CrossMark

# Sobolev-Hermite versus Sobolev nonparametric density estimation on ℝ

**Denis Belomestny**[1,2] · **Fabienne Comte**[3] ·
**Valentine Genon-Catalot**[3]

**Abstract** In this paper, our aim is to revisit the nonparametric estimation of a square integrable density $f$ on ℝ, by using projection estimators on a Hermite basis. These estimators are studied from the point of view of their mean integrated squared error on ℝ. A model selection method is described and proved to perform an automatic bias variance compromise. Then, we present another collection of estimators, of deconvolution type, for which we define another model selection strategy. Although the minimax asymptotic rates of these two types of estimators are mainly equivalent, the complexity of the Hermite estimators is usually much lower than the complexity of their deconvolution (or kernel) counterparts. These results are illustrated through a small simulation study.

✉ Fabienne Comte
  fabienne.comte@parisdescartes.fr

  Denis Belomestny
  denis.belomestny@uni-due.de

  Valentine Genon-Catalot
  valentine.genon-catalot@parisdescartes.fr

[1] Faculty of Mathematics, Duisburg-Essen University, Thea-Leymann-Str. 9, 45127 Essen, Germany

[2] National Research University Higher School of Economics, Shabolovka, 26, 119049 Moscow, Russia

[3] MAP5 UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité, 45 rue des Saints-Pères, 75 270 Paris Cedex 06, France

 Springer

## 1 Introduction

Consider an i.i.d. $n$-sample $X_1, \ldots, X_n$ from an unknown density $f$. The nonparametric estimation of $f$ has been the subject of such a huge number of contributions in the past decades that it is difficult to make an exhaustive list of references. Roughly speaking, there are two main approaches to estimate $f$, kernel or projection method. In the projection method which is our concern here, for $f$ belonging to $\mathbb{L}^2(\mathbb{R})$, considering an orthonormal basis in $\mathbb{L}^2(\mathbb{R})$, estimators are built by estimating a finite number of coefficients of the development of $f \in \mathbb{L}^2(\mathbb{R})$ on the basis. Fourier and wavelet bases, for instance, are commonly used. Bases of orthogonal polynomials are also used for compactly supported densities (see e.g., Donoho et al. 1996; Birgé and Massart 2007; Efromovich 1999; Massart 2007; Tsybakov 2009 for reference books). For densities with a non-compact support included in $\mathbb{R}^+$, recent contributions use bases composed of Laguerre functions (see e.g., Comte and Genon-Catalot 2015; Belomestny et al. 2016; Mabon 2017).

To our knowledge, for densities on $\mathbb{R}$, the use of a Hermite basis is only considered in Schwartz (1967) and Walter (1977). In this paper, we are going to revisit the nonparametric estimation of $f \in \mathbb{L}^2(\mathbb{R})$ by using projection estimators on a Hermite basis. To find the minimax asymptotic rates of convergence, authors generally assume that the unknown density belongs to a function space specifying some regularity properties of $f$. Here, we consider the Sobolev-Hermite spaces which are naturally associated with the Hermite basis and are defined in Bongioanni and Torrea (2006). It turns out that the Sobolev-Hermite space of regularity index $s$ is included in the classical Sobolev space with same index. Therefore, we are led to compare the performances of the projection estimators on the Hermite basis with those of the deconvolution estimators which are projection estimators on the sine cardinal basis. Deconvolution estimators have been widely studied mainly for observations with additive noise and also for direct observations (see e.g., Comte et al. 2008). The optimal $\mathbb{L}^2$-risk for density estimation on a Sobolev ball with regularity index $s$ is of order $O(n^{-2s/(2s+1)})$, see Schipper (1996), Efromovich (2008) and Efromovich (2009). For densities having a fifth-order moment belonging to a Sobolev Hermite ball with the same regularity index $s$, we obtain the same rate. Therefore, from the asymptotic point of view, no difference can be made between these two classes of estimators at least for non-heavy tailed densities. Apart from Sobolev spaces, we consider a class of Gaussian mixtures where Hermite-based estimators also achieve the minimax convergence rates. Finally, we study Hermite projection estimators in a different context, the estimation of the Lévy density of a Lévy process in the pure-jump case.

While most papers focus on deriving minimax convergence rates, the computational efficiency of the proposed estimator is not often considered. This issue is especially important for densities with a non-compact support. We prove that the Hermite estimators have usually a much lower complexity than the deconvolution estimators, resulting in a noteworthy computational gain.

The plan of the paper is as follows. In Sect. 2, we present the Hermite basis, and the $\mathbb{L}^2$-risk of the associated projection estimators is studied together with the possible orders for the variance term. A data-driven choice of the dimension is proposed and the associated estimator is proved to be realize adequately the bias–variance trade-off. In Sect. 3, results on deconvolution estimators are presented. Section 4 is devoted to the study of asymptotic rates of convergence. From this point of view, the two approaches of the previous sections are proved to be equivalent, except in some special cases. Then, we compare the complexity of the procedures and conclude that the Hermite method has a substantial advantage from this point of view. Section 4.6 is devoted to numerical simulation results and aims at illustrating the previous findings. In Sect. 5, the estimation of the Lévy density is considered in the same framework as Belomestny et al. 2015, Chapter "Adaptive estimation for Lévy processes." A short conclusion is delivered in Sect. 6, and proofs are gathered in Sect. 7.

## 2 Projection estimators on the Hermite basis

### 2.1 Hermite basis

Below, we denote by $\|.\|$ the $\mathbb{L}^2$-norm on $\mathbb{R}$ and by $\langle \cdot, \cdot \rangle$ the $\mathbb{L}^2$-scalar product. The Hermite polynomial of order $j$ is given, for $j \geq 0$, by:

$$H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j}(e^{-x^2}).$$

Hermite polynomials are orthogonal with respect to the weight function $e^{-x^2}$ and satisfy: $\int_{\mathbb{R}} H_j(x)H_\ell(x)e^{-x^2}dx = 2^j j!\sqrt{\pi}\delta_{j,\ell}$ (see e.g., Abramowitz and Stegun 1964). The Hermite function of order $j$ is given by:

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

The sequence $(h_j, j \geq 0)$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R})$. The density $f$ to be estimated can be developed in the Hermite basis $f = \sum_{j \geq 0} a_j(f)h_j$ where $a_j(f) = \int_{\mathbb{R}} f(x)h_j(x)dx = \langle f, h_j \rangle$.

We define $S_m = \text{span}(h_0, h_1, \ldots, h_{m-1})$ the linear space generated by the $m$ functions $h_0, \ldots, h_{m-1}$ and $f_m = \sum_{j=0}^{m-1} a_j(f)h_j$ the orthogonal projection of $f$ on $S_m$.

### 2.2 Hermite estimator and risk bound

Consider a sample $X_1, \ldots, X_n$ of i.i.d. random variables with density $f$, belonging to $\mathbb{L}^2(\mathbb{R})$. We define for each $m \geq 0$, $\hat{f}_m = \sum_{j=0}^{m-1} \hat{a}_j h_j$ a projection estimator of $f$, with $\hat{a}_j = n^{-1}\sum_{i=1}^{n} h_j(X_i)$, that is, an unbiased estimator of $f_m = \sum_{j=0}^{m-1} a_j(f)h_j$.

These estimators are considered in Schwartz (1967) and then in Walter (1977). As usual, the $\mathbb{L}^2$-risk is split into a variance and a square bias term. We give a more

accurate rate for the variance term than in the latter papers. Indeed, we have the classical decomposition

$$\mathbb{E}(\|\hat{f}_m - f\|^2) = \|f - f_m\|^2 + \sum_{j=0}^{m-1} \mathrm{Var}(\hat{a}_j) = \|f - f_m\|^2 + \frac{1}{n} \sum_{j=0}^{m-1} \mathrm{Var}(h_j(X_1))$$

$$\leq \|f - f_m\|^2 + \frac{V_m}{n}, \tag{2}$$

where

$$V_m = \int_{\mathbb{R}} \left( \sum_{j=0}^{m-1} h_j^2(x) \right) f(x) \mathrm{d}x = \mathbb{E} \left( \sum_{j=0}^{m-1} h_j^2(X_1) \right). \tag{3}$$

The infinite norm of $h_j$ satisfies (see Abramowitz and Stegun 1964; Szegö 1975, p. 242):

$$\|h_j\|_\infty \leq \Phi_0, \qquad \Phi_0 \simeq 1,086435/\pi^{1/4} \simeq 0.8160. \tag{4}$$

Therefore, we have $V_m \leq \Phi_0^2 m$, as usual for projection density estimator, see Massart (2007), Chapter 7. However, more precise properties of the Hermite functions provide refined bounds:

**Proposition 1** (i) *There exists constant c such that, for any density f and for any integer m,*

$$V_m \leq cm^{5/6}.$$

(ii) *If $\mathbb{E}|X|^5 < +\infty$, then there exists constant $c'$ such that for any integer m,*

$$V_m \leq c'm^{1/2}.$$

(iii) *Assume that there exists $K > 0$ with*

$$|f(x)| \leq g(x) := \alpha \frac{1}{(1+|x|)^a}, \text{ for } |x| \geq K \text{ and } \alpha > 0, a > 1.$$

*Then, there exists $c''$ such that, for m large enough, $V_m \leq c'' m^{\frac{a+2}{2(a+1)}}$.*

Proposition 1(i) shows that $V_m$ is at most of order $m^{5/6}$, a property obtained in Walter (1977). However, (ii)–(iii) show that this order can be improved depending on additional assumptions on $f$. At this point, it is worth stressing that, under the moment assumption of (ii), the rate of variance term $V_m/n$ is not $m/n$ as usual but $m^{1/2}/n$. This means that, in this approach, the role of the dimension is played by $m^{1/2}$. This fact, together with the regularity spaces introduced below to evaluate the rate of the bias term, allows to prove that the Hermite projection estimators is asymptotically equivalent to the sine cardinal estimators.

In the next paragraph, we make no assumption on the regularity properties of $f$. Moreover, because the variance order depends on assumptions on $f$, we do not want

to consider it as given unlike in most model selection strategies. Our proposal of data-driven $m$ leads to an estimator whose $\mathbb{L}^2$-risk automatically realizes the bias–variance trade-off in a non-asymptotic way without knowing the regularity of the function $f$ nor knowing the rate of the variance term.

## 2.3 Model selection

For model selection, we must estimate the bias and the variance term. Define $\mathcal{M}_n = \{1, \ldots, m_n\}$, where $m_n$ is the largest integer such that $m_n^{5/6} \leq n/\log(n)$ and set

$$
\hat{m} = \arg \min_{m \in \mathcal{M}_n} \{-\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m)\}, \quad \widehat{\text{pen}}(m) = \kappa \frac{\widehat{V}_m}{n}, \quad \widehat{V}_m = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{m-1} h_j^2(X_i),
$$
(5)

where $\kappa$ is a numerical constant. The quantity $-\|\hat{f}_m\|^2$ estimates $-\|f_m\|^2 = \|f - f_m\|^2 - \|f\|^2$, and we can ignore the (unknown) constant term $\|f\|^2$. Usually, the penalty is chosen equal to $\kappa \Phi_0^2 m/n$, which is the known upper bound of the variance term, where $\Phi_0$ is defined by (4). Here, we know that this rate is not the adequate one and the fact that the order of $V_m$ varies according to the assumptions on $f$ justifies that we rather use $\widehat{V}_m$, an unbiased estimator of $V_m$. We can prove the following result.

**Theorem 1** *Assume that $f$ is bounded and that $\inf_{a \leq x \leq b} f(x) > 0$ for some interval $[a, b]$. Then there exists $\kappa_0$ such that, for $\kappa \geq \kappa_0$, the estimator $\hat{f}_{\hat{m}}$ where $\hat{m}$ is defined by (5) satisfies*

$$
\mathbb{E}\left(\|\hat{f}_{\hat{m}} - f\|^2\right) \leq C \inf_{m \in \mathcal{M}_n} \left(\|f - f_m\|^2 + \kappa \frac{V_m}{n}\right) + \frac{C'}{n},
$$

*where $C$ is a numerical constant ($C = 4$ suits) and $C'$ is a constant depending on $\|f\|_\infty$.*

The estimator $\hat{f}_{\hat{m}}$ is adaptive in the sense that its risk bound achieves automatically the bias–variance compromise, up to a negligible term of order $O(1/n)$. It follows from the proof that $\kappa_0 = 8$ is possible. This value of $\kappa_0$ is certainly not optimal; finding the optimal theoretical value of $\kappa$ in the penalty is not an easy task, even in simple models (see for instance Birgé and Massart (2007) in a Gaussian regression model). This is why it is standard to calibrate the value $\kappa$ in the penalty by preliminary simulations, as we do in Sect. 4.6. Actually, the assumption $\inf_{a \leq x \leq b} f(x) > 0$ is due to the fact that the proof requires the condition

$$
\forall m \geq m_0, \ V_m \geq 1, \ \text{and } \forall a > 0, \ \sum_{m \in \mathcal{M}_n} e^{-a\sqrt{V_m}} \leq A < +\infty.
$$
(6)

Condition (6) holds, as we can prove:

**Proposition 2** *If $\inf_{a \leq x \leq b} f(x) > 0$ for some interval $[a, b]$, then, for $m$ large enough, $V_m \geq c'' m^{1/2}$ where $c''$ is a constant.*

## 3 Deconvolution estimators

As we want to compare the performances of projection estimators on the Hermite basis to those of projection estimators on the sine cardinal basis, we recall the definition of the latter estimators, i.e., the deconvolution estimators. Let $\varphi(x) = \sin(\pi x)/(\pi x)$ which satisfies $\varphi^*(t) = 1_{[-\pi,\pi]}(t)$, where $\varphi^*$ denotes the Fourier transform of $\varphi$. The functions $(\varphi_{\ell,j}(x) = \sqrt{\ell}\varphi(\ell x - j), j \in \mathbb{Z})$ constitute an orthonormal system in $\mathbb{L}^2(\mathbb{R})$. The space $\Sigma_\ell$ generated by this system is exactly the subspace of $\mathbb{L}^2(\mathbb{R})$ of functions having Fourier transforms with compact support $[-\pi\ell, \pi\ell]$. The orthogonal projection $\bar{f}_\ell$ of $f$ on $\Sigma_\ell$ satisfies $\bar{f}_\ell^* = f^* 1_{[-\pi\ell,\pi\ell]}$. Therefore,

$$\|f - \bar{f}_\ell\|^2 = \frac{1}{2\pi} \int_{|t| \geq \pi\ell} |f^*(t)|^2 dt. \tag{7}$$

The projection estimator $\widetilde{f}_\ell$ of $f$ is defined by:

$$\widetilde{f}_\ell(x) = \frac{1}{2\pi} \int_{-\pi\ell}^{\pi\ell} e^{-itx} \frac{1}{n} \sum_{k=1}^n e^{itX_k} dt = \frac{1}{n} \sum_{k=1}^n \frac{\sin(\pi\ell(X_k - x))}{\pi(X_k - x)}. \tag{8}$$

This expression corresponds to the fact that:

$$\bar{f}_\ell = \frac{1}{2\pi} \int_{-\pi\ell}^{\pi\ell} e^{-itx} f^*(t) dt = \sum_{j \in \mathbb{Z}} a_{\ell,j} \varphi_{\ell,j}(x), \qquad a_{\ell,j} = \langle f, \varphi_{\ell,j} \rangle.$$

Contrary to $\hat{f}_m$, the estimator $\widetilde{f}_\ell$ cannot be expressed as the corresponding sum with the estimated coefficients $\tilde{a}_{\ell,j} = \frac{1}{n} \sum_{k=1}^n \varphi_{\ell,j}(X_k)$ as this sum would be infinite and not defined. To compute it in concrete, one can use (8) or a truncated version

$$\widetilde{f}_\ell^{(n)}(x) = \sum_{|j| \leq K_n} \tilde{a}_{\ell,j} \varphi_{\ell,j}(x), \qquad \tilde{a}_{\ell,j} = \frac{1}{n} \sum_{k=1}^n \varphi_{\ell,j}(X_k),$$

which creates an additional bias but is comparable to the previous Hermite estimator. We give the results for $\widetilde{f}_\ell$ and $\widetilde{f}_\ell^{(n)}$.

**Proposition 3** *The estimator $\widetilde{f}_\ell$ satisfies*

$$\mathbb{E}(\|\widetilde{f}_\ell - f\|^2) \leq \|f - \bar{f}_\ell\|^2 + \frac{\ell}{n}.$$

*If moreover $M_2 = \int x^2 f^2(x) dx < +\infty$, then the estimator $\widetilde{f}_\ell^{(n)}$ satisfies*

$$\mathbb{E}\left(\|\widetilde{f}_\ell^{(n)} - f\|^2\right) \leq 2\|f - \bar{f}_\ell\|^2 + \frac{\ell}{n} + 4\frac{\ell^2(M_2 + 1)}{K_n}.$$

If $\ell \leq n$ and $K_n \geq n^2$, the last term is of order $O(\ell/n)$ and can be associated to the variance term $\ell/n$. Note that condition $K_n \geq n^2$ implies that the computation of a large number of coefficients is required for $\widetilde{f}_\ell^{(n)}$, for large $n$. In practice, we take $K_n$ even smaller than $n$ in order to keep reasonable computation times.

As in the previous case, we can define a data-driven choice of the cutoff parameter $\ell$ and build adaptive estimators:

$$\tilde{\ell} = \arg\min_{\ell \leq n} \left\{ -\|\widetilde{f}_\ell\|^2 + \tilde{\kappa}\frac{\ell}{n} \right\}, \quad \tilde{\ell}_n = \arg\min_{\ell \leq n} \left\{ -\|\widetilde{f}_\ell^{(n)}\|^2 + \tilde{\kappa}\frac{\ell}{n} \right\}, \qquad (9)$$

where $\tilde{\kappa}$ is a numerical constant. Note that

$$\|\widetilde{f}_\ell\|^2 = \frac{1}{n^2} \sum_{1 \leq j,k \leq n} \frac{\sin(\pi\ell(X_k - X_j))}{\pi(X_k - X_j)}, \quad \|\widetilde{f}_\ell^{(n)}\|^2 = \sum_{|j| \leq K_n} |\tilde{a}_{\ell,j}|^2.$$

We give the result for $\widetilde{f}_\ell^{(n)}$ only, as $\|\widetilde{f}_\ell^{(n)}\|^2$ is faster to compute if $K_n$ is chosen in a restricted range, $K_n \leq n$, see Sects. 4.5 and 4.6. The following result holds.

**Theorem 2** *If $K_n \geq n^2$ and $M_2 = \int x^2 f^2(x)dx < +\infty$, then there exists a numerical constant $\tilde{\kappa}_0$ such that, for $\tilde{\kappa} \geq \tilde{\kappa}_0$, the estimator $\widetilde{f}_{\tilde{\ell}_n}^{(n)}$ where $\tilde{\ell}_n$ is defined by (9) satisfies*

$$\mathbb{E}\left(\left\|\widetilde{f}_{\tilde{\ell}_n}^{(n)} - f\right\|^2\right) \leq C_1 \inf_{\ell \leq n} \left(\|f - f_\ell\|^2 + \tilde{\kappa}\frac{\ell}{n} + \frac{\ell(M_2+1)}{n}\right) + \frac{C_2}{n},$$

*where $C_1$ is a numerical constant and $C_2$ is a constant depending on $\|f\|_\infty$.*

For $\widetilde{f}_{\tilde{\ell}}$, an analogous risk bound may be obtained, without condition $M_2 < +\infty$ and without the term $\ell(M_2+1)/n$ in the bound. For Theorem 2, we refer to Comte et al. (2008), Proposition 5.1, p. 97.

## 4 Comparison of rates of convergence and discussion

In this section, we compute the rates of convergence that can be deduced from the optimization of the upper bounds of $\mathbb{L}^2$-risks. This requires to assess the rate of decay of the bias terms $\|f - f_m\|^2$ in the Hermite case, $\|f - \bar{f}_\ell\|^2$ in the deconvolution framework. The latter is usually obtained by assuming that the unknown density $f$ belongs to a Sobolev space. For the former, we consider the Sobolev-Hermite spaces which are naturally linked with the Hermite basis.

### 4.1 Sobolev and Sobolev-Hermite regularity

For $s > 0$, the Sobolev-Hermite space with regularity $s$ may be defined by:

$$W^s = \left\{ f \in \mathbb{L}^2(\mathbb{R}), \|f\|_{s,\text{sobherm}}^2 = \sum_{n \geq 0} n^s a_n^2(f) < +\infty \right\}, \qquad (10)$$

where $a_n(f) = \langle f, h_n \rangle$ is the $n$-th component of $f$ in the Hermite basis. We refer to Bongioanni and Torrea (2006) for a definition using operator theory. Let $\mathcal{F} = \left\{ \sum_{j \in J} a_j h_j, \ J \subset \mathbb{N}, \ \text{finite} \right\}$ be the set of finite linear combinations of Hermite functions and $C_c^\infty$ the set of infinitely derivable functions with compact support. The sets $C_c^\infty$ and $\mathcal{F}$ are dense in $W^s$. As the Fourier transform of $h_n$ satisfies

$$h_n^* = \sqrt{2\pi} i^n h_n, \tag{11}$$

$f \in W^s$ if and only if $f^* \in W^s$. We now describe $W^s$ when $s$ is integer. Let

$$A_+ f = f' + xf, \quad A_- f = -f' + xf.$$

The following result is proved in Bongioanni and Torrea (2006). For sake of clarity, we give a simplified proof.

**Proposition 4** *For s integer, the Sobolev-Hermite space $W^s$ is equal to:*

$$W^s = \left\{ f \in \mathbb{L}^2(\mathbb{R}), \ f \ \text{admits derivatives up to order } s, \right.$$

$$\left. \|f\|_{s,\text{sobherm}} = \sum_{\substack{j_1, \ldots, j_m \in \{-,+\}, \\ 1 \leq m \leq s}} \|A_{j_1} \ldots A_{j_m} f\| + \|f\| < +\infty \right\}.$$

*Moreover, the following statements are equivalent: for s integer,*

(1) $f \in W^s$,
(2) $f$ *admits derivatives up to order $s$ which satisfy $f, f', \ldots, f^{(s)}, x^{s-\ell} f^{(\ell)}, \ell = 0, \ldots, s - 1$ belong to $\mathbb{L}^2(\mathbb{R})$.*

The two norms $\|f\|_{s,\text{sobherm}}$ and $\|\!|f|\!\|_{s,\text{sobherm}}$ are equivalent.

Now, we recall the definition of usual Sobolev spaces. The Sobolev space with regularity index $s$ is defined by

$$\mathcal{W}^s = \left\{ f \in \mathbb{L}^2(\mathbb{R}), \ \|f\|_{s,\text{sob}}^2 = \int_{\mathbb{R}} (1 + t^{2s})|f^*(t)|^2 \mathrm{d}t < +\infty \right\}. \tag{12}$$

If $s$ is integer, then

$$\mathcal{W}^s = \left\{ f \in \mathbb{L}^2(\mathbb{R}), \ f \ \text{admits derivatives up to order } s \right.$$

$$\left. \text{such that } \|\!|f|\!\|_{s,\text{sob}}^2 = \|f\|^2 + \|f'\|^2 + \cdots + \|f^{(s)}\|^2 < +\infty \right\}.$$

The two norms $\|\!| \cdot |\!\|_{s,\text{sob}}$ and $\| \cdot \|_{s,\text{sob}}$ are equivalent. Therefore, for $s$ integer, $W^s \subset \mathcal{W}^s$. Moreover, the following properties are proved in Bongioanni and Torrea (2006): for all $s > 0$,

– $W^s \subsetneq \mathcal{W}^s$. If $f \in \mathcal{W}^s$ has compact support, then $f \in W^s$.
–

$$f \in W^s \Rightarrow x^s f \in \mathbb{L}^2(\mathbb{R}). \tag{13}$$

### 4.2 Rates of convergence

Now, we look at asymptotic rates of convergence. We first consider rates for Hermite projection estimators. We already studied the variance rate $V_m/n$ (see the bounds for $V_m$ in Proposition 1). If $f$ belongs to

$$W^s(L) = \left\{ f \in \mathbb{L}^2(\mathbb{R}), \sum_{n \geq 0} n^s a_n^2(f) \leq L \right\},$$

then $\|f - f_m\|^2 \leq Lm^{-s}$. Plugging this and the bounds of Proposition 1 in Inequality (2) gives the following rates of the $\mathbb{L}^2(\mathbb{R})$-risk.

**Proposition 5** *Assume that $f \in W^s(L)$ and consider the three cases (i), (ii), (iii) of Proposition* 1.
*Case* (i) *(general case). For $m_{\mathrm{opt}} = [n^{1/(s+(5/6))}]$, $\mathbb{E}(\|\hat{f}_{m_{\mathrm{opt}}} - f\|^2) \lesssim n^{-\frac{s}{s+(5/6)}}$.*
*Case* (ii). *For $m_{\mathrm{opt}} = [n^{1/(s+(1/2))}]$, $\mathbb{E}(\|\hat{f}_{m_{\mathrm{opt}}} - f\|^2) \lesssim n^{-\frac{s}{s+1/2}}$.*
*Case* (iii). *For $m_{\mathrm{opt}} = [n^{1/(s+(a+2)/(2(a+1)))}]$, $\mathbb{E}(\|\hat{f}_{m_{\mathrm{opt}}} - f\|^2) \lesssim n^{-\frac{s}{s+(a+2)/[2(a+1)]}}$.*

Case (ii) gives the best rate. In view of the constraint on $m_n$ in Theorem 1, the adaptive procedure reaches this best rate if $m_{\mathrm{opt}}^{5/6} \leq n/\log(n)$, that is $s > 1/3$. Note that the rate in case (iii) is strictly better than in case (i) as $(a+2)/(a+1) < 5/3$ as soon as $a > 1/2$. Cases (ii)–(iii) improve the results of Schwartz (1967) and Walter (1977). Now, we can compare the rates to those of projection estimators in the sine cardinal basis. The following result is deduced from Proposition 3 and (7).

**Proposition 6** *If*

$$f \in \mathcal{W}^s(R) = \left\{ f \in \mathbb{L}^2(\mathbb{R}), \|f\|_{s,\mathrm{sob}}^2 = \int_{\mathbb{R}} (1 + t^{2s}) |f^*(t)|^2 \mathrm{d}t \leq R \right\},$$

*and $\ell_{\mathrm{opt}} = n^{1/(2s+1)}$, we have $\mathbb{E}(\|\widetilde{f}_{\ell_{\mathrm{opt}}} - f\|^2) \lesssim n^{-2s/(2s+1)}$. If moreover $K_n \geq n^2$, $\mathbb{E}(\|\widetilde{f}_{\ell_{\mathrm{opt}}}^{(n)} - f\|^2) \lesssim n^{-2s/(2s+1)}$.*

In Schipper (1996), it is proved that this rate is minimax optimal (with exact Pinsker constant) on Sobolev balls (at least for an integer $s$), see also Efromovich (2009) for $s < 1/2$. Rigollet (2006) uses the blockwise Stein method to build an adaptive deconvolution estimator, which reaches the optimal rate with exact constant for any $s > 1/2$.

Let us compare results of Proposition 6 and of Proposition 5. As $W^s \subset \mathcal{W}^s$, see Sect. 4.1, the comparison is relevant. In case (i), we see that the estimator $\widetilde{f}_{\ell_{\mathrm{opt}}}$ has a

better rate than $\hat{f}_{m_{\text{opt}}}$. In case (ii), the estimators have the same rate. In case (iii), the estimator $\widetilde{f}_{\ell_{\text{opt}}}$ is slightly better than $\hat{f}_{m_{\text{opt}}}$. In view of case (ii), the Hermite method is competitive. Indeed, the moment condition for (ii) is not very strong. In this case, sine cardinal estimators with cutoff parameter $\ell$ and Hermite projection estimators on the space $S_m$ are asymptotically equivalent when $\ell = m^{1/2}$.

### 4.3 Rates of convergence in some special cases

When the density $f$ belongs to $W^s$ for all $s$, we must obtain directly the exact rate of decay of the bias term. This is possible for Gaussian and some related densities as one can make an exact computation of the coefficients $a_j(f)$. Let

$$f_\mu(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu)^2\right), \tag{14}$$

and

$$f_{p,\sigma}(x) = \frac{x^{2p}}{\sigma^{2p} C_{2p}} f_\sigma(x) \quad \text{with} \quad f_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad \text{and} \quad C_{2p} = \mathbb{E}X^{2p}, \tag{15}$$

for $X$ a standard Gaussian variable. The distribution $f_{p,\sigma}(x)\mathrm{d}x$ is equal to $\varepsilon G^{1/2}$ for $\varepsilon$ a symmetric Bernoulli variable, $G$ a Gamma$(p+(1/2), 1/(2\sigma^2))$ variable, independent of $\varepsilon$.

**Proposition 7** *Assume that $f = f_\mu$. Then for $m_{\text{opt}} = [(\log(n)/\log(2)) + e\mu^2]$, we have*

$$\mathbb{E}(\|\hat{f}_{m_{\text{opt}}} - f_\mu\|^2) \lesssim \sqrt{\log n}/n.$$

*Assume that $f = f_\sigma$. Then for $m_{\text{opt}} = [(\log n)/\lambda]$ where $\lambda = \log\left(\frac{\sigma^2+1}{\sigma^2-1}\right)^2$, we have*

$$\mathbb{E}(\|\hat{f}_{m_{\text{opt}}} - f_\sigma\|^2) \lesssim \sqrt{\log n}/n.$$

The same result holds for $f = f_{p,\sigma}$ or any finite mixture of such distributions. For $f = f_\sigma$, the estimator $\widetilde{f}_\ell$ satisfies,

$$\mathbb{E}(\|\widetilde{f}_\ell - f\|^2) \lesssim \frac{1}{\ell} \exp\left(-\ell^2/2\sigma^2\right) + n^{-1}\ell.$$

For $\ell_{\text{opt}} = \sigma\sqrt{2\log n}$, the rate of $\widetilde{f}_{\ell_{\text{opt}}}$ is $\sqrt{\log n}/n$. The rate is identical to the one obtained in Proposition 7. The result is analogous for $f = f_{p,\sigma}$.

Finally, the Cauchy density will provide a counter-example. Let

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

From Proposition 1, case (iii), we take $a = 2$ and obtain for the variance term $V_m \lesssim m^{2/3}$. Using Proposition 4, we check that $f \in W^1$, $f \notin W^2$. Moreover, by (13), $x^s f \notin W^s$ for $s \geq 3/2$. Therefore, $f \notin W^{3/2}$, so the best rate we can obtain is $n^{-s/s+(2/3)}$ with $s < 3/2$, for $m_{\mathrm{opt}} = [n^{1/(s+(2/3))}]$.

For the sine cardinal method, $f^*(t) = \exp(-|t|)$, so that $\|f - f_\ell\| \lesssim \exp(-2\pi\ell)$. Therefore, for $\ell_{\mathrm{opt}} = \log n/2\pi$, the estimator $\widetilde{f}_{\ell_{\mathrm{opt}}}$ has a risk with rate $\log n/n$. This is much better than for the Hermite estimator.

This discussion on rates of convergence points out the interest of the adaptive method. Indeed, it automatically realizes the bias–variance compromise and thus the previous rates are reached without any specific knowledge on $f$.

### 4.4 Rates of convergence for Gaussian mixtures

Kim (2014) provides optimal rates of convergence for estimating densities that are mean mixtures of normal distributions, that is for densities $f$ in the class

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

where $\phi$ denotes the standard normal density and $\mathcal{P}(\mathbb{R})$ the set of all probability measures on the real line. The minimax optimal rate for the mean square risk is $\sqrt{\log(n)}/n$. Moreover, the sine cardinal estimator $\widetilde{f}_\ell$ reaches the upper bound for the $\mathbb{L}^2$-risk on the class $\mathcal{F}$, for $\ell \propto \sqrt{\log(n)}$.

We study Hermite projection estimators for mean mixtures of Gaussian but also for variance mixtures. We consider, as suggested in Kim (2014), the subclass $\mathcal{F}_{\mathrm{sub}}(\mathbb{R}) = \cup_{C>0}\mathcal{F}_{\mathrm{sub}}(C)$,

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

where

$$\mathcal{P}_{\mathrm{sub}}(C) := \{\Pi \in \mathcal{P}(\mathbb{R}),\ \Pi(|u| > t) \leq C \exp(-t^2/C) \text{ for all positive } t\}.$$

**Proposition 8** *For $f \in \mathcal{F}_{\mathrm{sub}}(C)$ and $m_{\mathrm{opt}} = [\log(n)(eC + 1/\log(2))]$, we have*

$$\mathbb{E}(\|\hat{f}_{m_{\mathrm{opt}}} - f\|^2) \lesssim \sqrt{\log(n)}/n.$$

Now we define the class of variance mixtures that we consider: let $v > 1$,

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

In other words, $f \in \mathcal{G}(v)$ is the density of $\sigma X$ with $X \sim \mathcal{N}(0, 1), \sigma \sim \Pi, \sigma \in [1/v, v]$ with $\sigma$ and $X$ independent.

**Proposition 9** *For $f \in \mathcal{G}(v)$, let*

$$\rho_0 = \left(\frac{v^2 - 1}{v^2 + 1}\right)^2 < 1. \tag{16}$$

*For $m_{\text{opt}} = [\log(n)/|\log(\rho_0)|]$, we have*

$$\mathbb{E}(\|\hat{f}_{m_{\text{opt}}} - f\|^2) \lesssim \sqrt{\log(n)}/n.$$

In the class of variance mixtures of Gaussians, the lower bound rate is not known. However, in Ibragimov and Has'minskii (1980), Ibragimov (2001), the estimation of a density on a compact set $[a, b]$ which is analytical in the vicinity of $[a, b]$ is considered. The authors prove that the rate $\log(n)/n$, is optimal in this class. The restricted class given by $\mathcal{G}(v)$ considered in Proposition 9 gives a slightly improved rate, which is coherent with the result of Proposition 7.

### 4.5 Complexity

In this paragraph, we compare the Hermite and deconvolution estimators from another point of view: the computational efficiency.

Consider an estimator $\hat{f}_n$ of a function $f$ whose $\mathbb{L}^2$-risk can be evaluated on a ball $B(L)$ of some functional space. Define its complexity $\mathcal{C}_{\hat{f}_n}(\varepsilon)$ as the minimal cost of computing $\hat{f}_n$ at the observation points $X_1, \ldots, X_n$, given that

$$\sup_{f \in B(L)} \mathbb{E}(\|\hat{f}_n - f\|^2) \leq \varepsilon^2.$$

Let us compute the complexity of the estimate $\widetilde{f}_{\ell_{\text{opt}}}$ on the Sobolev ball $\mathcal{W}^s(L)$. As we need to evaluate the function $\frac{\sin(\pi \ell \cdot)}{\pi \cdot}$ at all points $(X_k - X_j)$, $1 \leq k, j \leq n$, the cost of computing $\widetilde{f}_{\ell_{\text{opt}}}$ is of order $n^2$. Thus $\varepsilon^2 \asymp n^{-2s/(2s+1)}$ yields $n \asymp \varepsilon^{-2-1/s}$ so that $\mathcal{C}_{\widetilde{f}_{\ell_{\text{opt}}}}(\varepsilon) \asymp \varepsilon^{-4-2/s}$ as $\varepsilon \to 0$. So even in the case of infinitely smooth densities, the complexity of the deconvolution estimate can not be (asymptotically) lower than $\varepsilon^{-4}$. A natural question is whether one can find an estimate with lower order of complexity. Note that the complexity would be the same for a kernel estimator on a ball of a Nikol'ski class with regularity $s$, see Tsybakov (2009), at least for kernels with a non-compact support used in Ibragimov and Has'minskii (1980).

For the truncated estimator $\widetilde{f}_{\ell_{\text{opt}}}^{(n)}$, the cost is of order $nK_n$: indeed, one must compute the $\varphi_{\ell, j}(X_i)$ for $i = 1, \ldots, n$ and $|j| \leq K_n$. Consequently, compared to the previous one, this estimate is competitive in term of computational cost as soon as $K_n < n$ (however, this choice would contradict Theorem 3.1 where $K_n \geq n^2$).

Now, let us look at the projection estimator $\hat{f}_{m_{\text{opt}}}$ for $f \in W^s(L)$. The cost of computing a projection estimator $\hat{f}_m$ at observation points $X_1, \ldots, X_n$ corresponds to the cost of computing $h_j(X_i)$ for $i = 1, \ldots, n$ and $j = 0, \ldots, m - 1$, i.e., is of order $nm$. Thus, we derive the following proposition.

**Table 1** Complexity for density estimation in different contexts

| Projection estimator on a compact set $A$ Besov ball of $B_{2,s,\infty}(A)$ | Projection estimator on $\mathbb{R}$ Sobolev-Hermite ball $W^s(L)$ | Deconvolution estimator on $\mathbb{R}$ Sobolev ball $\mathcal{W}^s(L)$ |
|---|---|---|
| $\varepsilon^{-2-2/s}$ | $\varepsilon^{-2-3/s}$ | $\varepsilon^{-4-2/s}$ |
|  | (best case $\alpha = 1/2$) |  |

**Proposition 10** *Assume that* $f \in W^s(L)$ *and consider the three cases (i), (ii), (iii) of Proposition* 1. *The complexity of the estimate* $\hat{f}_{m_{opt}}$ *is given by* $\mathcal{C}_{\hat{f}}(\varepsilon) \sim \varepsilon^{-2-\frac{2(\alpha+1)}{s}}$ *with* $\alpha = 5/6, 1/2, (a+2)/[2(a+1)],$, *respectively.*

*Proof of Proposition 10.* Taking $\varepsilon^2 \asymp n^{-2s/(2s+1)}$, hence $n \asymp \varepsilon^{-2-1/s}$, and the three values of $m_{opt}$ given Proposition 5 yield the result. $\qquad\square$

As can be seen, the complexity order of the Hermite-based estimate $\hat{f}_{m_{opt}}$ is lower than the complexity order of the deconvolution estimate $\widetilde{f}_{\ell_{opt}}$ provided $s > \alpha$. So in the case of densities with finite fifth moment already for $s > 1/2$, our approach leads to estimates with much lower complexity. The difference between the estimates $\hat{f}_{m_{opt}}$ and $\widetilde{f}_{\ell_{opt}}$ becomes especially pronounced in the limiting case $s \to \infty$, where $\mathcal{C}_{\hat{f}_{m_{opt}}}(\varepsilon) \asymp \varepsilon^{-2}$ while $\mathcal{C}_{\widetilde{f}_{\ell_{opt}}}(\varepsilon) \asymp \varepsilon^{-4}$ as $\varepsilon \to 0$, resulting in a huge computational gain.

For any projection estimator, the cost of computation if of order $nm_{opt}$ where $m_{opt}$ is the optimal dimension. In the case of a density with compact support $A$, if we evaluate the $\mathbb{L}^2$-risk of a projection estimator on a Besov ball of $B_{2,s,\infty}(A)$, we have $\varepsilon^2 \asymp n^{-2s/(2s+1)}$ with $m_{opt} \asymp n^{1/(2s+1)}$, thus a cost of order $\varepsilon^{-2-2/s}$, see Barron et al. (1999) for rates and definition of Besov spaces. All these results are summarized in Table 1.

## 4.6 Simulation illustrations

In this section, we propose a few illustrations of the previous theoretical findings. To that aim, we consider several densities, fitting different assumptions of our setting.

   (i) A Gaussian $\mathcal{N}(0, 1)$,
  (ii) A Gaussian $\mathcal{N}(0, \sigma^2)$, $\sigma = 0.5$,
 (iii) A mixed Gaussian density $0.4\mathcal{N}(-3, \sigma^2) + 0.6\mathcal{N}(3, \sigma^2)$, $\sigma = 0.5$,
 (iv) A Gamma $\gamma(3, 0.5)$ density,
  (v) A mixed Gamma $0.4\gamma(2, 1/2) + 0.6\gamma(16, 1/4)$
 (vi) A beta density $\beta(3, 3)$,
(vii) A beta density $\beta(3, 6)$,
(viii) Laplace density $f(x) = e^{-|x|}/2$,
 (ix) A Cauchy density, $f(x) = 5/[\pi(1 + (5x)^2)]$.

Density (i) is proportional to the first basis function $h_0$ and should be perfectly estimated in the Hermite procedure, densities (vi) and (vii) are compactly supported
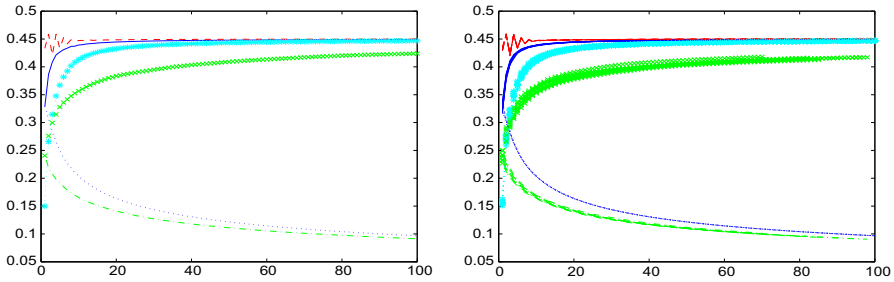
**Fig. 1** Left: $m \mapsto \widehat{V}_m / \sqrt{m}$ for 1 sample of densities (i) (blue line), (iv) (cyan stars), (vi) (red dashed), (ix) (green x marks) and $m \mapsto \widehat{V}_m / m^{5/6}$ for 1 sample of densities (i) (blue dashed) and (ix) (green dash-dot). Right: the same as previously for 10 samples color figure online

and density (ix) does not admit any moment (in particular no fifth moment, so it does not fit case (ii) of Proposition 1). Hermite functions are recursively computed via (34) and with normalization (1).

We plot in Fig. 1, the representation of $m \mapsto \widehat{V}_m / \sqrt{m}$ for 1 and 10 samples drawn from densities (i), (iv), (vi), (ix) (see (5)). It seems that the ratio is stable along the repetitions and converges to a fixed value, which is the same in the first three cases. On the contrary, $m \mapsto \widehat{V}_m / m^{5/6}$ given for (i) and (ix) seems to decrease and to tend to zero in any case. It is tempting to conclude from these plots that the order of $V_m$ is $O(m^{1/2})$ in a rather general case.

We have implemented the Hermite projection estimator $\hat{f}_{\hat{m}}$ with $\hat{m}$ given in (5), $\widetilde{f}_{\tilde{\ell}_n}^{(n)}$ with $\tilde{\ell}_n$ given by (9) and the kernel estimator given by the function `ksdensity` of Matlab. For the model selection steps of the first two estimators, the two constants $\kappa$ and $\tilde{\kappa}$ of the procedures have been both calibrated by preliminary simulations including other densities than the ones mentioned above (to avoid overfitting): the selected values were $\kappa = \tilde{\kappa} = 4$. We considered two sample sizes $n = 250$ and $n = 1000$, but as the sine cardinal procedure is rather slow, we only took $K_{250} = K_{1000} = 100$. The theoretical value $K_n = n^2$ is unreachable in practice (the computing time becomes much too large), and our choice of $K_n$ is consistent with the complexity considerations of Sect. 4.5.

For $m_n$, we should take $(n/\log(n))^{6/5}$, which is of order 100 for $n = 250$ and 400 for $n = 1000$. We took $m_{250} = m_{1000} = 200$ as a compromise. The cutoff $\ell\pi$ is selected among 100 equispaced values between 0 and 10. For each distribution, we present in Table 2 the MISE computed over 200 repetitions, together with the standard deviation. In the three cases, we provide also the mean (and standard deviations in parenthesis) of the selected dimension (Hermite), cutoff (Sine cardinal) or bandwidth (kernel).

We can see from the results of Table 2 that the Hermite and sine cardinal methods give very similar results, except for the $\mathcal{N}(0, 1)$ where the Hermite projection is much better as expected, as the procedure most of the time chooses $m = 1$. The kernel method seems globally less satisfactory. The noteworthy difference between the first two methods is the computation time: as the models are nested in the Hermite projection strategy, all coefficients can be computed once for all, and then, the dimension is

**Table 2** Results after 200 iterations of simulations of density (i)–(ix)

| $f$ | $n = 250$ | | | $n = 1000$ | | |
|---|---|---|---|---|---|---|
| | Hermite | Sin. Card. | Kernel | Hermite | Sin. Card. | Kernel |
| (i) | $0.5_{(1.4)}$ | $2.0_{(1.8)}$ | $2.9_{(2.1)}$ | $0.1_{(0.4)}$ | $0.6_{(0.5)}$ | $1.1_{(0.6)}$ |
| | $1.08_{(0.32)}$ | $0.77_{(0.08)}$ | $0.35_{(0.03)}$ | $1.08_{(0.53)}$ | $0.87_{(0.08)}$ | $0.27_{(0.01)}$ |
| (ii) | $4.7_{(4.7)}$ | $4.7_{(4.9)}$ | $6.0_{(4.2)}$ | $1.4_{(1.3)}$ | $1.4_{(1.4)}$ | $2.1_{(1.3)}$ |
| | $8.87_{(3.69)}$ | $1.46_{(0.21)}$ | $0.17_{(0.01)}$ | $11.6_{(5.1)}$ | $1.65_{(0.24)}$ | $0.13_{(0.005)}$ |
| (iii) | $4.9_{(2.6)}$ | $5.5_{(2.5)}$ | $24.2_{(14.5)}$ | $1.5_{(0.9)}$ | $1.5_{(0.9)}$ | $11.1_{(3.1)}$ |
| | $11.6_{(1.5)}$ | $1.28_{(0.13)}$ | $0.51_{(0.12)}$ | $14.3_{(2.0)}$ | $1.53_{(0.10)}$ | $0.39_{(0.04)}$ |
| (iv) | $5.6_{(3.4)}$ | $5.3_{(3.4)}$ | $4.6_{(2.9)}$ | $1.8_{(1.0)}$ | $1.8_{(1.0)}$ | $1.9_{(1.0)}$ |
| | $6.28_{(2.64)}$ | $1.25_{(0.19)}$ | $0.27_{(0.02)}$ | $11.9_{(4.5)}$ | $1.7_{(0.25)}$ | $0.21_{(0.01)}$ |
| (v) | $7.2_{(3.6)}$ | $6.6_{(2.8)}$ | $17.1_{(2.6)}$ | $2.4_{(0.9)}$ | $2.7_{(0.8)}$ | $10.2_{(1.2)}$ |
| | $15.1_{(2.0)}$ | $1.13_{(0.20)}$ | $0.74_{(0.07)}$ | $18.2_{(2.6)}$ | $1.67_{(0.27)}$ | $0.57_{(0.02)}$ |
| (vi) | $7.2_{(7.2)}$ | $7.3_{(7.4)}$ | $12.8_{(8.3)}$ | $3.2_{(2.6)}$ | $3.3_{(2.7)}$ | $4.8_{(2.7)}$ |
| | $46.5_{(10.5)}$ | $3.14_{(0.29)}$ | $0.07_{(0.005)}$ | $63.3_{(27.3)}$ | $3.6_{(0.65)}$ | $0.05_{(0.002)}$ |
| (vii) | $17.2_{(11.3)}$ | $19.3_{(15.6)}$ | $19.4_{(12.2)}$ | $5.8_{(3.1)}$ | $6.3_{(4.1)}$ | $7.0_{(4.0)}$ |
| | $104_{(25.5)}$ | $4.77_{(0.91)}$ | $0.05_{(0.004)}$ | $143_{(13.5)}$ | $5.89_{(0.85)}$ | $0.04_{(0.002)}$ |
| (viii) | $7.4_{(2.0)}$ | $6.7_{(3.0)}$ | $5.5_{(3.2)}$ | $2.6_{(0.8)}$ | $2.5_{(0.8)}$ | $2.3_{(1.1)}$ |
| | $2.5_{(2.96)}$ | $1.03_{(0.25)}$ | $0.36_{(0.04)}$ | $7.8_{(4.4)}$ | $1.42_{(0.32)}$ | $0.27_{(0.01)}$ |
| (ix) | $21.6_{(9.1)}$ | $21.5_{(9.3)}$ | $18.6_{(10.6)}$ | $6.9_{(2.9)}$ | $6.9_{(3.0)}$ | $7.6_{(3.9)}$ |
| | $65_{(25)}$ | $3.7_{(0.7)}$ | $0.10_{(0.01)}$ | $97_{(21)}$ | $4.71_{(0.77)}$ | $0.08_{(0.004)}$ |

For each density (i)–(ix), first line: MISE $\times$ 1000 with (std $\times$ 1000) in parenthesis; second line: mean of selected dimension (Hermite), cutoff (sine cardinal) or bandwidth (kernel) with std in parenthesis

selected. In the sine cardinal strategy, each time $\ell$ is changed, all the coefficients have to be recalculated. For instance, when the maximal dimension proposed $m_n$ is 50, and $K_n$ is 100, the elapsed times for 100 simulations is: for $n = 250$, around 0.5s for Hermite, 41s for sine cardinal; for $n = 1000$, around 1.2s for Hermite, 137s for sine cardinal, all times measured on the same personal computer to give an order of the difference. This is coherent with the lower complexity property of the Hermite method.

Table 2 also provides the selected dimensions, cutoffs and bandwidths. As could be expected, $\hat{m}$, $\tilde{\ell}$ vary in opposite way, compared to $\hat{h}$. Without surprise also, the selected dimensions and cutoffs increase when the sample size increases. What is remarkable is the values of the selected dimensions for $\beta$-distributions, which are very large. Globally, we can see that these values are very different from one distribution to the other. Contrary to the theoretical result, the Cauchy density is estimated with similar MISEs in the Hermite and sine cardinal methods.

In Fig. 2, density and 25 estimators are plotted for models (iii), (vii) and (ix). Risks and standard deviation for the 25 curves are given above each plot, together with the mean of the selected dimension, cutoff or bandwidth. The methods are comparable, even for the Cauchy distribution, except for the mixtures, where the kernel method fails. The first two lines illustrate the improvement obtained when increasing $n$. We
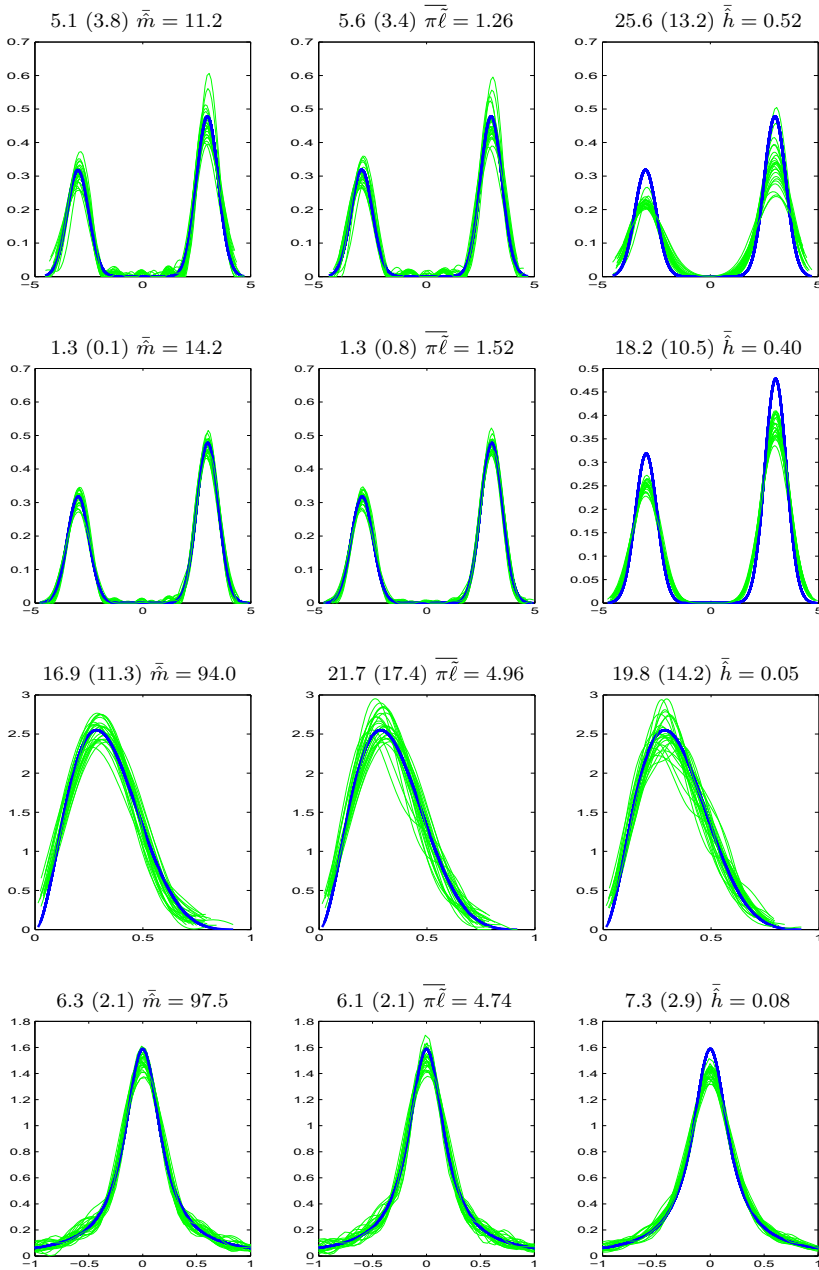
**Fig. 2** True density $f$ in bold blue for Model (iii) (first two lines), Model (vii) (third line) and Model (ix) (fourth line), together with 25 estimates (green/gray) with $n = 250$ (lines 1 and 3) or $n = 1000$ (lines 2 and 4). First column: Hermite; second column: Sine cardinal; third column: kernel. Above each plot: MISE $\times 1000$ and std $\times 1000$ in parenthesis, followed par the mean of selected dimensions, cutoffs and bandwidths (all means over the 25 samples) color figure online

note again that the selected dimensions in the Hermite method are possibly rather high (see the beta and the Cauchy densities). However, computation time remains very short.

## 5 Pure jump Lévy processes

We now look at projection Hermite estimators in a different context, namely the estimation of the Lévy density of a Lévy process.

Let $(L_t, t \geq 0)$ be a real-valued Lévy process, i.e., a process with stationary independent increments with characteristic function of the form:

$$\phi_t(u) = \mathbb{E}(\exp iuL_t) = \exp t\psi(u), \quad \psi(u) = \int_{\mathbb{R}} (e^{iux} - 1)n(x)\mathrm{d}x \qquad (17)$$

where we assume that the Lévy density $n(\cdot)$ satisfies

(H1) $\int_{\mathbb{R}} |x|n(x)\mathrm{d}x < \infty$.

Uner (H1) and (17), the process $(L_t)$ is of pure jump type, has no drift component, finite variation on compacts and satisfies $\mathbb{E}(|L_t|) < +\infty$ (see e.g., Bertoin 1996, Chap. 1). The distribution of $(L_t)$ is entirely specified by $n(\cdot)$, which describes the jumps behavior. We assume that the process is discretely observed with sampling interval $\Delta$ and set $(Z_k = Z_k^\Delta = L_{k\Delta} - L_{(k-1)\Delta}, k = 1, \ldots, n)$ which are independent, identically distributed random variables with common characteristic function $\phi_\Delta(u)$. In Belomestny et al. (2015), second chapter, "Adaptive estimation for Lévy processes," methods of estimation of the function

$$g(x) = xn(x)$$

in this context are presented and studied under the asymptotic framework of high-frequency data, i.e., the sampling interval $\Delta = \Delta_n$ tends to 0 while $n$ and the total length time of observations $n\Delta_n$ tend to infinity.

In here, we consider the same framework and propose estimators of $g$ using the Hermite functions basis. For simplicity, we omit the index $n$ in notations and denote $\Delta = \Delta_n$, $Z_k = Z_k^{\Delta_n}, k = 1, \ldots, n$. The following additional assumptions are required.

(H2) The function $g$ belongs to $\mathbb{L}^2(\mathbb{R})$.
(H3) $\int_{\mathbb{R}} |x|^7 n(x)\mathrm{d}x < \infty$.

Assumption (H2) is obviously needed for the projection method. Assumption (H3) implies that $\mathbb{E}|Z_1|^7 < +\infty$. Let $P_\Delta$ denote the distribution of $Z_1$. In the above reference, the following property is proved (Proposition 3.3, p. 84): the measure

$$\mu_\Delta(\mathrm{d}x) = \frac{1}{\Delta}xP_\Delta(\mathrm{d}x) = g_\Delta(x)\mathrm{d}x \qquad (18)$$

where $g_\Delta(x) = \mathbb{E}g(x - Z_1) = \int g(x - z)P_\Delta(dz)$ and $\mu_\Delta(dx)$ weakly converges to $g(x)dx$ as $\Delta \to 0$. In view of this property, the measure

$$\hat{\mu}_n(dx) = \frac{1}{n\Delta} \sum_{k=1}^{n} Z_k \delta_{Z_k}(dx) \tag{19}$$

will play the role of empirical measure for the estimation of $g$. By (H2), the function $g$ can be developed in the Hermite basis:

$$g = \sum_{j \geq 0} a_j(g)h_j, \qquad a_j(g) = \int g(x)h_j(x)dx = \langle g, h_j \rangle.$$

We define, for $m \geq 0$, the projection estimator $\hat{g}_m$ of $g$ on the space $S_m$ by:

$$\hat{g}_m = \sum_{j=0}^{m-1} \hat{a}_j h_j, \quad \hat{a}_j = \frac{1}{n\Delta} \sum_{k=1}^{n} Z_k h_j(Z_k). \tag{20}$$

Now, $\hat{a}_j$ is no more an unbiased estimator of $a_j(g)$. For $t$ a function, we set when it is well defined,

$$R(t) = \frac{1}{\Delta}\mathbb{E}(Z_1 t(Z_1)) - \int t(x)g(x)dx. \tag{21}$$

Thus, we have $\mathbb{E}\hat{a}_j = a_j(g) + R(h_j)$. The following holds:

**Proposition 11** *Assume that (H1)–(H3) hold. Consider for $m \geq 0$, the estimator $\hat{g}_m$ of $g$ and denote by $g_m$ the orthogonal projection of $g$ on $S_m$. Then*

$$\mathbb{E}\left(\|\hat{g}_m - g\|^2\right) \leq \|g - g_m\|^2 + \frac{V_m}{n\Delta} + \|g\|_1^2 \rho_{m,\Delta}, \tag{22}$$

*where*

$$V_m := \frac{1}{\Delta} \sum_{j=0}^{m-1} \mathbb{E}\left(Z_1^2 h_j^2(Z_1)\right) \leq c\left(\frac{\mathbb{E}(Z_1^2)}{\Delta}\sqrt{m} + \frac{\mathbb{E}(|Z_1|^7)}{\Delta}\log(m)\right), \tag{23}$$

*and*

$$\rho_{m,\Delta} = \begin{cases} \dfrac{C}{2\pi}\Delta^2 & \text{if } C := \int u^2|g^*(u)|^2du < +\infty, \\[2ex] 8\|g\|^2\Delta^2 m & otherwise \end{cases}$$

*($\|g\|_1$ denotes the $\mathbb{L}^1$-norm of $g$).*

First, let us recall some small sample properties of moments and absolute moments of $Z_1$, see e.g., Belomestny et al. (2015). Under (17), (H1) and (H3), it holds that

$$\frac{\mathbb{E}(Z_1^2)}{\Delta} = \int x^2 n(x)\mathrm{d}x + o(1) \quad \text{and} \quad \frac{\mathbb{E}(|Z_1|^7)}{\Delta} = \int |x|^7 n(x)\mathrm{d}x + o(1), \quad (24)$$

where $o(1)$ means that the term tends to 0 as $\Delta$ tends to 0 (see Proposition 3.1 and 3.2, pp. 82–83). Now, let us compare the projection Hermite estimators $(\hat{g}_m, m \geq 0)$ to the estimators studied in the latter reference. In Section 4.1, p. 87, a deconvolution estimator is studied, given by:

$$g_\ell^\dagger(x) = \frac{1}{n\Delta} \sum_{k=1}^n Z_k \frac{\sin(\pi \ell(Z_k - x))}{\pi(Z_k - x)}.$$

Looking at Proposition 4.3, p. 90 and Proposition 4.4, p. 91, we see that the sine cardinal estimators and the projection Hermite estimators are equivalent for $\ell = m^{1/2}$ with the same optimal rates of convergence. Then, in Section 4.2, p. 105, the estimation of $g1_A$ where $A$ is a compact subset of $\mathbb{R}$ is considered by a projection method on finite dimensional subspaces of $\mathbb{L}^2(A)$. Here, on the contrary, the Hermite method gives better results as can be seen from Proposition 4.6, p. 110. The difference lies in the additional bias term $\rho_{m,\Delta}$ which is smaller.

## 6 Concluding remarks

This paper is concerned with the nonparametric estimation of the density of an i.i.d. sample. Although there is an ocean of papers on this topic, it seems that the method developed here has not received yet much attention. Under the assumption that the unknown density belongs to $\mathbb{L}^2(\mathbb{R})$, we build and study projection estimators using an orthonormal basis composed of Hermite functions. Usually, for projection estimators, the variance term of the $\mathbb{L}^2$-risk is proportional to the dimension of the projection space. The special feature of the Hermite function basis is that the variance term is governed by the square root of the dimension. Moreover, we introduce specific regularity function spaces to evaluate the order of the bias term, namely the Sobolev Hermite spaces. This allows to prove that Hermite estimators are asymptotically equivalent to sine cardinal estimators. From the practical point of view, Hermite estimators are much faster to compute.

Adaptive estimators are studied, using an appropriate data-driven choice of the dimension.

This paper is only concerned with $\mathbb{L}^2$-risks, but $\mathbb{L}^p$-risks have also been studied by many authors (see the classical reference Donoho et al. 1996). Moreover, the $\mathbb{L}^1$-approach is especially developed in Devroye and Györfi (1985). In this setting, adaptive estimators have been constructed by Devroye and Lugosi (2001). The study of $\mathbb{L}^p$-risks with the Hermite approach would be an interesting field of further investigation.

## 7 Proofs

### 7.1 Proof of Propositions 1 and 2

We start by proving Proposition 1.

(i). The following bound comes from Szegö (1975, p. 242) where an expression of $C_\infty$ is given:

$$\forall x \in \mathbb{R}, \quad |h_j(x)| \leq C_\infty (j+1)^{-(1/12)}, \quad j = 0, 1, \ldots. \tag{25}$$

Therefore, $V_m \leq C_\infty^2 \sum_{j=0}^{m-1} (j+1)^{-(1/6)} \leq \frac{6}{5} C_\infty^2 m^{5/6}$.

(ii). Now, as in Walter (1977), we use the following expression for the Hermite function $h_n$ (see Szegö (1975, p. 248)):

$$h_j(x) = \lambda_j \cos\left((2j+1)^{1/2} x - \frac{j\pi}{2}\right) + \frac{1}{(2j+1)^{1/2}} \xi_j(x) \tag{26}$$

where $\lambda_j = |h_j(0)|$ if $j$ is even, $\lambda_j = |h_j^{'}(0)|/(2j+1)^{1/2}$ if $j$ is odd and

$$\xi_j(x) = \int_0^x \sin\left[(2j+1)^{1/2}(x-t)\right] t^2 h_j(t) \mathrm{d}t. \tag{27}$$

By the Cauchy–Schwarz inequality, $\xi_j^2(x) \leq \int_0^{|x|} t^4 \mathrm{d}t \int_0^{|x|} h_j^2(t) \mathrm{d}t \leq \frac{|x|^5}{5} \times \frac{1}{2}$. Moreover,

$$\lambda_{2j} = \frac{(2j)!^{1/2}}{2^j j! \pi^{1/4}}, \quad \lambda_{2j+1} = \lambda_{2j} \frac{\sqrt{2j+1}}{\sqrt{2j+3/2}}.$$

By the Stirling formula and its proof, $\lambda_{2j} \sim \pi^{-1/2} j^{-1/4}$, $\lambda_{2j+1} \sim \pi^{-1/2} j^{-1/4}$ and for all $j$, there exists constants $c_1, c_2$ such that, for all $j \geq 1$,

$$\frac{c_1}{\pi^{1/2} j^{1/4}} \leq \lambda_j \leq \frac{c_2}{\pi^{1/2} j^{1/4}}. \tag{28}$$

Therefore, $h_j^2(x) \leq 2 \frac{c_2^2}{\pi j^{1/2}} + \frac{1}{2j+1} \frac{|x|^5}{5}$. This yields:

$$\int h_j^2(x) f(x) \mathrm{d}x \leq 2 \frac{c_2^2}{\pi j^{1/2}} + \frac{1}{5(2j+1)} \mathbb{E}|X|^5,$$

which implies $V_m \lesssim m^{1/2}$.

Now, we study case (iii). The following bound for $h_j$ is given in Markett (1984, p. 190): There exist positive constants $C, \gamma$, independent of $x$ and $j$, such that, for $J = 2j+1$,

$$|h_j(x)| \leq C(J^{1/3} + |x^2 - J|)^{-1/4}, \quad x^2 \leq 2J,$$
$$\leq C \exp(-\gamma x^2), \quad x^2 > 2J.$$

Consider a sequence $(a_j)$ such that $a_j \to +\infty$, $a_j/\sqrt{j} \to 0$ with $J = 2j + 1$ large enough to ensure $\frac{a_J}{\sqrt{J}} \leq 1/\sqrt{2}$, $a_J \geq K$. As $\int h_j^2(x)dx = 1$, $a_J < \sqrt{J}$, $a_J \geq K$ and $g$ is decreasing,

$$\int h_j^2(x) f(x)dx \leq 2C\|f\|_\infty \int_0^{a_J} (J^{1/3} + J - x^2)^{-1/2}dx + g(a_J).$$

Set $x = (J^{1/3} + J)^{1/2}y$ in the integral. This yields:

$$\int_0^{a_J} \frac{dx}{\sqrt{J^{1/3} + J - x^2}} = \int_0^{a_J/(J^{1/3}+J)^{1/2}} \frac{dy}{\sqrt{1 - y^2}} = \text{Arcsin}\left(\frac{a_J}{(J^{1/3} + J)^{1/2}}\right) \leq 2\frac{a_J}{\sqrt{J}},$$

as for $0 \leq x \leq 1/\sqrt{2}$, $\text{Arcsin}\,x \leq 2x$. Now, we choose the sequence $(a_j)$ and consider $a_j = j^{1/(2(a+1))}$. We deduce $\int h_j^2(x) f(x)dx \lesssim j^{-a/(2(a+1))}$, which leads to

$$V_m \lesssim m^{\frac{a+2}{2(a+1)}}. \tag{29}$$

□

Now we turn to the proof of Proposition 2 and we look at the lower bound. We have, setting $c = \inf_{a \leq x \leq b} f(x)$, and using (26),

$$\int h_j^2(x) f(x)dx \geq c \int_a^b h_j^2(x)dx$$

$$\geq c\lambda_j^2 \int_a^b \cos^2\left((2j + 1)^{1/2}x - \frac{j\pi}{2}\right)dx$$

$$+ c\frac{2\lambda_j}{(2j + 1)^{1/2}} \int_a^b \cos\left((2j + 1)^{1/2}x - \frac{j\pi}{2}\right)\xi_j(x)dx.$$

We have $j^{-3/4}c_1/\sqrt{\pi} \leq \frac{2\lambda_j}{(2j+1)^{1/2}} \leq j^{-3/4}\sqrt{2/\pi}c_2$ and

$$\left|\int_a^b \cos\left((2j + 1)^{1/2}x - \frac{j\pi}{2}\right)\xi_j(x)dx\right| \leq \int_a^b \frac{|x|^{5/2}}{\sqrt{10}}dx := C.$$

Thus, the second term is lower bounded by $-Cj^{-3/4}c_1/\sqrt{\pi}$. For the first term, $\lambda_j^2 \geq j^{-1/2}c_1^2/\pi$ and

$$\int_a^b \cos^2\left((2j + 1)^{1/2}x - \frac{j\pi}{2}\right)dx = \frac{1}{2}(b - a) + \int_a^b \cos\left(2(2j + 1)^{1/2}x - j\pi\right)dx$$

$$= \frac{1}{2}(b - a) + O\left(\frac{1}{j^{1/2}}\right).$$

Therefore, $\int h_j^2(x) f(x) \mathrm{d}x \geq c j^{-1/2} c_1^2 / \pi \left[ \frac{b-a}{2} + O(\frac{1}{j^{1/2}}) \right] - C j^{-3/4} c_1 / \sqrt{\pi}$. Consequently, for $j$ large enough, $\int h_j^2(x) f(x) \mathrm{d}x \geq c' j^{-1/2}$. This implies, $V_m \geq c'' m^{1/2}$. $\qquad\square$

### 7.2 Proof of Theorem 1

Let $S_m$ be the space spanned by $\{h_0, \ldots, h_{m-1}\}$ and $B_m = \{t \in S_m, \|t\| = 1\}$. We have $\hat{f}_m = \arg\min_{t \in S_m} \gamma_n(t)$ where $\gamma_n(t) = \|t\|^2 - 2n^{-1} \sum_{i=1}^n t(X_i)$ and $\gamma_n(\hat{f}_m) = -\|\hat{f}_m\|^2$. Now, we write, for two functions $t, s \in \mathbb{L}^2(\mathbb{R})$ ,

$$\gamma_n(t) - \gamma_n(s) = \|t - f\|^2 - \|s - f\|^2 - 2\nu_n(t - s)$$

where

$$\nu_n(t) = \frac{1}{n} \sum_{i=1}^n [t(X_i) - \langle t, f \rangle].$$

Then, for any $m \in \mathcal{M}_n = \{1 \leq m \leq m_n\}$, $m_n \leq n / \log n$, and any $f_m \in S_m$,

$$\gamma_n(\hat{f}_{\hat{m}}) + \widehat{\mathrm{pen}}(\hat{m}) \leq \gamma_n(f_m) + \widehat{\mathrm{pen}}(m).$$

This yields $\|\hat{f}_{\hat{m}} - f\|^2 \leq \|f - f_m\|^2 + \widehat{\mathrm{pen}}(m) - \widehat{\mathrm{pen}}(\hat{m}) + 2\nu_n(\hat{f}_{\hat{m}} - f_m)$. We use that

$$2\nu_n(\hat{f}_{\hat{m}} - f_m) \leq 4 \sup_{t \in B_{m \vee \hat{m}}} \nu_n^2(t) + \frac{1}{4} \|\hat{f}_{\hat{m}} - f_m\|^2,$$

and some classical algebra to obtain:

$$\frac{1}{2} \|\hat{f}_{\hat{m}} - f\|^2 \leq \frac{3}{2} \|f - f_m\|^2 + \widehat{\mathrm{pen}}(m) + 4 \left( \sup_{t \in B_{m \vee \hat{m}}} \nu_n^2(t) - p(m \vee \hat{m}) \right)$$
$$+ (4p(m \vee \hat{m}) - \mathrm{pen}(\hat{m})) + (\mathrm{pen}(\hat{m}) - \widehat{\mathrm{pen}}(\hat{m})). \qquad (30)$$

We can choose $p(m)$ such that

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E} \left( \sup_{t \in B_{m \vee m'}} \nu_n^2(t) - p(m \vee m') \right)_+ \leq \frac{c}{n}. \qquad (31)$$

Indeed, for this, we apply the Talagrand Inequality (see Klein and Rio 2005):

$$\mathbb{E} \left( \sup_{t \in B_m} \nu_n^2(t) - 4H^2 \right)_+ \leq \frac{C_1}{n} \left( v^2 \mathrm{e}^{-C_2 \frac{nH^2}{v^2}} + \frac{M_1^2}{n} \mathrm{e}^{-C_3 \frac{nH}{M_1}} \right)$$

where $\mathbb{E}\left(\sup_{t\in B_m} v_n^2(t)\right) \leq \frac{V_m}{n} := H^2$, $\sup_{t\in B_m} \mathrm{Var}(t(X_1)) \leq \sup_{t\in B_m} \mathbb{E}(t^2(X_1)) \leq \|f\|_\infty := v^2$ and $\sup_{t\in B_m}\sup_x |t(x)| \leq \sqrt{\sup_x \sum_{j=0}^{m-1} h_j^2(x)} \leq C'_\infty m^{5/12} \leq C'_\infty \sqrt{n} := M_1$ (see (25)). Therefore, we obtain

$$\mathbb{E}\left(\sup_{t\in B_m} v_n^2(t) - 4\frac{V_m}{n}\right)_+ \leq \frac{C_1}{n}\left(\|f\|_\infty e^{-C'_2 V_m} + e^{-C'_3\sqrt{V_m}}\right).$$

Therefore, with the choice $p(m) = 4V_m/n$, (31) holds under condition (6) which is ensured by Proposition 2. Taking expectation in (30) yields

$$\frac{1}{2}\mathbb{E}(\|\hat{f}_{\hat{m}} - f\|^2) \leq \frac{3}{2}\|f - f_m\|^2 + \mathrm{pen}(m) + \mathbb{E}(4p(m\vee\hat{m}) - \mathrm{pen}(\hat{m}))$$
$$+ \mathbb{E}(\mathrm{pen}(\hat{m}) - \widehat{\mathrm{pen}}(\hat{m}))_+ + \frac{c}{n}. \tag{32}$$

Let us define

$$Y_i^{(m)} := \sum_{j=0}^{m-1} h_j^2(X_i), \quad \widehat{V}_m = \frac{1}{n}\sum_{i=1}^n Y_i^{(m)},$$

and the set inspired by Bernstein Inequality $\Omega :=$

$$\left\{\forall m \in \mathcal{M}_n, \frac{1}{n}\left|\sum_{i=1}^n\left(Y_i^{(m)} - \mathbb{E}\left(Y_i^{(m)}\right)\right)\right| \leq \sqrt{2V_m C''_\infty m^{5/6}\frac{\log(n)}{n}} + 4C''_\infty m^{5/6}\frac{\log(n)}{3n}\right\}$$

with $C''_\infty := (C'_\infty)^2$ and $C'_\infty$ is the constant appearing in $M_1$ above. We split the term to study in (32) as follows:

$$\mathbb{E}(\mathrm{pen}(\hat{m}) - \widehat{\mathrm{pen}}(\hat{m}))_+ \leq \mathbb{E}\left[(\mathrm{pen}(\hat{m}) - \widehat{\mathrm{pen}}(\hat{m}))_+\mathbf{1}_\Omega\right] + \mathbb{E}\left[(\mathrm{pen}(\hat{m}) - \widehat{\mathrm{pen}}(\hat{m}))_+\mathbf{1}_{\Omega^c}\right].$$

On $\Omega$,

$$|\widehat{V}_{\hat{m}} - V_{\hat{m}}| \leq \sqrt{2V_{\hat{m}} C''_\infty \hat{m}^{5/6}\log(n)/n} + 4C''_\infty \hat{m}^{5/6}\log(n)/(3n)$$
$$\leq \frac{1}{2}V_{\hat{m}} + \frac{7}{3}C''_\infty\frac{\hat{m}^{5/6}\log(n)}{n},$$

using that $2xy \leq x^2 + y^2$ applied to $\sqrt{2VA} = 2\sqrt{V/2}\sqrt{A} \leq V/2 + A$ with $V = V_{\hat{m}}$ and $A = C''_\infty\hat{m}^{5/6}\log(n)/n$. Thus, by definition of $\mathcal{M}_n$,

$$\mathbb{E}\left[(\mathrm{pen}(\hat{m}) - \widehat{\mathrm{pen}}(\hat{m}))_+\mathbf{1}_\Omega\right]_+ \leq \frac{1}{2}\mathbb{E}(\mathrm{pen}(\hat{m})) + \frac{c}{n}.$$

On the other hand, $\mathbb{E}\left[(\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m}))_+ \mathbf{1}_{\Omega^c}\right] \leq 2\kappa \mathbb{P}(\Omega^c)$. Now, by applying Bernstein inequality, we get

$$\mathbb{P}(\Omega^c) \leq \sum_{m \in \mathcal{M}_n} 2\mathrm{e}^{-2\log(n)} \leq \frac{c}{n}.$$

Indeed, we have $\mathbb{P}(|S_n/n| \geq \sqrt{2v^2 x/n} + bx/(3n)) \leq 2\mathrm{e}^{-x}$ for $S_n = \sum_{i=1}^n (U_i - \mathbb{E}(U_i))$, $\text{Var}(U_1) \leq v^2$, $|U_i| \leq b$. In our case $U_i = Y_i^{(m)}$ and $v^2 = V_m C_\infty'' m^{5/6}$, $b = C_\infty'' m^{5/6}$ and we took $x = 2\log(n)$.

So Eq. (32) becomes

$$\frac{1}{2}\mathbb{E}(\|\hat{f}_{\hat{m}} - f\|^2) \leq \frac{3}{2}\|f - f_m\|^2 + \text{pen}(m) + \mathbb{E}(4p(m \vee \hat{m}) - \text{pen}(\hat{m}))$$
$$+ \frac{1}{2}\mathbb{E}(\text{pen}(\hat{m})) + \frac{c}{n}$$
$$\leq \frac{3}{2}\|f - f_m\|^2 + \text{pen}(m) + \mathbb{E}\left(4p(m \vee \hat{m}) - \frac{1}{2}\text{pen}(\hat{m})\right) + \frac{c}{n}.$$

Now we note that, for $\kappa \geq 8 := \kappa_0$, $4p(m \vee \hat{m}) - \frac{1}{2}\text{pen}(\hat{m}) \leq \text{pen}(m)$. Finally, we get, for all $m \in \mathcal{M}_n$,

$$\mathbb{E}(\|\hat{f}_{\hat{m}} - f\|^2) \leq 3\|f - f_m\|^2 + 4\text{pen}(m) + \frac{c}{n},$$

which ends the proof.                                                                                                              $\square$

### 7.3 Proof of Proposition 3

The first inequality is standard. Let us study $\tilde{f}_\ell^{(n)}(x)$. We write that

$$\|\tilde{f}_\ell^{(n)} - f\|^2 = \|\tilde{f}_\ell^{(n)} - \mathbb{E}\tilde{f}_\ell^{(n)}\|^2 + \|\mathbb{E}(\tilde{f}_\ell^{(n)}) - f\|^2$$
$$\leq \|\tilde{f}_\ell^{(n)} - \mathbb{E}\tilde{f}_\ell^{(n)}\|^2 + 2\|\mathbb{E}(\tilde{f}_\ell^{(n)}) - \bar{f}_\ell\|^2 + 2\|\bar{f}_\ell - f\|^2.$$

The term $\|\bar{f}_\ell - f\|^2$ is the usual bias term. Moreover,

$$\mathbb{E}\left(\|\tilde{f}_\ell^{(n)} - \mathbb{E}\tilde{f}_\ell^{(n)}\|^2\right) = \sum_{|j| \leq K_n} \text{Var}(\tilde{a}_{\ell,j}) = \frac{1}{n} \sum_{|j| \leq K_n} \text{Var}(\varphi_{\ell,j}(X_1))$$
$$\leq \frac{1}{n} \sum_{|j| \leq K_n} \mathbb{E}[\varphi_{\ell,j}^2(X_1)] \leq \frac{\ell}{n}$$

because $\sum_{j \in \mathbb{Z}} |\varphi_{\ell,j}(x)|^2 \leq \ell$. This is the standard variance term order.

The new term is

$$\|\mathbb{E}(\widetilde{f}_\ell^{(n)}) - \bar{f}_\ell\|^2 = \sum_{|j| \geq K_n} |a_{\ell,j}|^2 \leq 2 \sup_j |j a_{\ell,j}|^2 \sum_{j > K_n} j^{-2} \leq \frac{2}{K_n} \sup_j |j a_{\ell,j}|^2. \tag{33}$$

We write that $j a_{\ell,j} = j\sqrt{\ell} \int \varphi(\ell x - j) f(x) \mathrm{d}x = \sqrt{\ell}(I_1 + I_2)$ where

$$I_1 = \ell \int x \varphi(\ell x - j) f(x) \mathrm{d}x, \quad I_2 = -\int (\ell x - j)\varphi(\ell x - j) f(x) \mathrm{d}x$$

and we bound $I_1$ and $I_2$.

$$|I_1| \leq \ell \sqrt{\int |\varphi(\ell x - j)|^2 \mathrm{d}x \int x^2 f^2(x) \mathrm{d}x} = \sqrt{\ell}\sqrt{M_2}, \quad \text{where } M_2 = \int x^2 f^2(x) \mathrm{d}x.$$

On the other hand, $|I_2| \leq \sup_{u \in \mathbb{R}} |u\varphi(u)| \int f(x)\mathrm{d}x \leq 1$. We obtain:

$$|j a_{\ell,j}| \leq \ell\sqrt{M_2} + \sqrt{\ell} \leq \ell(\sqrt{M_2} + 1).$$

Plugging this in (33), we find the bound: $\|\mathbb{E}(\widetilde{f}_\ell^{(n)}) - \bar{f}_\ell\|^2 \leq 4\ell^2(M_2 + 1)/K_n$. This term is $O(\ell/n)$ if $\ell \leq n$ and $K_n \geq n^2$. □

### 7.4 Proof of Proposition 4

Using the relations (see e.g., Abramowitz and Stegun 1964):

$$2x H_n(x) = H_{n+1}(x) + 2n H_{n-1}(x), \quad H_n'(x) = 2n H_{n-1}(x), \quad n \geq 1, \tag{34}$$

we get:

$$A_+ h_n = \sqrt{2n} h_{n-1}, \quad A_- h_n = \sqrt{2(n + 1)} h_{n+1}.$$

We deduce:

$$\sqrt{2} h_n' = \sqrt{n}\, h_{n-1} - \sqrt{n+1} h_{n+1}, \quad 2x\, h_n = \sqrt{2(n + 1)}\, h_{n+1} + \sqrt{2n}\, h_{n-1}, \tag{35}$$

Assume first that $f \in \mathbb{L}^2(\mathbb{R})$, $f$ admits derivatives up to order s, and for $j_1, \ldots, j_m \in \{-, +\}$ and $1 \leq m \leq s$, $A_{j_1} \ldots A_{j_m} f \in \mathbb{L}^2(\mathbb{R})$. We prove that $\sum_{n \geq 0} n^s a_n^2(f) < +\infty$. We do the proof only for $f$ compactly supported and refer to Bongioanni and Torrea (2006) otherwise.
For the proof, set $A_{-1} = A_-$, $A_{+1} = A_+$ so that, for $n - j \geq 0$, $A_j h_n = \sqrt{2(n + d_j)} h_{n-j}, d_j = 0$ if $j = 1, d_j = 1$ if $j = -1$. Thus, for $n - j_1 - j_2 - \cdots - j_m \geq 0$,

$$A_{j_1} \ldots A_{j_m} h_n = \sqrt{2(n + d_{j_1})} \times \cdots \times \sqrt{2(n + d_{j_m})} h_{n-j_1-j_2-\cdots-j_m}.$$

Now, for $f$ compactly supported,

$$\langle A_j f, h_n \rangle = \langle f, A_{-j} h_n \rangle = \sqrt{2(n + d_{-j})} \langle f, h_{n+j} \rangle.$$

Iterating yields, for $n + j_1 + j_2 + \cdots + j_m \geq 0$,

$$\langle A_{j_1} \ldots A_{j_m} f, h_n \rangle = \langle f, A_{-j_m} A_{-j_{m-1}} \ldots A_{-j_1} h_n \rangle$$
$$= \prod_{1 \leq k \leq m} \sqrt{2(n + d_{-j_k})} \langle f, h_{n+j_1+j_2+\cdots+j_m} \rangle.$$

Therefore, $\sum_{n \geq 0} (\langle A_{j_1} \ldots A_{j_m} f, h_n \rangle)^2 < +\infty$ is equivalent to

$$\sum_{n+j_1+j_2+\cdots+j_m \geq 0} n^m a_{n+j_1+j_2+\cdots+j_m}^2 (f) < \infty.$$

Now assume that $\sum_{n \geq 0} n a_n^2(f) < +\infty$. We have $f = \sum_{n \geq 0} a_n(f) h_n$. We can write for $n_1$ large enough:

$$\left| \sum_{n=n_1}^{n_1+n_2} a_n(f) h_n(x) \right| \leq \left( \sum_{n=n_1}^{n_1+n_2} n^{7/6} a_n^2(f) h_n^2(x) \sum_{n=n_1}^{n_1+n_2} n^{-7/6} \right)^{1/2} \leq C \sum_{n=n_1}^{n_1+n_2} n a_n^2(f).$$

Thus, the series for $f$ converges uniformly, $f$ is continuous and satisfies for all $x$, $f(x) = \sum_{n \geq 0} a_n(f) h_n(x)$. Therefore, we have:

$$f(y) - f(x) = \sum_{n \geq 0} a_n(f) \int_x^y h_n'(t) dt$$
$$= a_0(f)(h_0(x) - h_0(y)) + 2^{-1/2} \sum_{n \geq 1} a_n(f)$$
$$\times \int_x^y (\sqrt{n}\, h_{n-1}(t) - \sqrt{n+1} h_{n+1}(t)) dt.$$

Set $S_N(t) = \sum_{n=1}^{N} a_n(f)(\sqrt{n}\, h_{n-1}(t) - \sqrt{n+1} h_{n+1}(t))$ and $S(t) = \sum_{n \geq 1} a_n(f)$ $(\sqrt{n}\, h_{n-1}(t) - \sqrt{n+1} h_{n+1}(t))$. The function $S(t)$ is well defined by assumption and $S_N$ converges to $S$ in $\mathbb{L}^2(\mathbb{R})$. Therefore, as $N$ tends to infinity, $\int_x^y |S_N(t) - S(t)| dt \leq \sqrt{y - x} \|S_N - S\| \to 0$. We have proved that

$$f(y) - f(x) = a_0(f) \int_x^y h_0'(t) dt + \int_x^y S(t) dt.$$

Thus, $f$ is absolutely continuous and $f' = S$ belongs to $\mathbb{L}^2(\mathbb{R})$. Analogously, we prove that $xf$ belongs to $\mathbb{L}^2(\mathbb{R})$. Thus, $A_+ f$, $A_- f$ belong to $\mathbb{L}^2(\mathbb{R})$.

Next, by the same reasoning as above, using that $\sum_n n^2 a_n(f) < +\infty$ the series for $f'(t) = S(t)$ is uniformly convergent and $f'(t)$ is continuous. We proceed analogously to prove that $f'$ is absolutely continuous and that $xf'$ and $f''$ belong to $\mathbb{L}^2(\mathbb{R})$. Iterating

the reasoning, we obtain that $f$ admits continuous derivatives up to $s-1$ and that $f^{(s-1)}$ is absolutely continuous and that $f, f', \ldots, f^{(s)}, x^{k-m} f^{(k-m)}, m = 0, \ldots, s-1$ all belong to $\mathbb{L}^2(\mathbb{R})$. This shows that, for $j_1, \ldots, j_m \in \{-, +\}$, $1 \le m \le s$, $A_{j_1} \ldots A_{j_m} f$ belongs to $\mathbb{L}^2(\mathbb{R})$.    □

### 7.5 Proof of Proposition 7

Let $f_\mu$ be the density $\mathcal{N}(\mu, 1)$, then

$$a_j(f_\mu) = \langle f_\mu, h_j \rangle = \int_{\mathbb{R}} h_j(x) e^{-\frac{1}{2}(x-\mu)^2} \frac{\mathrm{d}x}{\sqrt{2\pi}} = \frac{c_j}{\sqrt{2\pi}} \int_{\mathbb{R}} H_j(x) e^{-\frac{x^2}{2} - \frac{1}{2}(x-\mu)^2} \mathrm{d}x$$

$$= \frac{c_j}{\sqrt{2\pi}} e^{-\frac{\mu^2}{4}} \int_{\mathbb{R}} H_j(x) e^{-\left(x-\frac{\mu}{2}\right)^2} \mathrm{d}x = \frac{c_j}{\sqrt{2\pi}} e^{-\frac{\mu^2}{4}} \int_{\mathbb{R}} H_j\left(v + \frac{\mu}{2}\right) e^{-v^2} \mathrm{d}v.$$

Now, we use the following formula, obtained by the Taylor formula and the recurrence relation $H_n' = 2n H_{n-1}$:

$$H_j(x + y) = \sum_{k=0}^{n} \binom{n}{k} (2x)^k H_{n-k}(y).$$

This yields

$$a_j(f_\mu) = \frac{c_j}{\sqrt{2\pi}} e^{-\frac{\mu^2}{4}} \sum_{k=0}^{j} \binom{j}{k} \mu^{j-k} \int_{\mathbb{R}} H_k(v) e^{-v^2} \mathrm{d}v$$

and $\int_{\mathbb{R}} H_k(v) e^{-v^2} \mathrm{d}v = (1/(c_k c_0)) \langle h_k, h_0 \rangle = 0$ if $k \ne 0$. Therefore, we get

$$a_j(f_\mu) = \frac{c_j}{\sqrt{2\pi}} \frac{\mu^j}{c_0^2} e^{-\frac{\mu^2}{4}} = \frac{\mu^j}{\sqrt{2\pi^{1/2} 2^j j!}} e^{-\frac{\mu^2}{4}}. \tag{36}$$

Then

$$\|f_\mu - (f_\mu)_m\|^2 = \sum_{j \ge m} a_j^2(f_\mu) = \frac{e^{-\mu^2/2}}{2\sqrt{\pi}} \sum_{j \ge m} \frac{\mu^{2j}}{2^j j!} = \frac{e^{-\mu^2/2}}{2\sqrt{\pi}} \sum_{j \ge 0} \frac{\mu^{2(j+m)}}{2^{j+m}(j+m)!}$$

$$\le \frac{1}{2\sqrt{\pi} m!} \left(\frac{\mu^2}{2}\right)^m \quad \text{as} \quad \frac{1}{(m+j)!} \le \frac{1}{m! j!}.$$

Using Stirling's formula, we get

$$\|f_\mu - (f_\mu)_m\|^2 \lesssim \frac{1}{2\pi \sqrt{2m}} \exp\left(-m \log\left(\frac{2m}{e\mu^2}\right)\right).$$

Therefore $m_{\mathrm{opt}} = [(\log(n)/\log(2)) + e\mu^2]$ yields $\|f_\mu - (f_\mu)_{m_{\mathrm{opt}}}\|^2 \lesssim 1/(n\sqrt{\log(n)})$. Combining with Proposition 1, we obtain the first result.

To prove the second result, we use the following proposition.

**Proposition 12** *Recall that $a_j(f) = \int f(x)h_j(x)\mathrm{d}x$. For $j \geq 0$, we have:*

$$a_{2j}(f_\sigma) = c_{2j}\left(\frac{1}{1+\sigma^2}\right)^{1/2}\frac{(2j)!}{j!}\left(\frac{\sigma^2-1}{\sigma^2+1}\right)^j, \quad a_{2j+1}(f_\sigma) = 0.$$

*For $n \geq p$, $j \geq 0$,*

$$|a_{2j}(f_{p,\sigma})| \leq C(p,\sigma^2)c_{2j}\frac{(2j)!}{(j-p)!}\left|\frac{\sigma^2-1}{\sigma^2+1}\right|^{j-p}, \quad a_{2j+1}(f_{p,\sigma}) = 0.$$

We can now deduce the risk of $\hat{f}_m$ when $f = f_\sigma$. We have:

$$a_{2j}^2(f_\sigma) \sim \pi^{-1}j^{-1/2}\frac{1}{1+\sigma^2}\left(\frac{\sigma^2-1}{\sigma^2+1}\right)^{2j}. \tag{37}$$

Therefore, setting $\lambda = \log\left[\left(\frac{\sigma^2+1}{\sigma^2-1}\right)^2\right]$ yields $\|f - f_m\|^2 \lesssim \frac{1}{\sqrt{m}}\exp(-\lambda m)$. Combining with Proposition 1, we obtain $\mathbb{E}(\|\hat{f}_m - f\|^2) \lesssim \frac{1}{\sqrt{m}}\exp(-\lambda m) + n^{-1}\sqrt{m}$, and thus Proposition 7. $\square$

*Proof of Proposition 12.* We first compute the coefficients of the centered Gaussian density. As Hermite polynomials of odd index are odd, the coefficients with odd index are null. We compute the coefficients with even index. Let

$$\bar{\sigma}^2 = (1 + \sigma^{-2})^{-1} = \frac{\sigma^2}{1+\sigma^2}. \tag{38}$$

Note that if $2\bar{\sigma}^2 = 1$, i.e., $\sigma^2 = 1$, the coefficients are null except for $n = 0$. We have

$$\int x^{2p} f_{\bar{\sigma}}(x)\mathrm{d}x = C_{2p}\bar{\sigma}^{2p} \quad \text{with} \quad C_{2p} = 3 \times 5 \times 7 \times \cdots \times (2p-1) = \frac{(2p)!}{2^p p!}.$$

Using that (see e.g., Lebedev 1972, formula (4.9.2), p. 60)

$$H_{2j}(x) = \sum_{k=0}^{j}\frac{(-1)^k(2j)!}{k!(2j-2k)!}(2x)^{2j-2k},$$

we obtain:

$$a_{2j}(f_\sigma) = (2j)! c_{2j} \frac{\bar\sigma}{\sigma} \sum_{k=0}^{j} \frac{(-1)^k}{k!(2j-2k)!} 2^{2j-2k} C_{2(j-k)} \bar\sigma^{2(j-k)}$$

$$= c_{2j} \frac{\bar\sigma}{\sigma} \frac{(2j)!}{j!} (2\bar\sigma^2 - 1)^j$$

$$= c_{2j} \left( \frac{1}{1+\sigma^2} \right)^{1/2} \frac{(2j)!}{j!} \left( \frac{\sigma^2 - 1}{\sigma^2 + 1} \right)^j.$$

Note that $|(\sigma^2 - 1)/(1 + \sigma^2)| < 1$. Analogously,

$$a_{2j}(f_{p,\sigma}) = \frac{(2j)!}{j!} c_{2j} \left( \frac{\bar\sigma}{\sigma} \right)^{2p+1} \sum_{k=0}^{j} \frac{(-1)^k j! \, 2^{2j-2k} \bar\sigma^{2(j-k)}}{k!(2j-2k)! C_{2p}} C_{2(j-k+p)}$$

$$= \frac{(2j)!}{j!} c_{2j} \left( \frac{\bar\sigma}{\sigma} \right)^{2p+1} \sum_{k=0}^{j} \frac{(-1)^k j!}{k!(j-k)!} (2\bar\sigma^2)^{j-k} \frac{C_{2(j-k+p)}}{C_{2(j-k)} C_{2p}}$$

$$= \frac{(2j)!}{j!} c_{2j} \left( \frac{\bar\sigma}{\sigma} \right)^{2p+1} \sum_{m=0}^{j} \frac{(-1)^{j-m} j!}{m!(j-m)!} (2\bar\sigma^2)^m \frac{C_{2(m+p)}}{C_{2m} C_{2p}}.$$

Now, we use the following result which is proved in Chaleyat-Maurel and Genon-Catalot (2006, Lemma 3.1, p. 1459):

$$\frac{C_{2(m+p)}}{C_{2m} C_{2p}} = \sum_{r=0}^{p} m(m-1)\dots(m-r+1) \binom{p}{r} \frac{2^r}{C_{2r}}.$$

After some computations, we get:

$$a_{2j}(f_{p,\sigma}) = \frac{(2j)!}{(j-p)!} c_{2j} \left( \frac{\bar\sigma}{\sigma} \right)^{2p+1} (2\bar\sigma^2 - 1)^{j-p} S_p \quad \text{where}$$

$$S_p = \sum_{r=0}^{p} \frac{(n-p)! \, p! \, 2^r}{(n-r)! \, j!(p-r)! C_{2r}} (2\bar\sigma^2)^r (2\bar\sigma^2 - 1)^{p-r}.$$

Therefore,

$$|S_p| \le c(p) \left( 2\bar\sigma^2 + |2\bar\sigma^2 - 1| \right)^p,$$

which allows to bound $|a_{2j}(f_{p,\sigma})|$ and ends the proof. $\qquad\square$

### 7.6 Proof of Proposition 8

Let $f \in \mathcal{F}_{\text{sub}}(C)$. We have from (36),

$$a_j(f) = \langle h_j, f \rangle = \mathbb{E}(h_j(X + M)) \text{ where } X \sim \mathcal{N}(0, 1), M \sim \Pi,$$
$$\text{and } X, M \text{ independent,}$$
$$= \mathbb{E}\left(\mathbb{E}(h_j(X + M)|M)\right) = \frac{1}{\sqrt{2\pi^{1/2}2^j j!}} \mathbb{E}\left(M^j e^{-\frac{M^2}{4}}\right).$$

Therefore

$$2\sqrt{\pi} \sum_{j \geq m} a_j^2(f) = \sum_{j \geq m} \frac{1}{j! 2^j} \mathbb{E}^2\left(M^j e^{-\frac{M^2}{4}}\right)$$

$$\leq \sum_{j \geq m} \frac{1}{j! 2^j} \mathbb{E}\left(M^{2j} e^{-\frac{M^2}{2}} \mathbf{1}_{M^2/2 \leq \lambda m}\right) + \mathbb{E}\left(\mathbf{1}_{M^2/2 > \lambda m}\right)$$

Let us look at the first term of the sum above.

$$\mathbb{E}\left(\sum_{j \geq m} \frac{M^{2j}}{j! 2^j} e^{-\frac{M^2}{2}} \mathbf{1}_{M^2/2 \leq \lambda m}\right) = \mathbb{E}\left(\left(\frac{M^2}{2}\right)^m e^{-\frac{M^2}{2}} \sum_{j \geq 0} \frac{1}{(j + m)!}\left(\frac{M^2}{2}\right)^j \mathbf{1}_{M^2/2 \leq \lambda m}\right)$$

$$\leq \mathbb{E}\left(\frac{\left(\frac{M^2}{2}\right)^m}{m!} e^{-\frac{M^2}{2}} \sum_{j \geq 0} \frac{1}{j!}\left(\frac{M^2}{2}\right)^j \mathbf{1}_{M^2/2 \leq \lambda m}\right)$$

$$= \mathbb{E}\left(\frac{\left(\frac{M^2}{2}\right)^m}{m!} \mathbf{1}_{M^2/2 \leq \lambda m}\right)$$

where we used that $1/(j + m)! \leq 1/(j! \, m!)$. Now, gathering the two terms again, we get

$$\sum_{j \geq m} a_j^2(f) \leq \frac{1}{2\pi^{1/2}}\left(\frac{(\lambda m)^m}{m!} + \Pi(|u| > \sqrt{2\lambda m})\right)$$

$$\leq \frac{1}{2\pi^{1/2}}\left(\frac{(\lambda m)^m}{m!} + C e^{-2\lambda m/C}\right).$$

By Stirling's formula

$$\frac{(\lambda m)^m}{m!} \sim (\lambda m)^m/(\sqrt{2\pi m}(m/e)^m) = (\lambda e)^m/\sqrt{2\pi m}.$$

We choose $\lambda = 1/(2e)$; thus, the decrease of the square bias term is exponentially fast. The choice $m_{\text{opt}} = [\mathtt{a} \log(n)]$ with $\mathtt{a} = eC + 1/\log(2)$. Combining this with Proposition 1 gives the rate $\sqrt{\log(n)/n}$ for the $\mathbb{L}^2$-risk of the estimator. □

### 7.7 Proof of Proposition 9

From Proposition 12 and formula (37), we have

$$\sum_{j \geq m} a_{2j}^2(f_\sigma) \lesssim \frac{1}{\pi} \frac{1}{1 + \sigma^2} \sum_{j \geq m} \frac{1}{\sqrt{j}} \left( \frac{\sigma^2 - 1}{\sigma^2 + 1} \right)^{2j}.$$

Now for $f \in \mathcal{G}(v)$, we get

$$\begin{aligned}
\|f - f_{2m}\|^2 &\lesssim \frac{1}{\pi} \sum_{j \geq m} \frac{1}{\sqrt{j}} \int_{1/v}^v \frac{1}{1 + u^2} \left( \frac{u^2 - 1}{u^2 + 1} \right)^{2j} d\Pi(u) \\
&\leq \frac{1}{\pi} \frac{1}{\sqrt{m}} \int_{1/v}^v \frac{1 + u^2}{4u^2} \left( \frac{u^2 - 1}{u^2 + 1} \right)^{2m} d\Pi(u) \\
&\leq \frac{1}{4\pi} \frac{1}{\sqrt{m}} \left( 1 + v^2 \right) \rho_0^m
\end{aligned}$$

where $\rho_0$ is given by (16). Therefore, choosing $m_{\mathrm{opt}} = [\log(n)/|\log(\rho_0)|]$ gives a squared bias of order $1/(n\sqrt{\log(n)})$ and a variance of order $\sqrt{\log(n)}/n$, thus a rate of order $\sqrt{\log(n)}/n$.

### 7.8 Proofs of Proposition 11

Using notation (21), we have:

$$\begin{aligned}
\mathbb{E}(\|\hat{g}_m - g\|^2) &= \mathbb{E}(\|\hat{g}_m - \mathbb{E}\hat{g}_m\|^2) + \|\mathbb{E}(\hat{g}_m) - g_m\|^2 + \|g_m - g\|^2 \\
&= \sum_{j=0}^{m-1} \mathrm{Var}(\hat{a}_j) + \sum_{j=0}^{m-1} R^2(h_j) + \|g_m - g\|^2.
\end{aligned}$$

First, $\sum_{j=0}^{m-1} R^2(h_j) = \sup_{t \in S_m, \|t\|=1} R^2(t)$ and the bound for this term follows from the following Lemma:

**Lemma 1** *Let $t \in S_m$ and assume that* (H1) *and* (H2) *hold.*
(1) *If $C := \int u^2 |g^*(u)|^2 du < +\infty$, then*

$$|R(t)| \leq \Delta \|t\| \|g\|_1 C^{1/2}/\sqrt{2\pi}.$$

(2) *Otherwise:*

$$|R(t)| \leq 2\sqrt{2} \|g\|_1 \|g\| \|t\| \Delta \sqrt{m}. \tag{39}$$

On the other hand, we have:

$$\sum_{j=0}^{m-1} \mathrm{Var}(\hat{a}_j) = \frac{1}{n} \sum_{j=0}^{m-1} \mathrm{Var}\left( \frac{1}{\Delta} Z_1 h_j(Z_1) \right) \leq \frac{1}{n\Delta} \mathcal{V}_m.$$

We need to bound $z^2 h_j^2(z)$. To that aim, we use relation (26) for $h_j$. We bound $\xi_j(x)$ given by (27) as in the proof of Proposition 1 by: $|\xi_j(x)|^2 \le |x|^5/10$, and using (28) we obtain for $j \ge 1$,

$$z^2 h_j^2(z) \le \frac{2c_2^2}{\pi} \frac{z^2}{\sqrt{j}} + \frac{2}{5} \frac{|z|^7}{2j+1}.$$

As a consequence

$$\frac{1}{\Delta} \mathbb{E}\left[ Z_1^2 h_j^2(Z_1) \right] \le 2c_2^2 \frac{\mathbb{E}(Z_1^2)}{\Delta} + \frac{2}{5(2j+1)} \frac{\mathbb{E}(|Z_1|^7)}{\Delta}.$$

The bound on $\mathcal{V}_m$ given in (23) follows.                                                              $\square$

*Proof of Lemma 1.* For case (1), we refer to Comte and Genon-Catalot (2009), Proposition 4.1, p. 4099.

For case (2), we have (see (21) and (18)) for $t \in S_m$,

$$R(t) = \frac{1}{\Delta} \mathbb{E} Z_1 t(Z_1) - \langle t, g \rangle = \int t(z) \mu_\Delta(dz) - \langle t, g \rangle = \int t(z) \mathbb{E} g(z - Z_1) dz - \langle t, g \rangle$$

$$= \mathbb{E}\left( \int t(z + Z_1) - t(z) \right) g(z) dz = \mathbb{E}\left( Z_1 \int_0^1 du \left( \int g(z) t'(z + u Z_1) dz \right) \right).$$

Thus,

$$|R(t)| \le \mathbb{E}|Z_1| \|g\| \|t'\|.$$

Under (17), (H1), we have

$$\mathbb{E}(|Z_1|) \le \Delta \int |x| n(x) dx = \Delta \|g\|_1, \tag{40}$$

(see Proposition 3.2, p. 83 in Belomestny et al. 2015). Now, by Lemma 2 below, $\|t'\| \le \sqrt{2m} \|t\|$ so the proof of Lemma 1 is complete.                           $\square$

**Lemma 2** $\forall m \ge 0$, $\forall t \in S_m$, $\|t'\|^2 \le 2m\|t\|^2$.

*Proof of Lemma 2.* A function $t \in S_m$ can be written $t = \sum_{j=0}^{m-1} a_j h_j$. Thus, $t' = \sum_{j=0}^{m-1} a_j h_j'$. We use $h_0'(x) = -h_1(x)/\sqrt{2}$ and Formula (35) to obtain:

$$\sqrt{2} t'(x) = -a_0 h_1(x) + \sum_{j=1}^{m-1} a_j (\sqrt{j} h_{j-1}(x) - \sqrt{j+1}\, h_{j+1}(x))$$

$$= a_1 h_0(x) + (\sqrt{2} a_2 - a_0) h_1(x) + \sum_{j=2}^{m-2} \left( \sqrt{j+1} a_{j+1} - \sqrt{j} a_{j-1} \right) h_j(x)$$

$$-a_{m-2}\sqrt{m-1}\, h_{m-1}(x) - a_{m-1}\sqrt{m}\, h_m(x).$$

This implies, for $m \geq 2$

$$\|t'\|^2 = \frac{1}{2}\left(a_1^2 + (\sqrt{2}a_2 - a_0)^2 + \sum_{j=2}^{m-2}\left(\sqrt{j+1}\,a_{j+1} - \sqrt{j}\,a_{j-1}\right)^2 + (m-1)a_{m-2}^2 + ma_{m-1}^2\right)$$

$$\leq \frac{1}{2}\left(a_1^2 + 2a_0^2 + 4a_2^2\right) + \frac{m}{2}\left(2\sum_{j=2}^{m-2}\left(a_{j+1}^2 + a_{j-1}^2\right) + a_{m-2}^2 + a_{m-1}^2\right)$$

$$\leq 2m\sum_{j=0}^{m-1} a_j^2 = 2m\|t\|^2.$$

For $m = 0, 1$, the same inequality holds obviously.            $\square$

## References

Abramowitz, M., Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. National Bureau of Standards Applied Mathematics Series, 55 for sale by the Superintendent of Documents. Washington, DC: U.S. Government Printing Office.

Barron, A., Birgé, L., Massart, P. (1999). Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, *113*, 301–413.

Belomestny, D., Comte, F., Genon-Catalot, V., Masuda, H., Reiss, M. (2015). *Lévy matters. IV. Estimation for discretely observed Lévy processes*. Lecture Notes in Mathematics, 2128. Lévy Matters. Cham: Springer.

Belomestny, D., Comte, F., Genon-Catalot, V. (2016). Nonparametric Laguerre estimation in the multiplicative censoring model. *Electronic Journal of Statistics*, *10*, 3114–3152.

Bertoin, J. (1996). *Lévy processes*. Cambridge tracts in mathematics, 121. Cambridge, NY: Cambridge University Press.

Birgé, L., Massart, P. (2007). Minimal penalties for Gaussian model selection. *Probability Theory and Related Fields*, *138*, 33–73.

Bongioanni, B., Torrea, J. L. (2006). Sobolev spaces associated to the harmonic oscillator. *Proceedings of the Indian Academy of Sciences*: *Mathematical Sciences*, *116*(3), 337–360.

Chaleyat-Maurel, M., Genon-Catalot, V. (2006). Computable infinite-dimensional filters with applications to discretized diffusion processes. *Stochastic Processes and Their Applications*, *116*, 1447–1467.

Comte, F., Genon-Catalot, V. (2009). Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Processes and Their Applications*, *119*, 4088–4123.

Comte, F., Genon-Catalot, V. (2015). Adaptive Laguerre density estimation for mixed Poisson models. *Electronic Journal of Statistics*, *9*, 1113–1149.

Comte, F., Dedecker, J., Taupin, M. L. (2008). Adaptive density deconvolution with dependent inputs. *Mathematical Methods of Statistics*, *17*, 87–112.

Devroye, L., Györfi, L. (1985). *Nonparametric density estimation. The L1 view*. Wiley series in probability and mathematical statistics: Tracts on probability and statistics. New York: Wiley.

Devroye, L., Lugosi, G. (2001). *Combinatorial methods in density estimation*. Springer series in statistics. New York: Springer.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., Picard, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics*, *24*, 508–539.

Efromovich, S. (1999). *Nonparametric curve estimation. Methods, theory, and applications*. Springer series in statistics. New York: Springer.

Efromovich, S. (2008). Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *The Annals of Statisticts*, *36*, 1127–1155.

Efromovich, S. (2009). Lower bound for estimation of Sobolev densities of order less 1/2. *Journal of Statistical Planning and Inference*, *139*, 2261–2268.

Ibragimov, I. (2001). Estimation of analytic functions. In M. de Gunst, C. Klaasen, A. van der Vaart (Eds.) *State of the art in probability and statistics (Leiden, 1999)*. Institute of mathematical statistics lecture notes—Monograph series 36 (pp. 359–383). Beachwood, OH: Institute of Mathematical Statistics.

Ibragimov, I. A., Has'minskii, R. Z. (1980). An estimate of the density of a distribution. *Studies in mathematical statistics, IV*. Zapiski Nauchnykh Seminarov Leningradskogo Otdeleniya Matematicheskogo Instituta imeni V. A. Steklova Akademii Nauk SSSR (LOMI) *98*, 61–85, 161–162 **(in Russian)**.

Kim, A. K. H. (2014). Minimax bounds for estimation of normal mixtures. *Bernoulli*, *20*, 1802–1818.

Klein, T., Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, *33*, 1060–1077.

Lebedev, N. N. (1972). *Special functions and their applications* (Revised edition, translated from the Russian and edited by Richard A. Silverman. Unabridged and corrected republication). New York: Dover Publications, Inc.

Mabon, G. (2017). Adaptive deconvolution on the nonnegative real line. *Scandinavian Journal of Statistics*: *Theory and Applications*, *44*, 707–740.

Markett, C. (1984). Norm estimates for $(C, \delta)$ means of Hermite expansions and bounds for $\delta_{\text{eff}}$. *Acta Mathematica Hungarica*, *43*, 187–198.

Massart, P. (2007). *Concentration inequalities and model selection.* Lectures from the 33rd summer school on probability theory held in Saint-Flour, July 6–23, 2003. With a foreword by Jean Picard. Lecture notes in mathematics, 1896. Berlin: Springer.

Rigollet, P. (2006). Adaptive density estimation using the blockwise Stein method. *Bernoulli*, *12*, 351–370.

Schipper, M. (1996). Optimal rates and constants in L2-minimax estimation of probability density functions. *Mathematical Methods of Statistics*, *5*, 253–274.

Schwartz, S. C. (1967). Estimation of a probability density by an orthogonal series. *The Annals of Mathematical Statistics*, *38*, 1261–1265.

Szegö, G. (1975). *Orthogonal polynomials* (4th ed.). American Mathematical Society, Colloquium Publications, Vol: XXIII. Providence, RI: American mathematical Society.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer series in statistics. New York: Springer.

Walter, G. G. (1977). Properties of Hermite series estimation of probability density. *The Annals of Statistics*, *5*, 1258–1264.