ESTIMATION OF CONVOLUTION IN THE MODEL WITH NOISE

C. CHESNEAU[1], F. COMTE[2], G. MABON[2,3], AND F. NAVARRO[4]

ABSTRACT. We investigate the estimation of the ℓ-fold convolution of the density of an unobserved variable X from n i.i.d. observations of the convolution model $Y = X + \varepsilon$. We first assume that the density of the noise ε is known and define nonadaptive estimators, for which we provide bounds for the mean integrated squared error (MISE). In particular, under some smoothness assumptions on the densities of X and ε, we prove that the parametric rate of convergence $1/n$ can be attained. Then we construct an adaptive estimator using a penalization approach having similar performances to the nonadaptive one. The price for its adaptivity is a logarithmic term. The results are extended to the case of unknown noise density, under the condition that an independent noise sample is available. Lastly, we report a simulation study to support our theoretical findings.

1. MOTIVATIONS

In actuarial or financial sciences, quantities of interest often involve sums of random variables. For example, in the individual risk model, the total amount of claims on a portfolio of insurance contracts is described by the sum of all claims on the individual policies. Consequently, it is natural to aim at estimating the density of sums of random variables, and estimating these directly from observations of the variables. A solution to this question is proposed in Chesneau et al. (2013). But the measures associated to the variables are also often not so precise, and clearly, measurement errors can occur. Thus, we consider the model

$$(1) \qquad Y_j = X_j + \varepsilon_j, \qquad j = 1, \dots, n,$$

where the Y_j's are i.i.d. with density f_Y, X_j's are i.i.d. with density f and ε_j's i.i.d. with density f_ε. We assume that X_j and ε_j are independent, and that f is unknown whereas, in a first step, f_ε is known. Note that our assumptions imply that $f_Y(u) = (f \star f_\varepsilon)(u)$, where $(h \star k)(t) = \int h(t-x)k(x)dx$ denotes the convolution product. Our aim is to estimate the ℓ-fold convolution of f for any integer $\ell \geq 1$, i.e. to estimate g_ℓ defined by

$$(2) \qquad g_\ell(u) = \underbrace{(f \star \cdots \star f)}_{\ell \text{ times}}(u).$$

A large and growing body of literature has investigated this problem without noise ($\varepsilon_j = 0$). In this case, various methods and results can be found in e.g. Frees (1994), Saavedra and Cao (2000), Ahmad and Fan (2001), Ahmad and Mugdadi (2003), Prakasa Rao (2004), Schick and Wefelmeyer (2004, 2007), Du and Schick (2007), Giné and Mason (2007), Nickl (2007), Nickl (2009) and Chesneau et al. (2013). However, the modeling of some complex phenomena implies that error disrupts the observation of X_j. Thus, model (1) with $\varepsilon_j \neq 0$ (the so-called "convolution model") has been widely considered and the estimation of f in this context is a well-known problem ("deconvolution estimator" corresponding to the case $\ell = 1$). See e.g. Caroll and Hall (1988), Devroye (1989), Fan (1991), Pensky and Vidakovic (1999), Fan and Koo (2002), Butucea (2004) Butucea and Tsybakov (2008), Comte et al. (2006), Delaigle and Gijbels (2006), Johannes (2009) and Meister (2009). The consideration of $\varepsilon_j \neq 0$ and $\ell \geq 1$ is a more general challenge with potential applications in many fields. A first approach has been investigated in Chesneau and Navarro (2014) via wavelet methods, with convincing numerical illustrations but sub-optimal theoretical results.

In this paper, we consider Model (1) with special emphasis on the influences of ℓ and ε_j on the estimation of g_ℓ. In a first part, we introduce non-adaptive estimators combining kernel methods and Fourier analysis. Then we determine a sharp upper bound of its mean integrated squared error (MISE) and exhibit fast rates of convergence according to different kinds of smoothness of f and f_ε (and suitable values of a tuning parameter). In particular, as proved in Saavedra and Cao (2000) and Chesneau et al. (2013) for the case $\varepsilon_j = 0$, we show that the parametric rate of convergence $1/n$ can be attained under some assumptions on ℓ, f and f_ε. In a second part, we develop an adaptive version of this estimator using a penalization approach. It has similar performance to the non-adaptive one, up to a logarithmic factor. This result significantly improves (Chesneau and Navarro, 2014, Proposition 5.1) in terms of rates of convergence and assumptions on the model.

In a second time, we assume that the distribution of the noise is unknown. Then, for identifiability purpose, some additional information on the errors has to be available. As in Neumann (1997), Comte and Lacour (2011) and Kappus and Mabon (2013), we consider the setting where an independent noise sample of size M is available. Other contexts may bring relevant information for estimation, as for example a repeated measurement setting, see Delaigle *et al.* (2008), Bonhomme and Robin (2010), Neumann (2007), Kappus and Mabon (2013); these settings are not considered here. We shall take advantage of the tools developed in Kappus (2014) and Kappus and Mabon (2013) to provide an adaptive strategy in this framework, which is more realistic but more difficult than the known noise density case.

The paper is structured as follows. We first study in Section 2 the case where the noise density is known. Our estimators are introduced in Section 2.2. Their performances in terms of MISE and rates of convergence are presented in Sections 2.3 and 2.4. The adaptive strategy is proved to reach an automatic bias-variance compromise up to negligible terms in Section 2.5. Section 3 is devoted to the case of unknown noise density: instead, to preserve identifiability of the problem, we assume that an independent noise sample is available. New estimates are proposed and studied: an adaptive procedure is proposed and proved to reach the best risk bound as possible, up to logarithmic terms. Simulation results are presented in Section 4. All proofs are relegated to Section 5.

2. Estimation with known noise density

2.1. **Notations and assumption.** For two real numbers a and b, we denote by $a \wedge b$ the minimum of a and b. For two functions φ and ψ belonging to $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, we denote by φ^* the Fourier transform defined by $\varphi^*(x) = \int \varphi(t)e^{-ixt}dt$, by $\|\varphi\|$ the $\mathbb{L}^2$-norm of φ,

$\|\varphi\|^2 = \int |\varphi(u)|^2 du$, by $\langle \varphi, \psi \rangle$ the scalar product $\langle \varphi, \psi \rangle = \int \varphi(u) \bar{\psi}(u) du$. We recall the inverse Fourier formula $\varphi(x) = (2\pi)^{-1} \int e^{iux} \varphi^*(u) du$, and that, for the convolution product $(\varphi \star \psi)(t) = \int \varphi(t-x) \psi(x) dx$, we have by Fourier transform $(\varphi \star \psi)^*(u) = \varphi^*(u) \psi^*(u)$. Lastly $\langle \varphi, \psi \rangle = (2\pi)^{-1} \langle \varphi^*, \psi^* \rangle$.

Moreover, we work all along the paper under the assumption

(A1) $f_\varepsilon^*(u) \neq 0, \forall u \in \mathbb{R}$.

2.2. Estimators when the noise density is known.
We define now our estimators, considering that the observations are drawn from model (1) and assuming that f_ε is known and satisfies **(A1)**.

The idea for this construction is based on the following relation. First under independence of X_j and ε_j in model (1), we have $f_Y = f \star f_\varepsilon$, which implies $f_Y^* = f^* f_\varepsilon^*$. Since in addition $g_\ell^*(u) = (f^*(u))^\ell$, we have, using **(A1)**,

$$g_\ell(x) = \frac{1}{2\pi} \int g_\ell^*(u) e^{iux} du = \frac{1}{2\pi} \int (f^*(u))^\ell e^{iux} du = \frac{1}{2\pi} \int \left(\frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right)^\ell e^{iux} du.$$

Then f_Y^* can easily be estimated by its empirical counterpart

$$\hat{f}_Y^*(u) = \frac{1}{n} \sum_{j=1}^n e^{-iuY_j}.$$

But a pure plug-in approach leads to integrability problems, this is why a cutoff has to be introduced, which can be seen as a truncated version of the empirical estimator: $\hat{f}_Y^*(u) \mathbf{I}_{[-\pi m, \pi m]}(u)$. Thus, we propose the estimator, $\forall m \in \mathbb{N}$,

$$(3) \qquad \hat{g}_{\ell,m}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left(\frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)} \right)^\ell e^{iux} du.$$

Naturally, the performance of $\hat{g}_{\ell,m}$ depends on the choice of m. We make $\hat{g}_{\ell,m}$ adaptive by considering the estimator: $\hat{g}_{\ell,\hat{m}}$, where $\hat{m}$ is the integer defined by

$$(4) \qquad \hat{m} = \arg \min_{m \in \{1 \dots, n\}} \left(-\|\hat{g}_{\ell,m}\|^2 + \mathrm{pen}_\ell(m) \right).$$

where, for numerical constant ϑ specified later,

$$(5) \qquad \mathrm{pen}_\ell(m) = \vartheta 2^{2\ell-1} \lambda_\ell^{2\ell}(m, \Delta_\ell(m)) \frac{\Delta_\ell(m)}{n^\ell}, \quad \Delta_\ell(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(u)|^{2\ell}} du$$

and

$$(6) \qquad \lambda_\ell(m, D) = \max \left\{ 2^{-1/\ell+1} \left(\log(1 + m^2 D) \right)^{1/2}, \frac{2^{2-1/(2\ell)}}{\sqrt{n}} \log(1 + m^2 D) \right\}.$$

2.3. Risk bound for $\hat{g}_{\ell,m}$.
Now we can prove the following risk bound for the MISE, a decomposition which explains our adaptive strategy and will allow to discuss rates.

Proposition 2.1. *Consider data from model (1), g_ℓ given by (2), and assume that **(A1)** is fulfilled and f is square integrable. Let $\hat{g}_{\ell,m}$ be defined by (3). Then all m in $\mathbb{N}$, we have*

$$(7) \qquad \mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq \frac{1}{2\pi} \int_{|u| \geq \pi m} |f^*(u)|^{2\ell} du + \frac{2^\ell - 1}{2\pi} \sum_{k=1}^\ell \binom{\ell}{k} \frac{C_k}{n^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2(\ell-k)}}{|f_\varepsilon^*(u)|^{2k}} du,$$

with $C_k = (72k)^{2k}(2k/(2k-1))^k$.

Inequality (7) deserves two comments. First, we note that this bound is non asymptotic, and so is the adaptive strategy inspired from it. Secondly, we can see the usual antagonism between the bias term $(\pi)^{-1} \int_{|u| \geq \pi m} |f^*(u)|^{2\ell} du$ which is smaller for large m and the variance term which gets larger when m increases.

Now, we can see how Inequality (7) allows us to explain the definition of the model selection given in (4). The bias term can also be written $\|g_\ell\|^2 - \|g_{\ell,m}\|^2$ with $g_{\ell,m}^*(u) = g_\ell^*(u)\mathbf{I}_{[-\pi m,\pi m]}(u)$; as $\|g_\ell\|^2$ is a constant, only $-\|g_{\ell,m}\|^2$ is estimated, by $-\|\hat{g}_{\ell,m}\|^2$. Usually, the variance is replaced by its explicit bound, given here by the second right-hand-side term of (7). It is noteworthy that here, the variance bound is the sum of ℓ terms and only the last term is involved in the penalty. This is rather unexpected and specific to our particular setting. These considerations explain the definition of (4) and the selection procedure.

2.4. **Rates of convergence for $\hat{g}_{\ell,m}$.** Now, from an asymptotic point of view, we can investigate the rates of convergence of our estimators according to the smoothness assumptions on f and f_ε. The following balls of function spaces are considered, for positive α, β,

$$\mathcal{A}(\beta, L) = \left\{ f \in \mathbb{L}^1 \cap \mathbb{L}^2(\mathbb{R}); \quad \int |f^*(u)|^2 \exp(2|u|^\beta) du \leq L, \quad \sup_{u \in \mathbb{R}} \left(|f^*(u)|^2 \exp(2|u|^\beta) \right) \leq L \right\}$$

and

$$\mathcal{S}(\alpha, L) = \left\{ f \in \mathbb{L}^1 \cap \mathbb{L}^2(\mathbb{R}); \quad \int |f^*(u)|^2 (1 + u^2)^\alpha du \leq L, \quad \sup_{u \in \mathbb{R}} \left(|f^*(u)|^2 (1 + u^2)^\alpha \right) \leq L \right\}.$$

As the noise density is assumed to be known, regularity assumptions on f_ε^* are standardly formulated in the following way, for positive a, b:

[OS(a)] f_ε^* is OS(a) if $\forall u \in \mathbb{R}$,

$$\frac{c_\varepsilon}{(1 + u^2)^a} \leq |f_\varepsilon^*(u)|^2 \leq \frac{C_\varepsilon}{(1 + u^2)^a}.$$

[SS(b)] f_ε^* is SS(b) if $\forall u \in \mathbb{R}$,

$$c_\varepsilon e^{-2|u|^b} \leq |f_\varepsilon^*(u)|^2 \leq C_\varepsilon e^{-2|u|^b}.$$

Note that the lower bound for f_ε^* is the one used to obtain upper bound of the risk.
These classical sets allow to compute rates, depending on the types of f and f_ε and to get the following consequence of Proposition 2.1.

Corollary 2.1. *Under the assumptions of Proposition 2.1, the following table exhibits the values of m and φ_n such that φ_n is the sharpest rates of convergence satisfying "there exists a constant $C_\ell^* > 0$ such that $\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq C_\ell^* \varphi_n$", according some smoothness assumptions on f and f_ε.*

	$f \in \mathcal{A}(\beta, L), f_\varepsilon$ OS(a)	$f \in \mathcal{S}(\alpha, L), f_\varepsilon$ SS(b)	$f \in \mathcal{S}(\alpha, L), f_\varepsilon$ OS(a)
m_{opt}	$\dfrac{1}{\pi}\left(\dfrac{\log(n)}{2\ell}\right)^{1/\beta}$	$\dfrac{1}{\pi}\left(\dfrac{\log(n)}{4\ell}\right)^{1/b}$	$n^{k_0/(2ak_0 + 2\alpha k_0 + 1)}$
φ_n	$\dfrac{1}{n}$	$(\log(n))^{-4\ell\alpha/b}$	$n^{-2\alpha\ell/(2a+2\alpha+1)} \vee n^{-1}$

where k_0 is the integer such that $a_{k_0-1} < 2\alpha < a_{k_0}$ where $a_0 = 0$, $a_k = (2ak + 1)/(\ell - k)$, $k = 1, \ldots, \ell - 1$.

The parametric rate of convergence $1/n$ is attained for the case $f \in \mathcal{A}(\beta, L), f_\varepsilon \ \mathrm{OS}(a)$, and in the case $f \in \mathcal{S}(\alpha, L), f_\varepsilon \ \mathrm{OS}(a)$ when $\ell > 1 + (a + 1/2)/\alpha$. To the best of our knowledge, Theorem 2.1 is the first result showing (sharp) rates of convergence for an estimator in the context of (1) for different types of smoothness on f and f_ε. The results correspond to classical deconvolution rates for $\ell = 1$ (see Comte et al. (2006)). Moreover, this completes and improves the rate of convergence determined in (Chesneau and Navarro, 2014, Proposition 5.1).

The case where both f and f_ε are super smooth is not detailed. For $\ell = 1$, the rates corresponding to this case, are implicit solutions of optimization equations (see Butucea and Tsybakov (2008)) and recursive formula are given in Lacour (2006). Explicit polynomial rates can be obtained, and also rates faster than logarithmic but slower than polynomial. We do not investigate them here for sake of conciseness. We can emphasize that the function f being unknown, its regularity is also unknown, and the asymptotic optimal choice of m can not be done in practice. This is what makes an automatic choice of m mandatory to obtain a complete estimation procedure. The automatic resulting trade-off avoids tedious computations of asymptotic rates and perform at best for all sample sizes.

2.5. **Risk bound on the adaptive estimator $\hat{g}_{\ell,\hat{m}}$.** Theorem 2.1 determines risk bound for the MISE of the estimator $\hat{g}_{\ell,\hat{m}}$.

Theorem 2.1. *Consider Model (1) under Assumption* **(A1)**. *Let $\hat{g}_{\ell,\hat{m}}$ be defined by (3)-(4) and define*

$$(8) \qquad \mathcal{V}_\ell(m) := \frac{2^\ell - 1}{\pi} \sum_{k=1}^\ell \binom{\ell}{k} C_k \frac{1}{n^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2(\ell-k)}}{|f_\varepsilon^*(u)|^{2k}} du.$$

Then, there exists ϑ_0 such that for any $\vartheta \geq \vartheta_0$, we have

$$\mathbb{E}(\|\hat{g}_{\ell,\hat{m}} - g_\ell\|^2) \leq 2^{2\ell} \min_{m \in \{1,\dots,n\}} \left(\|g - g_\ell\|^2 + \mathcal{V}_\ell(m) + \mathrm{pen}_\ell(m) \right) + \frac{c_\ell}{n^\ell}$$

where c_ℓ is a constant independent of n. Note that $\vartheta_0 = 1$ suits.

The result in Theorem 2.1 shows that the adaptive estimator $\hat{g}_{\ell,\hat{m}}$ performs an automatic trade-off between the bias and variance terms described in Proposition 2.1. Moreover, the bound on the risk given in Theorem 2.1 is non asymptotic, and holds for all sample size.

As a consequence, we can derive asymptotic results. Table 1 exhibits the sharpest rate of convergence φ_n satisfying "there exists a constant $C_\ell^{**} > 0$ such that $\mathbb{E}(\|\hat{g}_{\ell,\hat{m}} - g_\ell\|^2) \leq C_\ell^{**} \varphi_n$", depending on smoothness assumptions on f and f_ε.

	$f \in \mathcal{A}(\beta, L), f_\varepsilon \ \mathrm{OS}(a)$	$f \in \mathcal{S}(\alpha, L), f_\varepsilon \ \mathrm{SS}(b)$	$f \in \mathcal{S}(\alpha, L), f_\varepsilon \ \mathrm{OS}(a)$
φ_n	$\log(n)/n$	$(\log(n))^{-4\ell\alpha/b}$	$(n/\log(n))^{-2\alpha\ell/(2a+2\alpha+1)} \vee (\log(n)/n)$

TABLE 1. Examples of rates of convergence φ_n of the estimator $\hat{g}_{\ell,\hat{m}}$

Theorem 2.1 implies thus that $g_{\ell,\hat{m}}$ attains similar rates of convergence to $g_{\ell,m}$ with automatic "optimal" choices for m. The only difference is possibly a logarithmic term.

3. Estimation with unknown noise density

We consider now that f_ε is unknown, but still satisfying **(A1)**. For the problem to be identifiable, we assume that we have at hand a sample $\varepsilon_1', \dots, \varepsilon_M'$ of i.i.d. replications of ε_1, independent of the first sample.

3.1. **Risk bound for fixed** m. We have to replace in (3) the noise characteristic function f_ε^* by an estimator, and to prevent it from getting too small in the denominator. Thus, we propose the estimator, $\forall m \in \mathbb{N}$,

$$(9) \qquad \tilde{g}_{\ell,m}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left(\frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} \right)^\ell e^{iux} du,$$

where still $\hat{f}_Y^*(u) = n^{-1} \sum_{j=1}^n e^{-iuY_j}$ and now

$$(10) \qquad \tilde{f}_\varepsilon(u) = \begin{cases} \hat{f}_\varepsilon^*(u) & \text{if } \hat{f}_\varepsilon^*(u) \geq k_M(u) \\ k_M(u) \text{ otherwise} \end{cases} \qquad \text{where } \hat{f}_\varepsilon^*(u) = \frac{1}{M} \sum_{j=1}^M e^{-iu\varepsilon_j'},$$

and

$$(11) \qquad k_M(u) = \kappa s_M(u) M^{-1/2}.$$

The definition of s_M is simply 1 when considering one estimator in the collection, and more elaborated to deal with the adaptive case. To study the risk bound of the estimator, we recall the following key Lemma (see Neumann (1997), Comte and Lacour (2011) and Kappus (2014)).

Lemma 3.1. *Assume that* **(A1)** *holds. Let* $s_M(u) \equiv 1$ *in (11) for* $\tilde{f}_\varepsilon$ *defined by (10), then for all* $p \geq 1$, *there exists a constant* τ_p *such that,* $\forall u \in \mathbb{R}$,

$$\mathbb{E}\left(\left| \frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} \right|^{2p} \right) \leq \tau_p \left(\frac{1}{|f_\varepsilon^*(u)|^{2p}} \wedge \frac{(k_M(u))^{2p}}{|f_\varepsilon^*(u)|^{4p}} \right).$$

To study the MISE of the $\tilde{g}_{\ell,m}$, we write $\|\tilde{g}_{\ell,m} - g_\ell\|^2 = \|g_\ell - g_{\ell,m}\|^2 + \|\tilde{g}_{\ell,m} - g_{\ell,m}\|^2$, which starts the bias-variance decomposition. Then we can split the integrated variance term of $\tilde{g}_{\ell,m}$ into three terms

$$
\begin{aligned}
\|\tilde{g}_{\ell,m} - g_{\ell,m}\|^2 &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left| \left(\frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} - \frac{f_Y^*(u)}{f_\varepsilon^*(u)} + \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right)^\ell - \left(\frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right)^\ell \right|^2 du \\
&= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left| \sum_{k=1}^\ell \binom{\ell}{k} \left(\frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right)^{\ell-k} \left(\frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} - \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right)^k \right|^2 du \\
&\leq \frac{2^\ell - 1}{2\pi} \int_{-\pi m}^{\pi m} \sum_{k=1}^\ell \binom{\ell}{k} \left| \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2(\ell-k)} \left| \frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} - \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2k} du \leq \frac{2^\ell}{2\pi}(T_1 + T_2 + T_3)
\end{aligned}
$$

with

$$T_1 = \int_{-\pi m}^{\pi m} \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \left| \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2(\ell-k)} \left| \frac{\hat{f}_Y^*(u) - f_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2k} du$$

$$T_2 = \int_{-\pi m}^{\pi m} \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \left| \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2(\ell-k)} \left| (\hat{f}_Y^*(u) - f_Y^*(u)) \left(\frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} \right) \right|^{2k} du$$

$$T_3 = \int_{-\pi m}^{\pi m} \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \left| \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2(\ell-k)} \left| f_Y^*(u) \left(\frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} \right) \right|^{2k} du.$$

Clearly, T_1 is the term corresponding to known f_ε^* and we have $\mathbb{E}(T_1) \leq 3^{2\ell-1}\mathcal{V}_\ell$. Using the first bound in Lemma 3.1 and the independence of the samples, we have also $\mathbb{E}(T_2) \leq 3^{2\ell-1}\mathcal{V}_\ell$. The novelty comes from T_3, where we use the second bound given in Lemma 3.1:

$$\mathbb{E}(T_3) \leq \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \int_{-\pi m}^{\pi m} \frac{|f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2(\ell-k)}} \tau_k \frac{M^{-k}}{|f_\varepsilon^*(u)|^{4k}} du$$

$$\leq \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \tau_k \frac{1}{M^k} \int_{-\pi m}^{\pi m} \frac{|f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2(\ell+k)}} du$$

$$(12) \qquad = \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \tau_k \frac{1}{M^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2k}} du := \mathcal{W}_\ell.$$

Therefore, the following result holds.

Proposition 3.1. *Consider model* (1) *under* **(A1)**. *Let* $\tilde{g}_{\ell,m}$ *be defined by* (9)-(11) *with* $s_M(u) \equiv 1$. *Then there exists constants* $C_\ell^{**}, C_\ell^{***} > 0$ *(independent of* n*) such that*

$$\mathbb{E}(\|\tilde{g}_{\ell,m} - g_\ell\|^2) \leq \frac{1}{\pi} \int_{|u| \geq \pi m} |f^*(u)|^{2\ell} du + C_\ell^{**}\mathcal{V}_\ell + C_\ell^{***}\mathcal{W}_\ell,$$

where $\mathcal{V}_\ell$ *is defined by* (8) *and* $\mathcal{W}_\ell$ *by* (12).

We can see that if $M \geq n$, we have

$$\frac{1}{M^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2k}} du \leq \frac{1}{n^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2(\ell-k)}}{|f_\varepsilon^*(u)|^{2k}} du$$

and thus $\mathcal{W}_\ell \leq C_\ell^{(4)}\mathcal{V}_\ell$. Thus, for $M \geq n$, the risk of $\tilde{g}_{\ell,m}$ has the same order as the risk of $\hat{g}_{\ell,m}$ and the estimation of the noise characteristic function does not modify the rate of the estimator. The results given in Corollary 2.1 are thus still valid. Obviously, if the noise sample size M is smaller than n, it may imply deteriorations of the rates. Moreover, the above result is a generalization of Neumann (1997) and Comte and Lacour (2011), which both correspond to the case $\ell = 1$.

3.2. **Adaptive procedure.** Adaptation estimation here is a difficult task. We apply the methodology recently proposed by Kappus and Mabon (2013) in the context $\ell = 1$, improving the results of Comte and Lacour (2011). In particular the estimator $\tilde{f}_\varepsilon^*$ is now taken with s_M in (11) defined as follows. Let for a given $\delta > 0$,

$$\forall u \in \mathbb{R}, \quad w(u) = (\log(e + |u|))^{-\frac{1}{2}-\delta},$$

originally introduced in Neumann and Reiss (2009), and $\tilde{f}_\varepsilon^*$ given by (10)-(11) and by

$$(13) \qquad s_M(u) = (\log M)^{1/2} w^{-1}(u).$$

With this threshold, the bound given in Proposition 3.1 becomes

$$\mathbb{E}(\|\tilde{g}_{\ell,m} - g_\ell\|^2) \leq \frac{1}{\pi} \int_{|u| \geq \pi m} |f^*(u)|^{2\ell} du + \mathbb{V}_\ell$$

with

$$(14) \qquad \mathbb{V}_\ell = C_\ell^{**} \mathcal{V}_\ell + C_\ell^{****} \sum_{k=1}^\ell \binom{\ell}{k} 3^{2k-1} \tau_k \frac{\log^k(M)}{M^k} \int_{-\pi m}^{\pi m} \frac{w^{-2k}(u)|f^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2k}} du.$$

Clearly, the definition of s_M here implies logarithmic loss if we compare the last term above with $\mathcal{W}_\ell$ given by (12). Then using k_M defined by (11) and (13) and considering $\tilde{g}_{\ell,m}$ defined by (9)-(10), we propose the following adaptive strategy. We set first

$$\hat{\Delta}_\ell(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell}}{\left|\tilde{f}_\varepsilon^*(u)\right|^{2\ell}} du \quad \text{and} \quad \hat{\Delta}_\ell^f(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell}|\hat{f}_Y^*(u)|^{2\ell}}{\left|\tilde{f}_\varepsilon^*(u)\right|^{4\ell}} du,$$

as estimators of $\Delta_\ell(m)$ given by (5) and of

$$\Delta_\ell^f(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}} du.$$

We can now define the stochastic penalty associated with the adaptive procedure

$$\widehat{\text{qen}}_\ell(m) = \widehat{\text{qen}}_{\ell,1}(m) + \widehat{\text{qen}}_{\ell,2}(m)$$
$$= 2^{2\ell+1} \lambda_\ell^{2\ell}(m, \hat{\Delta}_\ell(m)) \frac{\hat{\Delta}_\ell(m)}{n^\ell} + 2^{2\ell+1} \kappa^{2\ell} \log^\ell(Mm) \frac{\hat{\Delta}_\ell^f(m)}{M^\ell}$$

which estimates the deterministic quantities

$$\text{qen}_\ell(m) = \text{qen}_{\ell,1}(m) + \text{qen}_{\ell,2}(m)$$
$$= 2^{2\ell+1} \lambda_\ell^{2\ell}(m, \Delta_\ell(m)) \frac{\Delta_\ell(m)}{n^\ell} + 2^{2\ell+1} \kappa^2 \log^\ell(Mm) \frac{\Delta_\ell^f(m)}{M^\ell}$$

where the weights $\lambda_\ell(m, D)$ are defined by (6). Then, we select the cutoff parameter $\tilde{m}$ as a minimizer of the following penalized criterion

$$(15) \qquad \tilde{m} = \arg \min_{m \in \{1,\dots,n\}} \left\{ -\|\tilde{g}_{\ell,m}\|^2 + \widehat{\text{qen}}_\ell(m) \right\}.$$

Theorem 3.1. *Consider Model (1) under* **(A1)**, *associated with an independent noise sample* $\varepsilon_1', \dots, \varepsilon_M'$. *Let the estimator* $\tilde{g}_{\ell,\tilde{m}}$ *be defined by (9) and (15). Then there are positive constants* C^{ad} *and* C *such that*

$$(16) \qquad \mathbb{E}\|g_\ell - \tilde{g}_{\ell,\tilde{m}}\|^2 \leq C^{ad} \inf_{m \in \mathcal{M}_n} \left\{ \|g_\ell - g_{\ell,m}\|^2 + \mathbb{V}_\ell(m) + \text{qen}_\ell(m) \right\} + \frac{C}{n^\ell} + \frac{C}{M^\ell},$$

where $\mathbb{V}_\ell$ *is defined by (14).*

This result is a generalization of Kappus and Mabon (2013), which corresponds to the case $\ell = 1$. As in the known density case, we do not use the whole variance for penalization, but only the last terms. Moreover, the final estimator automatically reaches the bias-variance compromise, up to logarithmic terms.
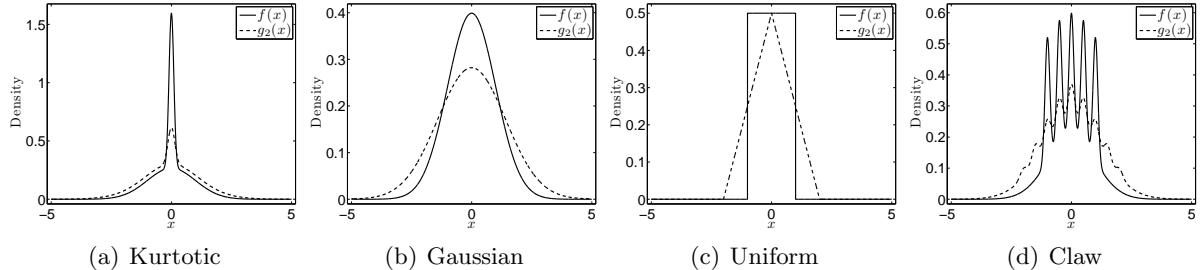
FIGURE 1. Test densities f (solid) and g_2 (dashed).

4. ILLUSTRATION

We now illustrate the theoretical results by a simulation study within the context described in Section 2 and Section 3. That is, we consider the problem of estimating the density g_ℓ, emphasizing the case $\ell = 2$, for both known and unknown types of noises. All simulations have been implemented under Matlab.

The performance of the proposed method is studied for four sets of test distributions for X_j representing different degrees of smoothness (see Figure 1),

(1) Kurtotic distribution: $\frac{2}{3}\mathcal{N}(0, 1) + \frac{1}{3}\mathcal{N}(0, (1/10)^2)$,
(2) Standard normal distribution: $\mathcal{N}(0, 1)$,
(3) Uniform distribution: $\mathcal{U}(-1, 1)$,
(4) Claw distribution: $\frac{1}{2}\mathcal{N}(0, 1) + \sum_{l=0}^{4} \mathcal{N}\left(l/2 - 1, (1/2)^2\right)$.

Note that the two Gaussian mixture densities are taken from Marron and Wand (1992).

We consider two types of error density f_ε, with the same variance $\sigma_\varepsilon^2 = 0.25$. The first one is Laplace distribution, with f_ε^* OS(2), and the second one is the Gaussian distribution with f_ε^* SS(2).

(a) Laplace error: $f_\varepsilon(x) = \frac{1}{2\sigma_\varepsilon}\exp\left(-\frac{|x|}{\sigma_\varepsilon}\right)$, $\quad f_\varepsilon^*(x) = \frac{1}{1 + \sigma_\varepsilon^2 x^2}$.

(b) Gaussian error: $f_\varepsilon(x) = \frac{1}{\sigma_\varepsilon\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma_\varepsilon^2}\right)$, $\quad f_\varepsilon^*(x) = \exp\left(-\frac{\sigma_\varepsilon^2 x^2}{2}\right)$,

In the case of an entirely known noise distribution, the constant ϑ in (5) is set to 1.8 for all the tests and in the unknown case the penalties are chosen according to Theorem 3.1. We propose the following penalty:

$$\widehat{\mathrm{qen}}_\ell(m) = \kappa_1 \lambda_\ell^{2\ell}(m, \hat{\Delta}_\ell(m))\frac{\hat{\Delta}_\ell(m)}{n^\ell} + \kappa_2 \kappa^{2\ell}\log^\ell(Mm)\frac{\hat{\Delta}_\ell^f(m)}{M^\ell}$$

with the associated constants penalties $\kappa_1 = \kappa_2 = 1$ and $\kappa = 1.8$. In both cases, the normalized sinc kernel was used throughout all experiments. As in Chesneau et al. (2013), the estimators proposed in this paper are well suited for FFT-based method. Thus, the resulting estimators $\hat{g}_{\ell,\hat{m}}$ and $\tilde{g}_{\ell,\tilde{m}}$ are simple to implement and fast which allows us to perform the penalized criterion in a reasonable time. For numerical implementation, we consider an interval $[a, b]$ that covers the range of the data and the density estimates were evaluated at $M = 2^r$ equally spaced points $t_i = a + (b - a)/M$, $i = 0, 1, \ldots, M - 1$, between a and b, with $r = 8$, $b = -a = 5$ and M is the number of discretization points. In each case, the grid of m values that we have considered consisted of 50 values from $0.5\hat{m}_0$ to $1.1\hat{m}_0$ where $\hat{m}_0 = n^{-\ell/5}$ denotes a pilot bandwidth (see e.g. Silverman (1986)).
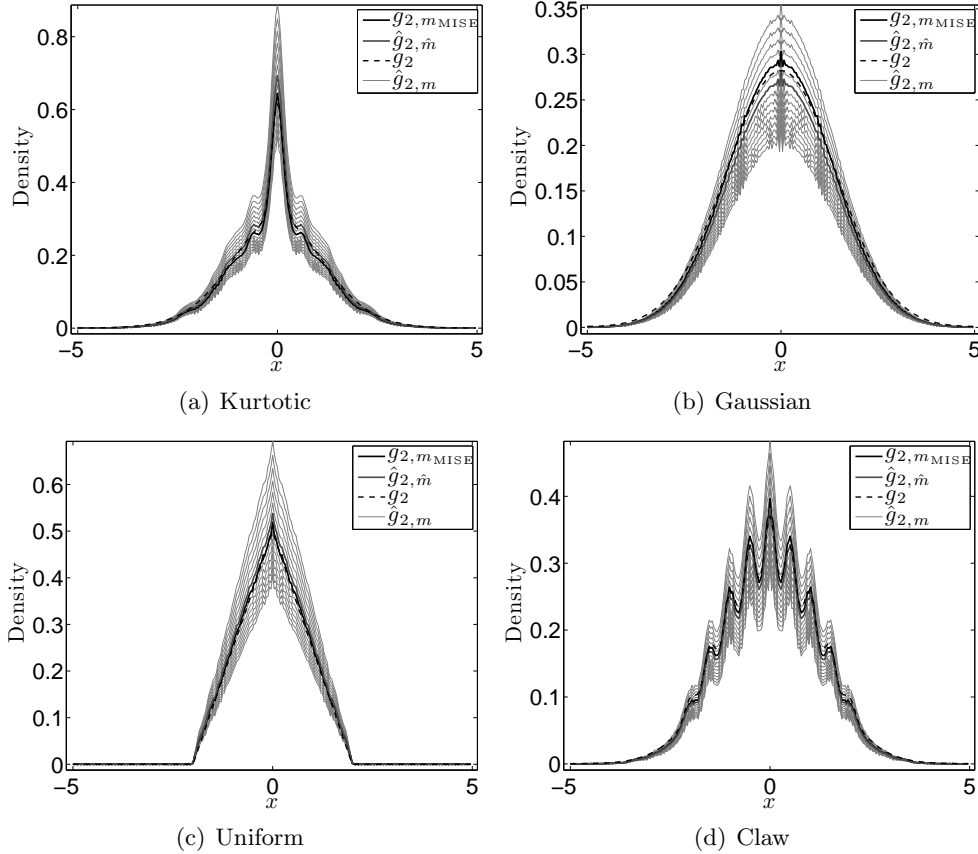
FIGURE 2. Laplace noise. True density (dotted), density estimates (gray) and sample of 20 estimates (thin gray) out of 50 proposed to the selection algorithm obtained with a sample of $n = 1000$ data.

In order to illustrate Theorem 2.1, we study the influence of the noise type on the numerical performances of the proposed estimator. For each density, samples of size $n = 1000$ were generated, with aim to estimate g_ℓ from $X_1, \ldots, X_n$ drawn from any of the test densities. The results are depicted in Figure 2 and Figure 3 for Laplace and Gaussian noise respectively. Figures 2-3 show the results of the numerical simulation of our adaptive estimator $\hat{g}_{\ell,m}$. Figure 4(a) contains a plot of the penalized criterion function versus the kernel bandwidth m and Figure 4(b) the estimated MISE as a function of m. We give this plot for kurtotic distribution and Laplace noise only, but the behaviour is the same in all the other cases. It is clear from Figure 4(a), that the value of $\hat{m}$ is the unambiguous minimizer of $-\|\hat{g}_{\ell,m}\|^2 + \text{pen}_\ell(m)$. We also see that $\hat{m}$ provides a result close to m_{MISE}: in the Laplace case, for the Kurtotic density, the bandwidth which minimizes $\text{MISE}(m)$ in this case is $m_{\text{MISE}} = 0.1306$ and $\hat{m} = 0.1257$. This holds true for all test densities. In practice, the minimum of the MISE is close to the minimum of the penalized loss function, thus supporting the choice dictated by our theoretical procedure. Therefore, the proposed estimator provides effective results for the four test densities.

We also compare the performance of $\hat{g}_{\ell,\hat{m}}$ with that of $\tilde{g}_{\ell,\tilde{m}}$. For each density, samples with size $n = 1000$ were generated and the MISE was approximated as an average of the Integrated Squared Error over 100 replications. Table 2 presents the MISE from 100 repetitions for the two
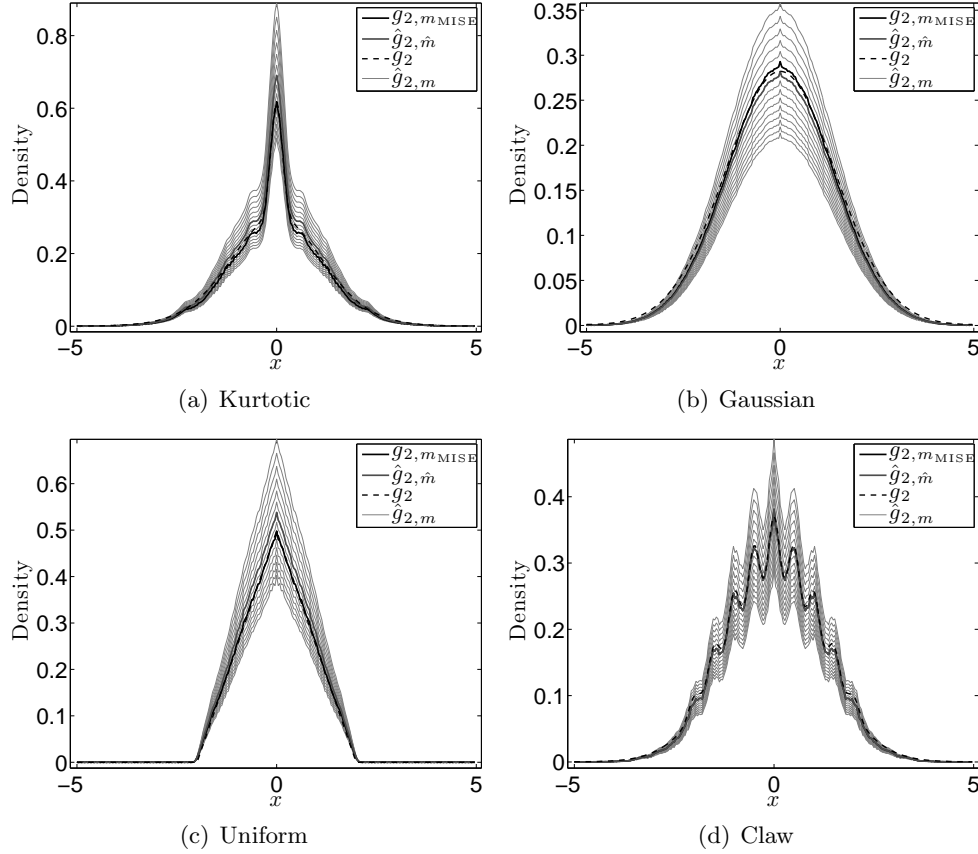
FIGURE 3. Gaussian noise. True density (dotted), density estimates (gray) and sample of 20 estimates (thin gray) out of 50 proposed to the selection algorithm obtained with a sample of $n = 1000$ data.

TABLE 2. $1000 \times$ MISE values from 100 replications of sample sizes $n = 1000$

	Noise			
	Gaussian		Laplace	
	known	unknown	known	unknown
Kurtotic	0.9200	0.4842	0.6678	0.4919
Gaussian	0.0285	0.0397	0.0350	0.0389
Uniform	1.4902	1.0301	1.1795	1.0300
Claw	0.1130	0.0494	0.0556	0.0485

types of errors. As noted in Comte and Lacour (2011), estimating the characteristic function of the noise reduces the risk compared to knowing it. Indeed, for virtually all cases, $\tilde{g}_{\ell,\tilde{m}}$ consistently showed lower risk than $\hat{g}_{\ell,\hat{m}}$, with the exception of the Gaussian density for which $\hat{g}_{\ell,\hat{m}}$ performs slightly better.
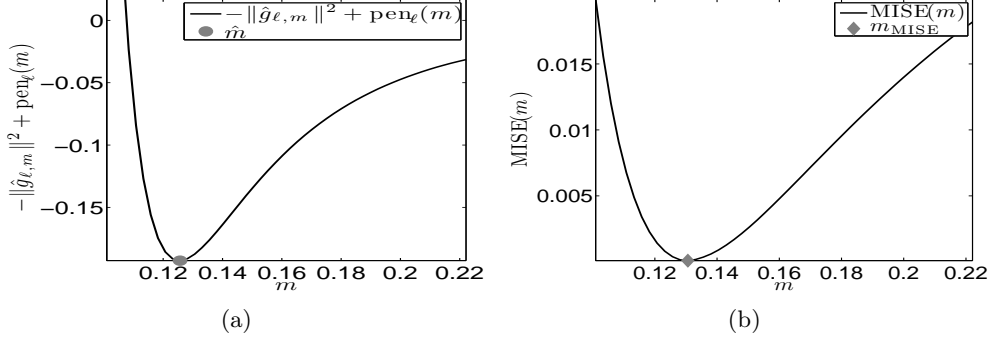
FIGURE 4. Laplace noise, Kurtotic density. (a): Graph of the function $-\|\hat{g}_{\ell,m}\|^2 + \mathrm{pen}_\ell(m)$ against the smoothing parameter m. (b): MISE(m). The gray diamond represents the global minimizer of MISE(m) and the gray circle represents the global minimizer of $-\|\hat{g}_{\ell,m}\|^2 + \mathrm{pen}_\ell(m)$.

5. PROOFS

5.1. Proof of Proposition 2.1. Upper bound on the MISE of $\hat{g}_{\ell,m}$.
Let us introduce the function $g_{\ell,m}$ defined by

$$g_{\ell,m}^*(u) = g_\ell^*(u)\mathbf{1}_{[-\pi m,\pi m]}(u) = (f^*(u))^\ell \mathbf{1}_{[-\pi m,\pi m]}(u) = \left(\frac{f_Y^*(u)}{f_\varepsilon(u)}\right)^\ell \mathbf{1}_{[-\pi m,\pi m]}(u).$$

The support of $(g_\ell - g_{\ell,m})^*$ being $[-\pi m, \pi m]^c$, disjoint from the support of $(g_{\ell,m} - \hat{g}_{\ell,m})^*$, we get

$$(17) \qquad \mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) = \|g_\ell - g_{\ell,m}\|^2 + \mathbb{E}(\|g_{\ell,m} - \hat{g}_{\ell,m}\|^2).$$

The Parseval identity yields

$$(18) \qquad \|g_\ell - g_{\ell,m}\|^2 = \frac{1}{2\pi}\int_{|u|\geq \pi m}|g_\ell^*(u)|^2 du = \frac{1}{2\pi}\int_{|u|\geq \pi m}|f^*(u)|^{2\ell}du.$$

The same argument gives

$$\mathbb{E}(\|g_{\ell,m} - \hat{g}_{\ell,m}\|^2) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{\mathbb{E}\left(|(\hat{f}_Y^*(u))^\ell - (f_Y^*(u))^\ell|^2\right)}{|f_\varepsilon^*(u)|^{2\ell}}du.$$

Using the binomial theorem and the inequality $|\sum_{i=1}^m b_i a_i|^2 \leq (\sum_{i=1}^m |b_i|)\sum_{i=1}^m |b_i||a_i|^2$, we get

$$
\begin{aligned}
\mathbb{E}\left(|(\hat{f}_Y^*(u))^\ell - (f_Y^*(u))^\ell|^2\right) &= \mathbb{E}\left(|(\hat{f}_Y^*(u) - f_Y^*(u) + f_Y^*(u))^\ell - (f_Y^*(u))^\ell|^2\right) \\
&= \mathbb{E}\left(\left|\sum_{k=1}^\ell \binom{\ell}{k}(\hat{f}_Y^*(u) - f_Y^*(u))^k (f_Y^*(u))^{\ell-k}\right|^2\right) \\
&\leq (2^\ell - 1)\sum_{k=1}^\ell \binom{\ell}{k}\mathbb{E}\left(|\hat{f}_Y^*(u) - f_Y^*(u)|^{2k}\right)|f_Y^*(u)|^{2(\ell-k)}.
\end{aligned}
$$

The Marcinkiewicz & Zygmund Inequality inequality applied to the n i.i.d. random variables $U_1,\dots,U_n$ with $U_j = e^{-iuY_j} - \mathbb{E}(e^{-iuY_1})$, $|U_j| \leq 2$ and the exponent $p = 2k$ (see Appendix)

yields

$$\mathbb{E}\left(|\hat{f}_Y^*(u) - f_Y^*(u)|^{2k}\right) \leq C_k \frac{1}{n^k},$$

with $C_k = (72k)^{2k}(2k/(2k-1))^k$. Therefore

$$\mathbb{E}\left(|(\hat{f}_Y^*(u))^\ell - (f_Y^*(u))^\ell|^2\right) \leq (2^\ell - 1) \sum_{k=1}^\ell \binom{\ell}{k} C_k \frac{1}{n^k} |f_Y^*(u)|^{2(\ell-k)}.$$

Since $f_Y^*(u) = f^*(u) f_\varepsilon^*(u)$, we obtain

(19)
$$\mathbb{E}(\|g_{\ell,m} - \hat{g}_{\ell,m}\|^2) \leq \frac{2^\ell - 1}{2\pi} \sum_{k=1}^\ell \binom{\ell}{k} C_k \frac{1}{n^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2(\ell-k)}}{|f_\varepsilon^*(u)|^{2k}} du.$$

Putting (17), (18) and (19) together, we get the desired result:

$$\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq \frac{1}{2\pi} \int_{|u| \geq \pi m} |f^*(u)|^{2\ell} du + \frac{2^\ell - 1}{2\pi} \sum_{k=1}^\ell \binom{\ell}{k} C_k \frac{1}{n^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2(\ell-k)}}{|f_\varepsilon^*(u)|^{2k}} du. \qquad \square$$

5.2. **Proof of Corollary 2.1.**

Case $f \in \mathcal{A}(\beta, L)$ and f_ε OS(a). The bias term is bounded by $L^\ell e^{-2\ell(\pi m)^\beta}$ since

$$\int_{|u| \geq \pi m} |f^*(u)|^{2\ell} du = \int_{|u| \geq \pi m} |f^*(u)|^2 e^{-2|u|^\beta} |f^*(u)|^{2(\ell-1)} e^{2(\ell-1)|u|^\beta} e^{-2\ell|u|^\beta} du$$

$$\leq L^\ell e^{-2\ell(\pi m)^\beta}$$

All the terms containing powers of $|f^*|$ in the numerator are integrable as f_ε is ordinary smooth. Therefore they are of order less than $1/n$. Then we have the general bound

$$\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq L^\ell e^{-2\ell(\pi m)^\beta} + \frac{1}{n^\ell} \int_{-\pi m}^{\pi m} \frac{du}{|f_\varepsilon(u)|^{2\ell}} + \frac{C}{n}$$

where $\int_{-\pi m}^{\pi m} du/|f_\varepsilon(u)|^{2\ell} = O(m^{2\ell a+1})$. Choosing $m = \pi^{-1}(\log(n)/(2\ell))^{1/\beta}$ gives the rate $1/n$.

Case $f \in \mathcal{S}(\alpha, L)$ and f_ε SS(b). Applying Lemma 6.2, we obtain

$$\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq L^\ell (\pi m)^{-2\ell\alpha} + \sum_{k=1}^\ell \binom{\ell}{k} C_k \frac{L^{\ell-k}}{c_\varepsilon n^k} \int_{-\pi m}^{\pi m} (1+u^2)^{-(\ell-k)} e^{2k|u|^b} du$$

$$\leq L^\ell (\pi m)^{-2\ell\alpha} + \sum_{k=1}^{\ell-1} \binom{\ell}{k} C_k \frac{L^{\ell-k}}{c_\varepsilon n^k} e^{2k(\pi m)^b} \int (1+u^2)^{-1} du + C \frac{1}{c_\varepsilon n^k} m^{-b} e^{2\ell(\pi m)^b}.$$

The choice $m = \pi^{-1}(\log(n)/(4\ell))^{1/b}$ makes all variance terms of order at most $O(1/\sqrt{n})$, so that it implies a rate $(\log(n))^{-4\ell\alpha/b}$, which is logarithmic, but faster when ℓ increases.

Case $f \in \mathcal{S}(\alpha, L), f_\varepsilon$ OS(a).
In this case, there exists a constant $K_\ell > 0$ such that

$$\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq K_\ell \left(m^{-2\alpha\ell} + \sum_{k=1}^{\ell-1} \frac{m^{(2ak-2\alpha(\ell-k)+1)_+} \log(m)^{\mathbf{I}_{2ak-2\alpha(\ell-k)+1=0}}}{n^k} + \frac{m^{2a\ell+1}}{n^\ell} \right).$$

Then α must be compared to the increasing sequence $a_k = (2ak+1)/(\ell-k)$.

First case: $2\alpha < a_1$. Then the trade-off between $m^{-2\alpha\ell}$ and $m^{2a-2\alpha(\ell-1)+1}/n$ implies the choice $m_{opt} = n^{1/(2a+2\alpha+1)}$ and the rate $n^{-2\alpha\ell/(2a+2\alpha+1)}$, where $2\alpha\ell/(2a+2\alpha+1) \in (0,1)$. In this situation, the terms following have orders

$$\sum_{k=2}^{\ell-1} \frac{m_{opt}^{2ak-2\alpha(\ell-k)+1}}{n^k} + \frac{m_{opt}^{2a\ell+1}}{n^\ell}$$

and we notice that

$$\frac{m_{opt}^{2ak-2\alpha(\ell-k)+1}}{n^k} = n^{\frac{2ak-2\alpha(\ell-k)+1}{2a+2\alpha+1}-k} = n^{\frac{-2\alpha\ell-k+1}{2a+2\alpha+1}} \leq n^{\frac{-2\alpha\ell}{2a+2\alpha+1}}$$

since $k \geq 2$. Therefore the rate is of order $n^{-2\alpha\ell/(2a+2\alpha+1)}$. Notice that making compromise using another variance term can be checked to give larger choice of m and thus worse rate, when inserted in the first variance term. Note also that the resulting rate is the usual rate of deconvolution to the power ℓ.

Second case: $2\alpha > a_1$. The term corresponding to $k=1$ is of order $O(1/n)$. Let $k_0 \geq 2$ be such that $a_{k_0-1} < 2\alpha < a_{k_0}$. Then the risk bound is

$$\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq C\left(m^{-2\alpha\ell} + \frac{1}{n} + \cdots + \frac{1}{n^{k_0-1}} + \sum_{k=k_0}^{\ell-1} \frac{m^{(2ak-2\alpha(\ell-k)+1)}}{n^k} + \frac{m^{2a\ell+1}}{n^\ell}\right).$$

Then the compromise is made between $^{-2\alpha\ell}$ and $m^{2ak_0-2\alpha(\ell-k_0)+1}/n^{k_0}$ and implies the choice $m_{opt} = n^{k_0/(2ak_0+2\alpha k_0+1)}$ and the rate $n^{-2\alpha\ell k_0/(2ak_0+2\alpha k_0+1)}$. But clearly

$$-\frac{2\alpha\ell k_0}{2ak_0+2\alpha k_0+1} \leq -1 \Leftrightarrow 2\alpha \geq \frac{2a}{\ell-1} + \frac{k_0-\ell+1}{k_0(\ell-1)} = \frac{2a}{\ell-1} + \frac{1}{\ell-1} - \frac{\ell-1}{k_0(\ell-1)}$$

and this last condition is fulfilled in our present case so that the rate is smaller than $1/n$.

In the same way, we can check that plugging the above m_{opt} in the other terms let them less than $1/n$. Namely

$$\frac{m_{opt}^{2a\ell+1}}{n^\ell} = n^{-\frac{2\alpha\ell k_0+\ell-k_0}{2ak_0+2\alpha k_0+1}} \leq n^{-1}, \quad \frac{m_{opt}^{2ak-2\alpha(\ell-k)+1}}{n^k} = n^{-\frac{2\alpha k_0\ell+k-k_0}{2ak_0+2\alpha k_0+1}} \leq n^{-1}, \text{ for } k=k_0+j, j \geq 1.$$

Consequently, we have

$$\text{if} \quad 2\alpha > \frac{2a}{\ell-1} + \frac{1}{\ell-1}, \text{ then} \quad \mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq Cn^{-1}.$$

Gathering both cases, we get that the rate for $f \in \mathcal{S}(\alpha, L), f_\varepsilon$ OS(a) is $n^{-2\alpha\ell/(2a+2\alpha+1)} \vee n^{-1}$. $\square$

5.3. **Proof of Theorem 2.1.** The result of Proposition 2.1 can be written

(20) $$\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq \|g_\ell - g_{\ell,m}\|^2 + \mathcal{V}_\ell(m),$$

where $\mathcal{V}_\ell(m)$ is defined by (8). Let us define the oracle

$$m^\star = \arg\min_{m \in \mathcal{M}_n} \left\{-\|g_{\ell,m}\|^2 + \mathcal{V}_\ell(m) + \text{pen}(m)\right\}$$

and note that, since $\|g_\ell - g_{\ell,m}\|^2 = \|g_\ell\|^2 - \|g_{\ell,m}\|^2$, we also have

$$m^\star = \arg\min_{m \in \mathcal{M}_n} \left\{\|g_\ell - g_{\ell,m}\|^2 + \mathcal{V}_\ell(m) + \text{pen}(m)\right\}.$$

Now, we start by writing that

(21) $$\|g_\ell - \hat{g}_{\ell,\hat{m}}\|^2 \leq 2\|g_\ell - \hat{g}_{\ell,m^\star}\|^2 + 2\|\hat{g}_{\ell,m^\star} - \hat{g}_{\ell,\hat{m}}\|^2.$$

i) We consider the random set $\Omega = \{\hat{m} \leq m^\star\}$, on which it holds that

$$
(22) \qquad \|\hat{g}_{\ell,m^\star} - \hat{g}_{\ell,\hat{m}}\|^2 \mathbf{I}_\Omega = (\|\hat{g}_{\ell,m^\star}\|^2 - \|\hat{g}_{\ell,\hat{m}}\|^2) \mathbf{I}_\Omega.
$$

Now, the definition of $\hat{m}$ implies that

$$
(23) \qquad -\|\hat{g}_{\ell,\hat{m}}\|^2 + \mathrm{pen}_\ell(\hat{m}) \leq -\|\hat{g}_{\ell,m^\star}\|^2 + \mathrm{pen}_\ell(m^\star),
$$

and thus

$$
\begin{aligned}
\|\hat{g}_{\ell,m^\star}\|^2 - \|\hat{g}_{\ell,\hat{m}}\|^2 &\leq \mathrm{pen}_\ell(m^\star) - \mathrm{pen}_\ell(\hat{m}) \\
&\leq \mathrm{pen}_\ell(m^\star).
\end{aligned}
$$
(24)

Then, (22) and (24) imply $\|\hat{g}_{\ell,m^\star} - \hat{g}_{\ell,\hat{m}}\|^2 \mathbf{I}_\Omega \leq \mathrm{pen}_\ell(m^\star)$. Plugging this in (21) yields

$$
\begin{aligned}
\mathbb{E}\left(\|g_\ell - \hat{g}_{\ell,\hat{m}}\|^2 \mathbf{I}_\Omega\right) &\leq 2\mathbb{E}(\|g_\ell - \hat{g}_{\ell,m^\star}\|^2) + 2\mathrm{pen}_\ell(m^\star) \\
&\leq 2(\|g_\ell - g_{\ell,m^\star}\|^2 + \mathcal{V}_\ell(m^\star)) + 2\mathrm{pen}_\ell(m^\star) \text{ with } (20), \\
&= 2 \min_{m \in \mathcal{M}_n} \left\{\|g_\ell - g_{\ell,m}\|^2 + \mathcal{V}_\ell(m) + \mathrm{pen}_\ell(m)\right\} \text{ by definition of } m^\star.
\end{aligned}
$$

Thus we have

$$
(25) \qquad \mathbb{E}\left(\|g_\ell - \hat{g}_{\ell,\hat{m}}\|^2 \mathbf{I}_\Omega\right) \leq 4 \min_{m \in \mathcal{M}_n} \left\{\|g_\ell - g_{\ell,m}\|^2 + \mathcal{V}_\ell(m) + \mathrm{pen}_\ell(m)\right\},
$$

which ends step **i)**.

ii) Now, we study the bound on Ω^c. Let us define, for $k > m$

$$
\mathrm{pen}_\ell(m,k) = 2^{2\ell} \lambda_\ell(m,k) \left(\Delta_\ell(k) - \Delta_\ell(m)\right),
$$

$$
\begin{aligned}
\lambda_\ell(m,k) = \max\Big\{ &2^{2\ell-1} \left(\log(1 + (k-m)^2(\Delta_\ell(k) - \Delta_\ell(m)))\right)^\ell, \\
&\frac{2^{4\ell-1}}{n^\ell} \left(\log(1 + (k-m)^2(\Delta_\ell(k) - \Delta_\ell(m)))\right)^{2\ell} \Big\}.
\end{aligned}
$$

First we write

$$
\begin{aligned}
\|\hat{g}_{\ell,m^\star} - \hat{g}_{\ell,\hat{m}}\|^2 \mathbf{I}_{\Omega^c} &= \left(\|\hat{g}_{\ell,m^\star} - \hat{g}_{\ell,\hat{m}}\|^2 - 2^{2\ell-1}\|g_{\ell,\hat{m}} - g_{\ell,m^\star}\|^2 - \mathrm{pen}_\ell(m^\star,\hat{m})\right)\mathbf{I}_{\Omega^c} \\
&\quad + \left(2^{2\ell-1}\|g_{\ell,\hat{m}} - g_{\ell,m^\star}\|^2 + \mathrm{pen}_\ell(m^\star,\hat{m})\right)\mathbf{I}_{\Omega^c} \\
&\leq \sup_{k \geq m^\star}\left\{\|\hat{g}_{\ell,m^\star} - \hat{g}_{\ell,k}\|^2 - 2^{2\ell-1}\|g_{\ell,k} - g_{\ell,m^\star}\|^2 - \mathrm{pen}_\ell(m^\star,k)\right\}_+ \mathbf{I}_{\Omega^c} \\
&\quad + 2^{2\ell-1}\|g_\ell - g_{\ell,m^\star}\|^2 + \sum_{k \geq m^\star} \mathrm{pen}_\ell(m^\star,k)\mathbf{I}_{k=\hat{m}}.
\end{aligned}
$$
(26)

Now, Inequality (23) writes

$$
\frac{1}{2}\mathrm{pen}_\ell(\hat{m}) \leq \|\hat{g}_{\ell,\hat{m}}\|^2 - \|\hat{g}_{\ell,m^\star}\|^2 + \mathrm{pen}_\ell(m^\star) - \frac{1}{2}\mathrm{pen}_\ell(\hat{m})
$$

and we notice that, if $k = \hat{m} > m^\star$, then $\|\hat{g}_{\ell,\hat{m}}\|^2 - \|\hat{g}_{\ell,m^\star}\|^2 = \|\hat{g}_{\ell,\hat{m}} - \hat{g}_{\ell,m^\star}\|^2$ and $\text{pen}_\ell(m^\star, k) \le 2\text{pen}_\ell(k)$, so that

$$
\begin{aligned}
\frac{1}{4}\text{pen}_\ell(m^\star, k) &\le \|\hat{g}_{\ell,k} - \hat{g}_{\ell,m^\star}\|^2 - \frac{1}{4}\text{pen}_\ell(m, k) + \text{pen}_\ell(m^\star) + \frac{1}{4}\text{pen}_\ell(m, k) - \frac{1}{2}\text{pen}_\ell(k) \\
&\le \left(\|\hat{g}_{\ell,k} - \hat{g}_{\ell,m^\star}\|^2 - \frac{1}{4}\text{pen}_\ell(m, k) - 2^{2\ell-1}\|g_{\ell,m^\star} - g_{\ell,k}\|^2 \right) \\
&\quad + \text{pen}_\ell(m^\star) + 2^{2\ell-1}\|g_{\ell,m^\star} - g_{\ell,k}\|^2 \\
&\le \sup_{k \ge m^\star} \left(\|\hat{g}_{\ell,k} - \hat{g}_{\ell,m^\star}\|^2 - \frac{1}{4}\text{pen}_\ell(m, k) - 2^{2\ell-1}\|g_{\ell,m^\star} - g_{\ell,k}\|^2 \right) \\
&\quad + \text{pen}_\ell(m^\star) + 2^{2\ell-1}\|g_{\ell,m^\star} - g_\ell\|^2.
\end{aligned}
$$

It follows that, on Ω^c,

$$
\begin{aligned}
\frac{1}{4} \sum_{k > m^\star} \text{pen}_\ell(m^\star, k) \mathbf{I}_{\hat{m}=k} &\le \sup_{k > m^\star} \left(\|\hat{g}_{\ell,k} - \hat{g}_{\ell,m^\star}\|^2 - \frac{1}{4}\text{pen}_\ell(m, k) - 2^{2\ell-1}\|g_{\ell,m^\star} - g_{\ell,k}\|^2 \right) \\
&\quad + \text{pen}_\ell(m^\star) + 2^{2\ell-1}\|g_{\ell,m^\star} - g_\ell\|^2.
\end{aligned}
$$

Inserting this in (26) implies

$$
\|\hat{g}_{\ell,m^\star} - \hat{g}_{\ell,\hat{m}}\|^2 \mathbf{I}_{\Omega^c} \le 3 \sup_{k \ge m^\star} \left\{ \|\hat{g}_{\ell,m^\star} - \hat{g}_{\ell,k}\|^2 - 2^{2\ell-1}\|g_{\ell,k} - g_{\ell,m^\star}\|^2 - \text{pen}_\ell(m^\star, k) \right\}_+ \mathbf{I}_{\Omega^c}
$$
$$
\tag{27} + 3\, 2^{2\ell-1}\|g_\ell - g_{\ell,m^\star}\|^2 + 2\text{pen}_\ell(m^\star).
$$

We can prove the following proposition, see Section 5.4.

Proposition 5.1. *Under the assumptions of Theorem 2.1, there exists a constant K_ℓ such that*

$$
\mathbb{E} \left(\sup_{k > m^\star} \left\{ \|\hat{g}_{\ell,m^\star} - \hat{g}_{\ell,k}\|^2 - 2^{2\ell-1}\|g_{\ell,k} - g_{\ell,m^\star}\|^2 - \text{pen}_\ell(m^\star, k) \right\}_+ \right) \le \frac{K_\ell}{n^\ell}.
$$

Then Inequality (27) and Proposition 5.1 imply

$$
\begin{aligned}
\mathbb{E}(\|\hat{g}_{\ell,m^\star} - \hat{g}_{\ell,\hat{m}}\|^2 \mathbf{I}_{\Omega^c}) &\le 3 \quad 2^{2\ell-1}\|g_\ell - g_{\ell,m^\star}\|^2 + 2\text{pen}_\ell(m^\star) + \frac{3K_\ell}{n^\ell} \\
&\le \quad 4[\|g_\ell - g_{\ell,m^\star}\|^2 + \mathcal{V}_\ell(m^\star) + \text{pen}_\ell(m^\star)] + \frac{3K_\ell}{n^\ell}
\end{aligned}
$$

which gives the result on Ω^c, by using again the definition of $m^\star$ and plugging this in (21). Gathering the two steps **i)** and **ii)** gives the result of Theorem 2.1. $\square$

The rates of convergence follow from those of Theorem 2.1 and Corollary 2.1.

5.4. **Proof of Proposition 5.1.** The proof of Proposition 5.1 relies on the following Lemma:

Lemma 5.1. *Let $X_1, \ldots, X_n$ be centered independent and identically distributed random variables, with $|X_i| \le b$, $\forall i$ a.s. and $\text{Var}(X_i) \le v^2$. Then, for $S_n = X_1 + \cdots + X_n$, and $H > 0$,*

$$
\mathbb{E} \left(\left|\frac{S_n}{n}\right|^p - H^p \right)_+ \le \frac{2^p \Gamma(p/2+1)}{n^{p/2}} \exp(-n2^{2/p-2} H^2/v^2) + \frac{2^p \Gamma(p+1)}{n^p} \exp(-n2^{1/p-2} H/b).
$$

Proof of Lemma 5.1. The proof of Lemma 5.1 relies on Bernstein Inequality, which gives

$$
\mathbb{P} \left(\left|\frac{S_n}{n}\right| \ge x \right) \le 2 \exp \left(-\frac{nx^2/2}{v^2 + bx} \right) \le 2 \left\{ \exp \left(-\frac{nx^2}{2v^2} \right) + \exp \left(-\frac{nx}{2b} \right) \right\}.
$$

Thus

$$\mathbb{E}\left(\left|\frac{S_n}{n}\right|^p - H^p\right)_+ \leq \int_0^{+\infty} \mathbb{P}\left(\left|\frac{S_n}{n}\right|^p - H^p \geq z\right) dz$$

$$\leq \int_0^{+\infty} \mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq [H^p + z]^{1/p}\right) dz$$

$$\leq 2\int_0^{+\infty}\left\{\exp\left(-\frac{n[H^p+z]^{2/p}}{2v^2}\right) + \exp\left(-\frac{n[H^p+z]^{1/p}}{2b}\right)\right\} dz$$

Now we use that by concavity inequality, we have that, for $q \geq 1$ and $x, y \geq 0$, $(x+y)^{1/q} \geq 2^{1/q-1}(x^{1/q} + y^{1/q})$, which implies for $p \geq 2$,

$$\mathbb{E}\left(\left|\frac{S_n}{n}\right|^p - H^p\right)_+ \leq 2\exp(-n2^{2/p-2}H^2/v^2)\int_0^{+\infty}\exp\left(-n\frac{2^{2/p-1}z^{2/p}}{2v^2}\right) dz$$

$$+2\exp(-n2^{1/p-2}H/c)\int_0^{+\infty}\exp\left(-n\frac{2^{1/p-1}z^{1/p}}{2b}\right) dz$$

$$= \frac{2^p\Gamma(p/2+1)}{n^{p/2}}\exp(-n2^{2/p-2}H^2/v^2) + \frac{2^p\Gamma(p+1)}{n^p}\exp(-n2^{1/p-2}H/b),$$

which is the announced result. $\square$.

Proof of Proposition 5.1. For $k > m^\star$, we have

$$\|\hat{g}_{\ell,k} - \hat{g}_{\ell,m^\star}\|^2 = \int_{\pi m^\star \leq |u| \leq \pi k}\left|\frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)}\right|^{2\ell} du$$

$$\leq \int_{\pi m^\star \leq |u| \leq \pi k}\frac{2^{2\ell-1}|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}}du + \int_{\pi m^\star \leq |u| \leq \pi k}\frac{2^{2\ell-1}|f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}}du$$

$$= \int_{\pi m^\star \leq |u| \leq \pi k}\frac{2^{2\ell-1}|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}}du + 2^{2\ell-1}\|g_{\ell,k} - g_{\ell,m^\star}\|^2.$$

Therefore

$$\mathbb{E}\left(\sup_{k>m^\star}\left\{\|\hat{g}_{\ell,m^\star} - \hat{g}_{\ell,k}\|^2 - 2^{2\ell-1}\|g_{\ell,k} - g_{\ell,m^\star}\|^2 - \text{pen}_\ell(m^\star, k)\right\}_+\right)$$

$$\leq \mathbb{E}\left(\sup_{k>m^\star}\left\{\int_{\pi m^\star \leq |u| \leq \pi k}\frac{2^{2\ell-1}|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}}du - \text{pen}_\ell(m^\star, k)\right\}_+\right)$$

$$\leq \sum_{k>m^\star}\mathbb{E}\left(\int_{\pi m^\star \leq |u| \leq \pi k}\frac{2^{2\ell-1}|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}}du - \text{pen}_\ell(m^\star, k)\right)_+$$

$$(28) \qquad = \sum_{k>m^\star}\mathbb{E}\left(\int_{\pi m^\star \leq |u| \leq \pi k}\frac{2^{2\ell-1}}{|f_\varepsilon^*(u)|^{2\ell}}\left(|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell} - \frac{\lambda_\ell(m^\star, k)}{n^\ell}\right)_+ du\right).$$

We apply now Lemma 5.1 with $b = v = 1$ and

$$H^{2\ell} \geq \frac{2^{2\ell-1}}{n^\ell}\left(\log(1 + (k - m^\star)^2(\Delta_\ell(k) - \Delta_\ell(m^\star)))\right)^\ell$$

in the first exponential term of the bound, and

$$H^{2\ell} \geq \frac{2^{4\ell-1}}{n^{2\ell}} \left(\log(1 + (k - m^\star)^2 (\Delta_\ell(k) - \Delta_\ell(m^\star))) \right)^{2\ell}$$

in the second term. We obtain

$$\mathbb{E} \left(|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell} - \frac{\lambda_\ell(m^\star, k)}{n^\ell} \right)_+ \leq \frac{2^{2\ell}\ell!}{n^\ell} \frac{1}{(k - m^\star)^2(\Delta_\ell(k) - \Delta_\ell(m^\star))}$$
$$+ \frac{2^{2\ell}(2\ell)!}{n^{2\ell}} \frac{1}{(k - m^\star)^2(\Delta_\ell(k) - \Delta_\ell(m^\star))}.$$

Plugging this into (28) yields

$$\mathbb{E} \left(\sup_{k > m^\star} \left\{ \|\hat{g}_{\ell,m^\star} - \hat{g}_{\ell,k}\|^2 - 2^{2\ell-1}\|g_{\ell,k} - g_{\ell,m^\star}\|^2 - \text{pen}_\ell(m^\star, k) \right\}_+ \right)$$
$$\leq 2^{2\ell-1} \left(\frac{2^{2\ell}\ell!}{n^\ell} + \frac{2^{2\ell}(2\ell)!}{n^{2\ell}} \right) \sum_{k > m^\star} \frac{1}{(k - m^\star)^2} \leq \frac{\pi^2}{6} \frac{2^{4\ell}(2\ell)!}{n^\ell}.$$

This ends the proof Proposition 5.1. □

5.5. **Proof of Theorem 3.1.** Let us first introduce some notations. For $k > m$, set

$$\hat{\Delta}(m, k) = \hat{\Delta}(k) - \hat{\Delta}(m) \quad \text{and} \quad \hat{\Delta}^f(m, k) = \hat{\Delta}^f(k) - \hat{\Delta}^f(m),$$

$$\widehat{\text{qen}}(m, k) := 2^{2\ell+1}\hat{\lambda}_1^2(m, k)\frac{\hat{\Delta}(m, k)}{n^\ell} + 2^{2\ell+1}\kappa^{2\ell} \log^\ell(M(k - m))\frac{\hat{\Delta}^f(m, k)}{M^\ell}$$

with

$$\hat{\lambda}_1(m, k) = \max \left\{ \sqrt{2^{-1/\ell+2} \log \left(1 + \hat{\Delta}(m, k)(k - m)^2 \right)}, \frac{2^{-1/2\ell+2}}{\sqrt{n}} \log \left(1 + \hat{\Delta}(m, k)(k - m)^2 \right) \right\}.$$

Now we can start the proof of Theorem 3.1. We denote by $m^\star$ the oracle cutoff defined by

$$m^\star = \arg \min_{m \in \mathcal{M}_n} \left\{ -\|g_{\ell,m}\|^2 + \mathbb{V}_\ell(m) + \text{qen}_\ell(m) \right\},$$

where $\mathbb{V}_\ell(m)$ is defined by (14). We have

$$(29) \qquad \|g_\ell - \tilde{g}_{\ell,\tilde{m}}\|^2 \leq 2 \|g_\ell - \tilde{g}_{\ell,m^\star}\|^2 + 2 \|\tilde{g}_{\ell,m^\star} - \tilde{g}_{\ell,\tilde{m}}\|^2$$

• Let us notice on the set $G = \{\tilde{m} \leq m^\star\}$:

$$\|\tilde{g}_{\ell,m^\star} - \tilde{g}_{\ell,\tilde{m}}\|^2 \mathbf{1}_G = \left(\|\tilde{g}_{\ell,m^\star}\|^2 - \|\tilde{g}_{\ell,\tilde{m}}\|^2 \right) \mathbf{1}_G.$$

Besides according to the definition of $\tilde{m}$, one has the following inequality:

$$(30) \qquad - \|\tilde{g}_{\ell,\tilde{m}}\|^2 + \widehat{\text{qen}}_\ell(\tilde{m}) \leq - \|\tilde{g}_{\ell,m^\star}\|^2 + \widehat{\text{qen}}_\ell(m^\star),$$

which implies $- \|\tilde{g}_{\ell,\tilde{m}}\|^2 \leq - \|\tilde{g}_{\ell,m^\star}\|^2 + \widehat{\text{qen}}_\ell(m^\star)$. Thus

$$\|\tilde{g}_{\ell,m^\star} - \tilde{g}_{\ell,\tilde{m}}\|^2 \mathbf{1}_G = \left(\|\tilde{g}_{\ell,m^\star}\|^2 - \|\tilde{g}_{\ell,\tilde{m}}\|^2 \right) \mathbf{1}_G \leq \widehat{\text{qen}}_\ell(m^\star).$$

Taking expectation, we apply the following Lemma

Lemma 5.2. *There exists a positive constant C such that for any arbitrary $m \in \mathcal{M}_n$*

$$(31) \qquad \mathbb{E} \left[\widehat{\text{qen}}_\ell(m) \right] \leq C\text{qen}_\ell(m).$$

It yields for some positive constant C

$$\mathbb{E}\left[\|g_\ell - \tilde{g}_{\ell,\tilde{m}}\|^2 \, \mathbf{1}_G\right] \leq 2\mathbb{E}\left[\|g_\ell - \tilde{g}_{\ell,m^\star}\|^2\right] + 2\mathbb{E}\left[\widehat{\mathrm{qen}}_\ell(m^\star)\right]$$

$$\leq 2\|g_\ell - g_{\ell,m^\star}\|^2 + 2\mathbb{V}_\ell(m^\star) + 2C\mathrm{qen}_\ell(m^\star).$$

We just proved the result on G

$$(32) \qquad \mathbb{E}\left[\|g_\ell - \tilde{g}_{\ell,\tilde{m}}\|^2 \mathbf{1}_G\right] \leq C \inf_{m \in \mathcal{M}_n}\left\{\|g_\ell - g_{\ell,m}\|^2 + \mathbb{V}_\ell(m) + \mathrm{qen}_\ell(m)\right\}.$$

• We now consider the set $G^c = \{\tilde{m} > m^\star\}$.

$$\|\tilde{g}_{\ell,\tilde{m}} - \tilde{g}_{\ell,m^\star}\|^2 \, \mathbf{1}_{G^c} = \left(\|\tilde{g}_{\ell,\tilde{m}} - \tilde{g}_{\ell,m^\star}\|^2 - 4\|g_{\ell,\tilde{m}} - g_{\ell,m^\star}\|^2 - \frac{1}{2}\widehat{\mathrm{qen}}(m^\star,\tilde{m})\right)\mathbf{1}_{G^c}$$

$$+ \left(4\|g_{\ell,\tilde{m}} - g_{\ell,m^\star}\|^2 + \frac{1}{2}\widehat{\mathrm{qen}}(m^\star,\tilde{m})\right)\mathbf{1}_{G^c}$$

$$\leq \sup_{\substack{k \geq m^\star \\ k \in \mathcal{M}_n}}\left\{\|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m^\star}\|^2 - 4\|g_{\ell,k} - g_{\ell,m^\star}\|^2 - \frac{1}{2}\widehat{\mathrm{qen}}(m^\star,k)\right\}_+$$

$$(33) \qquad\qquad + 4\|g_\ell - g_{\ell,m^\star}\|^2 + \frac{1}{2}\sum_{\substack{k \geq m^\star \\ k \in \mathcal{M}_n}}\widehat{\mathrm{qen}}(m^\star,k)\mathbf{1}\{\tilde{m} = k\}.$$

Let us first notice the following inequality

$$(34) \qquad\qquad\qquad \forall k > m, \quad \widehat{\mathrm{qen}}(m,k) \leq \widehat{\mathrm{qen}}_\ell(k).$$

Besides by definition of $\tilde{m}$ (see (15)), on the set $\{\tilde{m} = k\} \cap G^c$ and applying (30), we have

$$\frac{1}{2}\left(\widehat{\mathrm{qen}}_\ell(k) - \widehat{\mathrm{qen}}_\ell(m^\star)\right) \leq \|\tilde{g}_{\ell,\tilde{m}} - \tilde{g}_{\ell,m^\star}\|^2 - \frac{1}{2}\left(\widehat{\mathrm{qen}}_\ell(k) - \widehat{\mathrm{qen}}_\ell(m^\star)\right)$$

so that

$$\frac{1}{2}\widehat{\mathrm{qen}}_\ell(k) \leq \|\tilde{g}_{\ell,\tilde{m}} - \tilde{g}_{\ell,m^\star}\|^2 - \frac{1}{2}\widehat{\mathrm{qen}}_\ell(m^\star,k) + \frac{1}{2}\widehat{\mathrm{qen}}(m^\star)$$

$$\leq \left(\|\tilde{g}_{\ell,\tilde{m}} - \tilde{g}_{\ell,m^\star}\|^2 - 4\|g_{\ell,\tilde{m}} - g_{\ell,m^\star}\|^2 - \frac{1}{2}\widehat{\mathrm{qen}}(m^\star,k)\right) + 4\|g_{\ell,\tilde{m}} - g_{\ell,m^\star}\|^2 + \frac{1}{2}\widehat{\mathrm{qen}}_\ell(m^\star)$$

$$(35) \leq \left(\|\tilde{g}_{\ell,\tilde{m}} - \tilde{g}_{\ell,m^\star}\|^2 - 4\|g_{\ell,\tilde{m}} - g_{\ell,m^\star}\|^2 - \frac{1}{2}\widehat{\mathrm{qen}}(m^\star,k)\right) + 4\|g_\ell - g_{\ell,m^\star}\|^2 + \frac{1}{2}\widehat{\mathrm{qen}}_\ell(m^\star)$$

Now using Inequalities (34) and (35)

$$\frac{1}{2}\sum_{\substack{k \geq m^\star \\ k \in \mathcal{M}_n}}\widehat{\mathrm{qen}}(m^\star,k) \leq \sup_{\substack{k \geq m^\star \\ k \in \mathcal{M}_n}}\left\{\|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m^\star}\|^2 - 4\|g_{\ell,\tilde{m}} - g_{\ell,m^\star}\|^2 - \frac{1}{2}\widehat{\mathrm{qen}}(m^\star,k)\right\}_+$$

$$+ 4\|g_\ell - g_{\ell,m^\star}\|^2 + \frac{1}{2}\widehat{\mathrm{qen}}_\ell(m^\star).$$

From Inequality (33), we now have

$$\|\tilde{g}_{\ell,\tilde{m}} - \tilde{g}_{\ell,m^\star}\|^2 \, \mathbf{1}_{G^c} \leq 2\sup_{\substack{k \geq m^\star \\ k \in \mathcal{M}_n}}\left\{\|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m^\star}\|^2 - 4\|g_{\ell,k} - g_{\ell,m^\star}\|^2 - \frac{1}{2}\widehat{\mathrm{qen}}(m^\star,k)\right\}_+$$

$$+ 8\|g_\ell - g_{\ell,m^\star}\|^2 + \frac{1}{2}\widehat{\mathrm{qen}}_\ell(m^\star).$$

Taking expectation the first summand is negligible by applying the following Proposition.

Proposition 5.2. *Under (**A1**) and (**A2**), there is a positive constant C such that for any arbitrary $m \in \mathcal{M}_n$*

$$(36) \qquad \mathbb{E}\left[\sup_{\substack{k \geq m \\ k \in \mathcal{M}_n}} \left\{ \|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m}\|^2 - 2^{2\ell} \|g_{\ell,k} - g_{\ell,m}\|^2 - \frac{1}{2}\widehat{\mathrm{qen}}(m,k) \right\}_+ \right] \leq \frac{C}{n^\ell} + \frac{C}{M^\ell}.$$

Finally gathering the last result with (29), we have

$$(37) \qquad \mathbb{E}\left[\|g_\ell - \tilde{g}_{\ell,\tilde{m}}\|^2 \mathbf{1}_{G^c}\right] \leq C\left(\|g_\ell - g_{\ell,m^\star}\|^2 + \mathbb{V}_\ell(m^\star) + \mathrm{qen}(m^\star)\right) + C'\left(\frac{1}{n^\ell} + \frac{1}{M^\ell}\right)$$

This combining with (32) complete the proof. $\square$

5.6. **Proof of Lemma 5.2.** Before proving Lemma 5.2, we first need to prove two auxiliary lemmas. In the sequel, C will always denote some universal positive constant, but the value may vary from line to line.

Lemma 5.3. *For an estimator of f_ε^* defined by (10), assume $\kappa > \sqrt{c_1 p}$. Let $\tau \geq 2\kappa$ and $x \geq 1$. Then for some positive constant C*

$$\mathbb{P}\left[\exists u \in \mathbb{R} : |\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)| > \tau \left(\log(Mx)\right)^{1/2} w(u)^{-1} M^{-1/2}\right] \leq Cx^{-p}M^{-p}$$

Proof of Lemma 5.3. We write the decomposition

$$\left|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)\right| \leq \left|\tilde{f}_\varepsilon^*(u) - \hat{f}_\varepsilon^*(u)\right| + \left|\hat{f}_\varepsilon^*(u) - f_\varepsilon^*(u)\right| \leq 2k_M(u) + \left|\hat{f}_\varepsilon^*(u) - f_\varepsilon^*(u)\right|.$$

Using the previous inequality and applying Lemma 5.5 in Kappus (2014), we have

$$\mathbb{P}\left[\exists u \in \mathbb{R} : |\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)| > \tau \left(\log(Mx)\right)^{1/2} w(u)^{-1} M^{-1/2}\right]$$

$$\leq \mathbb{P}\left[\exists u \in \mathbb{R} : |\hat{f}_\varepsilon^*(u) - f_\varepsilon^*(u)| + 2k_M(u) > \tau \left(\log(Mx)\right)^{1/2} w(u)^{-1} M^{-1/2}\right]$$

$$\leq \mathbb{P}\left[\exists u \in \mathbb{R} : |\hat{f}_\varepsilon^*(u) - f_\varepsilon^*(u)| > (\tau - 2\kappa) \left(\log(Mx)\right)^{1/2} w(u)^{-1} M^{-1/2}\right]$$

$$\leq Cx^{-p}M^{-p}. \qquad \square$$

Proof of Lemma 5.2. For $q = 1/2$ or 1, using Cauchy-Schwarz's inequality, we have

$$\mathbb{E}\left[\log^q\left(1 + \hat{\Delta}(m)m^2\right)\hat{\Delta}(m)\right] \leq \sqrt{\mathbb{E}\left[\log^{2q}\left(1 + \hat{\Delta}(m)m^2\right)\right]\mathbb{E}\left[\hat{\Delta}^2(m)\right]}$$

Let $A_p(x) = \left|\dfrac{1}{\tilde{f}_\varepsilon^*(x)} - \dfrac{1}{f_\varepsilon^*(x)}\right|^{2p} \bigg/ \left(\dfrac{1}{|f_\varepsilon^*(x)|^{2p}} \wedge \dfrac{k_M^{2p}(x)}{|f_\varepsilon^*(x)|^{4p}}\right)$

$$\hat{\Delta}(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{w(u)^{-2}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} w(u)^{-2\ell}\left|\frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} + \frac{1}{f_\varepsilon^*(u)}\right|^{2\ell} du$$

$$\leq 2\Delta(m) + \frac{1}{\pi}\int_{-\pi m}^{\pi m} w(u)^{-2\ell}\left|\frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)}\right|^{2\ell} du$$

$$\leq 2\Delta(m) + 2\Delta(m)\sup_{u \in \mathbb{R}} A_\ell(u) \leq 2\Delta(m)(1 + \sup_{u \in \mathbb{R}} A_\ell(u))$$

and applying Lemma 4.1 in Kappus and Mabon (2013), we get $\mathbb{E}\left[\hat{\Delta}^2(m)\right] \leq 4C\Delta^2(m)$. By Jensen inequality (since log is concave)

$$\mathbb{E}\left[\log^{2q}\left(1+\hat{\Delta}(m)m^2\right)\right] \leq \log^{2q}\left(\mathbb{E}\left[1+\hat{\Delta}(m)m^2\right]\right) \leq \log^{2q}\left(1+\mathbb{E}\left[\hat{\Delta}(m)\right]m^2\right)$$

$$\leq \log^{2q}\left(1+2\Delta(m)\left(1+\mathbb{E}\left[\sup_{u\in\mathbb{R}}A_\ell(u)\right]\right)m^2\right)$$

$$\leq \log^{2q}\left(1+2\Delta(m)\left(1+C\right)m^2\right) \leq C\log^{2q}\left(1+\Delta(m)m^2\right)$$

So

$$\mathbb{E}\left[\log^q\left(1+\hat{\Delta}(m)m^2\right)\hat{\Delta}(m)\right] \leq C\log^q\left(1+\Delta(m)m^2\right)\Delta(m)$$

which means $\mathbb{E}[\widehat{\mathrm{qen}}_1(m)] \leq C\mathrm{qen}_1(m)$. Consider now $\widehat{\mathrm{qen}}_2(m)$. Another application of Lemma 4.1 in Kappus and Mabon (2013) yields

$$\frac{1}{M^\ell}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{w(u)^{-2\ell}|\hat{f}_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}}du\right]$$

$$\leq \frac{2}{M^\ell}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{w(u)^{-2\ell}|f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}}du\right] + \frac{2}{M^\ell}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{w(u)^{-2\ell}|\hat{f}_Y^*(u)-f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}}du\right]$$

$$\leq \frac{4}{M^\ell}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{w(u)^{-2\ell}|f^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}}du\right]\mathbb{E}\left[1+\sup_{u\in\mathbb{R}}A_{2\ell}(u)\right]$$

$$+\frac{2}{M^\ell}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{w(u)^{-2\ell}|\hat{f}_Y^*(u)-f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}}du\right]$$

$$\leq \frac{4}{M^\ell}\Delta^f(m)\mathbb{E}\left[1+\sup_{u\in\mathbb{R}}A_{2\ell}(u)\right] + \frac{2}{M^\ell}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{w(u)^{-2\ell}|\hat{f}_Y^*(u)-f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}}du\right]$$

We use the fact that $\tilde{f}_\varepsilon^*(u) \geq M^{-1/2}(\log M)^{1/2}w(u)^{-1}$, $w(u) \leq 1$ as well as the independence of $\hat{f}_Y^*$ and $\tilde{f}_\varepsilon^*$ to find

$$\frac{1}{M^\ell}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{w(u)^{-2\ell}|\hat{f}_Y^*(u)-f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}}du\right]$$

$$= \frac{1}{M^\ell}\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\mathbb{E}\left[|\hat{f}_Y^*(u)-f_Y^*(u)|^{2\ell}\right]\mathbb{E}\left[\frac{w(u)^{-2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}}du\right] \leq \frac{c_\ell}{n^\ell}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{w(u)^{-2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}}du\right]$$

where c_ℓ is the constant of the Rosenthal inequality. Applying the same arguments as for the bounding of $\mathbb{E}[\widehat{\mathrm{qen}}_1(m)]$, we get $\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{w(u)^{-2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}}du\right] \leq C\Delta(m)$. This completes the proof.
□

5.7. **Proof of Proposition 5.2.** For $k > m$, let us introduce the following notation : $A(m,k) := \{u \in \mathbb{R}, |u| \in [\pi m, \pi k]\}$. We need to prove auxiliary lemmas before proving Proposition 5.2.

Lemma 5.4. *There is a positive constant C such that for any arbitrary $m \in \mathcal{M}_n$*

$$(38) \qquad \mathbb{E}\left[\sup_{k\geq m}\left\{\frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u)-f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}}du - \frac{1}{2^{2\ell+1}}\widehat{\mathrm{qen}}_1(m,k)\right\}_+\right] \leq \frac{C}{n^\ell}$$

Proof of Lemma 5.4.

$$\mathbb{E}\left[\sup_{\substack{k \geq m \\ k \in \mathcal{M}_n}} \left\{\frac{1}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du - \frac{1}{2^{2\ell+1}} \widehat{\mathrm{qen}}_1(m,k)\right\}_+\right]$$

$$\leq \mathbb{E}\left[\sum_{\substack{k \geq m \\ k \in \mathcal{M}_n}} \mathbb{E}\left[\left\{\frac{1}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du - \frac{1}{2^{2\ell+1}} \widehat{\mathrm{qen}}_1(m,k)\right\}_+ \Big| \tilde{f}_\varepsilon^*\right]\right]$$

$$\leq \mathbb{E}\left[\sum_{\substack{k \geq m \\ k \in \mathcal{M}_n}} \frac{1}{2\pi} \int_{A(m,k)} \mathbb{E}\left[\left\{\frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} - \frac{1}{2^{2\ell+1}} \frac{\hat{\lambda}^2(m,k)}{n^\ell |\tilde{f}_\varepsilon^*(u)|^{2\ell}}\right\}_+ \Big| \tilde{f}_\varepsilon^*\right] du\right]$$

Now $\hat{f}_Y^*(u)/\tilde{f}_\varepsilon^*(u)$ (conditional on $\tilde{f}_\varepsilon^*(u)$) is the sum of n i.i.d. random variables with variance $v^2 \leq 1/|\tilde{f}_\varepsilon^*(u)|^2$ which are bounded by $2/\tilde{f}_\varepsilon^*(u)$. Thus Lemma 5.1 gives

$$\mathbb{E}\left[\left\{\frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} - \frac{1}{2^{2\ell+1}} \frac{\hat{\lambda}^2(m,k)}{n^\ell |\tilde{f}_\varepsilon^*(u)|^{2\ell}}\right\}_+ \Big| \tilde{f}_\varepsilon^*\right]$$

$$\leq 2^\ell \Gamma(2^{\ell-1}+1) \frac{n^{-\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} \exp\left(-\hat{\lambda}^2(m,k)\right) + 2^\ell \Gamma(2^\ell+1) \frac{n^{-2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} \exp\left(-n^{1/2}\hat{\lambda}(m,k)\right)$$

$$\leq 2^\ell \Gamma(2^{\ell-1}+1) \frac{n^{-\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} (k-m)^{-2} \hat{\Delta}(m,k)^{-1} + 2^\ell \Gamma(2^\ell+1) \frac{n^{-2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} (k-m)^{-2} \hat{\Delta}(m,k)^{-1}$$

where we used the fact that

$$\hat{\lambda}(m,k) \leq \max\left\{\sqrt{2^{-1/\ell+2} \log\left(1 + \hat{\Delta}(m,k)(k-m)^2\right)}, \frac{2^{-1/2\ell+2}}{\sqrt{n}} \log\left(1 + \hat{\Delta}(m,k)(k-m)^2\right)\right\}.$$

We have thus shown for a universal positive constant C that, for any $m, k \in \mathcal{M}_n$

$$\int_{A(m,k)} \mathbb{E}\left[\left\{\frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} - \frac{1}{2^{\ell+1}} \frac{\hat{\lambda}^2(m,k)}{n^\ell |\tilde{f}_\varepsilon^*(u)|^{2\ell}}\right\}_+ \Big| \tilde{f}_\varepsilon^*\right] du$$

$$\leq \frac{C}{n^\ell} (k-m)^{-2} \hat{\Delta}(m,k)^{-1} \int_{A(m,k)} \frac{du}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} \leq \frac{C}{n^\ell}(k-m)^{-2}.$$

Finally $\mathbb{E}\left[\sup_{k \geq m}\{(2\pi)^{-1} \int_{A(m,k)} (|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell})/|\tilde{f}_\varepsilon^*(u)|^{2\ell} du - 2^{-2\ell+1} \widehat{\mathrm{qen}}_1(m,k)\}_+\right] \leq C/n^\ell$, which ends the proof. $\square$

Proof of Proposition 5.2. Applying Plancherel's formula we get

$$\|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m}\|^2 = \frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u))|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}}du$$

$$= \frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u))|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}}\mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)|>|f_\varepsilon^*(u)|\}}du + \frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u))|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}}\mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)|\le|f_\varepsilon^*(u)|\}}du$$

$$:= S_1 + S_2$$

Let us consider S_1.

$$(39)\qquad S_1 \le \frac{2^{2\ell-1}}{\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}}du + \underbrace{\frac{2^{2\ell-1}}{\pi}\int_{A(m,k)}\frac{|f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}}du,}_{2^{2\ell}\|g_{\ell,k}-g_{\ell,m}\|^2}$$

now S_2,

$$S_2 = \frac{1}{2\pi}\int_{A(m,k)}|\hat{f}_Y^*(u)|^{2\ell}\left|\frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} + \frac{1}{f_\varepsilon^*(u)}\right|^{2\ell}\mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)|>|f_\varepsilon^*(u)|\}}du$$

$$\le \frac{2^{2\ell-1}}{\pi}\int_{A(m,k)}|\hat{f}_Y^*(u)|^{2\ell}\left|\frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)}\right|^{2\ell}\mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)|\le|f_\varepsilon^*(u)|\}}du$$

$$(40)\qquad +\frac{2^{2\ell-1}}{\pi}\int_{A(m,k)}\left|\frac{\hat{f}_Y^*(u))}{f_\varepsilon^*(u)}\right|^{2\ell}\mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)|\le|f_\varepsilon^*(u)|\}}du := S_{2,1} + S_{2,2}.$$

Yet for the second term of (40), we can notice

$$(41)\qquad S_{2,2} \le \frac{2^{4\ell-2}}{\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}}du + 2^{4\ell-1}\|g_{\ell,k} - g_{\ell,m}\|^2.$$

For the first term of (40) we can write

$$S_{2,1} = \frac{2^{2\ell-1}}{\pi}\int_{A(m,k)}|\hat{f}_Y^*(u)|^{2\ell}\frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}|f_\varepsilon^*(u)|^{2\ell}}\mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)|\le|f_\varepsilon^*(u)|\}}du$$

$$(42)\qquad \le \frac{2^{2\ell-1}}{\pi}\int_{A(m,k)}|\hat{f}_Y^*(u)|^{2\ell}\frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}}\mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)|\le|f_\varepsilon^*(u)|\}}du.$$

Let us introduce the set following set

$$(43)\qquad C(m,k) = \left\{\forall u\in\mathbb{R}: |\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^2 \le 4\kappa^2\log(M(k-m))w(u)^{-2}M^{-1}\right\}$$

On $C(m,k)$, the following inequalities can be deduced

$$\frac{1}{2\pi}\int_{A(m,k)}|\hat{f}_Y^*(u)|^{2\ell}\frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}}\mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)|\le|f_\varepsilon^*(u)|\}}du$$

$$\le 4^\ell\kappa^{2\ell}\log^\ell(M(k-m))M^{-\ell}\frac{1}{2\pi}\int_{A(m,k)}\frac{w(u)^{-2\ell}|\hat{f}_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}}\mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)|\le|f_\varepsilon^*(u)|\}}du$$

$$\le 4^\ell\kappa^{2\ell}\log^\ell(M(k-m))M^{-\ell}\hat{\Delta}^f(m,k) \le \frac{1}{2^{2\ell+1}}\widehat{\text{qen}}_2(m,k).$$

Gathering Equations (39), (41) and (42), we have

$$\|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m}\|^2 = \frac{2^{4\ell-2}}{\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du + 2^{4\ell-1} \|g_{\ell,k} - g_{\ell,m}\|^2$$

$$+ \frac{2^{2\ell-1}}{\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)\leq |f_\varepsilon^*(u)|\}} du \left[\mathbf{1}_{C(m,k)} + \mathbf{1}_{C(m,k)^c}\right]$$

$$\leq \frac{2^{4\ell-2}}{\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du + 2^{4\ell-1} \|g_{\ell,k} - g_{\ell,m}\|^2 + \frac{1}{2}\widehat{\text{qen}}_2(m,k)$$

$$(44) \qquad + \frac{2^{2\ell-1}}{\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)\leq |f_\varepsilon^*(u)|\}} du\, \mathbf{1}_{C(m,k)^c}.$$

Starting from (44), we can now write the following inequalities

$$\|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m}\|^2 - 2^{4\ell-1} \|g_{\ell,k} - g_{\ell,m}\|^2 - \frac{1}{2}\widehat{\text{qen}}(m,k) \leq \frac{2^{4\ell-2}}{\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du$$

$$-\frac{1}{2}\widehat{\text{qen}}_1(m,k) \quad + \frac{2^{2\ell-1}}{\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)\leq |f_\varepsilon^*(u)|\}} du\, \mathbf{1}_{C(m,k)^c}$$

$$\leq 2^{2\ell} \left\{ \frac{1}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du - \frac{1}{2^{2\ell+1}}\widehat{\text{qen}}_1(m,k) \right\}_+$$

$$+ \frac{2^{2\ell-1}}{\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)\leq |f_\varepsilon^*(u)|\}} du\, \mathbf{1}_{C(m,k)^c}.$$

Taking expectation we get

$$\mathbb{E}\left[\sup_{\substack{k\geq m \\ k\in\mathcal{M}_n}} \left\{ \|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m}\|^2 - 2^{4\ell-1} \|g_{\ell,k} - g_{\ell,m}\|^2 - \frac{1}{2}\widehat{\text{qen}}(m,k) \right\}_+ \right]$$

$$\leq \sum_{\substack{k\geq m \\ k\in\mathcal{M}_n}} \mathbb{E}\left[2^{2\ell} \left\{ \frac{1}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du - \frac{1}{2^{2\ell+1}}\widehat{\text{qen}}_1(m,k) \right\}_+ \right]$$

$$+ \frac{2^{2\ell}}{2\pi} \sum_{\substack{k\geq m \\ k\in\mathcal{M}_n}} \mathbb{E}\left[\int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)\leq |f_\varepsilon^*(u)|\}} du\, \mathbf{1}_{C(m,k)^c} \right].$$

We can notice that on $C(m,k)^c$ defined by (43) following Lemma 5.3 for $x = k - m$ and $p = 3\ell$, we have $\mathbb{P}\left[C(m,k)^c\right] \leq M^{-3\ell}(k-m)^{-3\ell}$ and we get

$$\sum_{\substack{k\geq m \\ k\in\mathcal{M}_n}} \mathbb{E}\left[\int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)\leq |f_\varepsilon^*(u)|\}} du\, \mathbf{1}_{C(m,k)^c} \right]$$

$$\leq \sum_{\substack{k\geq m \\ k\in\mathcal{M}_n}} \mathbb{E}\left[\int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{2^{2\ell}|f_\varepsilon^*(u)|^{2\ell}}{k_M^{4\ell}(u)} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)\leq |f_\varepsilon^*(u)|\}} du\, \mathbf{1}_{C(m,k)^c} \right]$$

$$\leq \sum_{\substack{k \geq m \\ k \in \mathcal{M}_n}} \mathbb{E}\left[\int_{A(m,k)} 2^{2\ell}\kappa^{-4\ell}(\log M)^{-2\ell} w(u)^{4\ell} M^{2\ell} du\, \mathbf{1}_{C(m,k)^c}\right]$$

$$\leq \sum_{\substack{k \geq m \\ k \in \mathcal{M}_n}} 2^{2\ell}\kappa^{-4\ell}(\log M)^{-2\ell} M^{2\ell}(k-m)\mathbb{P}\left[C(m,k)^c\right] \leq \frac{C}{M^\ell},$$

and applying Lemma 5.4, we have

$$\mathbb{E}\left[\sup_{\substack{k \geq m \\ k \in \mathcal{M}_n}}\left\{\|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m}\|^2 - 2^{4\ell-1}\|g_{\ell,k} - g_{\ell,m}\|^2 - \frac{1}{2}\widehat{\text{qen}}(m,k)\right\}_+\right] \leq \frac{C}{n^\ell} + \frac{C}{M^\ell}$$

This completes the proof.$\square$

6. Appendix

We recall the following result, which can be found in Shiryaev (1996) (inequality (26) p.498).

Lemma 6.1 (Marcinkiewicz & Zygmund Inequality). *Let n be a positive integer, $p > 1$ and $U_1, \dots, U_n$ be n zero mean independent random variables such that $\sup_{j \in \{1,\dots,n\}} \mathbb{E}(|U_j|^p) < \infty$. Then, for $B_p = (18p)^p(p/(p-1))^{p/2}$,*

$$\mathbb{E}(|\sum_{j=1}^n U_j|^p) \leq B_p \mathbb{E}(|\sum_{j=1}^n |U_j|^2|^{p/2}),$$

We also give the following useful lemma, see Lemma 2 p. 35 in Butucea and Tsybakov (2008). For two functions u, v, we denote $u(x) \lesssim v(x)$ if there exists a positive constant C not depending on x such that $u(x) \leq Cv(x)$ and $u(x) \approx v(x)$ if $u(x) \lesssim v(x)$ and $v(x) \lesssim u(x)$.

Lemma 6.2. *Consider c, s nonnegative real numbers, and γ a real such that $2\gamma > -1$ if $c = 0$ or $s = 0$. Then, for all $m > 0$,*

- $\int_{-m}^m (x^2 + 1)^\gamma \exp(c|x|^s)dx \approx m^{2\gamma+1-s}e^{cm^s}$,

and if in addition $2\gamma > 1$ if $c = 0$ or $s = 0$,

- $\int_m^\infty (x^2 + 1)^{-\gamma} \exp(-c|x|^s)dx \approx m^{-2\gamma+1-s}e^{-cm^s}$.

References

Ahmad, I.A. and Fan, Y. (2001). Optimal bandwidth for kernel density estimators of functions of observations. *Statist. Probab. Lett.*, 51, (3), 245-251.

Ahmad, I.A. and Mugdadi, A.R. (2003). Analysis of kernel density estimation of functions of random variables. *J. Nonparametric Statistics*, 15, 579-605.

Bonhomme, S. and Robin, J.-M. (2010) Generalized non-parametric deconvolution with an application to earnings dynamics. *Rev. Econom. Stud.* **77**, 491-533.

Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *The Canadian Journal of Statistics*, **32**, 181-192.

Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the signal density in a semiparametric convolution model. *Bernoulli*, 11, **2**, 309-340.

Butucea, C. and Tsybakov, A. B. (2008). Sharp optimality in density deconvolution with dominating bias. II. *Theory Probab. Appl.* **52**, 237-249.

Caroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83, 1184-1186.

Cherubini, U., Mulinacci, S. and Romagnoli, S. (2011) On the distribution of the (un)bounded sum of random variables. *Insurance Math. Econom.* 48, no. 1, 56-63.

Chesneau, C., Comte, F. and Navarro, F. (2013). Fast nonparametric estimation for convolutions of densities, *The Canadian Journal of Statistics*, 41, 4, 617-636.

Chesneau, C. and Navarro, F. (2014). On a plug-in wavelet estimator for convolutions of densities. *Journal of Statistical Theory and Practice*, (to appear).

Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006). Penalized contrast estimator for density deconvolution. *The Canadian Journal of Statistics*, 34, 431-452.

Comte, F. and Genon-Catalot, V. (2012). Convolution power kernels for density estimation. *Journal of Statistical Planning and Inference*, 142, 7, 1698-1715.

Comte, F. and Lacour, C. (2011) Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc. Ser.* B Stat. Methodol., 73, 601-627.

Delaigle, A. and Gijbels, I. (2006). Estimation of boundary and discontinuity points in deconvolution problems. *Statistica Sinica*, 16, 773 -788.

Delaigle, A., Hall, P. and Meister, A. (2008) On deconvolution with repeated measurements. *Ann. Statist.* **36**, 665-685.

Devroye, L. (1989). Consistent deconvolution in density estimation. *Canad. Journ. Statist.*, 17, 235-239.

Du, J. and Schick, A. (2007). Root-n consistency and functional central limit theorems for estimators of derivatives of convolutions of densities. *Internat. J. Statist. Management Systems*, 2, 67-87.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problem. *Ann. Statist.*, 19, 1257-1272.

Fan, J. and Koo, J.Y. (2002). Wavelet deconvolution. *IEEE transactions on information theory*, 48, 734-747.

Frees, E. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.*, 89, 17-525.

Giné, E. and Mason, D.M. (2007). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.*, 35, 1105-1145.

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39, 1608-1632.

Johannes, J. (2009) Deconvolution with unknown error distribution. *Ann. Statist.* **37**, 2301-2323.

Jones, M.C. and Lotwick, H.W. (1984). Remark AS R50. A remark on algorithm AS 176. *Appl. Statist.*, 33 120-122.

Kappus, J. (2014) Adaptive nonparametric estimation for Lévy processes observed at low frequency. *Stochastic Process. Appl.* 124, 730-758.

Kappus J. and Mabon G. (2013) Adaptive density estimation in deconvolution problems with unknown error distribution. Preprint hal-00915982.

Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Acad. Sci. Paris Ser. I Math.*, 342 (11), 877-882.

Lepski, O.V. (1990). On a problem of adaptive estimation in gaussian white noise. *Theory of Probability and its Applications*, 35, 454-466.

Marron, J.S. and Wand, M.P. (1992), Exact Mean Integrated Squared Error. *Ann. Statist.*, 20, 712–736.

Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics.* **193**. Lecture Notes in Statistics. Springer-Verlag, Berlin.

Mugdadi, A.R. and Ahmad, I. (2004). A Bandwidth Selection for Kernel Density Estimation of Functions of Random Variables. *Computational Statistics and Data Analysis*, 47,1, 49-62.

Neumann, M.H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist.*, 7, 307-330.

Neumann, M. (2007). Deconvolution from panel data with unknown error distribution. *Journal of Multivariate Analysis*, **98**, 1955-1968.

Neumann, M. and Reiss, M. (2009). Nonparametric estimation for Levy processes from low-frequency observations. *Bernoulli*, 15, 223-248.

Nickl, R. (2007) Donsker-type theorems for nonparametric maximum likelihood estimators. *Probab. Theory Related Fields*, 138, no. 3-4, 411-449.

Nickl, R. (2009). On convergence and convolutions of random signed measures. *J. Theoret. Probab.*, 22, 38-56.

Panjer, H.H. and Willmot, G.E. (1992). Insurance Risk Models. Society of Actuaries, Schaumburg.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, 27, 2033-2053.

Prakasa Rao, B.L.S. (2004). Moment inequalities for supremum of empirical processes of U-statistic structure and application to density estimation. *J.Iran. Statist. Soc.*, 3, 59-68.

Rosenthal, H.P. (1970). On the subspaces of $\mathbb{L}_p$ ($p \geq 2$) spanned by sequences of independent random variables. *Israel Journal of Mathematics*, 8, 273-303.

Saavedra, A. and Cao, R. (2000). On the estimation of the marginal density of a moving average process. *Canad. J. Statist.*, 28, 799-815.

Schick, A. and Wefelmeyer, W. (2004). Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Statist.*, 16, 925-935.

Schick, A. and Wefelmeyer, W. (2007). Root n consistent density estimators of convolutions in weighted $\mathbb{L}_1$-norms. *J. Statist. Plann. Inference*, 137, 1765–1774.

Shiryaev, A. N. (1996) *Probability.* Translated from the first (1980) Russian edition by R. P. Boas. Second edition. Graduate Texts in Mathematics, **95**. Springer-Verlag, New York.

Silverman B.W. (1982), Algorithm AS 176. Kernel density estimation using the fast Fourier transform. *Appl. Statist.*, 31, 93–99.

Silverman, B. W. (1986), Density estimation: for statistics and data analysis, Chapman & Hall.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation.* Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. Springer Series in Statistics. Springer, New York.

Zhang, C.-H. (2005). Estimation of sums of random variables: examples and information bounds. *Ann. Statist.*, 33, no. 5, 2022-2041.

Zijaeva, Z.T. (1975) The estimation of the density of a sum of random variables. (Russian) *Izv. Akad. Nauk. USSR Ser. Fiz.-Mat. Nauk*, no. 3, 13-15, 95.