Minimax estimation of the noise level and of the signal density in a semiparametric convolution model

Cristina Butucea[1] and Catherine Matias[2]

1. Modal'X, Université Paris X, Bât. G, 200, avenue de la République, 92 001 Nanterre Cedex, France and Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VI, 175 rue du Chevaleret, 75 013 Paris, France. E-mail: butucea@ccr.jussieu.fr

2. Laboratoire de Statistique et Génôme, La Génopole, Tour Evry 2, 523 pl. des Terrasses de l'Agora, 91 000 Evry, France. E-mail: matias@genopole.cnrs.fr. Research done while at Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VI, 175, rue du Chevaleret, 75 013 Paris, France.

Abstract

We consider a semiparametric convolution model of an unknown signal with supersmooth noise having unknown scale parameter. We construct a consistent estimation procedure for the noise level and prove that its rate is optimal in the minimax sense. For identifiability reasons, the noise has to be smoother than the signal in this problem. Two convergence rates are distinguished according to different smoothness properties for the signal. In one case the rate is sharp optimal, i.e. the asymptotic value of the risk is evaluated up to a constant. Moreover, we construct a consistent estimator of the signal, by using a plug-in method in the classical kernel estimation procedure. We establish that the estimation of the signal is deteriorated comparatively to the case of entirely known noise distribution. In fact, nonparametric rates of convergence are governed by the rate of estimation of the noise level. We also prove that those rates are minimax (or nearly minimax in a few specific cases). Simulation results bring new ideas on practical use of our estimation algorithms.

Key words and phrases: Analytic densities, deconvolution, minimax estimation, noise level, pointwise risk, semiparametric model, Sobolev classes.
AMS Classification: 62G07, 62G20, 62C20.

Let us consider the observations Y_i, $i = 1, \dots, n$ such that

$$Y_i = X_i + \sigma\varepsilon_i,$$

where X_i and ε_i are independent and identically distributed real valued random variables, the two sequences $\{X_i\}$ and $\{\varepsilon_i\}$ being independent of each other. Two components are unknown in this model: the common law of X_i having probability density f (with respect to the Lebesgue measure on $\mathbb{R}$) and characteristic function Φ and the scale parameter $\sigma > 0$. Variables ε_i have a known supersmooth distribution, that is a known density function f^ε having a Fourier transform Φ^ε such that, for large enough $|u|$,

$$be^{-|u|^s} \leq |\Phi^\varepsilon(u)| \leq Be^{-|u|^s}, \tag{1}$$

for some known $s > 0$ and fixed constants b, $B > 0$.

In the more classical deconvolution problem the noise is supposed entirely known (the law as well as the scale parameter). In this case, minimax rates of convergence are described in the literature for various associations of smoothness classes for the signal (Hölder, Sobolev, Besov or analytic functions) and global behaviours of the errors' law. Even if the noise law is entirely known, estimators behave differently whether the characteristic function of the noise decays polynomially asymptotically (ordinary smooth noise) or exponentially (supersmooth noise). There is a huge amount of literature since the work by Caroll & Hall (1988). For detailed reviews see recent papers by Pensky and Vidakovic (1999) and Butucea and Tsybakov (2002).

Here, the signal and the scale parameter of the noise are unknown. In a first part, we are interested in recovering the scale of the noise $\sigma > 0$. Indeed, the assumption of entirely known noise is rather unrealistic from a practical point of view. Therefore, evaluating the scale parameter of the noise can relax this restraining assumption. Moreover, in a second part, we use it as a preliminary step in the nonparametric deconvolution problem of estimating the signal when the scale is unknown.

The estimation of the scale parameter and of the signal was already considered by Matias (2002) in the case of Gaussian errors and a large collection of signal functions, signals "without Gaussian component". The estimators of the scale parameter were based on Fourier, respectively Laplace, transform and they were proven to be consistent over some subclasses. Lower bounds of order $1/\log n$ were found for both estimation problems. It was already noted there that estimation of the signal is more difficult (larger lower bounds of order $1/\log n$) when the scale parameter is unknown than in the classical deconvolution problem.

This problem has been formulated by Matias (2002) in relation with error-in-variables non-linear regression. More generally, in physics and biology one needs to study models with main data which are not directly observed but mixed with noise. We suggest to use deconvolution with unknown scale parameter for the noise.

A similar problem was considered by Lindsay (1986) for mixture of exponential families with applications to Bayesian statistics. Among other results, he considers an infinitely divisible mixing density, with unknown parameters which are recovered via least-squares estimation. This problem is similar to noise-level evaluation in our model where the main signal is in a parametric exponential family. Thus, our results extend this estimation to nonparametric main signal.

Similarly to Zhang (1990), we can regard this model as a mixing model of location families. Zhang (1990) consider location (in θ) families $f^{\varepsilon}(\cdot - \theta)$ with mixing density $f(\theta)$. The observations $Y_i, i = 1, \ldots, n$ have density $\int f^{\varepsilon}(\cdot - \theta) f(\theta) d\theta$ and the mixing density f is estimated. More generally, in our model, the location families have an unknown scale parameter σ: $f^{\varepsilon}((\cdot - \theta)/\sigma)/\sigma$ that we estimate together with the mixing density f.

In this paper, we propose a new estimation algorithm for the scale parameter, prove its consistency and compute the upper bounds of its mean squared error in several different setups. Moreover, we prove that these rates are optimal by giving the corresponding minimax lower bounds.

We solve the problem of estimating the noise level in two possible setups:

Assumption (A) *We suppose that the signal belongs to the class $\mathcal{A}(\alpha, r)$ of densities whose Fourier transform is not decaying asymptotically faster than an exponential*

$$|\Phi(u)| \geq ce^{-\alpha|u|^r}, \quad |u| \text{ large enough,}$$

with known parameters $\alpha > 0$ and $r \in (0, s)$, and some arbitrary constant $c > 0$;

respectively,

Assumption (B) *The signal is in the class $\mathcal{B}(\beta)$ of densities having Fourier transform not decaying asymptotically faster than a polynomial*

$$|\Phi(u)| \geq c\,|u|^{-\beta}, \quad |u| \text{ large enough,}$$

with known parameter $\beta > 1$ and an arbitrary constant $c > 0$.

Under one of these assumptions, the model becomes identifiable since, considering Φ^Y, the Fourier transform of the distribution of the observations, we get:

$$\frac{\log|\Phi^Y(u)|}{|u|^s} = \frac{\log|\Phi(u)|}{|u|^s} + \frac{\log|\Phi^\varepsilon(\sigma u)|}{|u|^s} \underset{|u|\to\infty}{\longrightarrow} -\sigma^s.$$

Let us remark that information on the nonparametric signal as well as on the unknown noise level must be retrieved from the same sample of Y_i's. We know that more regular (in the sense of smoother) is the noise, more difficult is to recover information on the signal, that means slower become the rates of convergence of the minimax estimators. On the contrary, more information will be left for the noise scale parameter. Indeed, our results agree to this heuristics and give faster rates when the noise is significantly smoother than the signal and slower rates when the noise is smooth but behaving similarly as the signal. We note that if the signal becomes smoother than the noise the parameter becomes non identifiable.

These rates are overall slower when compared to classical parametric estimation. This is not surprising in this semiparametric model where we distinguish apart a parametric component from a nonparametric unknown function. We note that the rates are sharp minimax under Assumption (A) and nearly sharp under Assumption (B).

In a second part, we study the rates of convergence for the estimation of the signal, in the presence of unknown noise level, regarded as a nuisance parameter. These rates are significantly slower than the rates obtained in classical deconvolution problem, and are governed by the rates of estimation of the noise level. The estimators attaining the upper bounds are classical kernel estimators where we plug-in the estimated value of the underlying parameter. In the risk, this implies a study of uniform empirical processes explaining the loss of performance of this estimator. The corresponding lower bounds show that this loss is unavoidable. As in the classical nonparametric minimax theory, we construct particular parameters as far apart as possible so that the resulting models be close in some distance (the χ^2 -distance in our case). Note that the choice of the parameters is natural for our problem and it was made on the simplest basis possible (see Section 4). Note also that these results agree with the lower bounds established by Matias (2002) in the case $s = 2$ and are more precise as we study here the rates of convergence when the signal belongs to the classes $\mathcal{A}(\alpha, r)$ (with $0 < r < s$) and $\mathcal{B}(\beta)$ ($\beta > 1$), whereas Matias's lower bounds concern every signal with non-gaussian component.

In the rest of the paper, we define the estimation method for the scale parameter and study its consistency, for the defined parameters $\alpha > 0$ and $s > r > 0$ under Assumption (A), respectively $\beta > 1$ and $s > 0$ under Assumption (B) (Sections 1.1 and 1.2).

Then, using a plug-in method combined with the natural kernel deconvolution technique, we construct an estimator of the signal and study its rates of convergence (Section 1.3). In order to establish its rates of convergence, we need to add assumptions on the least smoothness the estimated signal may have as in classical deconvolution problem. Moreover, because of the plug-in method (unknown parameter is replaced by its consistent estimator), we also need the expression of the whole function $\Phi^\varepsilon(\cdot)$. For simplicity of notations, we decide to fix the noise exactly distributed according to a stable law (described in Section 1.3) which corresponds to parameter s belonging to the interval $(0, 2]$. This assumption is not very restrictive since most encountered examples are in this range of parameters, and could be relaxed at this point.

Next, we prove the optimality from a minimax point of view of the constructed estimators (for the noise level as well as for the signal, Section 2). In order to construct optimal paths proving these lower bounds we need specific density functions with known behaviour and explicit Fourier transform. Then, we also deal with stable laws for the noise. Proofs are presented in Sections 3 and 4.

In Section 5, we implement the algorithm for estimating the noise level. We use our proofs of the upper bounds to suggest implementation technique. We find Monte-Carlo results close to the asymptotic theoretical rates.

1 Estimation methods

1.1 Noise level evaluation algorithm

The estimator we propose is defined implicitly via the following criterion. Let us first estimate the characteristic function of the observed variables $\Phi^Y(u) = \mathbb{E}\left[\exp(iuY)\right]$ by using the given sample:

$$\widehat{\Phi}_n^Y(u) = \frac{1}{n}\sum_{k=1}^{n} e^{iuY_k}.$$

Remark that in the following $\mathbb{P}$, $\mathbb{E}$ and Var denote, respectively, the probability $\mathbb{P}_{\sigma,f}$ the expectation $\mathbb{E}_{\sigma,f}$ and the variance $\mathrm{Var}_{\sigma,f}$ with respect to the probability when the underlying unknown parameters are $\sigma > 0$ and some signal f in the class.

Let us consider next the function

$$\widehat{F}_n(\tau,u) = \widehat{\Phi}_n^Y(u)\, e^{(\tau u)^s},$$

for $\tau, u > 0$ and fixed known $s > 0$. Our estimator $\widehat{\sigma}_n$ of σ is defined by

$$\widehat{\sigma}_n = \widehat{\sigma}_n(Y_1,\ldots,Y_n) = \inf_{\tau>0}\left\{|\widehat{F}_n(\tau,u_n)| \geq 1\right\}, \tag{2}$$

for some positive sequence $u_n \to \infty$ well-chosen, as described later.

This construction is based on the observation that $|\widehat{F}_n(\tau,u)|$ is an unbiased estimator of

$$|F(\tau,u)| = |\Phi^Y(u)| e^{(\tau u)^s} = O(1)|\Phi(u)| e^{(\tau^s-\sigma^s)u^s},$$

the last equality being valid for large enough $|u|$. This quantity converges, when $u \to \infty$, either to 0 when $\tau \leq \sigma$ or to ∞ when $\tau > \sigma$.

Note that, for the convergence of this estimation method it is sufficient to assume (1), but we prove the convergence of the plug-in estimator for f (Section 1.3) and of the lower bounds (Section 2) only when the noise has a stable distribution.

1.2 Consistency and rates

Theorem 1 *Fix $\sigma_0 > 0$, $\delta > 0$ arbitrarily small and a vicinity $\mathcal{V}(\sigma_0) = \mathcal{V}_\delta(\sigma_0) = (\sigma_0 - \delta; \sigma_0 + \delta)$. Under Assumption (A), consider the sequence of parameters $u_n = (\sigma_0 + 2\delta)^{-1}(\log n/2)^{1/s}$ and the rate*

$$\varphi_{n,\delta} = \frac{\alpha}{s}\frac{(\sigma_0+2\delta)^{1-r+s}}{(\sigma_0-\delta)^s}\left(\frac{\log n}{2}\right)^{r/s-1}.$$

Then, for all $\sigma \in \mathcal{V}(\sigma_0)$, for all $f \in \mathcal{A}(\alpha,r)$, and large enough n, we have

$$\mathbb{P}\left(|\widehat{\sigma}_n - \sigma| \geq \varphi_{n,\delta}\right) \leq O(1)\exp\left[\frac{2\alpha}{(\sigma_0+2\delta)^r}\left(\frac{\log n}{2}\right)^{r/s}\right]\left(\frac{1}{n}\right)^{1-\sigma^s/(\sigma_0+2\delta)^s}.$$

Moreover, consider the rate

$$\varphi_n = \frac{\alpha}{s\sigma_0^{r-1}}\left(\frac{\log n}{2}\right)^{r/s-1},$$

then

$$\limsup_{n\to\infty}\sup_{\sigma\in\mathcal{V}(\sigma_0)}\sup_{f\in\mathcal{A}(\alpha,r)}\varphi_n^{-2}\mathbb{E}(|\widehat{\sigma}_n-\sigma|^2) \leq 1+\delta.$$

Theorem 2 *Fix $\sigma_0 > 0$, $\delta > 0$ arbitrarily small and a vicinity $\mathcal{V}(\sigma_0) = \mathcal{V}_\delta(\sigma_0) = (\sigma_0 - \delta; \sigma_0 + \delta)$. Under Assumption (B), consider the sequence of parameters $u_n = (\sigma_0 + 2\delta)^{-1} (\log n/2)^{1/s}$, and the rate*

$$\psi_{n,\delta} = \frac{2\beta}{s^2} \frac{(\sigma_0 + 2\delta)^{s+1}}{(\sigma_0 - \delta)^s} \frac{\log \log n}{\log n}.$$

Then, for all $\sigma \in \mathcal{V}(\sigma_0)$, for all $f \in \mathcal{B}(\beta)$ and large enough n, we have

$$\mathbb{P}\left(|\widehat{\sigma}_n - \sigma| \geq \psi_{n,\delta} \right) \leq O(1) (\log n)^{2\beta/s} \left(\frac{1}{n} \right)^{1 - \sigma^s/(\sigma_0 + 2\delta)^s}.$$

Moreover, consider the rate

$$\psi_n = \frac{2\beta\sigma_0}{s^2} \frac{\log \log n}{\log n},$$

then

$$\limsup_{n\to\infty} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{f \in \mathcal{B}(\beta)} \psi_n^{-2} \mathbb{E}(|\widehat{\sigma}_n - \sigma|^2) \leq 1 + \delta.$$

The proofs can be found in Section 3.

1.3 Plug-in deconvolution density estimator

Our estimator $\hat{\sigma}_n$ of the noise level, defined by (2), leads to a natural estimator of the signal, using a kernel estimator combined with a plug-in method. In this part, we establish the rate of convergence of this estimator, which we prove to be optimal in Section 2. In the following, $C > 0$ denotes a large enough constant.

We consider noise ε having stable distribution denoted $S(1, s, \nu, \mu)$, with self-similarity index $s \in (0, 2]$, symmetry parameter $\nu \in [-1, 1]$ and location $\mu \in \mathbb{R}$. In our model the noise is multiplied by an unknown scale parameter $\sigma > 0$. By Zolotarev (1986), $\sigma\varepsilon$ has also a stable law whose explicit Fourier transform is given by

$$\Phi^\varepsilon(\sigma u) = \begin{cases} \exp\{-\sigma^s |u|^s (1 - i\nu \operatorname{sgn}(u) \tan(\pi s/2)) + iu\sigma\mu\} & , s \neq 1 \\ \exp\{-\sigma |u| (1 + i\nu \operatorname{sgn}(u) 2/\pi \log |u|) + iu\sigma(\mu - \nu 2/\pi \log \sigma)\} & , s = 1. \end{cases} \tag{3}$$

Note that $|\Phi^\varepsilon(\sigma u)| = e^{-\sigma^s |u|^s}$. Moreover, a sum of independent copies of a stable law with the same self-similarity index s is distributed as a stable law with the same parameter s. Indeed, for $\sigma_1, \sigma_2 > 0$,

$$\Phi^\varepsilon(\sigma_1 u) \Phi^\varepsilon(\sigma_2 u) = \Phi^\varepsilon(\sigma u) e^{iua},$$

for any values of the parameters s and ν, where $\sigma_1^s + \sigma_2^s = \sigma^s$ and

$$a = \begin{cases} \mu(\sigma_1 + \sigma_2 - \sigma) & , s \neq 1 \\ 2/\pi\nu(\sigma_1 \log(\sigma_1/\sigma) + \sigma_2 \log(\sigma_2/\sigma)) & , s = 1. \end{cases}$$

Define moreover the parameter

$$\tilde{s} = \begin{cases} s \vee 1 & \text{if } \mu \neq 0, \\ s & \text{if } \mu = 0. \end{cases}$$

This parameter will be useful as it is related to the behaviour in a vicinity of zero of the stable law $\Phi^\varepsilon(\sigma\cdot)$. Its role will be clearer in the proofs of the following theorems. Note that when the location parameter μ differs from 0, we can write the model as $Y = X + \sigma(\varepsilon_0 + \mu)$, with noise ε_0 having stable law located at 0, which means centered if it has finite expectation ($s \geq 1$). This expression shows that the role of the known location parameter μ can not be neglected, as the model does not simply write $Y = X + \sigma\varepsilon_0 + \mu$.

We will now describe the estimation procedure. Consider the kernel k_n defined by its Fourier transform

$$\Phi^{k_n}(u) = \Phi^\varepsilon \left(h_n^{-1}u\right)^{-1} 1_{|u|\leq 1}, \tag{4}$$

where h_n is some positive sequence of numbers decreasing to zero. The kernel estimator of the signal f is given by

$$\hat{f}_{n,\hat{\sigma}_n}(x) = \frac{1}{n\hat{\sigma}_n h_n} \sum_{i=1}^n k_n\left(\frac{Y_i - x}{\hat{\sigma}_n h_n}\right). \tag{5}$$

In order to get an upper-bound for the pointwise risk of our estimator, we need to restrict ourselves to signals belonging to bounded function spaces: analytic classes or Sobolev balls. Let us denote, respectively

$$\begin{aligned} \mathcal{S}(\alpha', R, L) &= \left\{f; f \text{ is a density and } \int |\Phi(u)|^2 e^{2\alpha'|u|^R} du \leq L^2\right\}, \\ \text{and} \quad W(\beta', L) &= \left\{f; f \text{ is a density and } \int |\Phi(u)|^2 (1 + |u|^{2\beta'}) du \leq L^2\right\}, \end{aligned}$$

where $\alpha', R, L > 0$ and $\beta' > 1/2$.

Three cases occur, distinguishing whereas the signal f belongs to $\mathcal{A}(\alpha, r) \cap \mathcal{S}(\alpha', R, L)$, which is nonempty for $R < r$ or $\{R = r \text{ and } \alpha' < \alpha\}$; or the signal f belongs to $\mathcal{A}(\alpha, r) \cap W(\beta', L)$; or it belongs to $\mathcal{B}(\beta) \cap W(\beta', L)$, which is nonempty when $\beta > \beta' + 1/2$. Note that in the third case, we automatically get that $\beta > 1$. Note also that the intersection $\mathcal{B}(\beta) \cap \mathcal{S}(\alpha', R, L)$ is always empty.

Theorem 3 *Under the assumptions and notations of Theorem 1, consider the kernel estimator $\hat{f}_{n,\hat{\sigma}_n}(x)$ defined by (4) and (5) with bandwidth*

$$h_n = \left[\frac{(\sigma_0+\delta)^R}{\alpha'}\left(1-\frac{r}{s}\right)\log\log n - \frac{(\sigma_0+\delta)^R}{\alpha'}\frac{1-R}{2R}\log\log\log n\right]^{-1/R}.$$

Then, for all σ in $\mathcal{V}(\sigma_0)$, for all f in $\mathcal{A}(\alpha,r)\cap\mathcal{S}(\alpha',R,L)$ (nonempty if $R<r$ or if $R=r$ and $\alpha'<\alpha$) and any x in $\mathbb{R}$, we have:

$$\limsup_{n\to\infty}\sup_{\sigma\in\mathcal{V}(\sigma_0)}\sup_{f\in\mathcal{A}(\alpha,r)\cap\mathcal{S}(\alpha',R,L)}\varphi_n^{-2}\mathbb{E}\left(|\hat{f}_{n,\hat{\sigma}_n}(x)-f(x)|^2\right)\le C<\infty.$$

Theorem 4 *Under the assumptions and notations of Theorem 1, consider the kernel estimator $\hat{f}_{n,\hat{\sigma}_n}(x)$ defined by (4) and (5) with bandwidth h_n precised later. Then, for all σ in $\mathcal{V}(\sigma_0)$, for all f in $\mathcal{A}(\alpha,r)\cap W(\beta',L)$ (with $\beta'>1/2$) and any x in $\mathbb{R}$, we have for v_n as described later:*

$$\limsup_{n\to\infty}\sup_{\sigma\in\mathcal{V}(\sigma_0)}\sup_{f\in\mathcal{A}(\alpha,r)\cap W(\beta',L)}v_n^{-2}\mathbb{E}\left(|\hat{f}_{n,\hat{\sigma}_n}(x)-f(x)|^2\right)\le C<\infty.$$

1) *If $\beta'>\tilde{s}+1/2$, take $h_n=(\log n)^{2(r/s-1)/(2\beta'-1)}$ and then $v_n=\varphi_n$.*

2) *If $\beta'<\tilde{s}+1/2$, and assuming that if $s=1$ then $\nu=0$ (symmetric noise), take $h_n=(\log n)^{(r/s-1)/\tilde{s}}$ and then $v_n=\varphi_n^{(2\beta'-1)/2\tilde{s}}$.*

Theorem 5 *Under the assumptions and notations of Theorem 2, consider the kernel estimator $\hat{f}_{n,\hat{\sigma}_n}(x)$ defined by (4) and (5) with bandwidth h_n precised later. Then, for all σ in $\mathcal{V}(\sigma_0)$, for all f in $\mathcal{B}(\beta)\cap W(\beta',L)$ (nonempty if $\beta>\beta'+1/2$), with $\beta'>1/2$, and any x in $\mathbb{R}$, we have for v_n as described later:*

$$\limsup_{n\to\infty}\sup_{\sigma\in\mathcal{V}(\sigma_0)}\sup_{f\in\mathcal{B}(\beta)\cap W(\beta',L)}v_n^{-2}\mathbb{E}\left(|\hat{f}_{n,\hat{\sigma}_n}(x)-f(x)|^2\right)\le C<\infty.$$

1) *If $\beta'>\tilde{s}+1/2$, take $h_n=(\log\log n/\log n)^{2/(2\beta'-1)}$ and then $v_n=\psi_n$.*

2) *If $\beta'<\tilde{s}+1/2$, take $h_n=(\log\log n/\log n)^{1/\tilde{s}}$ and then $v_n=\psi_n^{(2\beta'-1)/2\tilde{s}}$.*

The second parts of these two last theorems give a more precise result in the particular case of $\beta'<\tilde{s}+1/2$, and if $s=1$ then $\nu=0$ (symmetric noise).To all previous theorems correspond optimality results below (lower bounds for the maximal risks). Some other specific cases, that are not covered by lower bounds, are studied in Subsection 1.4.

When the noise distribution is $s-$supersmooth and entirely known, minimax estimation of the deconvolution signal belonging to $\mathcal{S}(\alpha', R, L)$ was done in Butucea and Tsybakov (2002). Sharp rates were obtained in particular for the case $R < s$ and they were faster than for the plug-in deconvolution estimator constructed here. Estimation of the Sobolev signal (belonging to $W(\beta, L)$) with $s-$supersmooth noise and entirely known has been done adaptively with faster rates too, by Goldenshluger (1999).

Here, we use the same kernel estimators as in Butucea and Tsybakov (2002) and plug, into the $\sigma-$dependent bandwidth, the preliminary estimator $\widehat{\sigma}_n$. Fortunately, the deconvolution kernel can be made free of σ and we finally obtain a kernel estimator with data-dependent bandwidth. Thus, we prove that the global estimation risk is at most that of the estimation of the noise level (the slowest).

The proofs can be found in Section 3. They are based on the convergence of $\widehat{\sigma}_n$ to σ. We evaluate the uniform risk for some parameter in a neighbourhood of σ using maximal inequalities for empirical processes in order to treat the uniform stochastic term. Next, we prove that the probability that $\widehat{\sigma}_n$ is outside the neighbourhood of σ is small enough to make this part of the risk even smaller. This idea was previously used by Butucea (2001) for a density estimator adaptive to the unknown smoothness of the density.

1.4 Specific cases

Particular cases to Theorems 4 and 5 are treated here in Table 1. Indeed, in some specific cases, we get a loss of order $\log\log n$ at some power in the rates of convergence of our estimator $\hat{f}_{n,\hat{\sigma}_n}$ due to the particular choice of stable distributions. The proofs of these results are immediate consequences of the expressions appearing in the term denoted by T_{11} in the respective proofs of Theorems 4 and 5 and are omitted.

Remember that the parameter $\tilde{s} = s$ if the noise is located at $\mu = 0$ but $\tilde{s} = s \vee 1$ if $\mu \neq 0$.

Note also that in the border case of $\beta' = \tilde{s} + 1/2$, optimal bandwidths coincide in Theorem 4, respectively in Theorem 5. The rates are lowered by $\log(1/h_n) = C\log\log n$ at some power, in this cases, where $C > 0$ is some constant.

2 Minimax optimality

We give in this section the minimax lower bounds for the estimation of the scale parameter σ and of the signal density f.

	$\beta' = \tilde{s} + 1/2$		$\beta' < \tilde{s} + 1/2$	
$(s = 1, \nu \neq 0)$	$\varphi_n^2 \log^3\left(\frac{1}{h_n}\right)$	$\psi_n^2 \log^3\left(\frac{1}{h_n}\right)$	$\varphi_n^2 \log^2\left(\frac{1}{h_n}\right)$	$\psi_n^2 \log^2\left(\frac{1}{h_n}\right)$
$s \neq 1$ or $(s = 1, \nu = 0)$	$\varphi_n^2 \log\left(\frac{1}{h_n}\right)$	$\psi_n^2 \log\left(\frac{1}{h_n}\right)$		

Table 1: *Upper bounds for the quadratic pointwise risk when* $f \in \mathcal{A}(\alpha, r) \cap W(\beta', L) \parallel f \in \mathcal{B}(\beta) \cap W(\beta', L)$*, respectively.*

We first present the results concerning parametric and nonparametric estimation under Assumption (A) (Section 2.1) and then under Assumption (B) (Section 2.2). Under Assumption (B) we assume $\beta > 1$ in order to be able to choose corresponding Fourier transforms in $\mathbb{L}_1(\mathbb{R})$. Note that it was already the case when estimating the signal over the class $\mathcal{B}(\beta) \cap W(\beta', L)$. In the following, $c > 0$ denotes some convenient fixed constant.

2.1 Optimality under Assumption (A)

Theorem 6 *Under Assumption (A) and notations of Theorem* 1*, for all* $\sigma_0 > 0$*, we have, for any neighbourhood* $\mathcal{V}(\sigma_0)$

$$\liminf_{n\to\infty} \inf_{\sigma_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in\mathcal{A}(\alpha,r)} \varphi_n^{-2}\mathbb{E}(|\sigma_n - \sigma|^2) \geq 1,$$

where the infimum is taken over arbitrary estimators σ_n *of* σ*.*

Theorem 7 *Under Assumption (A) and notations of Theorems 1 and 3, for all* $\sigma_0 > 0$ *and for any neighbourhood* $\mathcal{V}(\sigma_0)$ *of* σ_0*, we have*

$$\liminf_{n\to\infty} \inf_{f_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in\mathcal{A}(\alpha,r)\cap\mathcal{S}(\alpha',R,L)} \varphi_n^{-2}\mathbb{E}\left(|f_n(x) - f(x)|^2\right) \geq c > 0,$$

where the infimum is taken over arbitrary estimators f_n *of* f*.*

Theorem 8 *Under Assumption (A) and notations of Theorems 1 and 4, for all* $\sigma_0 > 0$ *and for any neighbourhood* $\mathcal{V}(\sigma_0)$ *of* σ_0*, we have, if* $\beta' > \tilde{s} + 1/2$

$$\liminf_{n\to\infty} \inf_{f_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in\mathcal{A}(\alpha,r)\cap W(\beta',L)} \varphi_n^{-2}\mathbb{E}\left(|f_n(x) - f(x)|^2\right) \geq c > 0$$

and, if $1/2 < \beta' \leq \tilde{s} + 1/2$

$$\liminf_{n\to\infty} \inf_{f_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in\mathcal{A}(\alpha,r)\cap W(\beta',L)} \varphi_n^{-(2\beta'-1)/\tilde{s}}\mathbb{E}\left(|f_n(x) - f(x)|^2\right) \geq c > 0$$

where the infimum is taken over arbitrary estimators f_n *of* f*.*

2.2 Optimality under Assumption (B)

Theorem 9 *Under Assumption (B) and notations of Theorem 2, we have, for any $\beta > 1$ and for for any neighbourhood $\mathcal{V}(\sigma_0)$*

$$\liminf_{n\to\infty} \inf_{\sigma_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in\mathcal{B}(\beta)} \psi_n^{-2}\mathbb{E}(|\sigma_n - \sigma|^2) \geq \left(1 - \frac{|s-1|}{2\beta}\right)^2,$$

where the infimum is taken over arbitrary estimators σ_n of σ.

Remark that in this case, the rate is minimax but the associated constant is slightly smaller than the optimal one (see Theorem 2).

Theorem 10 *Under Assumption (B) and notation of Theorems 2 and 5, for $\beta > \beta'+1/2$, for all $\sigma_0 > 0$ and for any neighbourhood $\mathcal{V}(\sigma_0)$ of σ_0, we have, if $\beta' > \tilde{s}+1/2$*

$$\liminf_{n\to\infty} \inf_{f_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in\mathcal{B}(\beta)\cap W(\beta',L)} \psi_n^{-2}\mathbb{E}\left(|f_n(x) - f(x)|^2\right) \geq c > 0$$

and, if $1/2 < \beta' \leq \tilde{s} + 1/2$

$$\liminf_{n\to\infty} \inf_{f_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in\mathcal{B}(\beta)\cap W(\beta',L)} \psi_n^{-(2\beta'-1)/\tilde{s}}\mathbb{E}\left(|f_n(x) - f(x)|^2\right) \geq c > 0$$

where the infimum is taken over arbitrary estimators f_n of f.

2.3 Tools

The proofs of all these theorems are based on suitable choices of two models with convenient parameters being as far from each other as possible, such that the convolution models are close in χ^2-distance.

The following proposition is the main tool in the proof of our lower bounds and it can be found in Butucea and Tsybakov (2002). The notation $\chi^2(P, Q)$ denotes the χ^2-distance between the probabilities P and Q

$$\chi^2(P,Q) = \begin{cases} \int \left(\frac{dP}{dQ} - 1\right)^2 dQ & \text{if } P \ll Q \\ +\infty & \text{otherwise.} \end{cases}$$

Proposition 1 *Let $\mathcal{P}_\Theta = \{\mathbb{P}_\theta; \theta \in \Theta\}$ be a family of models. Assume that there exists θ_1 and θ_2 in Θ with $|\theta_2 - \theta_1| \geq 2s_n > 0$ such that the probability measures $\mathbb{P}_1 = \mathbb{P}_{\theta_1}$ and $\mathbb{P}_2 = \mathbb{P}_{\theta_2}$ satisfy*

$$\mathbb{P}_1 \ll \mathbb{P}_2 \quad \text{and} \quad \chi^2(\mathbb{P}_1^{\otimes n}, \mathbb{P}_2^{\otimes n}) \leq K^2 < \infty.$$

Then we have

$$\liminf_{n\to\infty} \inf_{\hat{\theta}_n} s_n^{-2} \max\{\mathbb{E}_1(|\hat{\theta}_n - \theta_1|^2), \mathbb{E}_2(|\hat{\theta}_n - \theta_2|^2)\} \geq (1-K)^2(1-\sqrt{K})^2,$$

where the infimum is over any estimator $\hat{\theta}_n$ of the underlying parameter and this bound is actually arbitrary close to 1 *for K small enough.*

Now, the previous lower bounds are established by the construction, in each different case (signal in $\mathcal{A}(\alpha, r)$ or $\mathcal{B}(\beta)$), of two particular models $\mathbb{P}_1 = \mathbb{P}_{\sigma_1, f_1}$ and $\mathbb{P}_2 = \mathbb{P}_{\sigma_2, f_2}$, with χ^2 distance converging to zero. Since we have

$$\begin{aligned}
\sup_\sigma \sup_f s_n^{-2}\mathbb{E}(|\hat{\sigma}_n - \sigma|^2) &\geq s_n^{-2} \max\{\mathbb{E}_1(|\hat{\sigma}_n - \sigma_1|^2), \mathbb{E}_2(|\hat{\sigma}_n - \sigma_2|^2)\} \\
\sup_\sigma \sup_f s_n^{-2}\mathbb{E}(|\hat{f}_n(x) - f(x)|^2) &\geq s_n^{-2} \max\{\mathbb{E}_1(|\hat{f}_n(x) - f_1(x)|^2), \mathbb{E}_2(|\hat{f}_n(x) - f_2(x)|^2)\}
\end{aligned}$$

respectively, for the particular models constructed in the proof, Proposition 1 entails the results. Note that the different rates of convergence s_n correspond to half of the distance between the parameters σ_1 and σ_2, when estimating the scale; and the parameters $f_1(x)$ and $f_2(x)$, when estimating the signal.

3 Proofs-upper bounds

Lemma 1 *For any $\sigma > 0$ and f in $\mathcal{A}(\alpha, r)$ (respectively $\mathcal{B}(\beta)$) in the described model we have for all $\tau, u > 0$*

$$\mathbb{E}\left(\widehat{F}_n(\tau, u)\right) = F(\tau, u) \text{ and } \operatorname{Var}\left(\widehat{F}_n(\tau, u)\right) \leq \frac{e^{2(\tau u)^s}}{n}.$$

Proof of Theorem 1.
Consider the probability of the event $\{|\widehat{\sigma}_n - \sigma| \geq \varphi_{n,\delta}\}$ and split it into two terms:

$$\mathbb{P}(|\widehat{\sigma}_n - \sigma| \geq \varphi_{n,\delta}) = \mathbb{P}(\widehat{\sigma}_n \geq \sigma + \varphi_{n,\delta}) + \mathbb{P}(\widehat{\sigma}_n \leq \sigma - \varphi_{n,\delta}) = T_1 + T_2.$$

By definition of the estimator $\widehat{\sigma}_n$, we bound the first term:

$$T_1 \leq \mathbb{P}(|\widehat{F}_n(\sigma + \varphi_{n,\delta}, u_n)| \leq 1) \leq \mathbb{P}(|F(\sigma + \varphi_{n,\delta}, u_n)| \leq 1 + M) + \Delta_M, \tag{6}$$

for some arbitrary $M > 0$ and Δ_M defined as:

$$\Delta_M = \mathbb{P}(|\widehat{F}_n(\sigma + \varphi_{n,\delta}, u_n) - F(\sigma + \varphi_{n,\delta}, u_n)| \geq M).$$

Note that:

$$\Delta_M \leq \frac{1}{M^2}\mathbb{E}\left(\left|\widehat{F}_n(\sigma+\varphi_{n,\delta},u_n)-F(\sigma+\varphi_{n,\delta},u_n)\right|^2\right) = \frac{1}{M^2}\mathrm{Var}\left(\widehat{F}_n(\sigma+\varphi_{n,\delta},u_n)\right).$$

But Lemma 1 leads to:

$$\Delta_M \leq \frac{e^{2(\sigma+\varphi_{n,\delta})^s u_n^s}}{nM^2}. \tag{7}$$

Note also that:

$$\begin{aligned}
\mathbb{P}\left(|F(\sigma+\varphi_{n,\delta},u_n)| \leq 1+M\right) &= \mathbb{P}\left(\left|\Phi^Y(u_n)\right|\exp\{(\sigma+\varphi_{n,\delta})^s u_n^s\} \leq 1+M\right)\\
&= \mathbb{P}\left(|\Phi(u_n)|\exp\left\{[(\sigma+\varphi_{n,\delta})^s-\sigma^s]u_n^s\right\} \leq 1+M\right)\\
&= \mathbb{P}\left(|\Phi(u_n)|\exp\{s\varphi_{n,\delta}\sigma^{s-1}u_n^s(1+o(1))\} \leq 1+M\right).
\end{aligned}$$

With no loss of generality, we have restricted here oourselves to the case $|\Phi^\varepsilon| = e^{-|u|^s}$, for large enough $|u|$. A slight adaptation in the following choice of the parameter M is needed in a more general context. Since Assumption (A) ensures that for large enough n, $|\Phi(u_n)| \geq c\exp\{-\alpha u_n^r\}$, we get that

$$\mathbb{P}\left(|F(\sigma+\varphi_{n,\delta},u_n)| \leq 1+M\right) \leq \mathbb{P}\left(c\exp\{-\alpha u_n^r + s\varphi_{n,\delta}\sigma^{s-1}u_n^s(1+o(1))\} \leq 1+M\right).$$

With our choice of the parameters u_n and $\varphi_{n,\delta}$, we have:

$$\lim_{n\to\infty}\left(-\alpha u_n^r + s\varphi_{n,\delta}\sigma^{s-1}u_n^s(1+o(1))\right) = +\infty,$$

then we choose

$$M = \frac{1}{2}c\exp\{-\alpha u_n^r + s\varphi_{n,\delta}\sigma^{s-1}u_n^s\},$$

and get that $\mathbb{P}\left(|F(\sigma+\varphi_{n,\delta},u_n)| \leq 1+M\right)$ is null for large enough n. Combining this with (6), (7) and the choice of M, we get that, for large enough n,

$$\begin{aligned}
T_1 &\leq \frac{4}{nc^2}\exp\{2(\sigma+\varphi_{n,\delta})^s u_n^s + 2\alpha u_n^r - 2s\sigma^{s-1}\varphi_{n,\delta}u_n^s\}\\
&\leq \frac{4}{nc^2}\exp\{2\sigma^s u_n^s + 2\alpha u_n^r + o(\varphi_{n,\delta}u_n^s)\},
\end{aligned}$$

which converges to zero with our choice of the parameters u_n and $\varphi_{n,\delta}$.
Consider now the second term:

$$T_2 = \mathbb{P}\left(\widehat{\sigma}_n \leq \sigma-\varphi_{n,\delta}\right) \leq \mathbb{P}\left(\left|\widehat{F}_n(\sigma-\varphi_{n,\delta},u_n)\right| \geq 1\right),$$

by definition of the estimator $\widehat{\sigma}_n$; and note that:

$$T_2 \leq \mathbb{E}\left|\widehat{F}_n(\sigma-\varphi_{n,\delta},u_n)\right|^2 = \mathrm{Var}\left(\widehat{F}_n(\sigma-\varphi_{n,\delta},u_n)\right) + |F(\sigma-\varphi_{n,\delta},u_n)|^2.$$

Since

$$|F(\sigma-\varphi_{n,\delta},u_n)| = |\Phi(u_n)|\exp\{-\sigma^s u_n^s + (\sigma-\varphi_{n,\delta})^s u_n^s\} \leq \exp\{-s\sigma^{s-1}\varphi_{n,\delta}u_n^s(1+o(1))\},$$

and using that by Lemma 1,

$$\operatorname{Var}\left(\widehat{F}_n(\sigma-\varphi_{n,\delta},u_n)\right) \leq \frac{\exp\{2(\sigma-\varphi_{n,\delta})^s u_n^s\}}{n},$$

our choice of the parameters u_n and $\varphi_{n,\delta}$ gives that T_2 converges also to zero as n tends to infinity and even faster than the upper bound of T_1. In conclusion, the quantity

$$\begin{aligned}
\mathbb{P}(|\widehat{\sigma}_n-\sigma| \geq \varphi_{n,\delta}) &\leq \frac{4(1+o(1))}{nc^2}\exp\{2\sigma^s u_n^s + 2\alpha u_n^r + o(\varphi_{n,\delta}u_n^s)\} \\
&\leq O(1)\exp\left[\log n\left(-1+\left(\frac{\sigma}{\sigma_0+2\delta}\right)^s + \frac{\alpha}{(\sigma_0+2\delta)^{r-s}}\left(\frac{\log n}{2}\right)^{r/s-1}\right)\right]
\end{aligned}$$

converges to zero as n tends to infinity. Moreover, note that for all σ in $\mathcal{V}(\sigma_0)$, and large enough n,

$$\begin{aligned}
\mathbb{E}(|\widehat{\sigma}_n-\sigma|^2) &= \int_0^{+\infty}\mathbb{P}(|\widehat{\sigma}_n-\sigma|^2 \geq t)dt \\
&= \int_0^{\varphi_{n,\delta}^2}\mathbb{P}(|\widehat{\sigma}_n-\sigma|^2 \geq t)dt + \int_{\varphi_{n,\delta}^2}^{2(\sigma_0+\delta)^2}\mathbb{P}(|\widehat{\sigma}_n-\sigma|^2 \geq t)dt \\
&\leq \varphi_{n,\delta}^2 + 2(\sigma_0+\delta)^2\mathbb{P}(|\widehat{\sigma}_n-\sigma| \geq \varphi_{n,\delta}).
\end{aligned}$$

By the previous statement, the second term on the right-hand side is negligible in front of $\varphi_{n,\delta}^2$. Finally, $\varphi_{n,\delta}/\varphi_n \to 1$, when $\delta \to 0$ and we get the desired result. ■

Proof of Theorem 2
The beginning of the proof follows the same lines and we establish that

$$\mathbb{P}(|\widehat{\sigma}_n-\sigma| \geq \psi_{n,\delta}) = \mathbb{P}(\widehat{\sigma}_n \geq \sigma+\psi_{n,\delta}) + \mathbb{P}(\widehat{\sigma}_n \leq \sigma-\psi_{n,\delta}) = T_1+T_2,$$

with

$$T_1 \leq \mathbb{P}(|F(\sigma+\psi_{n,\delta},u_n)| \leq 1+M) + \frac{e^{2(\sigma+\psi_{n,\delta})^s u_n^s}}{nM^2}.$$

Assumption (B) ensures that for large enough n,

$$|\Phi(u_n)| \geq c\,|u_n|^{-\beta},$$

so that we get:

$$\mathbb{P}(|F(\sigma+\psi_{n,\delta},u_n)| \leq 1+M) \leq \mathbb{P}\left(c\exp\{-\beta\log u_n + s\psi_{n,\delta}\sigma^{s-1}u_n^s(1+o(1))\} \leq 1+M\right).$$

With our choice of the parameters u_n and $\psi_{n,\delta}$, we have:

$$\lim_{n\to\infty} -\beta\log u_n + s\psi_{n,\delta}\sigma^{s-1}u_n^s(1+o(1)) = +\infty,$$

then we choose

$$M = \frac{c}{2}\exp\{s\psi_{n,\delta}\sigma^{s-1}u_n^s - \beta\log u_n\},$$

and get that $\mathbb{P}(|F(\sigma+\psi_{n,\delta},u_n)| \leq 1+M)$ is null for large enough n. Combining with the bound on T_1 we get that, for large enough n,

$$\begin{aligned} T_1 &\leq \frac{4}{nc^2}\exp\{2(\sigma+\psi_{n,\delta})^s u_n^s - 2s\sigma^{s-1}\psi_{n,\delta}u_n^s + 2\beta\log u_n\} \\ &\leq \frac{4}{c^2}\exp\{-\log n + 2\sigma^s u_n^s + 2\beta\log u_n + o(\psi_{n,\delta}u_n^s)\}, \end{aligned}$$

which converges to zero with our choice of the parameters u_n and $\psi_{n,\delta}$.
The rest of the proof is exactly the same as in the preceding theorem. ∎

Proof of Theorem 3.
Fix σ in $\mathcal{V}(\sigma_0)$ and f in $\mathcal{A}(\alpha,r)\cap\mathcal{S}(\alpha',R,L)$. Denote the following neighbourhood of σ by $\mathcal{U}(\sigma) = (\sigma-\varphi_{n,\delta};\sigma+\varphi_{n,\delta})$. The idea of the proof is that $\widehat{\sigma}_n$ being convergent to σ we study separately the uniform behaviour of the kernel estimator when $\widehat{\sigma}_n$ is in a neighbourhood of the true value or not. For the first part we use the bias-variance decomposition and treat the uniform variance with maximal inequality for empirical processes. Then we prove that the small probability of $\widehat{\sigma}_n$ being outside the neighbourhood makes the global estimation risk even smaller. We split the risk of our estimator into two terms:

$$\begin{aligned} \mathbb{E}\left(|\hat{f}_{n,\hat{\sigma}_n}(x)-f(x)|^2\right) &= \mathbb{E}\left(|\hat{f}_{n,\hat{\sigma}_n}(x)-f(x)|^2 1_{\hat{\sigma}_n\in\mathcal{U}(\sigma)}\right) \\ &\quad + \mathbb{E}\left(|\hat{f}_{n,\hat{\sigma}_n}(x)-f(x)|^2 1_{\hat{\sigma}_n\notin\mathcal{U}(\sigma)}\right) = T_1+T_2. \end{aligned} \tag{8}$$

We consider the first term:

$$\begin{aligned} T_1 &\leq \mathbb{E}\left(\sup_{\tau\in\mathcal{U}(\sigma)}|\hat{f}_{n,\tau}(x)-f(x)|^2\right) \\ &\leq 2\sup_{\tau\in\mathcal{U}(\sigma)}\left|\mathbb{E}\hat{f}_{n,\tau}(x)-f(x)\right|^2 + 2\mathbb{E}\left(\sup_{\tau\in\mathcal{U}(\sigma)}|\hat{f}_{n,\tau}(x)-\mathbb{E}\hat{f}_{n,\tau}(x)|^2\right) \\ &\leq 2T_{11}+2T_{12}. \end{aligned} \tag{9}$$

The term T_{11} is the maximal bias term over $\mathcal{U}(\sigma)$. Note that

$$\mathbb{E}\hat{f}_{n,\tau}(x)=\frac{1}{\tau h_n}\int k_n\left(\frac{u-x}{\tau h_n}\right)f^Y(u)du=\frac{1}{2\pi}\int \Phi^{k_n}(\tau h_n t)e^{-ixt}\Phi(t)\Phi^\varepsilon(\sigma t)dt,$$

(remember that E is a shorted notation for $\mathrm{E}_{\sigma,f}$ the expectation when the unknown parameters are σ and f) so that we get:

$$\begin{aligned}
T_{11} &= \sup_{\tau\in\mathcal{U}(\sigma)}\left|\frac{1}{2\pi}\int e^{-ixt}\Phi(t)\left(\Phi^\varepsilon(\tau t)^{-1}\Phi^\varepsilon(\sigma t)1_{|t|\le 1/(\tau h_n)}-1\right)dt\right|^2\\
&\le \frac{1}{4\pi^2}\left(\int|\Phi(t)|^2e^{2\alpha'|t|^R}dt\right)\sup_{\tau\in\mathcal{U}(\sigma)}\int_{|t|\le 1/(\tau h_n)}e^{-2\alpha'|t|^R}|\Phi^\varepsilon(\tau t)^{-1}\Phi^\varepsilon(\sigma t)-1|^2dt\\
&\quad+\sup_{\tau\in\mathcal{U}(\sigma)}\frac{1}{4\pi^2}\left(\int_{|t|>1/(\tau h_n)}|\Phi(t)|dt\right)^2.
\end{aligned}$$

By assumption f belongs to $\mathcal{S}(\alpha',R,L)$ so that:

$$\left(\int_{|t|>1/(\tau h_n)}|\Phi(t)|dt\right)^2\le L^2\int_{|t|>1/(\tau h_n)}e^{-2\alpha'|t|^R}dt\le\frac{L^2\tau^{1-R}h_n^{1-R}}{\alpha' R}\exp\left(\frac{-2\alpha'}{\tau^R h_n^R}\right)(1+o(1)),$$

so that

$$\begin{aligned}
T_{11} &\le \frac{L^2}{4\pi^2}\sup_{\tau\in\mathcal{U}(\sigma)}\int_{|t|\le 1/(\tau h_n)}e^{-2\alpha'|t|^R}\left|\Phi^\varepsilon(\tau t)^{-1}\Phi^\varepsilon(\sigma t)-1\right|^2dt\\
&\quad+\frac{L^2\sigma^{1-R}}{4\pi^2\alpha' R}h_n^{1-R}\exp\left(\frac{-2\alpha'}{\sigma^R h_n^R}\right)(1+o(1)).
\end{aligned}$$

According to Formula (3), the quantity $\Phi^\varepsilon(\tau t)^{-1}\Phi^\varepsilon(\sigma t)$ equals

$$\begin{cases}\exp\{(\tau^s-\sigma^s)|t|^s(1-i\nu\mathrm{sgn}(t)\tan(\pi s/2))-it\mu(\tau-\sigma)\} & \text{if } s\neq 1\\[2ex] \exp\{(\tau-\sigma)|t|(1+i\nu\mathrm{sgn}(t)\tfrac{2}{\pi}\log|t|)-it\mu(\tau-\sigma)+it\nu\tfrac{2}{\pi}(\tau\log\tau-\sigma\log\sigma)\} & \text{if } s=1\end{cases}$$

Write $\tau=\sigma+a$ with $|a|\le\varphi_{n,\delta}$ and $|t|\le 1/(\tau h_n)$ such that $a|t|^s=o(1)$. We get that

$$\Phi^\varepsilon(\tau t)^{-1}\Phi^\varepsilon(\sigma t)=\begin{cases}\exp\{sa\sigma^{s-1}|t|^s(1+o(1))(1-i\nu\mathrm{sgn}(t)\tan(\pi s/2))-it\mu a\} & \text{if } s\neq 1\\ \exp\{(a|t|(1+i\nu\mathrm{sgn}(t)2/\pi\log|t|)-it\mu a+it\nu\tfrac{2}{\pi}a)(1+o(1))\} & \text{if } s=1,\end{cases}$$

which leads to

$$|\Phi^\varepsilon(\tau t)^{-1}\Phi^\varepsilon(\sigma t)-1|=\begin{cases}O(1)\varphi_{n\delta}|t|^s+O(1)\varphi_{n\delta}\mu|t|=O(1)\varphi_{n\delta}|t|^{\tilde{s}} & \text{if } s\neq 1\\ O(1)\varphi_{n\delta}|t|(1+\nu\log|t|) & \text{if } s=1.\end{cases}\qquad(10)$$

Returning to the upper-bound on T_{11}, we get that

$$T_{11} \ \leq \ O(1)\left(\int |t|^{2\tilde{s}}(1+\nu\log|t|)^2 e^{-2\alpha'|t|^R}dt\right)\varphi_{n,\delta}^2 + O(1)h_n^{1-R}\exp\left(\frac{-2\alpha'}{\sigma^R h_n^R}\right).$$

The bandwidth h_n is the largest possible such that T_{11} is not larger than the inevitable (large enough) loss of $\varphi_{n,\delta}^2$. We see later that all other terms in the decomposition of T_1 and T_2 in (8) are much smaller, because $R \leq r < s$. We get:

$$T_{11} \leq O(\varphi_{n,\delta}^2). \tag{11}$$

Return now to inequality (9) and consider the second term:

$$T_{12} = \mathbb{E}(\sup_{\tau\in\mathcal{U}(\sigma)}|\hat{f}_{n,\tau}(x) - \mathbb{E}\hat{f}_{n,\tau}(x)|^2) = \frac{1}{n}\mathbb{E}(\sup_{\tau\in\mathcal{U}(\sigma)}|\mathbb{G}(k_{n,\tau,x})|^2),$$

where $\mathbb{G}$ is the empirical process associated to the measure $\mathbb{P} = \mathbb{P}_{\sigma,f}$, which means that $\mathbb{G}(g) = n^{-1/2}\sum_{i=1}^n(g(Y_i) - \mathbb{P}g)$ and the function $k_{n,\tau,x}$ equals $(\tau h_n)^{-1}k_n(\frac{\cdot - x}{\tau h_n})$.

Now we use a maximal inequality to control the norm of the empirical process. The following notation can be found in more details in van der Vaart and Wellner (1996). We consider the class of functions $\mathcal{F}_n$ defined by $\{k_{n,\tau,x}; \tau \in \mathcal{U}(\sigma)\}$. Note that this class has an envelope function equal to the constant

$$K = \frac{\|k_n\|_\infty}{\tau h_n} \leq \frac{1}{2\pi\tau h_n}\int_{|u|\leq 1} e^{h_n^{-s}|u|^s}du = \frac{1}{\tau\pi s}h_n^{s-1}e^{1/h_n^s}(1+o(1)).$$

The complexity of this family stands in its entropy defined through the bracketing numbers for this class. Theorem 2.7.11 in van der Vaart and Wellner (1996) applies in our context as there exists some positive constant C_1 (depending only on the fixed point x) such that for all τ_1, τ_2 in $\mathcal{U}(\sigma)$,

$$|k_{n,\tau_1,x}(u) - k_{n,\tau_2,x}(u)| \leq C_1 h_n^{-1}e^{1/h_n^s}|\tau_1 - \tau_2|.$$

This theorem asserts that the bracketing numbers for the class $\mathcal{F}_n$ (that means the minimal number of brackets of size ϵ needed to cover $\mathcal{F}_n$) are controlled by the covering numbers of $\mathcal{U}(\sigma)$ (i.e the minimal number of balls of radius ϵ needed to cover $\mathcal{U}(\sigma)$):

$$N_{[\,]}(2\epsilon; \mathcal{F}_n; L_2(Q)) \leq N\left(\frac{\epsilon}{C_1}h_n e^{-1/h_n^s}; \mathcal{U}(\sigma); |\cdot|\right),$$

where Q is any probability measure. But it is easy to bound the covering numbers for $\mathcal{U}(\sigma)$

$$N\left(\frac{\epsilon}{C_1}h_n e^{-1/h_n^s}; \mathcal{U}(\sigma); |\cdot|\right) \leq \frac{2C_1}{\epsilon}h_n^{-1}e^{1/h_n^s}\varphi_{n,\delta}.$$

So that we obtain the following control on the bracketing numbers for the class $\mathcal{F}_n$

$$N_{[\,]}(2\epsilon;\mathcal{F}_n;L_2(Q)) \le \frac{2C_1}{\epsilon} h_n^{-1} e^{1/h_n^s} \varphi_{n,\delta},$$

and the control on the covering numbers for $\mathcal{F}_n$

$$N(\epsilon;\mathcal{F}_n;L_2(Q)) \le \frac{2C_1}{\epsilon} h_n^{-1} e^{1/h_n^s} \varphi_{n,\delta}. \tag{12}$$

Let us define the entropy of this class by the formula:

$$J(1,\mathcal{F}_n) = \sup_Q \int_0^1 \{1+\log N(\epsilon K,\mathcal{F}_n,L_2(Q))\}^{1/2} d\epsilon, \tag{13}$$

where the supremum is taken over all discrete probability measure Q. Then, Theorem 2.14.1 in van der Vaart and Wellner (1996) applies and gives that

$$\mathbb{E}\left[\left(\sup_{\tau\in\mathcal{U}(\sigma)} |\mathbb{G}(k_{n,\tau,x})|\right)^2\right] \le cK^2 J(1,\mathcal{F}_n)^2,$$

where c is an absolute constant. Combining with the definition of the entropy (13), with inequality (12) and the expression of K, we obtain that there exists some constant κ such that

$$\mathbb{E}\left[\left(\sup_{\tau\in\mathcal{U}(\sigma)} |\mathbb{G}_n(k_{n,\tau,x})|\right)^2\right] \le \kappa h_n^{2(s-1)} e^{2/h_n^s} \left|\log(\varphi_{n.\delta}) - s\log(h_n)\right| (1+o(1)).$$

Returning to the definition of the term T_{12}, we get:

$$T_{12} = \frac{O(1)}{n} h_n^{s-2} e^{2/h_n^s} \log\log n. \tag{14}$$

Combining inequalities (9), (11) and (14), and the definition of the rate $\varphi_{n,\delta}$ we get:

$$T_1 \le O(\varphi_{n,\delta}^2). \tag{15}$$

Return to the expression of the risk (8) and consider the second term:

$$\begin{aligned} T_2 &= \mathbb{E}\left(\left|\hat{f}_{n,\hat{\sigma}_n}(x) - f(x)\right|^2 1_{\hat{\sigma}_n\notin\mathcal{U}(\sigma)}\right) \\ &\le 2\mathbb{E}\left[\left(\|\hat{f}_{n,\hat{\sigma}_n}\|_\infty^2 + \|f\|_\infty^2\right) 1_{\hat{\sigma}_n\notin\mathcal{U}(\sigma)}\right] \end{aligned}$$

But we know that:

$$\|\hat{f}_{n,\hat{\sigma}_n}\|_\infty^2 \le \frac{1}{|\hat{\sigma}_n|^2} \times \frac{h_n^{2(s-1)} e^{2/h_n^s}}{\pi^2 s^2}(1+o(1)),$$

and that

$$\mathbb{E}\left(\frac{1}{|\hat{\sigma}_n|^2}1_{\hat{\sigma}_n\notin\mathcal{U}(\sigma)}\right)=\frac{1}{\sigma^2}\mathbb{P}(\hat{\sigma}_n\notin\mathcal{U}(\sigma))(1+r),$$

where

$$\begin{aligned}|r| &= \sigma^2P(\hat{\sigma}_n\notin\mathcal{U}(\sigma))^{-1}\left|\mathbb{E}\left[\left(\hat{\sigma}_n^{-2}-\sigma^{-2}\right)1_{\hat{\sigma}_n\notin\mathcal{U}(\sigma)}\right]\right| \\ &\le \sigma^2\mathbb{E}|\hat{\sigma}_n^{-2}-\sigma^{-2}|^2 \\ &\le \sigma^{-2}\mathbb{E}\left[\hat{\sigma}_n^{-4}\left|\hat{\sigma}_n^2-\sigma^2\right|^2\right].\end{aligned}$$

But $\hat{\sigma}_n$ converges in probability to $\sigma>0$, so that $\hat{\sigma}_n^{-4}$ is bounded in probability and finally:

$$|r|=O(1)\ \mathbb{E}|\hat{\sigma}_n^2-\sigma^2|^2=o(1).$$

We get:

$$T_2\le 2\left(\frac{h_n^{2(s-1)}e^{2/h_n^s}}{\pi^2s^2\sigma^2}(1+o(1))+\|f\|_\infty^2\right)\mathbb{P}(\hat{\sigma}_n\notin\mathcal{U}(\sigma)).$$

As f belongs to $\mathcal{A}(\alpha,r)$, result of Theorem 1 gives that

$$T_2\le O(1)h_n^{2(s-1)}e^{2/h_n^s}\exp\left(\frac{2\alpha}{(\sigma_0+2\delta)^r}\left(\frac{\log n}{2}\right)^{r/s}\right)\left(\frac{1}{n}\right)^{1-\sigma^s/(\sigma_0+2\delta)^s}. \tag{16}$$

Combining inequalities (8), (15) and (16), we get the desired result. ■

Proof of Theorem 4.

Fix σ in $\mathcal{V}(\sigma_0)$ and f in $\mathcal{A}(\alpha,r)\cap W(\beta',L)$. We only sketch the proof as it follows the same lines and notation than the proof of Theorem 3. Here, the term T_{11} writes

$$\begin{aligned}T_{11} &= \sup_{\tau\in\mathcal{U}(\sigma)}\left|\frac{1}{2\pi}\int e^{-ixt}\Phi(t)\left(\Phi^\varepsilon(\sigma t)\Phi^\varepsilon(\tau t)^{-1}1\!\!1_{|t|\le 1/(\tau h_n)}-1\right)dt\right|^2 \\ &\le \frac{1}{4\pi^2}\left(\int|\Phi(t)|^2(1+|t|^{2\beta'})dt\right)\sup_{\tau\in\mathcal{U}(\sigma)}\int_{|t|\le 1/(\tau h_n)}(1+|t|^{2\beta'})^{-1}|\Phi^\varepsilon(\sigma t)\Phi^\varepsilon(\tau t)^{-1}-1|^2dt \\ &\quad+\sup_{\tau\in\mathcal{U}(\sigma)}\frac{1}{4\pi^2}\left(\int_{|t|>1/(\tau h_n)}|\Phi(t)|dt\right)^2.\end{aligned}$$

But here f belongs to $W(\beta',L)$ so that:

$$\left(\int_{|t|>1/(\tau h_n)}|\Phi(t)|dt\right)^2\le L^2\int_{|t|\ge 1/(\tau h_n)}(1+|t|^{2\beta'})^{-1}dt\le\frac{2}{2\beta'-1}L^2\tau^{2\beta'-1}h_n^{2\beta'-1}(1+o(1)).$$

In the same way as we established the bound (11) using the expressions given in equality (10) we get

$$T_{11} \leq O(1)\varphi_{n,\delta}^2 \int_{|t|\leq 1/(\tau h_n)} \frac{|t|^{2\tilde{s}}(1+\nu \log|t| 1\!\!1_{s=1})^2}{1+|t|^{2\beta'}} dt + O(1) h_n^{2\beta'-1}.$$

In case $\beta' > \tilde{s}+1/2$, we bound $\int_{|t|\leq 1/(\tau h_n)} |t|^{2\tilde{s}}(1+\nu\log|t|)^2/\left(1+|t|^{2\beta'}\right) dt$, by the constant limit. The choice of the bandwidth $h_n = (\log n)^{2(r/s-1)/(2\beta'-1)}$ is the largest such that

$$T_{11} \leq O(\varphi_{n,\delta}^2) = O\left(1\right)\left(\log n\right)^{2(r/s-1)}.$$

In case $\beta' < \tilde{s}+1/2$, and when $s=1$ then $\nu=0$, we evaluate the rate of divergence of the integral in the bound of T_{11} and obtain a global slower rate of convergence. Here,

$$T_{11} \leq O\left(1\right) h_n^{2\beta'-2\tilde{s}-1}\varphi_{n,\delta}^2 + O\left(1\right) h_n^{2\beta'-1}.$$

This bound is optimised for $h_n = (\log n)^{(r/s-1)/\tilde{s}}$ giving the global rate of order

$$T_{11} \leq O\left(1\right)\left(\varphi_{n,\delta}\right)^{(2\beta'-1)/\tilde{s}} = O\left(1\right)\left(\log n\right)^{(r/s-1)(2\beta'-1)/\tilde{s}}.$$

More generally, when $\beta' \leq \tilde{s}+1/2$, we can evaluate the rate of divergence of the remaining integral in every cases, which gives the more precise results presented in Subsection 1.4. The rates obtained differ from the previous one only by powers of $\log\log n$.

The controls of the terms T_{12} and T_2 remain valid and they are much smaller than T_{11} giving the global rate, and we automatically get the different results. ■

Proof of Theorem 5.
Fix σ in $\mathcal{V}(\sigma_0)$ and f in $\mathcal{B}(\beta)\cap W(\beta',L)$. The first part of the proof of Theorem 4 applies and we we just replace $\varphi_{n,\delta}$ by $\psi_{n,\delta}$

$$T_{11} \leq O(1)\psi_{n,\delta}^2 \int_{|t|\leq 1/(\tau h_n)} \frac{|t|^{2\tilde{s}}(1+\nu \log|t| 1\!\!1_{s=1})^2}{1+|t|^{2\beta'}} dt + O(1) h_n^{2\beta'-1}.$$

Following the same discussion, in case $\beta' > \tilde{s}+1/2$, we bound the previous integral by its constant limit and choose $h_n = (\log\log n/\log n)^{2/(2\beta'-1)}$ the largest possible such that

$$T_{11} \leq O\left(\psi_{n,\delta}^2\right) = O\left(1\right)\left(\frac{\log\log n}{\log n}\right)^2.$$

In the other cases, a loss in rate is inevitable. When $\beta' < \tilde{s}+1/2$ and if $s=1$ then consider only $\nu=0$, we get

$$T_{11} \leq O\left(1\right) h_n^{2\beta'-2\tilde{s}-1}\psi_{n,\delta}^2 + O\left(1\right) h_n^{2\beta'-1}.$$

The optimal bandwidth is $h_n = (\log\log n/\log n)^{1/\tilde{s}}$ giving a slower risk of order

$$T_{11} \leq O\left((\psi_{n,\delta})^{(2\beta'-1)/\tilde{s}} \right) = O\left(1\right) \left(\frac{\log\log n}{\log n} \right)^{(2\beta'-1)/\tilde{s}}.$$

More generally, evaluating the rate of divergence of the integral appearing in the bound of T_{11} gives the remaining cases.

The control of the term T_{12} remains valid. The control of the last term, T_2 follows the same lines till we get that

$$T_2 \leq 2\left(\|f\|_\infty^2 + \frac{h_n^{2(s-1)} e^{2/h_n^s}}{\pi^2 s^2 \sigma^2}(1+o(1)) \right) \mathbb{P}(\hat{\sigma}_n \notin \mathcal{U}\left(\sigma\right)).$$

As f belongs to $\mathcal{B}(\beta)$, result of Theorem 2 gives that

$$T_2 \leq O(1) h_n^{2(s-1)} e^{2/h_n^s} (\log n)^{2\beta/s} \left(\frac{1}{n} \right)^{1-\sigma^s/(\sigma_0+2\delta)^s},$$

and we get the desired results. Indeed, under the respective hypothesis with chosen bandwidths, T_{12} and T_2 converge to 0 faster than T_{11}. ■

4 Proofs-lower bounds

We construct two models having parameters far enough from each other, which are nevertheless close enough in χ^2 - distance. They are used all through the proofs of the lower bounds with suitable choices of signals and scale parameters, under Assumptions (A) and (B), respectively.

Let us fix the scale parameter σ_0 and a symmetric density f_1 in the needed class having Fourier transform Φ_1. The first model has signal density f_1 and scale parameter $\sigma_1 = (1+t)^{1/s}\sigma_0$. In this model, observations $Y_1, \dots, Y_n$ have density $f_1^Y(x) = f_1 * [f^\varepsilon(\cdot/\sigma_1)/\sigma_1](x)$ and Fourier transform $\Phi_1^Y(u) = \Phi_1(u)\Phi^\varepsilon(\sigma_1 u)$. Recall that the noise has stable density of parameters $S\left(1, s, \nu, \mu\right)$.

Consider next a perturbation of this model, having scale parameter $\sigma_2 = (1-t)^{1/s}\sigma_0$ and a signal density f_2 defined by its Fourier transform

$$\Phi_2(u) = \Phi_1(u) \left[\Phi^\varepsilon((2t)^{1/s}\sigma_0 u) e^{-iu\delta_t} k^\star\left(\frac{u}{M}\right) + \left(1 - k^\star\left(\frac{u}{M}\right)\right) \right], \qquad (17)$$

for the auxiliary function k having Fourier transform $k^\star$, M being some sequence of positive numbers and the real valued function δ_t is defined by the relations:

$$\delta_t = \begin{cases} \mu\left(\sigma_2 + (2t)^{1/s}\,\sigma_0 - \sigma_1\right), & \text{if } s \neq 1 \\ -\nu\sigma_0\frac{2}{\pi}\left((1-t)\log\left(\frac{1-t}{2}\right) + (1+t)\log\left(\frac{1+t}{2}\right)\right), & \text{if } s = 1. \end{cases} \tag{18}$$

Indeed, by (3) and a simple computation,

$$\Phi^\varepsilon(\sigma_2 u)\Phi^\varepsilon((2t)^{1/s}\sigma_0 u) = \Phi^\varepsilon(\sigma_1 u)e^{iu\delta_t}.$$

We denote in this case $f_2^Y(x) = f_2 * [f^\varepsilon(\cdot/\sigma_2)/\sigma_2](x)$ and $\Phi_2^Y(u) = \Phi_2(u)\Phi^\varepsilon(\sigma_2 u)$.

This construction is actually based on Fourier transforms $\Phi_{1,2}$ which have the same behaviour for large values of u, so that they belong to the same class of signals. Moreover, the resulting models $\Phi_{1,2}^Y$ coincide (in absolute value) on a large interval around 0 in order to get close models in χ^2-distance. By Proposition 1, the rate of convergence is given, for small t, by

$$|\sigma_1 - \sigma_2| = \frac{2\sigma_0}{s}\, t\,(1 + o(1)),$$

when we estimate the scale parameter σ, and by the difference

$$|f_1(x) - f_2(x)| = \frac{1}{2\pi}\left|\int e^{-ixu}\left(\Phi_1 - \Phi_2\right)(u)\,du\right|$$

when we are interested in pointwise estimation of $f(x)$.

Let us proceed to the proof of the lower bounds via some auxiliary result.

Lemma 2 *Let g and h be two nonnegative functions such that g has a unique mode and h is such that $\int h(x)dx = c > 0$. Then the convolution product $g * h$ satisfies:*

$$g * h(x) \geq \frac{c}{2}\min\left\{g(x+A), g(x-A)\right\},$$

for some large enough $A > 0$.

Remark that if g is symmetric the lower bound becomes $g\left(|x| + A\right)c/2$.

Proof of Lemma 2.
It is immediate to note that for some $A > 0$ large enough $\int_{-A}^{A} h(u)du \geq c/2$ and

$$g * h(x) \geq \int_{-A}^{A} g(x-u)h(u)du \geq \min\left\{g(x+A), g(x-A)\right\}\int_{-A}^{A} h(u)du,$$

which concludes the proof. ∎

Throughout this section C denotes a positive constant which values may change along the lines.

4.1 Signals in the class $\mathcal{A}(\alpha, r)$

We now particularize the choice of the function f_1 to deal with the case of signals belonging to $\mathcal{A}(\alpha, r)$.

Lemma 3 *Consider the function $\Phi_1(u) = e^{-\alpha|u|^r}$ which is the Fourier transform of a symmetric stable density f_1 in the class $\mathcal{A}(\alpha, r)$. There exists a kernel k such that*

a) k is an even function,

b) the Fourier transform $k^\star$ has a support included in $[-2; 2]$,

c) for all u in $[-1; 1]$, $k^\star(u) = 1$.

d) $k^\star$ is four times continuously differentiable on $\mathbb{R}$ (i.e. $\mathcal{C}^4$).

Consider the function Φ_2 defined by (17). Then Φ_2 is the Fourier transform of a density f_2 included in $\mathcal{A}(\alpha, r)$ for all large enough M and small enough $t > 0$.

Proof of Lemma 3.
Without loss of generality, we assume that $\sigma_0 = 1$.

Let us construct a function k with desired properties. Consider the function $g(x) = \sin x/(\pi x)$, with $g^\star(u) = 1_{|u|\leq 1}$. Next we consider successive convolutions of $g^\star$ with itself, say $g^{\star 32}$ having support on $[-32, 32]$ and being 4-times continuously differentiable, corresponding to a positive density function $g^{32}(x)$. Let us rescale this function $G^\star(u) = g^{\star 32}(u/32)/32$ and finally split $G^\star$ to get what we need, as follows

$$k^\star(u) = \begin{cases} 1, & |u| \leq 1; \\ G^\star(u-1), & u \in [1, 2]; \\ G^\star(u+1), & u \in [-2, -1]. \end{cases}$$

Remember that f_2 denotes the function

$$f_2(x) = \frac{1}{2\pi} \int e^{-iux} \Phi_2(u)\, du.$$

Since $\Phi_2(0) = 1$, we know that $\int f_2(x)dx = 1$. Our purpose is to establish that f_2 is a positive function (and then a density function). The fact that f_2 belongs to the class $\mathcal{A}(\alpha, r)$ is a direct consequence of the construction of Φ_2 (see equation (17)), since the kernel k has a Fourier transform boundedly supported.

The argument for the positivity of f_2 consists in two steps. First, we prove that the uniform distance $\|f_2 - f_1\|_\infty$ converges to zero as t tends to zero and M tends to infinity, and then (f_1 being strictly positive, see Zolotarev (1986) Remark 4 after Theorem 2.2.3), for all fixed compact K in $\mathbb{R}$, small enough t and large enough M, we get $f_2(x) > 0$ for all x in K. The second step is to establish that for large enough $|x|$, we have

$$f_2(x) \geq \frac{C}{|x|^{r+1}} + \frac{O(1)}{|x|^3}, \tag{19}$$

for some constant $C > 0$, and since $r < s \leq 2$, we conclude that, for large enough $|x|$, the function f_2 is positive.

Let us establish the first step. Note that

$$\Phi_2(u) = \Phi_1(u)\Phi^\varepsilon((2t)^{1/s}u)e^{-iu\delta_t} + \Phi_1(u)\left(1 - k^\star\left(\frac{u}{M}\right)\right)\left(1 - \Phi^\varepsilon((2t)^{1/s}u)e^{-iu\delta_t}\right),$$

and consequently

$$\begin{aligned} f_2(x) = & \left[f_1 * \left(\frac{1}{(2t)^{1/s}} f^\varepsilon\left(\frac{\cdot}{(2t)^{1/s}}\right)\right)\right](x + \delta_t) \\ & + \frac{1}{2\pi}\int e^{-iux}\Phi_1(u)\left(1 - k^\star\left(\frac{u}{M}\right)\right)\left(1 - \Phi^\varepsilon((2t)^{1/s}u)e^{-iu\delta_t}\right) du. \end{aligned} \tag{20}$$

Using that the kernel k satisfies:

$$\left|1 - k^\star\left(\frac{u}{M}\right)\right| \leq 1_{|u| \geq M},$$

the second term in the right hand side of the previous equality is bounded by

$$\left|\frac{1}{2\pi}\int e^{-iux}\Phi_1(u)\left(1 - k^\star\left(\frac{u}{M}\right)\right)\left(1 - \Phi^\varepsilon((2t)^{1/s}u)e^{-iu\delta_t}\right) du\right| \leq \frac{1}{\pi}\int_{|u| \geq M} e^{-\alpha|u|^r} du$$

which is $O\left(M^{1-r}e^{-\alpha M^r}\right)$, uniformly in x, and then converges to zero as M tends to infinity. Now, let us denote by f_t^ε the scale transformed function $(2t)^{-1/s} f^\varepsilon((2t)^{-1/s}\cdot)$ so that the first term in (20) is the convolution between the continuous and bounded function f_1 with f_t^ε, combined with a translation by δ_t. We get that

$$\begin{aligned} \|f_1 * f_t^\varepsilon(\cdot + \delta_t) - f_1\|_\infty & \leq \|f_1 * f_t^\varepsilon - f_1\|_\infty + \|f_1 * f_t^\varepsilon - f_1 * f_t^\varepsilon(\cdot + \delta_t)\|_\infty \\ & = o(1) + \|f_1 * f_t^\varepsilon - f_1 * f_t^\varepsilon(\cdot + \delta_t)\|_\infty, \end{aligned}$$

as t tends to zero, by properties of approximate convolution identities. Now using that $\delta_t \to 0$ (see definition on Equations (18)) and that f_t^ε (and then the convolution product too) is continuously differentiable, with uniformly bounded derivative, we have

$$\|f_1 * f_t^\varepsilon - f_1 * f_t^\varepsilon(\cdot + \delta_t)\|_\infty = O(\delta_t) = o(1),$$

as t tends to zero. Returning to (20), we get that:

$$\|f_2 - f_1\|_\infty = o(1), \quad \text{as } M \to \infty \text{ and } t \to 0.$$

Denote by Ψ the function

$$\Psi(u) = \Phi_1(u) \left(1 - k^\star\left(\frac{u}{M}\right)\right)(1 - \Phi^\varepsilon((2t)^{1/s}u)e^{-iu\delta_t}),$$

in such a way that according to (20)

$$f_2(x) = (f_1 * f_t^\varepsilon)(x + \delta_t) + \frac{1}{2\pi}\int e^{-iux}\Psi(u)du.$$

Using that Ψ is three times continuously differentiable, identically equal to zero on $[-2M; 2M]$ and vanishes at infinity such as its derivatives, an integration by parts establishes that:

$$\left|\int e^{-iux}\Psi(u)du\right| = \left|\int e^{-iux}\frac{\Psi^{(3)}(u)}{(-ix)^3}du\right| = \frac{O(1)}{|x|^3},$$

since we have $\|\Psi^{(3)}\|_1 < \infty$. It means that

$$f_2(x) = (f_1 * f_t^\varepsilon)(x + \delta_t) + \frac{O(1)}{|x|^3}.$$

Now we apply Lemma 2 with the densities f_1 and f_t^ε, the first one being a symmetric function, with unique mode in zero, which gives

$$f_2(x) \geq \frac{1}{2}f_1(|x| + A) + \frac{O(1)}{|x|^3},$$

for some large enough $A > 0$. Since $\Phi_1(u) = e^{-\alpha|u|^r}$ with $r < 2$, we know that the asymptotic behaviour of $f_1(x)$ is $C/|x|^{r+1}$ for some positive constant C (if $r = 1$, this is the Cauchy distribution, if $r \neq 1$, see Zolotarev (1986) Equations (2.4.8) and (2.5.4)). Finally, we get equality (19) and conclude the proof. ■

Proof of Theorem 6.
Consider, for arbitrary small $\epsilon > 0$, the sequences of positive numbers

$$t = t_n = \sqrt{1-\epsilon}\left(\frac{\alpha}{\sigma_0^r}\left(\frac{\log n}{2}\right)^{r/s-1} - \frac{7-s}{s}\frac{\log\log n}{\log n}\right) \text{ and } M = M_n = \left(\frac{\log n}{2\sigma_0^s}\right)^{1/s}. \tag{21}$$

According to Lemma 3, the densities f_1 and f_2 constructed in this lemma with the preceding choice of parameters belong to the class $\mathcal{A}(\alpha, r)$ (for large enough n), and then applying Proposition 1, we need to control the distance $\chi^2(\mathbb{P}_1^{\otimes n}, \mathbb{P}_2^{\otimes n}) = n\chi^2(f_1^Y, f_2^Y)$. Write:

$$n\chi^2(f_1^Y, f_2^Y) = n \int \frac{|f_1^Y(y) - f_2^Y(y)|^2}{f_1^Y(y)} dy,$$

and use Lemma 2 and the relation $f_1^Y(x) = f_1 * [f^\varepsilon(\cdot/\sigma_1)/\sigma_1](x)$ to bound this expression:

$$n\chi^2(f_1^Y, f_2^Y) \leq n \int \frac{|f_1^Y(y) - f_2^Y(y)|^2}{f_1(|y| + A)} dy,$$

for some large enough A. Now, split this integral into two terms and use that f_1 is a strictly positive function, with behaviour $O(1/|x|^{r+1})$ at infinity, to get that:

$$n\chi^2(f_1^Y, f_2^Y) \leq nC_1 \int_{|y| \leq A} |f_1^Y(y) - f_2^Y(y)|^2 dy + nC_2 \int_{|y|>A} |y|^{r+1} |f_1^Y(y) - f_2^Y(y)|^2 dy \tag{22}$$

for some positive constants C_1 and C_2. Consider the first term on the right-hand side:

$$T_1 = nC_1 \int_{|y| \leq A} |f_1^Y(y) - f_2^Y(y)|^2 dy \leq nC_1 \int |f_1^Y(y) - f_2^Y(y)|^2 dy = \frac{nC_1}{2\pi} \|\Phi_1^Y - \Phi_2^Y\|_2^2.$$

By definition,

$$\Phi_2^Y(u) = \Phi_2(u)\Phi^\varepsilon(\sigma_2 u) = \Phi_1(u) \left[\Phi^\varepsilon((2t)^{1/s}\sigma_0 u) e^{-iu\delta_t} k^\star\left(\frac{u}{M}\right) + 1 - k^\star\left(\frac{u}{M}\right)\right] \Phi^\varepsilon(\sigma_2 u),$$

and

$$\Phi_1^Y(u) = \Phi_1(u)\Phi^\varepsilon(\sigma_1 u) = \Phi_1(u)\Phi^\varepsilon((2t)^{1/s}\sigma_0 u)\Phi^\varepsilon(\sigma_2 u) e^{-iu\delta_t},$$

so that we get:

$$\begin{aligned} |\Phi_1^Y(u) - \Phi_2^Y(u)| &= \left|\Phi_1(u)\Phi^\varepsilon(\sigma_2 u)\left(1 - k^\star\left(\frac{u}{M}\right)\right)\left(1 - \Phi^\varepsilon((2t)^{1/s}\sigma_0 u)e^{-iu\delta_t}\right)\right| \\ &\leq 2e^{-\alpha|u|^r - (1-t)\sigma_0^s|u|^s} 1\!\!1_{|u|>M}. \end{aligned} \tag{23}$$

Returning to the first term T_1:

$$T_1 \leq \frac{2nC_1}{\pi} \int_{|u|>M} e^{-2\alpha|u|^r - 2(1-t)\sigma_0^s|u|^s} du = O\left(nM^{1-s} e^{-2\alpha M^r - 2(1-t)\sigma_0^s M^s}\right).$$

But $M = (\log n/2\sigma_0^s)^{1/s}$ and by our choice of t given in (21) we have $T_1 = o(1)$.

Let us deal with the second term appearing on the right-hand side in (22):

$$T_2 = nC_2 \int_{|y|>A} |y|^{r+1} |f_1^Y(y) - f_2^Y(y)|^2 dy \leq nC_2 \int |y|^4 |f_1^Y(y) - f_2^Y(y)|^2 dy.$$

By Parseval's equality and since $\left(\Phi_1^Y - \Phi_2^Y\right)(u)$ is $\mathcal{C}^4$ on its support: $\{|u| \geq M\}$
$$T_2 \leq \frac{nC_2}{2\pi} \int \left|(\Phi_1^Y - \Phi_2^Y)''(u)\right|^2 du,$$
and according to the expression (23) of the difference $\Phi_1^Y - \Phi_2^Y$, we bound this term by:
$$T_2 \leq nC_2' \int_{|u|\geq M} |u|^6 e^{-2\alpha|u|^r - 2(1-t)\sigma_0^s |u|^s} du = O\left(nM^{7-s} e^{-2\alpha M^r - 2(1-t)\sigma_0^s M^s}\right),$$
and we conclude exactly in the same way that $T_2 = o(1)$.

Then, using Proposition 1, we get
$$\inf_{\hat\sigma_n} \sup_{f,\sigma} \varphi_n^{-2} E\left[|\hat\sigma_n - \sigma|^2\right] \geq (1-\epsilon) \frac{\inf_{\hat\sigma_n} \max_{i=1,2} E\left[|\hat\sigma_n - \sigma_i|^2\right]}{(\sigma_0 t/s)^2} \geq 1 - \epsilon,$$
for arbitrary small $\epsilon > 0$, hence the theorem. ■

Proof of Theorem 7.
The proof uses the same construction as for the Theorem 6. So, we use the same notations and reasoning starting with Lemma 3. We apply again Proposition 1 for functions f_1 and f_2. Indeed, $\Phi_1(u) = \exp(-\alpha|u|^r)$ and thus f_1 belongs to $\mathcal{A}(\alpha, r) \cap \mathcal{S}(\alpha', R, L)$ if $R < r$ or if $R = r$ and $\alpha' < \alpha$. We already saw that f_2 is in $\mathcal{A}(\alpha, r)$. Let us remark that $|\Phi_2(u)| \leq |\Phi_1(u)|$ and then f_2 belongs to $\mathcal{S}(\alpha', R, L)$, too.

As we already checked that $n\chi^2(f_1^Y, f_2^Y) = o(1)$, when $n \to \infty$, for t and M given by (21) it is enough (by Proposition 1) to get a lower bound of $|f_1(x) - f_2(x)|$. Without loss of generality we can evaluate:
$$\begin{aligned} |f_1(0) - f_2(0)| &= \frac{1}{2\pi} \left| \int (\Phi_1(u) - \Phi_2(u))\, du \right| \\ &= \frac{1}{2\pi} \left| \int \Phi_1(u) k^\star\left(\frac{u}{M}\right) \left(1 - \Phi^\varepsilon((2t)^{1/s}\sigma_0 u) e^{-iu\delta_t}\right) du \right|. \end{aligned}$$
Using the definition of the characteristic functions of stable laws, we get that the real part
$$\mathcal{R}e(\Phi^\varepsilon((2t)^{1/s}\sigma_0 u) e^{-iu\delta_t}) = e^{-2t\sigma_0^s|u|^s} \cos(R(t,u)),$$
for some function $R(t,u)$. This leads to the lower bound
$$\begin{aligned} |f_1(0) - f_2(0)| &\geq \frac{1}{2\pi} \left| \int e^{-\alpha|u|^r} k^\star\left(\frac{u}{M}\right) [1 - e^{-2t\sigma_0^s|u|^s} \cos(R(t,u))] du \right| \\ &\geq \frac{1}{2\pi} \int e^{-\alpha|u|^r} k^\star\left(\frac{u}{M}\right) (1 - e^{-2t\sigma_0^s|u|^s}) du \\ &\geq \frac{1}{2\pi} \left(\int_{|u|\leq 1} e^{-\alpha|u|^r} (1 - e^{-2t\sigma_0^s|u|^s}) du + \int_{1<|u|\leq M} e^{-\alpha|u|^r} (1 - e^{-2t\sigma_0^s}) du \right), \end{aligned}$$

as $k^\star$ is a positive function and for large enough M. Finally, write that both terms above are of order $O(t)$:

$$\begin{aligned} |f_1(0) - f_2(0)| &\geq \frac{1}{2\pi} \int_{|u|\leq 1} e^{-\alpha|u|^r} 2t\sigma_0^s |u|^s du (1 + o(1)) + \frac{2t\sigma_0^s}{2\pi} \int_{1<|u|\leq M} e^{-\alpha|u|^r} du \\ &\geq Ct \geq C (\log n)^{r/s-1}, \end{aligned}$$

which achieves the proof. ■

Proof of Theorem 8.
The same construction of functions f_1 and f_2 remains valid for the model. Indeed, the signals are analytic so that they belong to the Sobolev class $W(\beta', L)$ as well. Thus we get the lower rate of convergence φ_n whatever the value of β' is. But this rate is too small (this bound is too low) in the case $\beta' \leq \tilde{s} + 1/2$, where the optimal rate is $\varphi_n^{(2\beta'-1)/2\tilde{s}}$.

In order to solve the case $\beta' \leq \tilde{s} + 1/2$, consider as main signal $f_{1,h}$ in (27) with t given in (21), which fulfils Assumption (A) as well. The same lines as at the end of the proof of Theorem 10 apply in this case to give the result. ■

4.2 Signals in the class $\mathcal{B}(\beta)$

Those proofs will follow the same lines as the ones concerning signals in $\mathcal{A}(\alpha, r)$. We choose a new function f_1 belonging to the class $\mathcal{B}(\beta)$ such that the resulting function f_2 (defined via its Fourier transform Φ_2 and equation (17)) also belongs to the set $\mathcal{B}(\beta)$.

Lemma 4 *Let $\Phi_1(u) = (1/2)(1+u^2)^{-\beta/2} + (1/2)e^{-|u|/2}$, with $\beta > 1$. This function is the Fourier transform of some density f_1 in the class $\mathcal{B}(\beta)$. Use the kernel k constructed in Lemma 3 and define the function Φ_2 by Equation (17). Then Φ_2 is the Fourier transform of a density f_2 included in $\mathcal{B}(\beta)$ for small enough $t > 0$ and large enough M.*

Proof of Lemma 4.
First, let us prove that the function f_1 defined by

$$f_1(x) = \tfrac{1}{2\pi} \int e^{-iux} \Phi_1(u) du = \tfrac{1}{2}(g_1(x) + g_2(x))$$
$$\text{where} \quad g_1(x) = \tfrac{1}{2\pi} \int \tfrac{e^{-iux}}{(1+u^2)^{\beta/2}} du \quad \text{and} \quad g_2(x) = \tfrac{1}{\pi(1+x^2)},$$

is a positive and integrable function, and then is a density, as by Parseval's equality $\int f_1(x)dx = \Phi_1(0) = 1$. Indeed, g_2 is the density of the Cauchy law, and we have

$$g_1(x) = \frac{1}{\pi}\int_0^{+\infty} \frac{\cos(ux)}{(1+u^2)^{\beta/2}}du.$$

Using Formulae 3.771.2, 8.432.3 and 8.334.3 in Gradshteyn and Ryzhik (2000), we get that for any $x \geq 0$, g_1 is given by

$$g_1(x) = (\Gamma(\beta/2))^{-2}\left(\frac{x}{2}\right)^{\beta-1}\int_1^{+\infty} e^{-xt}(t^2-1)^{\beta/2-1}dt,$$

and then is a positive function on $\mathbb{R}^+$, which is also an even function. Moreover, according to Formulae 3.771.2 and 8.451.6 in (Gradshteyn & Ryzhik 2000), we have

$$g_1(x) \underset{+\infty}{\sim} Cx^{\beta/2-1}e^{-x},$$

for some positive constant C, and then is an integrable function. This establishes that f_1 is a density function on $\mathbb{R}$.

The rest of the proof follows the same lines as the one for Lemma 3. Establishing the positivity of f_2, the first step involves a bound on the quantity

$$\left|\frac{1}{2\pi}\int e^{-iux}\Phi_1(u)\left(1-k^\star\left(\frac{u}{M}\right)\right)\left(1-\Phi^\varepsilon((2t)^{1/s}u)e^{-iu\delta_t}\right)du\right|$$
$$\leq \frac{1}{2\pi}\int_{|u|\geq M}(1+u^2)^{-\beta/2}du + \frac{1}{2\pi}\int_{|u|\geq M} e^{-|u|/2}du$$

which is $O(M^{-\beta+1})$ and converges to zero, uniformly in x, as M tends to infinity (and for $\beta > 1$). The second step is proved exactly in the same way, as the asymptotic behaviour of f_1 is given by the Cauchy density g_2 and it is of order $\pi^{-1}x^{-2}$. ■

Proof of Theorem 9.

Here, the notations are the one used in Lemma 4, and the proof follows the same lines as for the proof of Theorem 6. Indeed, for arbitrary small $\epsilon > 0$, consider the parameters

$$t = \sqrt{1-\epsilon}\frac{2\beta - |s-1|}{s}\frac{\log\log n}{\log n} \quad \text{and} \quad M = (\log n/2\sigma_0^s)^{1/s}, \tag{24}$$

and the functions f_1 and f_2 corresponding to this choice, in Lemma 4.

According to this lemma, the functions f_1 and f_2 belong to $\mathcal{B}(\beta)$ (for large enough n). The control of the χ^2 distance between the laws induced by f_1 and f_2 is established exactly in the same way as in Theorem 6, where now the asymptotic behaviour of the function f_1 is $O(x^{-2})$. The first term T_1 is controlled by $O(n)\left\|\Phi_1^Y-\Phi_2^Y\right\|_2^2$ and

$$\left|\left(\Phi_1^Y-\Phi_2^Y\right)(u)\right| \leq O(1)\frac{1_{|u|>M}}{(1+u^2)^{\beta/2}}\exp\left(-(1-t)\,\sigma_0^s\,|u|^s\right).$$

This gives

$$T_1 = O(n)\,M^{-2\beta+1-s}e^{-2\sigma_0^s M^s+2t\sigma_0^s M^s}. \tag{25}$$

On the other hand, $T_2 = O(n)\int_{|y|>A}|y|^2\,|f_1^Y(y)-f_2^Y(y)|^2dy = O(n)\left\|\left(\Phi_1^Y-\Phi_2^Y\right)'\right\|_2^2$. We write first

$$\left(\Phi_1^Y-\Phi_2^Y\right)(u) = \Phi_1(u)\left(\Phi^\varepsilon(\sigma_1 u)-\Phi^\varepsilon(\sigma_2 u)\right)\left(1-k^*\left(\frac{u}{M}\right)\right),$$

to see that the function is continuously differentiable on its support $\{|u|>M\}$. Now,

$$\begin{aligned}\left|\left(\Phi_1^Y-\Phi_2^Y\right)'(u)\right| &\leq O(1)\,|\Phi_1(u)|\left|\left(\Phi^\varepsilon(\sigma_1 u)-\Phi^\varepsilon(\sigma_2 u)\right)'\right|\left(1-k^*\left(\frac{u}{M}\right)\right)\\ &\leq O(1)|u|^{-\beta+(s-1)_+}e^{-(1-t)\sigma_0^s|u|^s}1_{|u|>M},\end{aligned}$$

where a_+ denotes the positive part of a real a. Then

$$T_2 = O(n)\int_{|u|>M}|u|^{-2\beta+2(s-1)_+}e^{-2(1-t)\sigma_0^s|u|^s}du = O(M^{-2\beta+2(s-1)_++1-s}e^{-2(1-t)\sigma_0^s M^s}) \tag{26}$$

From (25) and (26) we deduce that

$$n\chi^2(f_1^Y,f_2^Y) \leq O(n)\,M^{-2\beta+|s-1|}e^{-2(1-t)\sigma_0^s M^s}.$$

Finally, the χ^2-distance goes to 0 when $n\to\infty$, for M and t in (24). Thus

$$\begin{aligned}\inf_{\hat\sigma_n}\sup_{f,\sigma}\psi_n^{-2}E\left[|\widehat{\sigma}_n-\sigma|^2\right] &\geq (1-\epsilon)\left(1-\frac{|s-1|}{2\beta}\right)^2\frac{\inf_{\hat\sigma_n}\max_{i=1,2}E\left[|\widehat{\sigma}_n-\sigma_i|^2\right]}{(\sigma_0 t/s)^2}\\ &\geq (1-\epsilon)\left(1-\frac{|s-1|}{2\beta}\right)^2,\end{aligned}$$

for arbitrary small $\epsilon>0$ and this ends the proof. ■

Proof of Theorem 10.

We use the construction in Theorem 9. As in the previous proof, $n\chi^2(f_1^Y,f_2^Y)$ goes to 0 when $n \to \infty$, for M and t in (24). Then, we need to bound from below, as in (7):

$$\begin{aligned}
|f_1(0) - f_2(0)| &= \frac{1}{2\pi}\left|\int \Phi_1(u)k^\star\left(\frac{u}{M}\right)\left(1 - \Phi^\varepsilon((2t)^{1/s}u)e^{-iu\delta_t}\right)du\right| \\
&\geq \frac{1}{2\pi}\int k^\star\left(\frac{u}{M}\right)\frac{1 - e^{-2t\sigma_0^s|u|^s}}{(1+u^2)^{\beta/2}}du \\
&\geq Ct\int_{|u|\leq 1}\frac{|u|^s}{(1+u^2)^{\beta/2}}du + C\int_{1<|u|\leq M}\frac{tdu}{(1+u^2)^{\beta/2}} \\
&\geq Ct \geq C\log n \log n/\log n
\end{aligned}$$

and the integrals are convergent whenever $\beta > \beta' + 1/2$.

This bound is too low if $\beta' \leq \tilde{s} + 1/2$. For this particular case consider as main signal

$$\begin{aligned}
f_{1,h}(x) &= f_1(x)(1 - h^{\beta'+1/2}) + h^{\beta'-1/2}g_1(x/h), \\
\Phi_{1,h}(u) &= \Phi_1(u)(1 - h^{\beta'+1/2}) + h^{\beta'+1/2}\Phi_{g_1}(hu)
\end{aligned} \tag{27}$$

where $h = t^{1/\tilde{s}} \to 0$ when $n \to \infty$, t is given by (24) and g_1 is defined by its Fourier transform in Lemma 4.

Then $f_{1,h}$ is a density having the needed properties as f_1. Moreover, define $f_{2,h}$ and its Fourier transform $\Phi_{2,h}$ through $\Phi_{1,h}$ and Equation (17). Then, as in the proof of Theorem 7, we get

$$\begin{aligned}
|f_{1,h}(0) - f_{2,h}(0)| &\geq \frac{1}{2\pi}\int \Phi_{1,h}(u)k^\star\left(\frac{u}{M}\right)(1 - e^{-2t\sigma_0^s|u|^s})du \\
&\geq \frac{1}{2\pi}\int_{|u|\leq M}\frac{h^{\beta'+1/2}}{(1+(hu)^2)^{\beta/2}}(1 - e^{-2t\sigma_0^s|u|^s})du \\
&\geq Ct\int_{|u|\leq 1/h}\frac{|u|^s h^{\beta'+1/2}}{(1+(hu)^2)^{\beta/2}}du + C\int_{1/h<|u|\leq M}\frac{h^{\beta'+1/2}}{(1+(hu)^2)^{\beta/2}}du \\
&\geq Ct\int_{|v|\leq 1}\frac{|v|^s h^{\beta'-s-1/2}}{(1+v^2)^{\beta/2}}dv + C\int_{1<|v|\leq hM}\frac{h^{\beta'-1/2}}{(1+v^2)^{\beta/2}}dv,
\end{aligned}$$

where the second integral converges as $\beta > 1$ and $hM \to \infty$, when $n \to \infty$. The first term is of order $h^{\beta'+\tilde{s}-s-1/2}$, which is always smaller than $h^{\beta'-1/2}$. We finally get

$$|f_{1,h}(0) - f_{2,h}(0)| \geq Ch^{\beta'-1/2} = C\psi_n^{(\beta'-1/2)/\tilde{s}}.$$

■

5 Simulation results

For practical implementation we actually use an immediate consequence of Theorem 1 (respectively Theorem 2). This is a global version of the minimax upper bounds, where the unknown parameter is supposed to belong to some interval and the estimation algorithm is based only on a strict upper bound Σ for the true unknown parameter σ.

Corollary 1 *Suppose σ is in some bounded set Θ, $\sigma > 0$ and $\sup\{\sigma, \sigma \in \Theta\} < \Sigma$. Under Assumption (A) consider*

$$u_n = \left(\frac{\log n}{2\Sigma^s}\right)^{r/s} \text{ and } \varphi_n(\Sigma) = \frac{\alpha}{s\sigma^{s-1}}\left(\frac{\log n}{2\Sigma^s}\right)^{r/s-1},$$

then we have for all $\sigma \in \Theta$ and $f \in \mathcal{A}(\alpha, r)$

$$\limsup_{n\to\infty} \sup_{\sigma\in\Theta} \sup_{f\in\mathcal{A}(\alpha,r)} \varphi_n^{-2}(\Sigma)\, \mathbb{E}(|\widehat{\sigma}_n - \sigma|^2) \leq 1.$$

Corollary 2 *Suppose σ is in some bounded set Θ, $\sigma > 0$ and $\sup\{\sigma, \sigma \in \Theta\} < \Sigma$. Under Assumption (B) consider*

$$u_n = \left(\frac{\log n}{2\Sigma^s}\right)^{r/s} \text{ and } \psi_n(\Sigma) = \frac{2\beta\Sigma^s}{s^2\sigma^{s-1}}\frac{\log\log n}{\log n},$$

then we have for all $\sigma \in \Theta$ and $f \in \mathcal{B}(\beta)$

$$\limsup_{n\to\infty} \sup_{\sigma\in\Theta} \sup_{f\in\mathcal{B}(\beta)} \psi_n^{-2}(\Sigma)\, \mathbb{E}(|\widehat{\sigma}_n - \sigma|^2) \leq 1.$$

We simulated a sample of size $n = 1000$ of X having a Cauchy distribution ($\alpha = 1$, $r = 1$) as a signal convoluted with a Gaussian noise ($s = 2$) normalised such that its characteristic function becomes $\Phi^\varepsilon(u) = \exp(-u^2)$. For $\sigma = 1$, we construct the sample $Y = X + \sigma\varepsilon$. We compute the estimator $\widehat{\sigma}_n$ described in Section 1.1 for different values of $\Sigma \in \{1.5, 2, 10, 20, 100\}$. It is an algorithm starting with a very small value $\tau = 0.05$ which tests whether the estimated function $\widehat{F}_n(\tau, u_n)$ is less than 1. If this condition is satisfied then τ is incremented by *step* $= 0.05$ (unless specified), otherwise it stops and provides $\widehat{\sigma}_n^{(1)} = \tau$. We actually remark that the issue of the algorithm strongly depends on the initial value of the upper bound Σ we use at the very beginning. Therefore, we reinitialise the estimation algorithm with $\Sigma = \widehat{\sigma}_n^{(1)}$ and obtain a second estimator $\widehat{\sigma}_n^{(2)}$ and so on. In a very few steps (3 to 7 steps) the estimator doesn't vary anymore, as can be seen in Table 2.

$\hat{\sigma}_n$	$\Sigma = 1.5$	$\Sigma = 2$	$\Sigma = 10$	$\Sigma = 20$	$\Sigma = 100$
1^{st} iteration	1.3	1.45	2.45	3.2	7.4
2^{nd} iteration	1.2	1.3	1.6	1.7	2.25
3^{rd} iteration	1.15	1.2	1.35	1.4	1.55
4^{th} iteration	1.15	1.15	1.25	1.25	1.35
5^{th} iteration	1.15	1.15	1.2	1.2	1.25
6^{th} iteration	1.15	1.15	1.15	1.15	1.2
7^{th} iteration	1.15	1.15	1.15	1.15	1.15

Table 2: *Values of* $\hat{\sigma}_n$*:* f *is Cauchy,* $s = 2$*,* $\sigma = 1$ *and step* $= 0.05$.

Indeed, in the proof of Theorem 1 (respectively Theorem 2) wee see that with much higher probability we overestimate the true σ. Moreover, the estimator is convergent and we expect with high probability to get closer and closer to the true value and attain the local minimax rate of convergence φ_n.

Next, for a Monte Carlo study, we simulated $m = 50$ samples of size $n = 1000$ of X having a Cauchy distribution ($\alpha = 1$, $r = 1$) as a signal convoluted with a Gaussian noise ($s = 2$) normalised such that its characteristic function becomes $\Phi^{\varepsilon}(u) = \exp(-u^2)$. We compute $Y = X + \sigma\varepsilon$, for different values of $\sigma \in \{0.1, 0.5, 1\}$. (Note that for large values of σ estimation is very good, so we don't study them in detail. Moreover, noise level is often expected to be small, in practice.) Then we estimate σ on each sample respectively, reinitialising the procedure each time as it was previously described. We give here the mean square error over $m = 50$ samples:

$$MC = \frac{1}{m} \sum_{k=1}^{m} \left(\widehat{\sigma}_{n,k} - \sigma\right)^2.$$

We compare this value to the minimax (theoretical) rate of convergence φ_n^2 of $E_{\sigma,f}\left[(\widehat{\sigma}_n - \sigma)^2\right]$ in Table 3. Remark that for $r = 1$ the minimax rate of convergence φ_n^2 doesn't depend on the true σ and it's value in this setting is $\varphi_n^2 = 0.0723824$.

Remark that for small values of σ we can refine our results by starting with the closest Σ possible and by decreasing the value of the *step* (see the case $\sigma = 0.1$ in Table 3).

As for the case (B) we considered another $m = 50$ samples of size $n = 1000$ of a Laplace distribution (having Fourier transform: $\Phi(u) = (1 + u^2)^{-1}$, $\beta = 2$) and a Gaussian noise ($s = 2$). For $Y = X + \sigma\varepsilon$ with $\sigma = 1$, we obtain estimators $\widehat{\sigma}_n$ ranging from 0.8 to 1.45 ($\Sigma = 1.5$). The results are presented in Table 4.

	$\Sigma = 1.5$ (step $= 0.05$)	$\Sigma = 1$ (step $= 0.01$)	$\Sigma = 0.5$ (step $= 0.02$)
$\sigma = 0.1$	MC $= 0.2908$ $\varphi_n^2 = 0.07238$	MC $= 0.232528$ $\varphi_n^2 = 0.07238$	MC $= 0.188488$ $\varphi_n^2 = 0.07238$
$\sigma = 0.5$	MC $= 0.12415$ $\varphi_n^2 = 0.07238$	MC $= 0.09611$ $\varphi_n^2 = 0.07238$	
$\sigma = 1$	MC $= 0.0963$ $\varphi_n^2 = 0.07238$		

Table 3: *MC square risk versus* φ_n^2*: f is Cauchy, s=2, $m = 50$ samples of size* $n = 1000$.

	$\Sigma = 1.5$		$\Sigma = 10$		$\Sigma = 15$
$\sigma = 1$	MC$= 0.06295$ $\psi_n^2 = 0.0782763$	$\sigma = 5$	MC$= 0.4296$ $\psi_n^2 = 1.95961$	$\sigma = 10$	MC$= 1.748$ $\psi_n^2 = 7.82763$

Table 4: *MC square risk: f is Laplace, s=2, $m = 50$ samples of size $n = 1000$ and step*$= 0.05$.

A last example is constituted of Laplace distribution ($\beta = 2$) convoluted with a Cauchy law ($s = 1$) in Table 5.

	$\Sigma = 1.5$	$\Sigma = 1$
$\sigma = 0.5$	MC $=0.8947$ $\psi_n^2 = 0.3133105$	MC $= 0.4609$ $\psi_n^2 = 0.3133105$
$\sigma = 1$	MC $= 0.93475$ $\psi_n^2 = 1.25242$	

Table 5: *MC square risk versus* ψ_n^2*: f is Laplace, s=1, $m = 50$ samples of size* $n = 1000$.

Remark again that there is way to improve a lot the global estimation by choosing the closest upper bound possible for the unknown estimated σ.

As we can expect from the minimax rates of convergence, the estimation is better under Assumption (B) than under Assumption (A). The results are good and we indicated practical ways for improving the estimation.

Acknowledgements *The authors are grateful to A. Tsybakov and to F. Ruymgaart for initialising this work.*

References

Butucea, C. (2001). Exact adaptive pointwise estimation on Sobolev classes of densities., *ESAIM, Probab. Stat.* **5**: 1–31. (electronic).

Butucea, C. & Tsybakov, A. B. (2002). Fast asymptotics in density deconvolution. Manuscript.

Carroll, R. J. & Hall, P. (1988). Optimal rates of convergence for deconvolving a density, *J. Amer. Statist. Assoc.* **83**(404): 1184–1186.

Goldenshluger, A. (1999). On pointwise adaptive nonparametric deconvolution, *Bernoulli* **5**(5): 907–925.

Gradshteyn, I. & Ryzhik, I. (2000). *Table of integrals, series, and products.*, San Diego, CA: Academic Press., Sixth ed. Translated from the Russian.

Lindsay, B. G. (1986). Exponential family mixture models (with least-squares estimators), *Ann. Statist.* **14**(1): 124–137.

Matias, C. (2002). Semiparametric deconvolution with unknown noise variance, *ESAIM, Probab. Stat.* **6**: 271–292. (electronic).

Pensky, M. & Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution, *Ann. Statist.* **27**(6): 2033–2053.

van der Vaart, A. W. & Wellner, J. A. (1996). *Weak convergence and empirical processes*, Springer-Verlag, New York. With applications to statistics.

Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions, *Ann. Statist.* **18**(2): 806–831.

Zolotarev, V. (1986). *One-dimensional stable distributions. Transl. from the Russian by H. H. McFaden, ed. by Ben Silver.*, Translations of Mathematical Monographs, Vol. 65. Providence, Rhode Island: American Mathematical Society. IX, 284 p. .