.53

TABLE 5.2. Empirical integrated mean squared errors computed from $(100 \times \mathbb{E}\|\hat{f}_m - f\|^2)$ over 100 independent simulations for $n = 100, 250, 500, 1000$.

		$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
f	Noise	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.
	Gaussian		1.95	1.27	5.67	5.00	5.01	5.11	2.41
Cauchy		4.07	1.07	2.45	0.79	2.43	0.70	1.40	0.52
Laplace		1.47	1.40	1.13	1.34	1.12	1.02	1.04	0.89
Gamma		0.67	0.88	0.66	0.73	0.82	0.49	1	0.37
M-Gaussian		1.26	2.17	1.45	2.24	1.17	1.68	0.95	1.15

TABLE 5.3. Ratio of the risks obtained in Comte and Lacour (2011) divided by those of Table 5.2.

6. Concluding remarks

We proposed a projection estimator of the density of X in the convolution model (1.1), relying on the Hermite basis. The estimator has the advantage to be kernel-free, as the integral is over the entire real line and not truncated as in the previous works by Comte and Genon-Catalot (2018). The method provides a parsimonious description of the function under estimation: indeed the function is relevantly estimated thanks to a small number of coefficients. This has also the advantage of making the method numerically fast and convenient. We prove a bound on the quadratic risk in the independent and β -dependent cases which shows that the relevant parameter is not the dimension m but rather \sqrt{m} . A data driven estimator is proposed: the model can be automatically chosen and the resulting estimator reaches optimal rates in most cases. We also provide numerical simulation results, and the comparison with Comte and Lacour (2011) ensures the good performances of our method.

7. Proofs

7.1. *Proof of Proposition 3.1.* We start by the part (i). For \hat{f}_m given by (2.6), we have:

$$\mathbb{E} \left[\|\hat{f}_m - f\|^2 \right] = \|f - f_m\|^2 + \mathbb{E} \left[\|\hat{f}_m - f_m\|^2 \right] = \|f - f_m\|^2 + \sum_{j=0}^{m-1} \text{Var}(\hat{a}_j). \quad (7.1)$$

Now with the definition of \hat{a}_j given by (2.6) we have

$$\begin{aligned} \text{Var}(\hat{a}_j) &= \text{Var} \left(\frac{(-i)^j}{\sqrt{2\pi n}} \int_{\mathbb{R}} \sum_{k=1}^n e^{iuZ_k} \frac{\varphi_j(u)}{f_{\varepsilon}^*(u)} du \right) \\ &= \frac{1}{2\pi n} \text{Var} \left((-i)^j \int_{\mathbb{R}} e^{iuZ_1} \frac{\varphi_j(u)}{f_{\varepsilon}^*(u)} du \right) \\ &\leq \frac{1}{2\pi n} \mathbb{E} \left[\left| (-i)^j \int_{\mathbb{R}} e^{iuZ_1} \frac{\varphi_j(u)}{f_{\varepsilon}^*(u)} du \right|^2 \right]. \end{aligned}$$

Plugging this in (7.1) yields

$$\mathbb{E} \left[\|\hat{f}_m - f\|^2 \right] \leq \|f - f_m\|^2 + \frac{1}{2\pi n} \sum_{j=0}^{m-1} \mathbb{E} \left[\left| \int_{\mathbb{R}} e^{iuZ_1} \frac{\varphi_j(u)}{f_{\varepsilon}^*(u)} du \right|^2 \right].$$

Using $|a + b|^2 \leq 2|a|^2 + 2|b|^2$, we deduce

$$\begin{aligned} \mathbb{E} \left[\sum_{j=0}^{m-1} \left| \int_{\mathbb{R}} e^{iuZ_1} \frac{\varphi_j(u)}{f_{\varepsilon}^*(u)} du \right|^2 \right] &\leq 2\mathbb{E} \left[\sum_{j=0}^{m-1} \left| \int_{|u| > \sqrt{lm}} e^{iuZ_1} \frac{\varphi_j(u)}{f_{\varepsilon}^*(u)} du \right|^2 \right] \\ &\quad + 2\mathbb{E} \left[\sum_{j=0}^{m-1} \left| \int_{|u| \leq \sqrt{lm}} e^{iuZ_1} \frac{\varphi_j(u)}{f_{\varepsilon}^*(u)} du \right|^2 \right]. \end{aligned}$$

We evaluate the two right-hand side terms of the previous inequality. By Bessel inequality we have, for the last term:

$$\begin{aligned} \mathbb{E} \left[\sum_{j=0}^{m-1} \left| \int_{|u| \leq \sqrt{lm}} e^{iuZ_1} \frac{\varphi_j(u)}{f_{\varepsilon}^*(u)} du \right|^2 \right] &= \mathbb{E} \left[\sum_{j=0}^{m-1} \left| \left\langle \frac{e^{iZ_1 \bullet}}{f_{\varepsilon}^*} \mathbb{1}_{|\bullet| \leq \sqrt{lm}}, \varphi_j \right\rangle \right|^2 \right] \\ &\leq \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_{\varepsilon}^*(u)|^2}. \end{aligned} \quad (7.2)$$

Moreover, let $\psi_j(u) = \frac{\varphi_j(u)}{f_{\varepsilon}^*(u)} \mathbb{1}_{|u| > \sqrt{lm}}$, we get for the other term

$$\begin{aligned} \mathbb{E} \left[\sum_{j=0}^{m-1} \left| \int_{|u| > \sqrt{lm}} e^{iuZ_1} \frac{\varphi_j(u)}{f_{\varepsilon}^*(u)} du \right|^2 \right] &= \sum_{j=0}^{m-1} \int_{\mathbb{R}} \left| \int_{|u| > \sqrt{lm}} e^{iuz} \frac{\varphi_j(u)}{f_{\varepsilon}^*(u)} du \right|^2 f_Z(z) dz \\ &\leq \|f_Z\|_{\infty} \sum_{j=0}^{m-1} \int_{\mathbb{R}} \left| \int_{|u| > \sqrt{lm}} e^{iuz} \frac{\varphi_j(u)}{f_{\varepsilon}^*(u)} du \right|^2 dz \end{aligned}$$

$$= \|f_Z\|_\infty \sum_{j=0}^{m-1} \|\psi_j^*\|^2 = 2\pi \|f_Z\|_\infty \sum_{j=0}^{m-1} \|\psi_j\|^2. \tag{7.3}$$

Putting (7.2), (7.3) in (7.1), we have the part (i).

Let us prove the part (ii). We have, using (2.3), that:

$$\sum_{j=0}^{m-1} \int_{|u|>\sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du \leq c_1 \sum_{j=0}^{m-1} \int_{|u|>\sqrt{lm}} (1+u^2)^\gamma |\varphi_j(u)|^2 e^{\mu|u|^\delta} du.$$

By (2.2), we have $|\varphi_j(x)| < Ce^{-\xi x^2}$ if $|x| \geq \sqrt{2j+1}$, for $j \in \{0, \dots, m-1\}$. Thus it is in particular true for $|x| \geq \sqrt{lm}$, with $l \geq 2$. Therefore, for $j \leq m-1$,

$$\begin{aligned} \int_{|u|>\sqrt{lm}} (1+u^2)^\gamma |\varphi_j(u)|^2 e^{\mu|u|^\delta} du &\leq C^2 \int_{|u|>\sqrt{lm}} (1+u^2)^\gamma e^{-2\xi u^2} e^{\mu|u|^\delta} du \\ &\leq C^2 e^{-\xi lm} \int_{\mathbb{R}} (1+u^2)^\gamma e^{-\xi u^2} e^{\mu|u|^\delta} du. \end{aligned}$$

And $\int_{\mathbb{R}} (1+u^2)^\gamma e^{-\xi u^2} e^{\mu|u|^\delta} du < +\infty$ if $\delta < 2$ or if $\delta = 2, \mu < \xi$, which corresponds to our assumptions. Therefore: $\sum_{j=0}^{m-1} \int_{|u|>\sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du = \mathcal{O}(me^{-\xi lm})$. Hence the result. \square .

7.2. *Proof of Proposition 3.2.* By (3.1) and (3.2), we have:

$$\mathbb{E}[\|\hat{f}_m - f\|^2] \leq \|f - f_m\|^2 + \frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} + \frac{c}{n}.$$

Using Lemma 2 in Comte and Lacour (2011) p.8, we have

$$\int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} \asymp m^{\gamma + \frac{1-\delta}{2}} e^{\mu l^{\frac{\delta}{2}} m^{\frac{\delta}{2}}}. \tag{7.4}$$

We denote for two functions u and v , $u(x) \asymp v(x)$, if $u(x) \lesssim v(x)$ and $v(x) \lesssim u(x)$.

From Belomestny et al. (2019) the bias term is exponentially small (see Proposition 7, 8 and 9), thus, the rate of convergence is given by the order of variance term. As f_ε is ordinary smooth, $\delta = 0$ in (7.4) and replacing m by $m_{opt} = \lceil \log(n)/C_1 \rceil$, with C_1 is given in Proposition 3.2, we have the result. \square

7.3. *Proof of Proposition 3.4.* As in the i.i.d. case, we have the bias-variance decomposition given by (7.1). Now,

$$\begin{aligned} \text{Var}(\hat{a}_j) &= \text{Var} \left(\frac{(-i)^j}{\sqrt{2\pi n}} \int_{\mathbb{R}} \sum_{k=1}^n e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(t)} du \right) \\ &= \frac{1}{2\pi n^2} \sum_{k=1}^n \text{Var} \left((-i)^j \int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right) \\ &\quad + \frac{1}{2\pi n^2} \sum_{1 \leq k, l \leq n, k \neq l} \text{Cov} \left((-i)^j \int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du, (-i)^j \int_{\mathbb{R}} e^{iuZ_l} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right). \end{aligned}$$

As $\text{Var}(X) \leq \mathbb{E}|X|^2$, it comes

$$\begin{aligned} \mathbb{E} \left[\|\hat{f}_m - f\|^2 \right] &\leq \|f - f_m\|^2 + \frac{1}{2\pi n} \sum_{j=0}^{m-1} \mathbb{E} \left[\left| \int_{\mathbb{R}} e^{iuZ_1} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right|^2 \right] \\ &+ \frac{1}{2\pi n^2} \sum_{j=0}^{m-1} \sum_{1 \leq k, l \leq n, k \neq l} \text{Cov} \left((-i)^j \int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du, (-i)^j \int_{\mathbb{R}} e^{iuZ_l} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right). \end{aligned} \quad (7.5)$$

The first two right hand side terms are the same as in the independent case and are dealt with as in Proposition 3.1. We compute the covariance term. First,

$$\begin{aligned} \text{Cov} \left((-i)^j \int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du, (-i)^j \int_{\mathbb{R}} e^{iuZ_l} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right) \\ = \mathbb{E} \left[\int_{\mathbb{R}} \int_{\mathbb{R}} e^{i(uZ_k - vZ_l)} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} \frac{\varphi_j(v)}{f_\varepsilon^*(-v)} dudv \right] \\ - \mathbb{E} \left[\int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right] \mathbb{E} \left[\int_{\mathbb{R}} e^{-ivZ_l} \frac{\varphi_j(v)}{f_\varepsilon^*(-v)} dv \right]. \end{aligned} \quad (7.6)$$

The first expectation is equal to

$$\begin{aligned} \mathbb{E} \left[\int_{\mathbb{R}} \int_{\mathbb{R}} e^{i(uZ_k - vZ_l)} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} \frac{\varphi_j(v)}{f_\varepsilon^*(-v)} dudv \right] \\ = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{E} \left[e^{i(uX_k + u\varepsilon_k - vX_l - v\varepsilon_l)} \right] \frac{\varphi_j(u)}{f_\varepsilon^*(u)} \frac{\varphi_j(v)}{f_\varepsilon^*(-v)} dudv \\ = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{E} \left[e^{i(uX_k - vX_l)} \right] \varphi_j(u) \varphi_j(v) dudv, \end{aligned} \quad (7.7)$$

and the second to:

$$\mathbb{E} \left[\int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right] \mathbb{E} \left[\int_{\mathbb{R}} e^{-ivZ_l} \frac{\varphi_j(v)}{f_\varepsilon^*(-v)} dv \right] = \left| \int_{\mathbb{R}} f^*(u) \varphi_j(u) du \right|^2. \quad (7.8)$$

Thus, from (7.6), (7.7) and (7.8) we deduce

$$\begin{aligned} \text{Cov} \left((-i)^j \int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du, (-i)^j \int_{\mathbb{R}} e^{iuZ_l} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right) \\ = \text{Cov} \left(\int_{\mathbb{R}} e^{iuX_k} \varphi_j(u) du, \int_{\mathbb{R}} e^{iuX_l} \varphi_j(u) du \right). \end{aligned}$$

As a consequence

$$\sum_{1 \leq k, l \leq n, k \neq l} \text{Cov} \left(\int_{\mathbb{R}} e^{iuX_k} \varphi_j(u) du, \int_{\mathbb{R}} e^{iuX_l} \varphi_j(u) du \right) \leq \text{Var} \left(\sum_{k=1}^n \int_{\mathbb{R}} e^{iuX_k} \varphi_j(u) du \right).$$

Using [Viennet \(1997\)](#)'s covariance inequality and equality (2.1), we have

$$\text{Var} \left(\sum_{k=1}^n \int_{\mathbb{R}} e^{iuX_k} \varphi_j(u) du \right) = \text{Var} \left(\sum_{k=1}^n \varphi_j^*(X_k) \right) \leq 8\pi n \int_{\mathbb{R}} b(u) \varphi_j(u)^2 f(u) du, \quad (7.9)$$

with $b = \sum_{k=0}^n b_k$ and b_k , a sequence of measurable functions such that $b_0 = 1$, $\int b_k(u) f(u) du = \beta_k$ (see Theorem 2.1 in [Viennet, 1997](#) given here in Appendix).

Lemma 7.1. *Under the assumptions and notations of Proposition 3.4, there exists a constant $c^* > 0$ depending on $\mathbb{E}[|X_1|^{2q/3}]$ and $\sum_{k=0}^{+\infty} (k+1)^{p-1} \beta_k < +\infty$ such that:*

$$\int_{\mathbb{R}} b(x) \varphi_j^2(x) f(x) dx \leq \frac{c^*}{\sqrt{j}}, \quad \forall j \geq 1.$$

By Lemma 7.1 and (7.9), we deduce

$$\begin{aligned} & \sum_{j=0}^{m-1} \text{Var} \left(\sum_{k=1}^n \int_{\mathbb{R}} e^{iuX_k} \varphi_j(u) du \right) \\ & \leq 8\pi n \left[\int_{\mathbb{R}} b(u) \varphi_0^2(u) f(u) du + \sum_{j=1}^{m-1} \int_{\mathbb{R}} b(u) \varphi_j(u)^2 f(u) du \right] \end{aligned} \tag{7.10}$$

$$\leq 8\pi n \left[\phi_0^2 \sum_{k \geq 0} \beta_k + \sum_{j=1}^{m-1} \frac{c^*}{\sqrt{j}} \right]. \tag{7.11}$$

Using (7.11), Proposition 3.1 and in view of (7.5), we obtain the announced result \square .

7.3.1. *Proof of Lemma 7.1.* To prove this lemma, we first use the decomposition formula of the Hermite basis in the Laguerre basis (see Comte and Genon-Catalot, 2018, Lemma 8.4, p. 287) given by:

$$\varphi_{2k}(x) = (-1)^k \sqrt{x/2} \psi_k^{(-1/2)}(x^2/2), \quad \varphi_{2k+1}(x) = (-1)^k \sqrt{x/2} \psi_k^{(1/2)}(x^2/2), \quad x \geq 0$$

where $(\psi_k^{(\delta)})_{k \geq 0}$ is the Laguerre function with index $\delta > -1$ defined from the Laguerre polynomial $(L_k^{(\delta)})_{k \geq 0}$ with index $\delta > -1$ and degree k given by:

$$\psi_k^{(\delta)}(x) = 2^{\frac{\delta+1}{2}} \left(\frac{k!}{\Gamma(k+\delta+1)} \right)^{1/2} L_k^{(\delta)}(2x) x^{\frac{\delta}{2}} e^{-x}, \quad L_k^{(\delta)}(x) = \frac{1}{k!} e^x x^{-\delta} \frac{d^k}{dx^k} (x^{\delta+k} e^{-x}).$$

Note that $(\psi_k^{(\delta)})_{k \geq 0}$ is an orthonormal basis on $\mathbb{L}^2(\mathbb{R}^+)$. Next, using the asymptotic formula of Askey and Wainger (1965) recalled in Section A.2, we get a bound of $(\psi_k^{(\delta)})_{k \geq 0}$, for k large enough. We distinguish two cases depending on the parity of j and we study only the first term of the following decomposition:

$$\int_{\mathbb{R}} b(x) \varphi_j^2(x) f(x) dx = \int_0^\infty b(x) \varphi_j^2(x) f(x) dx + \int_0^\infty b(-x) \varphi_j^2(x) f(-x) dx,$$

since $(\varphi_j)_{j \geq 0}$ is even for j even and odd for j odd. The study of the other term is similar and its bound is the same as the one on the first term.

For j even, $j = 2k$, we have:

$$\int_0^\infty b(x) \varphi_j^2(x) f(x) dx = \frac{1}{2} \int_0^\infty x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 f(x) b(x) dx := \sum_{l=1}^6 J_l,$$

where J_l are integrals on disjoint domains specified below, see also Section A.2. Setting $\nu = 4k + 1$, we have six terms to evaluate.

$$J_1 = \frac{1}{2} \int_0^{1/\sqrt{\nu}} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 b(x) f(x) dx$$

$$\begin{aligned}
&\leq \frac{C}{2} \int_0^{1/\sqrt{\nu}} x \left[(x^2\nu)^{-1/4} \right]^2 b(x)f(x)dx \leq \frac{C}{2\sqrt{\nu}} \int_{\mathbb{R}} b(x)f(x)dx \leq \frac{C}{2\sqrt{\nu}} \sum_{k \geq 0} \beta_k. \\
J_2 &= \frac{1}{2} \int_{1/\sqrt{\nu}}^{\sqrt{\nu/2}} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 b(x)f(x)dx \\
&\leq \frac{C}{2} \int_{1/\sqrt{\nu}}^{\sqrt{\nu/2}} x \left[(x^2\nu)^{-1/4} \right]^2 b(x)f(x)dx \leq \frac{C}{2\sqrt{\nu}} \sum_{k \geq 0} \beta_k. \\
J_3 &= \frac{1}{2} \int_{\sqrt{\nu/2}}^{(\nu-\nu^{1/3})^{1/2}} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 b(x)f(x)dx \\
&\leq \frac{C}{2} \int_{\sqrt{\nu/2}}^{(\nu-\nu^{1/3})^{1/2}} x \left(\nu^{-1/4}(\nu-x^2)^{-1/4} \right)^2 b(x)f(x)dx \\
&= \frac{C}{2} \int_{\sqrt{\nu/2}}^{(\nu-\nu^{1/3})^{1/2}} x^{1/3} x^{2/3} \nu^{-1/2} (\nu-x^2)^{-1/2} b(x)f(x)dx \\
&\leq \frac{C}{2\sqrt{\nu}} \int_{\mathbb{R}} |x|^{2/3} b(x)f(x)dx.
\end{aligned}$$

Using the Hölder inequality, we have

$$\begin{aligned}
\int_{\mathbb{R}} |x|^{2/3} b(x)f(x)dx &\leq \left(\int_{\mathbb{R}} |x|^{2q/3} f(x)dx \right)^{1/q} \left(\int_{\mathbb{R}} b^p(x)f(x)dx \right)^{1/p} \\
&= \mathbb{E} \left[|X_1|^{2q/3} \right]^{1/q} \mathbb{E} [b(X_1)^p]^{1/p},
\end{aligned}$$

with $\frac{1}{p} + \frac{1}{q} = 1$. By Lemma 4.2 in [Viennet \(1997\)](#), page 481, we have:

$$\mathbb{E} [b(X_1)^p] \leq p \sum_{k \geq 0} (k+1)^{p-1} \beta_k.$$

It comes: $J_3 \leq \frac{C}{2\sqrt{\nu}} \mathbb{E} [|X_1|^{2q/3}]^{1/q} (p \sum_{k \geq 0} (k+1)^{p-1} \beta_k)^{1/p}$.

$$\begin{aligned}
J_4 &= \frac{1}{2} \int_{(\nu-\nu^{1/3})^{1/2}}^{(\nu+\nu^{1/3})^{1/2}} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 b(x)f(x)dx \\
&\leq \frac{C}{2} \int_{(\nu-\nu^{1/3})^{1/2}}^{(\nu+\nu^{1/3})^{1/2}} x (\nu^{-1/3})^2 b(x)f(x)dx \\
&\leq \frac{C}{2} \int_{(\nu-\nu^{1/3})^{1/2}}^{(\nu+\nu^{1/3})^{1/2}} x^{1/3} x^{2/3} \nu^{-2/3} b(x)f(x)dx \leq \frac{C}{\sqrt{\nu}} \int_{\mathbb{R}} |x|^{2/3} b(x)f(x)dx.
\end{aligned}$$

By the same computation as for J_3 we deduce: $J_4 \leq \frac{C}{\sqrt{\nu}} \mathbb{E} [|X_1|^{2q/3}]^{1/q} (p \sum_{k \geq 0} (k+1)^{p-1} \beta_k)^{1/p}$.

$$\begin{aligned}
J_5 &= \frac{1}{2} \int_{(\nu+\nu^{1/3})^{1/2}}^{\sqrt{3\nu/2}} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 b(x)f(x)dx \\
&\leq \frac{C}{2} \int_{(\nu+\nu^{1/3})^{1/2}}^{\sqrt{3\nu/2}} x^{1/3} x^{2/3} \left(\nu^{-1/4} (x^2-\nu)^{-1/4} e^{-\gamma_1 \nu^{-1/2} (x^2-\nu)^{3/2}} \right)^2 b(x)f(x)dx
\end{aligned}$$

$$\begin{aligned} &\leq \frac{C}{2} \int_{(\nu+\nu^{1/3})^{1/2}}^{\sqrt{3\nu/2}} \nu^{-1/2} x^{1/3} (x^2 - \nu)^{-1/2} e^{-2\gamma_1 \nu^{-1/2} (x^2 - \nu)^{3/2}} x^{2/3} b(x) f(x) dx \\ &\leq \frac{C}{\sqrt{\nu}} \int_{\mathbb{R}} |x|^{2/3} b(x) f(x) dx. \end{aligned}$$

Again by the Hölder inequality we get:

$$J_5 \leq \frac{C}{\sqrt{\nu}} \mathbb{E} \left[|X_1|^{2q/3} \right]^{1/q} \left(p \sum_{k \geq 0} (k+1)^{p-1} \beta_k \right)^{1/p}.$$

Finally, it holds

$$\begin{aligned} J_6 &= \frac{1}{2} \int_{\sqrt{3\nu/2}}^{\infty} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 b(x) f(x) dx \leq \frac{C}{2} \int_{\sqrt{3\nu/2}}^{\infty} x e^{-\gamma_2 x^2} b(x) f(x) dx \\ &\leq C' e^{-3\frac{\gamma_2 \nu}{4}} \int_{\mathbb{R}} b(x) f(x) dx = C' e^{-3\frac{\gamma_2 \nu}{4}} \mathbb{E} [b(X_1)] \leq C' e^{-3\frac{\gamma_2 \nu}{4}} \sum_{k \geq 0} \beta_k. \end{aligned}$$

For j odd, $j = 2k + 1$, and setting $\nu = 4k + 3$, we have:

$$\int_0^{\infty} b(x) \varphi_{2k+1}^2(x) f(x) dx = \frac{1}{2} \int_0^{\infty} x \left(\psi_k^{(1/2)}(x^2/2) \right)^2 f(x) b(x) dx := \sum_{l=1}^6 K_l.$$

Only the first term changes, thus, we just compute K_1 and the other terms are such that the bounds coincide with the case where j is even for $l = 2, \dots, 6$.

$$\begin{aligned} K_1 &= \frac{1}{2} \int_0^{1/\sqrt{\nu}} x \left(\psi_k^{(1/2)}(x^2/2) \right)^2 b(x) f(x) dx \leq \frac{C}{2} \int_0^{1/\sqrt{\nu}} x \left[(x^2 \nu)^{1/4} \right]^2 b(x) f(x) dx \\ &\leq \frac{C}{2\sqrt{\nu}} \sum_{k \geq 0} \beta_k \end{aligned}$$

By gathering all these inequalities according to the parity of j , we have the announced result.

7.4. *Proof of Theorem 4.1.* By definition of \hat{m} , we have: $\gamma_n(\hat{f}_{\hat{m}}) + \text{pen}(\hat{m}) \leq \gamma_n(f_m) + \text{pen}(m)$. Moreover, for two functions s, t in $\mathbb{L}^2(\mathbb{R})$, $\gamma_n(t) - \gamma_n(s) = \|t - f\|^2 - \|s - f\|^2 - 2\nu_n(t - s)$, where

$$\nu_n(t) = \frac{1}{n} \sum_{k=1}^n (\phi_t(Z_k) - \langle t, f \rangle),$$

where ϕ_t is defined in (4.1). Thus, for m any element of \mathcal{M}_n , we have

$$\|\hat{f}_{\hat{m}} - f\|^2 \leq \|f_m - f\|^2 + \text{pen}(m) + 2\nu_n(\hat{f}_{\hat{m}} - f_m) - \text{pen}(\hat{m})$$

As the function $t \mapsto \nu_n(t)$ is linear, we deduce

$$\begin{aligned} \|\hat{f}_{\hat{m}} - f\|^2 &\leq \|f_m - f\|^2 + \text{pen}(m) + 2\|\hat{f}_{\hat{m}} - f_m\| \nu_n \left(\frac{\hat{f}_{\hat{m}} - f_m}{\|\hat{f}_{\hat{m}} - f_m\|} \right) - \text{pen}(\hat{m}) \\ &\leq \|f_m - f\|^2 + \text{pen}(m) + 2\|\hat{f}_{\hat{m}} - f_m\| \sup_{t \in S_m + S_{\hat{m}}, \|t\|=1} \nu_n(t) - \text{pen}(\hat{m}). \end{aligned} \tag{7.12}$$

For all $x, y \geq 0$ we have: $2xy \leq x^2/4 + 4y^2$, therefore, we obtain

$$2\|\hat{f}_{\hat{m}} - f_m\| \sup_{t \in S_m + S_{\hat{m}}, \|t\|=1} \nu_n(t) \leq \frac{1}{4}\|\hat{f}_{\hat{m}} - f_m\|^2 + 4 \sup_{t \in S_m + S_{\hat{m}}, \|t\|=1} (\nu_n(t))^2. \quad (7.13)$$

Now, $\|\hat{f}_{\hat{m}} - f_m\|^2 \leq 2\|\hat{f}_{\hat{m}} - f\|^2 + 2\|f_m - f\|^2$ and plugging this and (7.13) in (7.12), we have

$$\frac{1}{2}\|\hat{f}_{\hat{m}} - f\|^2 \leq \frac{3}{2}\|f_m - f\|^2 + \text{pen}(m) + 4 \sup_{t \in S_m + S_{\hat{m}}, \|t\|=1} (\nu_n(t))^2 - \text{pen}(\hat{m}). \quad (7.14)$$

We decompose the empirical process $\nu_n(t)$ in two processes. We set $m^* = \hat{m} \vee m$. For $t \in S_{m^*}$, we have using Plancherel-Parseval

$$\begin{aligned} \nu_n(t) &= \frac{1}{n} \sum_{k=1}^n (\phi_t(Z_k) - \langle t, f \rangle) \\ &= \frac{1}{n} \sum_{k=1}^n \left(\frac{1}{2\pi} \int_{|u| \leq \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuZ_k} du - \mathbb{E} \left[\frac{1}{2\pi} \int_{|u| \leq \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuZ_k} du \right] \right) \\ &\quad + \frac{1}{n} \sum_{k=1}^n \left(\frac{1}{2\pi} \int_{|u| > \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuZ_k} du - \mathbb{E} \left[\frac{1}{2\pi} \int_{|u| > \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuZ_k} du \right] \right) \\ &= \frac{1}{n} \sum_{k=1}^n (\phi_{t,1}(Z_k) - \mathbb{E}[\phi_{t,1}(Z_k)]) + \frac{1}{2\pi} \int_{|u| > \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} (\hat{f}_Z^*(u) - f_Z^*(u)) du, \end{aligned} \quad (7.15)$$

with $\phi_{t,1}(x) = \frac{1}{2\pi} \int_{|u| \leq \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iux} du$. Therefore, we write $\nu_n(t) = \nu_{n,1}(t) + \nu_{n,2}(t)$ where $\nu_{n,1}(t) = \frac{1}{n} \sum_{k=1}^n (\phi_{t,1}(Z_k) - \mathbb{E}[\phi_{t,1}(Z_k)])$ and $\nu_{n,2}(t) = \frac{1}{2\pi} \int_{|u| > \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} (\hat{f}_Z^*(-u) - f_Z^*(-u)) du$. Using that $(\nu_{n,1}(t) + \nu_{n,2}(t))^2 \leq 2(\nu_{n,1}(t))^2 + 2(\nu_{n,2}(t))^2$ and by (7.14), (7.15) we deduce

$$\begin{aligned} \frac{1}{2}\|\hat{f}_{\hat{m}} - f\|^2 &\leq \frac{3}{2}\|f_m - f\|^2 + \text{pen}(m) + 8 \sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,1}(t))^2 \\ &\quad + 8 \sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2 - \text{pen}(\hat{m}). \end{aligned}$$

We introduce the function $p(m, m') = \frac{\kappa}{8} \frac{\Delta(m \vee m')}{n}$ if f_ε is ordinary smooth or super smooth with $\delta \leq 1/2$ and $p(m, m') = 2\kappa(1 + \varepsilon(m, m')) \frac{\Delta(m \vee m')}{8n}$ otherwise, where $\varepsilon(m, m')$ is given below, which verifies $8p(m, m') \leq \text{pen}(m) + \text{pen}(m')$. We obtain:

$$\begin{aligned} \|\hat{f}_{\hat{m}} - f\|^2 &\leq 3\|f_m - f\|^2 + 4\text{pen}(m) + 16 \sum_{m' \in \mathcal{M}_n} \left(\sup_{t \in S_{m \vee m'}, \|t\|=1} (\nu_{n,1}(t))^2 - p(m, m') \right)_+ \\ &\quad + 16 \sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2. \end{aligned}$$

By taking expectation, we get

$$\begin{aligned} \mathbb{E} \left[\|\hat{f}_{\hat{m}} - f\|^2 \right] &\leq 3\|f_m - f\|^2 + 4\text{pen}(m) + 16\mathbb{E} \left[\sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2 \right] \\ &\quad + 16 \sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in S_{m \vee m'}, \|t\|=1} (\nu_{n,1}(t))^2 - p(m, m') \right)_+ \right]. \end{aligned}$$

The two followings lemmas lead to the result of Theorem 4.1:

Lemma 7.2. *There exists a constant Σ_1 such that*

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in S_{m \vee m'}, \|t\|=1} (\nu_{n,1}(t))^2 - p(m, m') \right)_+ \right] \leq \frac{\Sigma_1}{n}.$$

Lemma 7.3. *There exists a constant Σ_2 such that*

$$\mathbb{E} \left[\sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2 \right] \leq \frac{\Sigma_2}{n}.$$

Using Lemmas 7.2 and 7.3, we have the result choosing $C = 4$ and $C' = 16(\Sigma_1 + \Sigma_2)$. \square

7.4.1. *Proof of Lemma 7.2.* To prove this lemma, we use Talagrand’s inequality given in Appendix A.3, and compute H^2 , M_1 , v defined there. Denote by $m'' = m \vee m'$. We start by computing H^2 . As the map $t \mapsto \nu_{n,1}(t)$ is linear, for $t = \sum_{j=0}^{m''-1} a_j \varphi_j$ such that $\|t\| = 1$, we have

$$(\nu_{n,1}(t))^2 = \left(\sum_{j=0}^{m''-1} a_j \nu_{n,1}(\varphi_j) \right)^2 \leq \sum_{j=0}^{m''-1} a_j^2 \sum_{j=0}^{m''-1} \nu_{n,1}(\varphi_j)^2 = \sum_{j=0}^{m''-1} \nu_{n,1}(\varphi_j)^2.$$

Therefore,

$$\begin{aligned} \mathbb{E} \left[\left(\sup_{t \in S_{m''}, \|t\|=1} (\nu_{n,1}(t))^2 \right) \right] &\leq \mathbb{E} \left[\sum_{j=0}^{m''-1} \nu_{n,1}(\varphi_j)^2 \right] = \sum_{j=0}^{m''-1} \frac{1}{n} \text{Var}(\phi_{\varphi_j,1}(Z_1)) \\ &\leq \frac{1}{n} \sum_{j=0}^{m''-1} \mathbb{E} [|\phi_{\varphi_j,1}(Z_1)|^2]. \end{aligned}$$

It comes using (7.2) that,

$$\mathbb{E} \left[\sum_{j=0}^{m''-1} |\phi_{\varphi_j}(Z_1)|^2 \right] = \frac{1}{(2\pi)^2} \mathbb{E} \left[\sum_{j=0}^{m''-1} \left| \int_{|u| \leq \sqrt{lm''}} \frac{\varphi_j^*(u) e^{-iuZ_1}}{f_\varepsilon^*(-u)} du \right|^2 \right] \leq \frac{\Delta(m'')}{n} := H^2. \tag{7.16}$$

Now we look for M_1 . Using Cauchy-Schwarz inequality and Parseval’s theorem

$$\begin{aligned} |\phi_{t,1}(x)| &= \frac{1}{2\pi} \left| \int_{|u| \leq \sqrt{lm''}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iux} du \right| \leq \frac{1}{2\pi} \int_{|u| \leq \sqrt{lm''}} \left| \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iux} \right| du \\ &\leq \frac{1}{2\pi} \sqrt{\int |t^*(u)|^2 du \int_{|u| \leq \sqrt{lm''}} \frac{du}{|f_\varepsilon^*(-u)|^2}} \\ &= \frac{1}{2\pi} \sqrt{2\pi \|t\|^2 \int_{|u| \leq \sqrt{lm''}} \frac{du}{|f_\varepsilon^*(-u)|^2}} \leq \sqrt{\Delta(m'')}. \end{aligned}$$

Thus, it follows

$$\sup_{t \in S_m + S_{m'}, \|t\|=1} \|\phi_{t,1}\|_\infty \leq \sqrt{\Delta(m'')} := M_1.$$

The case of v is more tedious,

$$\begin{aligned} \text{Var}(\phi_{t,1}(Z_1)) &\leq \mathbb{E} \left[|\phi_{t,1}(Z_1)|^2 \right] = \frac{1}{2\pi} \int \left| \int_{|u| \leq \sqrt{lm''}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuz} du \right|^2 f_Z(z) dz \\ &= \frac{1}{2\pi} \iiint \frac{t^*(u)}{f_\varepsilon^*(-u)} \frac{t^*(-v)}{f_\varepsilon^*(v)} e^{-i(u-v)z} f_Z(z) \mathbb{1}_{|u| \leq \sqrt{lm''}} \mathbb{1}_{|v| \leq \sqrt{lm''}} dudvdz \\ &= \frac{1}{2\pi} \iint \frac{t^*(u)}{f_\varepsilon^*(-u)} \frac{t^*(-v)}{f_\varepsilon^*(v)} f_Z^*(v-u) \mathbb{1}_{|u| \leq \sqrt{lm''}} \mathbb{1}_{|v| \leq \sqrt{lm''}} dudv \\ &\leq \frac{1}{2\pi} \iint \left| \frac{t^*(u)}{f_\varepsilon^*(-u)} \right|^2 |f_Z^*(v-u)| \mathbb{1}_{|u| \leq \sqrt{lm''}} \mathbb{1}_{|v| \leq \sqrt{lm''}} dudv \\ &\leq \frac{1}{2\pi} \int |f_Z^*(z)| dz \int \left| \frac{t^*(u)}{f_\varepsilon^*(-u)} \right|^2 \mathbb{1}_{|u| \leq \sqrt{lm''}} du. \end{aligned}$$

Using the Cauchy-Schwarz inequality and Parseval's theorem we have:

$$\int |f_Z^*(z)| dz = \int |f^*(z) f_\varepsilon^*(z)| dz \leq 2\pi \|f_\varepsilon\| \cdot \|f\|.$$

Thus, we get: $\text{Var}(\phi_{t,1}(Z_1)) \lesssim \int \left| \frac{t^*(u)}{f_\varepsilon^*(-u)} \right|^2 \mathbb{1}_{|u| \leq \sqrt{lm''}} du$. We consider separately two cases.

(1) **Ordinary smooth case:** In this case, we have by (7.16) and by (7.4) that $H^2 \asymp \frac{m''^{\gamma+1/2}}{n}$. Moreover,

$$\begin{aligned} \text{Var}(\phi_{t,1}(Z_1)) &\leq \int |t^*(u)|^2 (1+t^2)^\gamma \mathbb{1}_{|u| \leq \sqrt{lm''}} du \leq (1+l^\gamma m''^\gamma) \int |t^*(u)|^2 du \\ &= 2\pi(1+l^\gamma m''^\gamma) \|t\|^2 = 2\pi(1+l^\gamma m''^\gamma). \end{aligned}$$

We can set $v = cm''^\gamma$, with $c > 0$. Thus, using Talagrand's inequality we have:

$$\mathbb{E} \left[\left(\sup_{t \in S_{m''}, \|t\|=1} (\nu_{n,1}(t))^2 - p(m, m') \right)_+ \right] \lesssim [U(m'') + V(m'')], \tag{7.17}$$

with $p(m, m') = \frac{\kappa}{8} \frac{\Delta(m'')}{n} = \frac{\kappa}{8} H^2 \geq 2(1+2\varepsilon)H^2$, we take $\kappa_0 = 17$, $\varepsilon = 1/2$, and

$$U(m'') = \frac{v}{n} \exp \left(-\frac{K_1}{2} \frac{nH^2}{v} \right) = \frac{cm''^\gamma}{n} \exp \left(-\frac{K_1}{2} n \frac{m''^{\gamma+\frac{1}{2}}}{cm''^\gamma} \right) \lesssim \frac{m''^\gamma}{n} e^{-\frac{\kappa_1}{2c} m''^{\frac{1}{2}}},$$

$$\begin{aligned} V(m'') &= \frac{M_1^2}{C(\varepsilon)^2 n^2} \exp \left(-K_1' C(\varepsilon) \frac{1}{\sqrt{2}} \frac{nH}{M_1} \right) = C_1 \frac{\Delta(m'')}{n^2} \exp \left(-C_2 n \frac{\sqrt{\frac{\Delta(m'')}{n}}}{\sqrt{\Delta(m'')}} \right) \\ &\lesssim \frac{1}{n} e^{-C_2 \sqrt{n}}, \end{aligned}$$

because for $m \in \mathcal{M}_n$, $\Delta(m) \leq n$. Therefore, we deduce by (7.17) that:

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in S_{m''}, \|t\|=1} (\nu_{n,1}(t))^2 - p(m, m') \right)_+ \right] \lesssim \sum_{m' \in \mathcal{M}_n} [U(m'') + V(m'')].$$

As

$$\begin{aligned} \sum_{m'} U(m'') &\lesssim \frac{1}{n} \sum_{m'} m''^\gamma e^{-\frac{K_1}{2c} \sqrt{m''}} \\ &= \frac{1}{n} \left[\sum_{m'=0}^m m''^\gamma e^{-\frac{K_1}{2c} \sqrt{m''}} + \sum_{m'=m}^{n^2} m''^\gamma e^{-\frac{K_1}{2c} \sqrt{m''}} \right] \\ &= \frac{1}{n} \left[m^{\gamma+1} e^{-\frac{K_1}{2c} \sqrt{m}} + \sum_{m'=m}^{+\infty} m''^\gamma e^{-\frac{K_1}{2c} \sqrt{m''}} \right] \leq \frac{C'_1}{n}, \end{aligned}$$

and

$$\sum_{m' \in \mathcal{M}_n} V(m'') \lesssim \frac{1}{n} \sum_{m' \in \mathcal{M}_n} e^{-C_2 \sqrt{n}} = \frac{1}{n} |\mathcal{M}_n| e^{-C_2 \sqrt{n}} \lesssim n e^{-C_2 \sqrt{n}} \leq \frac{C''_1}{n}.$$

We deduce that

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in S_{m''}, \|t\|=1} (\nu_{n,1}(t))^2 - p(m, m') \right)_+ \right] \leq \frac{\Sigma_1}{n}, \quad \Sigma_1 = C'_1 + C''_1. \quad (7.18)$$

(2) **Super smooth case:** In this case the order of H^2 is given by (7.4): $H^2 \asymp \frac{m''^{\frac{1-\delta}{2}} e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}}}{n}$,

$$\begin{aligned} \text{Var}(\phi_{t,1}(Z_1)) &\leq c_1 \int |t^*(u)|^2 e^{\mu|u|^\delta} \mathbf{1}_{|u| \leq \sqrt{lm''}} du \leq c_1 e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}} \int |t^*(u)|^2 du \\ &= 2\pi c_1 e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}} \|t\|^2 \lesssim e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}} = v. \end{aligned}$$

We use Talagrand's inequality again, we must compute $U(m'')$ and $V(m'')$.

$$\begin{aligned} U(m'') &= \frac{v}{n} \exp\left(-K_1 \varepsilon \frac{nH^2}{v}\right) = \frac{ce^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}}}{n} \exp\left(-K_1 \varepsilon n \frac{m''^{\frac{1-\delta}{2}} e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}}}{e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}}}\right) \\ &\lesssim \frac{1}{n} e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}}}, \end{aligned}$$

$$\begin{aligned} V(m'') &= \frac{M_1^2}{C^2(\varepsilon)n^2} \exp\left(-K'_1 C(\varepsilon) \sqrt{\varepsilon} \frac{nH}{M_1}\right) = \frac{\Delta(m'')}{C^2(\varepsilon)n^2} \exp\left(-K'_1 C(\varepsilon) \sqrt{\varepsilon} \sqrt{n}\right) \\ &\leq \frac{1}{C^2(\varepsilon)n} \exp\left(-K'_1 C(\varepsilon) \sqrt{\varepsilon} \sqrt{n}\right). \end{aligned}$$

• **Study of $\sum_{m' \in \mathcal{M}_n} U(m'')$:** we have

$$\sum_{m' \in \mathcal{M}_n} U(m'') \lesssim \frac{1}{n} \sum_{m' \in \mathcal{M}_n} e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}}}.$$

We are going to study this term according the value of δ .

(i) **Case $0 < \delta < 1/2$:** In this case $\delta/2 < (1-\delta)/2$. Thus the choice $\varepsilon = 1$ implies that $m e^{\mu \frac{\delta}{2} m^{\frac{\delta}{2}} - K_1 \varepsilon m^{\frac{1-\delta}{2}}}$ is bounded by a constant independent of m' , and $e^{\mu \frac{\delta}{2} m'^{\frac{\delta}{2}} - K_1 \varepsilon m'^{\frac{1-\delta}{2}}}$ is integrable in m' . We deduce that:

$$\frac{1}{n} \sum_{m' \in \mathcal{M}_n} e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}}} =$$

$$\begin{aligned}
&= \frac{1}{n} \left[\sum_{m'=1}^m e^{\mu l^{\delta/2} m'^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}}} + \sum_{m'=m}^{n^2} e^{\mu l^{\delta/2} m'^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}}} \right] \\
&\leq \frac{1}{n} \left[m e^{\mu l^{\delta/2} m^{\frac{\delta}{2}} - K_1 \varepsilon m^{\frac{1-\delta}{2}}} + \sum_{m' \in \mathcal{M}_n} e^{\mu l^{\delta/2} m'^{\frac{\delta}{2}} - K_1 \varepsilon m'^{\frac{1-\delta}{2}}} \right] \\
&\leq \frac{C_1''}{n}.
\end{aligned}$$

(ii) **Case** $\delta \geq 1/2$: We choose ε such that $\mu l^{\delta/2} m''^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}} = -\mu l^{\frac{\delta}{2}} m''^{\frac{\delta}{2}}$, that is $\varepsilon = \frac{2\mu l^{\delta/2}}{K_1} m''^{\delta - \frac{1}{2}}$. This implies

$$\frac{1}{n} \sum_{m' \in \mathcal{M}_n} e^{\mu l^{\delta/2} m'^{\frac{\delta}{2}} - K_1 \varepsilon m'^{\frac{1-\delta}{2}}} = \frac{1}{n} \sum_{m' \in \mathcal{M}_n} e^{-\mu l^{\delta/2} m'^{\frac{\delta}{2}}} \leq \frac{1}{n} \sum_{m'} e^{-\mu l^{\delta/2} m'^{\frac{\delta}{2}}} \leq \frac{C_1''}{n}.$$

In the all cases, we have : $\sum_{m' \in \mathcal{M}_n} U(m'') \leq \frac{C_1''}{n}$.

• **Study of** $\sum_{m' \in \mathcal{M}_n} V(m'')$

As $|\mathcal{M}_n| = \mathcal{O}(n^2)$ and for all choice of ε in the study of $U(m'')$, we have $C(\varepsilon) = 1$, $\varepsilon \geq 1$. Thus, it follows

$$\sum_{m' \in \mathcal{M}_n} V(m'') \leq \frac{|\mathcal{M}_n|}{C^2(\varepsilon)n} \exp(-K_1' C(\varepsilon) \varepsilon \sqrt{n}) \leq \frac{n}{C^2(\varepsilon)} \exp(-K_1' C(\varepsilon) \sqrt{\varepsilon} \sqrt{n}) \leq \frac{C_1'}{n}.$$

Therefore, (7.18) holds and the result of Lemma 7.2 is proven. \square

7.4.2. Proof of Lemma 7.3. Here $m^* = m \vee \hat{m}$. Using the Cauchy-Schwarz inequality for $t = \sum_{j=0}^{m^*-1} a_j \varphi_j$ such that $\|t\|^2 = \sum_{j=0}^{m^*-1} a_j^2 = 1$, we have:

$$\begin{aligned}
\nu_{n,2}(t)^2 &= \frac{1}{(2\pi)^2} \left(\int_{|u| > \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} (\hat{f}_Z^*(-u) - f_Z^*(-u)) du \right)^2 \\
&\leq \frac{1}{(2\pi)^2} \left(\sum_{j=0}^{m^*-1} \left| \int_{|u| > \sqrt{lm^*}} \frac{\varphi_j^*(u)}{f_\varepsilon^*(-u)} (\hat{f}_Z^*(-u) - f_Z^*(-u)) du \right|^2 \right).
\end{aligned}$$

By (2.1)-(2.2) and using the Cauchy-Schwarz inequality, we have:

$$\begin{aligned}
&\sum_{j=0}^{m^*-1} \left| \int_{|u| > \sqrt{lm^*}} \frac{\varphi_j^*(u)}{f_\varepsilon^*(-u)} (\hat{f}_Z^*(u) - f_Z^*(u)) du \right|^2 \\
&= 2\pi \sum_{j=0}^{m^*-1} \left| \int_{|u| > \sqrt{lm^*}} \frac{\varphi_j(u)}{f_\varepsilon^*(-u)} (\hat{f}_Z^*(-u) - f_Z^*(-u)) du \right|^2 \\
&\lesssim \sum_{j=0}^{m^*-1} \left(\int_{|u| > \sqrt{lm^*}} \frac{|\hat{f}_Z^*(-u) - f_Z^*(-u)|}{|f_\varepsilon^*(-u)|} |\varphi_j(u)| du \right)^2 \\
&\lesssim \sum_{j=0}^{m^*-1} \left(\int_{|u| > \sqrt{lm^*}} \frac{|\hat{f}_Z^*(-u) - f_Z^*(-u)|}{|f_\varepsilon^*(-u)|} e^{-\xi u^2} du \right)^2
\end{aligned}$$

$$\begin{aligned} &\lesssim \sum_{j=0}^{m^*-1} \left(\int_{|u|>\sqrt{lm^*}} \frac{|\hat{f}_Z^*(-u) - f_Z^*(-u)|^2}{|f_\varepsilon^*(-u)|^2} e^{-\xi u^2} du \right) \\ &\quad \times \int_{|u|>\sqrt{lm^*}} e^{-\xi u^2} du. \end{aligned}$$

As $\int_{|u|>\sqrt{lm^*}} e^{-\xi u^2} du \leq ce^{-\xi m^*}$ and the function $x \mapsto xe^{-\xi x}$ reaches its maximum $(1/\xi)e^{-1}$ in $x = 1/\xi$, it implies $\nu_{n,2}(t)^2 \lesssim \int_{\mathbb{R}} \frac{|\hat{f}_Z^*(-u) - f_Z^*(-u)|^2}{|f_\varepsilon^*(-u)|^2} e^{-\xi u^2} du$. Therefore,

$$\mathbb{E} \left[\sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2 \right] \lesssim \int_{\mathbb{R}} \frac{\mathbb{E} \left[|\hat{f}_Z^*(-u) - f_Z^*(-u)|^2 \right]}{|f_\varepsilon^*(-u)|^2} e^{-\xi u^2} du.$$

Now, we have

$$\mathbb{E} \left[|\hat{f}_Z^*(-u) - f_Z^*(-u)|^2 \right] = \text{Var}[\hat{f}_Z^*(-u)] = \frac{1}{n} \text{Var}[e^{-iuZ_1}] = \frac{1}{n} (1 - |f_Z^*(-u)|^2) \leq \frac{1}{n}.$$

Thus, by this last inequality we deduce

$$\mathbb{E} \left[\sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2 \right] \lesssim \frac{1}{n} \int_{\mathbb{R}} \frac{1}{|f_\varepsilon^*(-u)|^2} e^{-\xi u^2} du.$$

If f_ε is ordinary smooth, the integral is convergent and the previous bound is of order $1/n$. Assume now f_ε super smooth, we have by (2.3):

$$\mathbb{E} \left[\sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2 \right] \lesssim \frac{1}{n} \int_{\mathbb{R}} e^{\mu|u|^\delta} e^{-\xi u^2} du \leq \frac{\Sigma_2}{n},$$

if $\delta < 2$, or if $\delta = 2$, and $\mu < \xi$. This gives the announced result.

Appendix A. Some inequalities

A.1. *Covariance inequality (Viennet, 1997)*. Let $(X_i)_{i \in \mathbb{Z}}$ be a strictly stationary absolutely process with β -missing sequence $(\beta_k)_{k \geq 0}$. Then, there exists a sequence of measurable functions $(b_k)_{k \geq 0}$, with $b_0 \equiv 1$, $0 \leq b_k \leq 1$, $\mathbb{E}_P[b_k] = \beta_k$ such that for any measurable function f in $\mathbb{L}^2(P)$ and any positive integer n , we have

$$\text{Var} \left(\sum_{i=1}^n f(X_i) \right) \leq 4n \int b(x) f^2(x) dP(x),$$

where $b = \sum_{k=0}^n b_k$ is such that $\mathbb{E}_P(b^p) \leq p \sum_{k \geq 0} (k+1)^{p-1} \beta_k$, for $1 \leq p < +\infty$ (see Lemma 4.2 in Viennet, 1997 p. 481).

A.2. *Asymptotic Askey and Wainger formula*. From Askey and Wainger (1965), we have for $\nu = 4k + 2\delta + 2$, and k large enough

$$|\psi_k^{(\delta)}(x/2)| \leq C \begin{cases} a) (x\nu)^{\delta/2} & \text{if } 0 \leq x \leq 1/\nu \\ b) (x\nu)^{-1/4} & \text{if } 1/\nu \leq x \leq \nu/2 \\ c) \nu^{-1/4}(\nu-x)^{-1/4} & \text{if } \nu/2 \leq x \leq \nu - \nu^{1/3} \\ d) \nu^{-1/3} & \text{if } \nu - \nu^{1/3} \leq x \leq \nu + \nu^{1/3} \\ e) \nu^{-1/4}(x-\nu)^{-1/4} e^{-\gamma_1 \nu^{-1/2}(x-\nu)^{3/2}} & \text{if } \nu + \nu^{1/3} \leq x \leq 3\nu/2 \\ f) e^{-\gamma_2 x} & \text{if } x \geq 3\nu/2 \end{cases}$$

where γ_1 and γ_2 are positive and fixed constants.

A.3. *Talagrand's inequality.* Let $(X_i)_{1 \leq i \leq n}$ be independent real random variables, \mathcal{F} a class at most countable of measurable functions and $\nu_n(f) = \frac{1}{n} \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)])$ for all $f \in \mathcal{F}$. We assume there exist three strictly positive constants M_1 , H , v such that:

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M_1, \quad \mathbb{E}[\sup_{f \in \mathcal{F}} |\nu_n(f)|] \leq H, \quad \text{and} \quad \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \text{Var}(f(X_i)) \leq v.$$

Then, for $\varepsilon > 0$,

$$\begin{aligned} & \mathbb{E} \left[\left(\sup_{f \in \mathcal{F}} |\nu_n^2(f)| - 2(1 + 2\varepsilon)H^2 \right)_+ \right] \\ & \leq \frac{4}{K_1} \left(\frac{v}{n} \exp(-K_1 \varepsilon \frac{nH^2}{v}) + \frac{49M_1^2}{K_1 C^2(\varepsilon) n^2} \exp(-K_1' C(\varepsilon) \sqrt{\varepsilon} \frac{nH}{M_1}) \right) \end{aligned}$$

where $C(\varepsilon) = (\sqrt{1 + \varepsilon} - 1) \wedge 1$, $K_1 = 1/6$ and K_1' a universal constant. The Talagrand inequalities has been proven in Talagrand (1996), reworded by Ledoux (1995/97). This version is given in Klein and Rio (2005).

Acknowledgements

The author thanks F. Comte and C. Duval for helpful advices and kind proof-readings.

References

- Abramowitz, M. and Stegun, I. A. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. U.S. Government Printing Office, Washington, D.C. (1964). [MR0167642](#).
- Askey, R. and Wainger, S. Mean convergence of expansions in Laguerre and Hermite series. *Amer. J. Math.*, **87**, 695–708 (1965). [MR182834](#).
- Belomestny, D., Comte, F., and Genon-Catalot, V. Sobolev-Hermite versus Sobolev nonparametric density estimation on \mathbb{R} . *Ann. Inst. Statist. Math.*, **71** (1), 29–62 (2019). [MR3898425](#).
- Bongioanni, B. and Torrea, J. L. Sobolev spaces associated to the harmonic oscillator. *Proc. Indian Acad. Sci. Math. Sci.*, **116** (3), 337–360 (2006). [MR2256010](#).
- Butucea, C. Deconvolution of supersmooth densities with smooth noise. *Canad. J. Statist.*, **32** (2), 181–192 (2004). [MR2064400](#).
- Butucea, C. and Tsybakov, A. B. Sharp optimality in density deconvolution with dominating bias. II. *Teor. Veroyatn. Primen.*, **52** (2), 336–349 (2007). [MR2742504](#).
- Carroll, R. J. and Hall, P. Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, **83** (404), 1184–1186 (1988). [MR997599](#).
- Comte, F., Dedecker, J., and Taupin, M. L. Adaptive density deconvolution with dependent inputs. *Math. Methods Statist.*, **17** (2), 87–112 (2008). [MR2429122](#).
- Comte, F. and Genon-Catalot, V. Laguerre and Hermite bases for inverse problems. *J. Korean Statist. Soc.*, **47** (3), 273–296 (2018). [MR3840862](#).

- Comte, F. and Lacour, C. Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **73** (4), 601–627 (2011). [MR2853732](#).
- Fan, J. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, **19** (3), 1257–1272 (1991). [MR1126324](#).
- Fan, J. Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.*, **21** (2), 600–610 (1993). [MR1232507](#).
- Indritz, J. An inequality for Hermite polynomials. *Proc. Amer. Math. Soc.*, **12**, 981–983 (1961). [MR132852](#).
- Kappus, J. and Mabon, G. Adaptive density estimation in deconvolution problems with unknown error distribution. *Electron. J. Stat.*, **8** (2), 2879–2904 (2014). [MR3299125](#).
- Klein, T. and Rio, E. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, **33** (3), 1060–1077 (2005). [MR2135312](#).
- Ledoux, M. On Talagrand’s deviation inequalities for product measures. *ESAIM Probab. Statist.*, **1**, 63–87 (1995/97). [MR1399224](#).
- Lukacs, E. *Characteristic functions*. Hafner Publishing Co., New York (1970). [MR0346874](#).
- Mabon, G. Adaptive deconvolution on the non-negative real line. *Scand. J. Stat.*, **44** (3), 707–740 (2017). [MR3687970](#).
- Meister, A. *Deconvolution problems in nonparametric statistics*, volume 193 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin (2009). ISBN 978-3-540-87556-7. [MR2768576](#).
- Pensky, M. and Vidakovic, B. Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, **27** (6), 2033–2053 (1999). [MR1765627](#).
- Talagrand, M. New concentration inequalities in product spaces. *Invent. Math.*, **126** (3), 505–563 (1996). [MR1419006](#).
- Viennet, G. Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*, **107** (4), 467–492 (1997). [MR1440142](#).