Quadratic functional estimation in view of minimax goodness-of-fit testing from noisy data

Cristina Butucea

[1]Laboratoire de Probabilités et Modèles Aléatoires (UMR CNRS 7599), Université Paris VI, 4, pl.Jussieu, Boîte courrier 188, 75252 Paris, France

[2]Modal'X, Université Paris X, 200, avenue de la République 92001 Nanterre Cedex, France, e-mail: butucea@ccr.jussieu.fr

Abstract

We consider the convolution model $Y_i = X_i + \varepsilon_i$, $i = 1, \ldots, n$ of i.i.d. random variables X_i having common unknown density f are observed with an additive i.i.d. noise, independent of X's. We assume that the density f belongs to a smoothness class, has a characteristic function described either by a polynomial $|u|^{-\beta}$, $\beta > 1/2$ (Sobolev class) or by an exponential $\exp(-\alpha|u|^r)$, $\alpha,\ r > 0$ (called supersmooth), as $|u| \to \infty$. The noise density is supposed to be known and such that its characteristic function decays either as $|u|^{-s}$, $s > 0$ (polynomial noise) or as $\exp(-\gamma|u|^s)$, $s,\ \gamma > 0$ (exponential noise), as $|u| \to \infty$.

We study the problems of estimating the quadratic functional $\int f^2$ and use this estimator for the goodness-of-fit test in L_2 distance, from noisy observations, in all possible combinations of the previous setups.

We construct an estimator of $\int f^2$ based on the deconvolution kernel. When the unknown density is smoother enough than the noise density, we prove that this estimator is $n^{-1/2}$ consistent, asymptotically normal and efficient (for the variance we compute). Otherwise, we give nonparametric minimax upper bounds for the same estimator. For the goodness-of-fit test, we prove minimax upper bounds for a test statistic derived from the previous estimator. Surprisingly, in the case of supersmooth densities and polynomial noise we obtain parametric $n^{-1/2}$ minimax rate of testing.

Finally, we give an approach unifying the proof of nonparametric minimax lower bounds. We prove them for Sobolev densities and polynomial noise, for Sobolev densities and exponential noise and for supersmooth densities with exponential noise such that $r < s$. Note that in these last two setups we obtain exact testing constants associated to the asymptotic minimax rates.

Mathematical Subject Classification 62F12, 62G05, 62G10, 62G20,

Keywords Asymptotic efficiency, convolution model, exact constant in nonparametric tests, goodness-of-fit tests, infinitely differentiable functions, quadratic functional estimation, minimax tests, Sobolev classes.

1 Introduction

We consider the **convolution model**,

$$Y_i = X_i + \varepsilon_i,\ i = 1, \ldots, n \tag{1}$$

where all observations are independent. We denote the common unknown density of X_i, $i = 1, \ldots, n$ by f, having given smoothness. Let $\Phi(u) = \int e^{ixu} f(x)dx$ denote its characteristic function. We observe only the Y_i, $i = 1, \ldots, n$. The noise is supposed i.i.d. having known probability density g.

We consider the following nonparametric classes for the underlying density, which is always supposed to belong to $L_1 \cap L_2$. A **Sobolev class** is defined by

$$W(\beta, L) = \left\{ f : \mathbb{R} \to \mathbb{R}_+, \mathcal{C}^\beta, \text{ density function, } \int |\Phi(u)| \, |u|^\beta \, du < \infty, \right. \\ \left. \frac{1}{2\pi} \int |\Phi(u)|^2 \, |u|^{2\beta} \, du \leq L \right\}, \tag{2}$$

with the smoothness $\beta > 1/2$ and radius $L > 0$.

A class of **supersmooth densities** is defined by

$$S(\alpha, r, L) = \left\{ f : \mathbb{R} \to \mathbb{R}_+, \mathcal{C}^\infty, \text{ density function, } \frac{1}{2\pi} \int |\Phi(u)|^2 \exp(2\alpha |u|^r) \, du \leq L \right\}, \tag{3}$$

for α, r, L positive constants.

Classes $S(\alpha, r, L)$ of infinitely derivable functions appeared e.g. in Lepski and Levit [24]. Note that for $r > 1$ these are analytic functions, for $r = 1$ they are analytic on a strip of size 2α around the real axis.

Let the noise be i.i.d. with probability density g and characteristic function Φ^g and the resulting observations have common density $p = f \star g$ and characteristic function $\Phi^p = \Phi \cdot \Phi^g$. We also consider noise having non null Fourier transform, $\Phi^g(u) \neq 0$, $\forall\ u \in \mathbb{R}$. Typically two different behaviours are distinguished in nonparametric estimation:

polynomially smooth (or polynomial) noise

$$|\Phi^g(u)| \sim |u|^{-s},\ |u| \to \infty,\ s > 1; \tag{4}$$

exponentially smooth (or supersmooth or exponential) noise

$$|\Phi^g(u)| \sim \exp(-\gamma |u|^s),\ |u| \to \infty,\ \gamma, s > 0. \tag{5}$$

We consider here nonparametric minimax goodness-of-fit tests from noisy data, that is for a given density f_0 in the smoothness class $W(\beta, L)$, respectively $S(\alpha, r, L)$, decide whether

$$\begin{aligned} H_0 &: \ f = f_0 \\ H_1(\mathcal{C}, \psi_n) &: \ f \text{ in the smoothness class }, \|f - f_0\|_2^2 \geq \mathcal{C}\psi_n^2, \end{aligned}$$

from observations $Y_1, \ldots, Y_n$, for some fixed $\mathcal{C} > 0$ and $\psi_n > 0$.

Definition 1 *For a given $0 < \gamma < 1$, a test statistic Δ_n^* is said to attain the testing rate ψ_n over the smoothness class if there exists $\mathcal{C}^* > 0$ such that*

$$\limsup_{n \to \infty} \left\{ P_{f_0}[\Delta_n^* = 1] + \sup_{f \in H_1(\mathcal{C}, \psi_n)} P_f[\Delta_n^* = 0] \right\} \leq \gamma, \tag{6}$$

for all $\mathcal{C} > \mathcal{C}^*$. The rate ψ_n is called minimax rate of testing, if there exists $\mathcal{C}_* > 0$ and

$$\liminf_{n \to \infty} \inf_{\Delta_n} \left\{ P_{f_0}[\Delta_n = 1] + \sup_{f \in H_1(\mathcal{C}, \psi_n)} P_f[\Delta_n = 0] \right\} \geq \gamma, \tag{7}$$

for all $0 < \mathcal{C} < \mathcal{C}_*$, where the inf *is taken over all test procedures Δ_n.*

Moreover, if $\mathcal{C}^* = \mathcal{C}_*$ we call ψ_n exact (or sharp) minimax rate of testing.

We recall that the usual procedure is to construct the test statistic Δ_n^* such that (6) holds, also called the upper bound of the testing rate and then prove the minimax optimality of this procedure, i.e. the lower bounds in (7). If the test procedure does not depend on the smoothness of the unknown functions (which may vary in some interval), it is called adaptive to the smoothness and ψ_n is minimax adaptive rate.

In the convolution model (1), the problem of nonparametric estimation of deconvolution density f was intensively studied over the past two decades. Densities belonging to Hölder or Sobolev classes are known to be estimated at reasonably fast rates when mixed with polynomial noise and logarithmic, slow rates when mixed with exponential noise (see Carroll and Hall [7], Fan [11], [12] and [13], etc.).

Classes of supersmooth densities were first considered in the convolution model by Pensky and Vidakovic [31], who computed rates of convergence, adaptive to the smoothness, of wavelets estimators and noticed that faster rates can still be expected in this problem. Comte and Taupin [8] used model selection for adaptive estimation of the deconvolution density. Rate minimax optimality of a kernel estimator and lower bounds for the pointwise risk over such classes was proven by Butucea [4] for polynomial noise (and nearly parametric rate) and optimality in the rate and in the constant in the case $r < s$, by Butucea and Tsybakov [6], for exponential noise.

In this paper, in order to surpass difficulties of estimation we address different issues and principally the goodness-of-fit test from noisy data in L_2 norm. To our knowledge this is the first time testing was performed from data contaminated with errors. Minimax and adaptive theory of testing was extensively developed in density model when direct observations are available, but also for regression and Gaussian white noise model. For nonparametric minimax rates in goodness-of-fit testing in different setups we refer to Ingster [19] and references therein, Ermakov [9] and [10]. Exact minimax rates were found, see e.g. Lepski and Tsybakov [26] for regression model in pointwise and sup-norm distances. First adaptive rates were given by Spokoiny [34]. For a complete review of the literature we refer to Ingster and Suslina [20].

To our knowledge, exact minimax rates of testing for supersmooth functions are known only in the Gaussian white noise model, see Pouet [32], in the case $r = 1$, with pointwise and sup-norm distances. These results are more related to pointwise estimation of the analytic function than to our results in L_2 distance and noisy observations hereafter.

A very original approach is the problem of goodness-of-fit to a parametric composite null hypothesis as in Pouet [33], Gayraud and Pouet [15]; composite null hypothesis plus adaptation to the smoothness in Fromont and Laurent [14] and Gayraud and Pouet [16]. Other developments concern non-asymptotic minimax rates for the mean of a sequence of Gaussian variables by Baraud [1]. In view of numerous practical applications of testing we expect the same problem in the context of data contaminated with errors to find similar extensive use in applied problems.

Here, the goodness-of-fit problem is considered in L_2 distance, that is, we reject the null hypothesis for densities f far enough from the density f_0 under H_0, where "far" is measured by $\|f - f_0\|_2$. This distance depends on n and it corresponds to the rate of testing. As we can expect, testing problem is easier than the estimation problem, i.e. the testing rates are faster as they appear in Table 2. One of the most surprising results is that minimax L_2 testing can be performed at parametric rate $n^{-1/2}$ for supersmooth densities and polynomial noise (though deconvolution rate is known to be less by a power of logarithm than $n^{-1/2}$).

Another remark concerns setups where densities and noise have similar smoothness properties. For Sobolev densities and polynomial noise, we have one rate of testing, slower than $n^{-1/2}$ but faster than the deconvolution estimation rate, as it was already noticed in testing problem with direct observations. On the contrary, for supersmooth densities and exponential noise a change in the rate is observed like in deconvolution density estimation (see Butucea and Tsybakov [6]).

We actually give exact minimax rates of testing in setups with densities less regular than the noise: Sobolev densities and exponential noise, supersmooth densities less smooth than the corresponding exponential noise ($r < s$).

The natural test statistic in this context is an estimator of $\int (f - f_0)^2$, where f_0 is given, from noisy data. Here we study also optimal estimators d_n^2 for $d^2 := \int f^2$, where f is the deconvolution density in the model (1).

Definition 2 *An estimator d_n^2 of d^2 is said to attain the rate φ_n over the smoothness class $W(\beta, L)$, respectively $S(\alpha, r, L)$, if there exists a constant $C > 0$ such that*

$$\limsup_{n\to\infty} \sup_f \varphi_n^{-1} E_f[|d_n^2 - d^2|] \leq C \tag{8}$$

and this rate is called minimax if no other estimator attains better rates uniformly over the class

$$\liminf_{n\to\infty} \inf_{\hat{d}_n^2} \sup_f \varphi_n^{-1} E_f[|\hat{d}_n^2 - d^2|] \geq c, \tag{9}$$

for some $c > 0$, depending only on fixed known parameters, where the supremum is taken over all densities in the smoothness class and the infimum over all estimators $\hat{d}_n^2$.

In the case where parametric $n^{-1/2}$ rate is attained we prove the asymptotic efficiency Cramer-Rao bound of the estimator (also called efficient estimator).

Definition 3 *An estimator d_n^2 of $d^2 = \int f^2$ is asymptotically normally distributed with asymptotic variance $\mathcal{W} = \mathcal{W}(f)$ if*

$$\sqrt{n}\left(d_n^2 - d^2\right) \xrightarrow{d} N\left(0, \mathcal{W}\left(f\right)\right).$$

Moreover, it attains the asymptotic efficiency Cramer-Rao bound if for any f_0 in the Sobolev class $W(\beta, L)$, respectively in $S(\alpha, r, L)$, and a family of shrinking neighbourhoods of f_0: $\mathcal{V}(f_0)$

$$\inf_{\mathcal{V}(f_0)} \liminf_{n\to\infty} \sup_{f\in\mathcal{V}(f_0)} nE_f\left[(\hat{d}_n^2 - d^2)^2\right] \geq \mathcal{W}(f_0),$$

for any other estimator $\hat{d}_n^2$ of d^2.

In direct estimation, when data $X_1, \ldots, X_n$ are available, it is well established that parametric rates could be achieved for smooth enough densities belonging e.g. to the Hölder class. Lower bounds for slower rates were found by Bickel and Ritov [2] for smoothnesses less than $1/4$. In this context, Laurent [23] gave efficient estimation at parametric rate, Birgé and Massart [3] proved nonparametric lower bounds for estimating more general quadratic functionals. The study of general functionals was completed by Kerkyacharian and Picard [21] for minimax rates and Tribouley [36] for adaptive estimation. In the context of regression model and Gaussian sequence model, Nemirovski [30] found necessary conditions for existence of asymptotically efficient estimators of less smooth functionals, one or two times continuously differentiable.

In the convolution model, linear functionals were estimated in a minimax setup by Matias and Taupin [29]. Finally, the estimator d_n^2 of $\int f^2$ considered here was partly studied by Butucea [5] for proving asymptotic normality of the integrated square error (ISE) for kernel estimator in the convolution model.

In this paper, we give minimax results in the setups on the nonparametric side ("regime") and efficiency constant in the sense of the theory by Ibragimov and Khas'minskii [18] and Khoshevnik and Levit [22] for asymptotically normal, $n^{-1/2}$-consistent estimator (see Table 1).

Though we deal with both testing and quadratic functional estimation problems, these are essentially different problems. Indeed, we know in the case of direct observations that for other L_p norms ($p \neq 2$) the rates for minimax testing are different from minimax rates when estimating the L_p norm of a density. Compare e.g. Spokoiny [35] for testing rates and Lepski, Nemirovski and Spokoiny [25] for L_p-norm estimation.

The structure of the paper is as follows. In Section 2 we introduce the estimator d_n^2 of $\int f^2$ and indicate the choice of bandwidth in order to prove either upper bounds in the minimax sense, or its asymptotic normality and efficiency, according to different setups. In Section 3 we deal with the goodness-of-fit testing problem and introduce the test statistic. For each setup, we define the tuning parameters and compute the minimax upper bounds for testing rates. Finally, in Section 4, we describe the approach unifying the proof of minimax nonparametric lower bounds from Sections 2 and 3 and prove them for nonparametric setups of Sobolev classes of densities and exponential, respectively polynomial noise, and for the bias dominated setup of supersmooth densities less smooth than the exponentially smooth noise ($r < s$). Finally, some auxiliary results appear in the Appendix.

2 Estimation of $\int f^2$ in the convolution model

In the described model, we consider the problem of estimating $d^2 = \|f\|_2^2$, from available observations $(Y_i)_{i=1,\ldots,n}$, where the density f of observations $(X_i)_{i=1,\ldots,n}$ is unknown. Let us denote the deconvolution kernel K_n defined via its Fourier transform as

$$\Phi^{K_n}(u) = \left(\Phi^g\left(\frac{u}{h}\right)\right)^{-1} \Phi^K(u), \tag{10}$$

where $K(x) = \sin(x)/(\pi x)$ is such that $\Phi^K(u) = I_{[|u|\le 1]}$ and the bandwidth $h = h_n \to 0$, when $n \to \infty$ will be specified later.

Define d_n^2 a bias-reduced estimator of d^2 by

$$d_n^2 = \frac{1}{n(n-1)} \sum_{k\neq j=1}^{n} \int K_{n,h}(x - Y_k) K_{n,h}(x - Y_j)\, dx. \tag{11}$$

In the sequel, we shall denote the L_2 scalar product of two functions M and N by $\langle M, N\rangle = \int M(x)N(x)dx$ and by $\overline{M}$ the complex conjugate of M.

Definition 4 *Let d_n^2 in (11) be the estimator of d^2, having bandwidth $h > 0$. We call the bias and the variance of this estimator, respectively:*

$$B(d_n) \triangleq |E_f[d_n^2] - d^2| \text{ and } V(d_n) \triangleq E_f\left[|d_n^2 - E_f[d_n^2]|^2\right].$$

2.1 Sobolev densities and polynomial noise

We shall study in detail the case where the underlying density f belongs to a Sobolev class $W(\beta, L)$, with $\beta > 1/2$, defined in (2) and the noise is s-polynomial as defined in (4).

Proposition 1 *If f is a fixed density in the Sobolev class $W(\beta, L)$, the estimator d_n^2 in* (11) *with bandwidth $h > 0$ is such that*

$$
\begin{aligned}
B(d_n) &\leq Lh^{2\beta} \\
V(d_n) &= \frac{4\Omega_g^2(f)}{n}(1+o(1)) + \frac{2\|p\|_2^2}{n^2 h^{4s+1}} \frac{(1+o(1))}{\pi(4s+1)}, \quad \text{if } \beta \geq s; \\
V(d_n) &= \frac{O(1)}{nh^{2(s-\beta)}} + \frac{2\|p\|_2^2}{n^2 h^{4s+1}} \frac{1+o(1)}{\pi(4s+1)}, \quad \text{if } \beta < s.
\end{aligned}
$$

where $\Omega_g(f) \geq 0$ is defined later on, in (12)*, $o(1) \to 0$ and $h \to 0$, when $n \to \infty$.*

In order to define $\Omega_g(f)$, let us see that for any f in the Sobolev class $W(\beta, L)$ and g a noise density satisfying (4), we have $\Phi/\overline{\Phi^g}$ a continuous function which is absolutely and quadratically integrable (see Lemma 7). Then we can define the function

$$F(y) = \frac{1}{2\pi} \int e^{-iyu} \frac{\Phi(u)}{\overline{\Phi^g}(u)} du,$$

which is uniformly continuous function, but it is not necessarily a density function. It is known (see Lukacs [28]) that if both characteristic functions Φ and $\overline{\Phi^g}$ are analytic around 0 then their quotient cannot be the characteristic function of any distribution function. Nevertheless, this function is bounded and its L_2 norm is uniformly bounded over densities f in the Sobolev class by M^F depending only on β, L and the fixed given density g.

Define:

$$\Omega_g^2(f) = \int |F(y)|^2 p(y) dy - \left(\int f^2(x) dx \right)^2 = E_f[|F(Y)|^2] - (E_f[F(Y)])^2. \quad (12)$$

Indeed, $E_f[F(Y)]$ is a real number, since:

$$\|f\|_2^2 = \frac{1}{2\pi} \langle \Phi, \overline{\Phi} \rangle = \frac{1}{2\pi} \langle \Phi^p, \frac{\overline{\Phi}}{\Phi^g} \rangle = \langle p, F \rangle = E_f[F(Y)].$$

Remark 1 Note that (12) says that

$$4\Omega_g^2(f) = 4V(F(Y)).$$

This is heuristically similar to the results by Laurent [23] for direct estimation of $\int f^2$ where $4V(f(X)) = 4\int f^3 - 4(\int f^2)^2$ appears. Moreover, Bickel and Ritov [2] estimate with direct observation $\int (f^{(s)})^2$, where $f^{(s)}$ is the s-derivative of f and s a nonnegative integer . They note that via integration by parts, we can write $\int (f^{(s)})^2 = E_f[(-1)^s f^{(2s)}(X)]$. Then they get nonparametric rates as soon as the smoothness of the density is larger than $2s + 1/4$ and parametric rate for smoothness less than $2s + 1/4$ with asymptotic efficiency constant $4V(f^{(2s)}(X))$.

In Theorems 1 and 2 we describe the same change of "regime" when $\beta \geq s + 1/4$, respectively $\beta < s + 1/4$. Similarities between deconvolution with s-polynomial noise and derivative of order s have been noticed before. Indeed, we actually estimate here $\int f^2 = E_f[F(Y)]$, where $F \star g = f$, whenever the function F exists and F replaces the s-derivative of the function f.

Proof of Proposition 1. Let us note that

$$\begin{aligned} E_f[d_n^2] &= E_f[\langle K_{n,h}(\cdot - Y_1), K_{n,h}(\cdot - Y_2)\rangle] \\ &= \|K_{n,h} \star p\|_2^2 = \|K_h \star f\|_2^2 = \frac{1}{2\pi}\int \Phi^K(hu)|\Phi(u)|^2 du. \end{aligned} \tag{13}$$

By Plancherel formula in equation (13):

$$B(d_n^2) = \frac{1}{2\pi}\left|\int (\Phi^K(hu) - 1)|\Phi(u)|^2 du\right| \leq \frac{1}{2\pi}\int_{|u|>1/h} (h|u|)^{2\beta}|\Phi(u)|^2 du \leq Lh^{2\beta}.$$

As for the variance let us write first:

$$\begin{aligned} d_n^2 - E_f[d_n^2] &= \frac{1}{n(n-1)}\sum_{k\neq j}^{n}\langle K_{n,h}(\cdot - Y_k) - K_h \star f, K_{n,h}(\cdot - Y_j) - K_h \star f\rangle \\ &+ \frac{2}{n}\sum_{k=1}^{n}\langle K_{n,h}(\cdot - Y_k) - K_h \star f, K_h \star f\rangle = S_1 + S_2, \quad say. \end{aligned}$$

Variables in S_1 are uncorrelated to the variables in S_2 and all of them are centered. Thus, $V(d_n^2) = E_f[|S_1|^2] + E_f[|S_2|^2]$. We have

$$\begin{aligned} E_f[|S_1|^2] &= \frac{2}{n(n-1)}\left(E_f[|\langle K_{n,h}(\cdot - Y_1) - K_h \star f, K_{n,h}(\cdot - Y_2) - K_h \star f\rangle|^2]\right) \\ &= \frac{2}{n(n-1)}\left(E_f[|\langle K_{n,h}(\cdot - Y_1), K_{n,h}(\cdot - Y_2)\rangle|^2] - \|K_h \star f\|_2^4\right) \\ &= \frac{2\|p\|_2^2}{\pi(4s+1)}\frac{1+o(1)}{n^2h^{4s+1}}, \end{aligned} \tag{14}$$

where indeed, $\|K_h \star f\|_2^4 = \|f\|_2^4(1+o(1))$ by the bias computations and this term is negligible with respect to the first one. Similarly to Butucea [5], we have

$$\begin{aligned} & E_f[|\langle K_{n,h}(\cdot - Y_1), K_{n,h}(\cdot - Y_2)\rangle|^2] \\ &= \frac{1}{h^2}\int\int\left|\int K_n\left(\frac{x-u}{h}\right)K_n\left(\frac{x-v}{h}\right)\frac{dx}{h}\right|^2 p(u)p(v)dudv \\ &= \frac{1}{h}\int\int\frac{1}{h}\left|\int K_n\left(z + \frac{v-u}{h}\right)K_n(z)\,dz\right|^2 p(u)p(v)dudv \\ &= \frac{1}{h}\int\int\frac{1}{h}\left|M_n\left(\frac{v-u}{h}\right)\right|^2 p(u)p(v)dudv = T, say, \end{aligned}$$

where $M_n(x) = \int K_n(z+x)K_n(z)dz$. Finally, use the fact that p is at least $(\beta+s-1/2)$ - Lipschitz continuous (Lemma 6)

$$
\begin{aligned}
&\left|T - \frac{1}{h}\|p\|_2^2\|M_n\|_2^2\right| \\
\leq\ & \frac{1}{h}\left|\int\int\left(\frac{1}{h}\left|M_n\left(\frac{v-u}{h}\right)\right|^2 p(u) - p(v)\int|M_n|^2\right)dup(v)dv\right| \\
\leq\ & \frac{1}{h}\int\int|M_n(x)|^2|p(v+hx)-p(v)|dxp(v)dv \\
\leq\ & \frac{1}{h}\int\left(\int_{|hx|\leq\epsilon}|M_n(x)|^2 L\epsilon^{\beta+s-1/2}dx + \int_{|x|>\epsilon/h} 2M^Y|M_n(x)|^2dx\right)p(v)dv \\
\leq\ & \frac{1}{h}o(\|M_n\|_2^2),
\end{aligned}
$$

where we chose $\epsilon \to 0$ such that $\epsilon/h \to \infty$ so that

$$
T = \frac{\|p\|_2^2\|M_n\|_2^2}{h}(1+o(1)). \tag{15}
$$

By Plancherel formula, $\|M_n\|_2^2 = \int|\Phi^{K_n}(u)\Phi^{K_n}(-u)|^2du = (\pi(4s+1)h^{4s})^{-1}(1+o(1))$. Note that we should again split the integration domain and evaluate the dominant term in the previous integral. Replace this in (15) in order to get (14).

On the other hand, let us deal now with:

$$
\begin{aligned}
E_f[|S_2|^2] &= \frac{4}{n}\left(E_f[|\langle K_{n,h}(\cdot - Y_1), K_h \star f\rangle|^2] - \|K_h\star f\|_2^4\right) \\
&= \frac{4}{n}\left(E_f\left[\frac{1}{2\pi}\left|\int e^{iuY_1}\frac{\Phi^K(hu)}{\Phi^g(u)}\overline{\Phi}(u)du\right|^2\right] - \|f\|_2^4(1+o(1))\right). \quad (16)
\end{aligned}
$$

Use Lemma 7 and Lebesgue convergence theorem to see that, if $\beta \geq s$, there exists a function

$$
F(y) = \frac{1}{2\pi}\int e^{iuy}\frac{\overline{\Phi}(u)}{\Phi^g(u)}du = \lim_{h\to 0}\frac{1}{2\pi}\int_{|u|\leq 1/h} e^{iuy}\frac{\overline{\Phi}(u)}{\Phi^g(u)}du,
$$

which is uniformly continuous, bounded such that $\|F\|_2^2 = \|\overline{\Phi}/\Phi^g\|_2^2/(2\pi)$. Note also that F is a limit of real-valued functions. Indeed, write

$$
\frac{1}{2\pi}\int_{|u|\leq 1/h} e^{iuy}\frac{\overline{\Phi}(u)}{\Phi^g(u)}du = \frac{1}{2\pi}\int e^{iuy}\frac{\Phi^K(hu)}{|\Phi^g(u)|^2}\overline{\Phi}^p(u)du,
$$

and note that $\Phi^K(hu)/|\Phi^g(u)|^2$ is a symmetric integrable function and $\overline{\Phi}^p(u)$ is the Fourier transform of $p(-y)$. Thus the integral is the convolution of two-real valued functions. The limit of real-valued functions, F is real-valued as well.

Thus, we obtain in (16):

$$E_f[|S_2|^2] = \frac{4}{n}(E_f[|F(Y)|^2] - \|f\|_2^2)(1+o(1)) = \frac{4\Omega_g^2(f)}{n}(1+o(1)). \qquad (17)$$

Together with (14) we get the variance for the case $\beta \geq s$.

In case $\beta < s$, go back to (16):

$$\begin{aligned}
& E_f\left[\frac{1}{2\pi}\left|\int e^{iuY_1}\frac{\Phi^K(hu)}{\Phi^\varepsilon(u)}\overline{\Phi}(u)du\right|^2\right] \\
\leq\ & \frac{M^Y}{2\pi}\left(\int_{|u|\leq 1/h}\left|\frac{\overline{\Phi}(u)}{\Phi^\varepsilon(u)}\right|du\right) \leq O(1)\left(\int_{M\leq|u|\leq 1/h}|u|^s|\Phi(u)|du\right)^2 \\
\leq\ & O(1)\left(h^{\beta-s}\int_{M\leq|u|\leq 1/h}|u|^\beta|\Phi(u)|du\right)^2 \leq O(1)h^{2(\beta-s)}. \qquad (18)
\end{aligned}$$

So, from (14), (16) and (18) we get the variance for $\beta < s$. ■

An easy consequence of Proposition 1 is that if the underlying unknown density is smoother enough than the noise ($\beta > s + 1/4$) our parameter can be estimated at parametric rate. We establish next asymptotic normality and a Cramer-Rao type of asymptotic efficiency bound.

Theorem 1 *If $\beta > s+1/4$, the estimator d_n^2 defined in* (11) *with bandwidth $h = h_*$ such that*

$$n^{-\frac{1}{4s+1}} \ll h_* \ll n^{-\frac{1}{4\beta}},$$

is asymptotically normally distributed estimator of d^2, i.e.

$$\sqrt{n}\left(d_n^2 - d^2\right) \xrightarrow{d} N\left(0, 4\Omega_g^2(f)\right).$$

Moreover, it attains the asymptotic efficiency Cramer-Rao bound.

Proof. Let us decompose the risk of the estimator as follows

$$E_f[|d_n^2 - d^2|] \leq B(d_n) + \sqrt{V(d_n)} \leq Lh^{2\beta} + \frac{2\Omega_g(f)}{\sqrt{n}}(1+o(1)),$$

and then use Proposition 1. Indeed, if $\beta > s+1/4$ and if $n^{-1}h^{-(4s+1)} \ll 1$ we see that $4\Omega_g^2(f)/n(1+o(1))$ is the dominant term in the variance. Let us take $h = o(n^{-1/(4\beta)})$ such that the bias be infinitely smaller, $Lh^{2\beta} \ll 2\Omega_g(f)/n$. So,

$$\sqrt{n}(d_n^2 - d^2) = \sqrt{n}(d_n^2 - E_f[d_n^2]) + \sqrt{n}B(d_n).$$

The second term of the sum in the right-hand side term tends to 0 and the asymptotic normality of the first term can be deduced from Butucea [5]. It is in this case a classical central limit theorem for U-statistics of order 1.

For the Cramer-Rao bound, we follow the lines of proof in Laurent [23]. Similar results were given by Bickel and Ritov [2] following the theory by Ibragimov and Hasminski [18] and Khoshevnik and Levit [22]. A first step of the proof is to compute the Fréchet derivative of the functional $\int f^2 = \int F \cdot p$ at likelihood $p_0 = f_0 \star g$:

$$\int F \cdot p - \int F_0 \cdot p_0 = \int 2F_0(p - p_0) + \int (F - F_0)(p - p_0)$$

and $\int (F - F_0)(p - p_0) = o(\|p - p_0\|_2)$, when $\|p - p_0\|_2 \to 0$. Next, consider the space orthogonal to the square root of the likelihood $\sqrt{p_0}$,

$$H = \{k : \int k\sqrt{p_0} = 0\}$$

and the projection operator unto this space:

$$P_{H(p_0)}(k) = k - (\int k\sqrt{p_0})\sqrt{p_0}.$$

Write $K_n = K = T'(p_0)\sqrt{p_0} = P_{H(p_0)}(k)$ as $\langle g, k \rangle$. Then the minimal variance is given by $\|g\|_2^2$.

Here, $T'(p_0)k = \int 2F_0 k$, then

$$\begin{aligned} K &= \int 2F_0\sqrt{p_0}(k - (\int k\sqrt{p_0})\sqrt{p_0}) \\ &= \int (2F_0\sqrt{p_0})k - \left(\int 2F_0 p_0\right) \int \sqrt{p_0} k. \end{aligned}$$

So, finally,

$$\|g\|_2^2 = 4\int |F_0|^2 p_0 - \left(\left|\int 2F_0 p_0\right|\right)^2 = 4V_{f_0}(F_0(Y)).$$

■

In the following theorem we compute the rate on the nonparametric side ($1/2 < \beta \leq s + 1/4$). We prove in Section 3 that this rate is optimal in the minimax approach under the following additional assumption on the noise distribution.

Assumption (P) The distribution of the polynomial noise in (4) is such that Φ^g is at least 3 times continuously differentiable. Moreover there exist $A_1,\ A_2 > 1$, $u_0,\ u_1,\ u_2 > 0$ large enough such that

$$|\Phi^g(u)| \geq u_0,\ \forall |u| \leq A_1 \text{ and } |(\Phi^g)^{(k)}(u)| \leq \frac{u_k}{|u|^{s+k}}, \quad \text{for } k = 1, 2,\ \forall |u| \geq A_2.$$

Theorem 2 *If $1/2 < \beta \leq s + 1/4$, the estimator d_n^2 of d^2 defined in (11) with bandwidth h_* satisfies the upper bound (8) for the rate φ_n, where*

$$h_* = n^{-\frac{2}{4\beta+4s+1}}, \quad \varphi_n = n^{-\frac{4\beta}{4\beta+4s+1}}.$$

Moreover, under Assumption (P) this rate is minimax.

Proof of (8). If $1/2 < \beta \le s+1/4$, $\|p\|_2^2/(\pi(4s+1)n^2h_*^{4s+1})$ is the dominant term in the variance no matter whether $\beta \ge s$ or $\beta < s$. The bandwidth h_* minimizes the bias plus the variance. The upper bound of the normalized mean error is less than $C = \max\{L, \sqrt{2M^p/(\pi(4s+1)}\}$, see Lemma 6. ■

Remark 2 In view of Butucea [5], we get asymptotic normality of the estimator d_n^2 defined in (11) in the case $1/2 < \beta < s+1/4$, for h such that

$$h = o(1)n^{-\frac{1}{4s+1}} \text{ and } h = o(1)n^{-\frac{2}{4\beta+4s+1}}$$

(i.e. the variance tends to 0 and the bias is infinitely smaller than the variance, when $n \to \infty$),

$$nh^{2s+1/2}\left(d_n^2(h) - d^2\right) \xrightarrow{d} N\left(0, \frac{2\|p\|_2^2}{\pi(4s+1)}\right).$$

We will actually use the following type of result for the testing problem

$$nh_*^{2s+1/2}\left(d_n^2 - E_f[d_n^2]\right) \xrightarrow{d} N\left(0, \frac{2\|p\|_2^2}{\pi(4s+1)}\right),$$

for the optimal bandwidth h_* defined in Theorem 2.

2.2 Supersmooth densities and polynomial noise

In the case of supersmooth densities, always smoother than the polynomial noise, we can always define the function F as the inverse Fourier transform of $\Phi/\overline{\Phi^g}$. Next Theorem gives us the right choice of the bandwidth so that d_n^2 be an asymptotically normal and efficient estimator.

Theorem 3 *The estimator d_n^2 defined in (11) with bandwidth h_* such that*

$$h_* \ll \left(\frac{\log n}{4\alpha}\right)^{-1/r}$$

is asymptotically normally distributed and it attains the asymptotic efficiency Cramer-Rao bound $4\Omega_g^2$, (see Definition 3).

Proof. In this case, the bias changes,

$$B(d_n) \le L\exp\left(-\frac{2\alpha}{h^r}\right).$$

The variance is strongly dependent on the noise distribution, so very little is changed. In this case, the underlying density is always much more regular than the noise, so the function F always exists in this setup. So, we can put together (14) and (17)

$$V(d_n) = \frac{4\Omega_g^2(f)}{n}(1+o(1)) + \frac{2\|p\|_2^2}{n^2h^{4s+1}}\frac{(1+o(1))}{\pi(4s+1)}.$$

It is obvious that we need to choose $h_* = o(1)(\log n/(4\alpha))^{-1/r}$, in order to have the squared bias infinitely smaller than the dominant term of the variance $1/n$. ■

2.3 Sobolev densities and exponential noise

In this setup the noise is much smoother so estimation is always difficult, i.e. at nonparametric slower rates. We prove the lower bounds (9), under the following additional assumption, which is not very restrictive.

Assumption (E) The exponential noise distribution in (5) has a continuously differentiable Fourier transform such that

$$|(\Phi^g)'(u)| \leq O(1)|u|^{\mathcal{A}} \exp(-\gamma|u|^s),$$

for large enough $|u|$ and some fixed constant $\mathcal{A} \in R$.

Theorem 4 *The estimator d_n^2 of d^2 defined in* (11) *with bandwidth h_* satisfies the upper bound* (8) *for the rate φ_n, where*

$$h_* = \left(\frac{\log n}{2\gamma} - \frac{2\beta+1}{2\gamma s}\log\frac{\log n}{2\gamma}\right)^{-1/s}, \quad \varphi_n = L\left(\frac{\log n}{2\gamma}\right)^{-\frac{2\beta}{s}}.$$

Moreover, under Assumption (E) this rate is minimax.

Proof of (8). In this case, the bias is the same as in Proposition 1, $B(d_n) \leq Lh^{2\beta}$. As for the variance, we can still write $V(d_n) = E_f[|S_1|^2] + E_f[|S_2|^2]$, but both terms are different now, since they are highly dependent on the noise distribution. We still have,

$$E_f[|S_1|^2] = \frac{2+o(1)}{n(n-1)}\frac{\|p\|_2^2}{h}\|M_n\|_2^2,$$

see (15). Now,

$$\|M_n\|_2^2 = \frac{h}{2\pi}\int_{|h|\leq 1/h} e^{4\gamma|u|^s}du = \frac{h^s}{4\pi\gamma s}e^{\frac{4\gamma}{h^s}}(1+o(1)),$$

as $h \to 0$, and then

$$E_f[|S_1|^2] = \frac{(2+o(1))\|p\|_2^2}{4\pi\gamma s}\frac{h^{s-1}}{n^2}e^{\frac{4\gamma}{h^s}}. \tag{19}$$

The other term, can never be of parametric order anymore, the function F never exists in this setup. Indeed, recall that

$$\begin{aligned} E_f[|S_2|^2] &\leq \frac{c_1}{n}E_f\left[\left|\int e^{iuY_1}\frac{\Phi^K(hu)}{\Phi^g(u)}\overline{\Phi}(u)du\right|^2\right] \qquad (20) \\ &\leq \frac{c_1}{n}\left(\int\left|\frac{\Phi^K(hu)}{\Phi^g(u)}\overline{\Phi}(u)\right|du\right)^2 \end{aligned}$$

and this integral does not check the Lebesgue convergence theorem anymore. We can compute the rate of divergence, giving a loss in the rate, via Cauchy-Schwarz

$$\left(\int \left|\frac{\Phi^K(hu)}{\Phi^g(u)}\overline{\Phi}(u)\right| du\right)^2 \leq \int |\Phi(u)|^2|u|^{2\beta}du \int_{|u|\leq 1/h} |u|^{-2\beta}e^{2\gamma|u|^s}du \leq c_2 h^{2\beta+s-1}e^{\frac{2\gamma}{h^s}}. \tag{21}$$

From (19), (20) and (21), we get

$$V(d_n^2) \leq C_1\frac{h^{s-1}}{n^2}e^{\frac{4\gamma}{h^s}}(1+o(1)) + C_2\frac{h^{2\beta+s-1}}{n}e^{\frac{2\gamma}{h^s}}, \tag{22}$$

where $c_1, c_2, C_1, C_2 > 0$ are some constants.

As in Theorem 2 we actually select the bandwidth by minimizing an upper bound of the error:

$$E_f[|d_n^2 - d^2|] \leq \left(E_f[|d_n^2 - d^2|^2]\right)^{1/2} \leq \left(B^2(d_n) + V_f[d_n]\right)^{1/2}.$$

The optimality of this upper bound is proven by the corresponding lower bounds. Now, we consider

$$h_* = \arg\inf_{h>0}\left(L^2h^{4\beta} + c_2\frac{h^{2\beta+s-1}}{n}\exp\left(\frac{2\gamma}{h^s}\right)\right)$$

then h_* is a solution of the equation

$$h_*^{2\beta+1} = \frac{c}{n}\exp\left(\frac{2\gamma}{h_*^s}\right)(1+o(1)). \tag{23}$$

This proves that the bias is infinitely larger than the variance and gives announced h_* and rate of order of the bias φ_n. (If we suppose that the first term on the right-hand side of (22) is dominant, we get a contradiction). ■

2.4 Supersmooth densities and exponential noise

In this setup unknown densities and noise densities are both exponentially smooth. Note that minimax rates in the nonparametric "regime" are faster than any logarithm but slower than any polynomial of n.

Theorem 5 *If $r > s$ or, if $r = s$ and $\alpha > \gamma$, the estimator d_n^2 defined in* (11) *with bandwidth h_* such that*

$$h_* \ll \left(\frac{\log n}{4\alpha}\right)^{-1/r}$$

is asymptotically normally distributed and it attains the asymptotic efficiency Cramer-Rao bound $4\Omega_g^2$.

If $r < s$ the same estimator d_n^2 of d^2 with bandwidth h_* satisfies the upper bounds (8) *for the rate φ_n, where*

$$h_* \text{ solution of } h_*^{r-1-(r-1)_+/2}\exp\left(\frac{2\alpha}{h_*^r} + \frac{2\gamma}{h_*^s}\right) = cn(1+o(1)), \quad \varphi_n = L\exp\left(-\frac{2\alpha}{h_*^r}\right).$$

Moreover, unde Assumption (E) this rate is minimax.

Proof. We skip the proof of asymptotic efficiency. If $r < s$, we know the bias is $B(d_n^2) \le L \exp(-2\alpha/h^r)$ and the variance writes also $V_f[d_n^2] = E_f[|S_1|^2] + E_f[|S_2|^2]$. Furthermore, $E_f[|S_1|^2]$ is the same as in (19).

As in (20), we need to study

$$\left| \int_{|u| \le 1/h} e^{iuY_1} \frac{\overline{\Phi}(u)}{\Phi^g(u)} du \right|^2 .$$

If $r > s$ or, if $r = s$ and $\alpha > \gamma$, this integral is bounded by a constant depending only on α, r, L and the noise density g. So, the variance

$$V_f[T_n^{*2}] = E_f[|S_2|^2](1 + o(1)) = \frac{4\Omega_g^2(f)}{n}(1 + o(1)),$$

as soon as $h^{s-1}n^{-1}\exp(4\gamma/h^s) = o(1)$. For the bandwidth we chose in the theorem, this holds and proves this case.

If $r < s$,

$$\begin{aligned} \left| \int_{|u| \le 1/h} e^{iuY_1} \frac{\overline{\Phi}(u)}{\Phi^g(u)} du \right|^2 &\le \int e^{2\alpha|u|^r} |\Phi(u)|^2 du \int_{|u| \le 1/h} e^{-2\alpha|u|^r} e^{2\gamma|u|^s} du \\ &\le c_1 h^{s-1} \exp\left(\frac{2\gamma}{h^s} - \frac{2\alpha}{h^r} \right). \end{aligned}$$

Thus,

$$V_f[d_n^2] \le c_1 \frac{h^{s-1}}{n} \exp\left(\frac{2\gamma}{h^s} - \frac{2\alpha}{h^r} \right) + c_2 \frac{h^{s-1}}{n^2} \exp\left(\frac{4\gamma}{h^s} \right),$$

where $c_1, c_2, \ldots$ are some positive constants. As in Theorems 2 and 4 we find the optimal bandwidth by minimizing an upper bound of the error

$$h_* = \arg \inf_{h > 0} \left(L^2 \exp\left(-\frac{4\alpha}{h^r} \right) + c_1 \frac{h^{s-1}}{n} \exp\left(\frac{2\gamma}{h^s} - \frac{2\alpha}{h^r} \right) + c_2 \frac{h^{s-1}}{n^2} \exp\left(\frac{4\gamma}{h^s} \right) \right).$$

When we minimize the sum of the bias and of the first term in the variance, we find that h_* is solution of the equation

$$h_*^{r-1} \exp\left(\frac{2\alpha}{h_*^r} + \frac{2\gamma}{h_*^s} \right) = cn(1 + o(1)).$$

It implies that the first term in the variance is dominant over the second, if $r < 1$, meaning, moreover that

$$L^2 \exp\left(-\frac{4\alpha}{h_*^r} \right) = h_*^{r-s} c_2 \frac{h_*^{s-1}}{n} \exp\left(\frac{2\gamma}{h_*^s} - \frac{2\alpha}{h_*^r} \right)$$

i.e. the bias is infinitely larger than the variance for $r < s$ for the optimal h_*.

If we minimize the sum of the bias and the second term in the variance, we find that optimal bandwidth h_* verifies

$$h_*^{(r-1)/2} \exp\left(\frac{2\alpha}{h_*^r} + \frac{2\gamma}{h_*^s}\right) = cn(1 + o(1)).$$

The second term of the variance is dominant if $r \geq 1$ and in this case also the bias is dominant over the variance for $r < s$ and the optimal h_*, respectively, the bias is of the same order as the dominant term in the variance, if $r = s$ and $\alpha < \gamma$. This finishes the proof of the Theorem. ■

Remark 3: More upper bounds Let us add upper bounds of the estimation risk in cases not included in the Theorem. We put them aside since we do not provide corresponding lower bounds in these cases.

In the case $r = s$ and $\alpha < \gamma$, the same choice of the bandwidth holds as in the case $r < s$ of the preceding Theorem. The bias is in this case of the same order as the dominant term in the variance (not larger than the variance).

If $r = s$ and $\alpha = \gamma$,

$$V_f[d_n^2] \leq \frac{c_1}{nh} + c_2 \frac{h^{s-1}}{n^2} \exp\left(\frac{4\gamma}{h^s}\right).$$

In this case, the first term in the upper bound of the variance is always dominating over the second and when we minimize the sum of this term and of the bias we get an optimal h_* solution of

$$h_*^{r-1} \exp\left(\frac{4\alpha}{h_*^r}\right) = cn(1 + o(1)),$$

giving a fast rate of convergence of the order of the variance:

$$\varphi_n^2 = c_3 \frac{(\log n)^r}{n}.$$

3 Goodness-of-fit tests

Let us construct a test statistic from noisy data. It is natural to suggest as a test statistic T_n^*, the square root of the optimal estimator of the quadratic functional $\|f - f_0\|_2^2$:

$$T_n^{*2} = \frac{1}{n(n-1)} \sum_{k \neq j} \langle K_{n,h}(\cdot - Y_k) - f_0, K_{n,h}(\cdot - Y_j) - f_0 \rangle,$$

where $h > 0$, $h \to 0$ and $K_{n,h} = 1/hK_n(\cdot/h)$, for the same K_n defined in (10).

Define the test procedure

$$\Delta_n^* = \begin{cases} 1 & T_n^{*2} > \mathcal{C}^* t_n^2 \\ 0 & T_n^{*2} \leq \mathcal{C}^* t_n^2 \end{cases} \tag{24}$$

for a constant $\mathcal{C}^* > 0$ and some threshold $t_n > 0$ depending on the setup.

Density \ Noise	Polynomial: $\vert u\vert^{-s}$	Exponential: $\exp(-\gamma\vert u\vert^s)$
$f \in W(\beta, L)$ $\beta > 1/2$	$\beta < s + 1/4$: $O(1)n^{-\frac{4\beta}{4\beta+4s+1}}$ $\beta \geq s + 1/4$: $2\Omega_g n^{-\frac{1}{2}}$	$O(1)\left(\log n/(2\gamma)\right)^{-\frac{2\beta}{s}}$
$f \in S(\alpha, r, L)$	$2\Omega_g n^{-\frac{1}{2}}$	$r < s$: $O(1)\exp\left(-2\alpha/h_*^r\right)$ $(r > s)$ or $(r = s,\ \alpha > \gamma)$: $2\Omega_g n^{-\frac{1}{2}}$

where h_* is solution of $h_*^{r-1-(r-1)_+/2}\exp\left(\frac{2\alpha}{h_*^r} + \frac{2\gamma}{h_*^s}\right) = n(1 + o(1))$.

Table 1: Estimation rates of d^2 from noisy data

3.1 Sobolev densities and polynomial noise

Though two rates were attainable in the same setup for estimating d^2, only one minimax rate of testing is possible. This phenomenon is similar to the case of testing with direct observations.

Theorem 6 *The test procedure Δ_n^* defined in* (24) *for the threshold t_n attains the rate ψ_n and, under Assumption (P), ψ_n is a minimax rate of testing over the class $W(\beta, L)$, where*

$$h = h_* = n^{-\frac{2}{4\beta+4s+1}},\ t_n = \psi_n = n^{-\frac{2\beta}{4\beta+4s+1}}.$$

Proof of upper bounds (6)**.** Let us bound from above successively the first and second type error. Note that, for a fixed density $f_0 \in W(\beta, L)$:

$$E_{f_0}[T_n^{*2}] = \|K_h \star f_0 - f_0\|_2^2 = Lh^{2\beta}o(1),$$

similarly to the proof of Proposition 1. In order to compute the variance let us write as follows

$$\begin{aligned} T_n^{*2} - E_{f_0}[T_n^{*2}] &= \frac{1}{n(n-1)}\sum_{k\neq j}\langle K_{n,h}(\cdot - Y_k) - f_0, K_{n,h}(\cdot - Y_j) - f_0\rangle - \|K_h \star f_0 - f_0\|_2^2 \\ &= \frac{1}{n(n-1)}\sum_{k\neq j}\langle K_{n,h}(\cdot - Y_k) - K_h \star f_0, K_{n,h}(\cdot - Y_j) - K_h \star f_0\rangle \\ &\quad + \frac{2}{n}\sum_{k=1}^{n}\langle K_{n,h}(\cdot - Y_k) - K_h \star f_0, K_h \star f_0 - f_0\rangle. \end{aligned}$$

Note that, for all $k = 1, \ldots, n$,

$$
\begin{aligned}
&\langle K_{n,h}(\cdot - Y_k) - K_h \star f_0, K_h \star f_0 - f_0 \rangle \\
= & \frac{1}{2\pi} \int \left(\Phi^{K_n}(hu) e^{iuY_k} - \Phi^K(hu)\Phi_0(u) \right) \left(\Phi^K(hu) - 1 \right) \Phi_0(u) du \\
= & \frac{1}{2\pi} \int \Phi^K(hu) \left(\Phi^K(hu) - 1 \right) \left(e^{iuY_k}/\Phi^g(hu) - \Phi_0(u) \right) \Phi_0(u) du = 0,
\end{aligned}
$$

due to the support of Φ^K. Finally,

$$
V_{f_0}[T_n^{*2}] = E_{f_0}[|S_1|^2] = \frac{S\|p_0\|_2^2}{n^2 h^{4s+1}}(1 + o(1)),
$$

where $S = 2/(\pi(4s + 1))$. So the first type error can be written as follows

$$
\begin{aligned}
P_{f_0}[T_n^{*2} \geq \mathcal{C}^* t_n^2] &= P_{f_0}[T_n^{*2} - E_{f_0}[T_n^{*2}] \geq \mathcal{C}^* t_n^2 - c_1 h^{2\beta}] \\
&= O(1) \frac{n^{-2} h^{-(4s+1)}}{(\mathcal{C}^* t_n^2 - c_1 h^{2\beta})^2} \leq \frac{\alpha}{2}
\end{aligned}
$$

since, for $h = h_*$

$$
\frac{1}{nh^{2s+1/2}} \leq O(1) \max\{t_n^2, h^{2\beta}\}. \tag{25}
$$

For the second type error, consider a density f in $H_1(\mathcal{C}, \psi_n)$. Then, $E_f[T_n^{*2}] = \|K_h \star f - f_0\|_2^2$. The bias can be bounded from above as follows

$$
\begin{aligned}
B[T_n^{*2}] &= |\|K_h \star f - f_0\|_2^2 - \|f - f_0\|_2^2| \\
&= |\|K_h \star f\|_2^2 - \|f\|_2^2 - 2\langle K_h \star f - f, f_0 \rangle| \\
&\leq \frac{1}{2\pi} \int_{|u|>1/h} |\Phi(u)|^2 du + \frac{2}{2\pi} \int_{|u|>1/h} |\Phi(u)| \cdot |\Phi_0(u)| du \\
&\leq Lh^{2\beta}(1 + o(1)),
\end{aligned}
$$

since $\int_{|u|>1/h} |u|^{2\beta} |\Phi_0(u)|^2 du = o(1)$, for the fixed density f_0. In order to evaluate the variance, let us decompose as follows

$$
\begin{aligned}
T_n^{*2} - E_f[T_n^{*2}] &= \frac{1}{n(n-1)} \sum_{k \neq j} \langle K_{n,h}(\cdot - Y_k) - f_0, K_{n,h}(\cdot - Y_j) - f_0 \rangle - \|K_h \star f - f_0\|_2^2 \\
&= \frac{1}{n(n-1)} \sum_{k \neq j} \langle K_{n,h}(\cdot - Y_k) - K_h \star f, K_{n,h}(\cdot - Y_j) - K_h \star f \rangle \\
&\quad + \frac{2}{n} \sum_{k=1}^{n} \langle K_{n,h}(\cdot - Y_k) - K_h \star f, f - f_0 \rangle \\
&= S_1(f) + S_2(f - f_0), \text{say.}
\end{aligned}
$$

As in Proposition 1, the last two terms are uncorrelated, so $V_f[T_n^{*2}] = E_f[|S_1(f)|^2] + E_f[|S_2(f-f_0)|^2]$. Similar computation lead to

$$\begin{aligned} V_f[T_n^{*2}] &\leq \frac{C_1}{n^2 h^{4s+1}} + \frac{(4+o(1))\Omega_g^2(f-f_0)}{n} I(\beta > s) + \frac{C_3}{nh^{2(s-\beta)}} I(\beta \leq s) \\ &\leq \frac{C_1}{n^2 h^{4s+1}} + \frac{C_2\|f-f_0\|_2^2}{n} I(\beta > s) =^{Def} v_n^2. \end{aligned}$$

Indeed, let us see that whenever $\beta > s$, we find $M > 0$ large enough such that

$$\begin{aligned} \Omega_g^2(f-f_0) &\leq \int (F(y)-F_0(y))^2 p(y)dy \leq \|F-f_0\|_\infty^2 \leq \frac{1}{4\pi^2} \int \left| \frac{\Phi(u)-\Phi_0(u)}{\Phi^g(u)} \right|^2 du \\ &\leq \int_{|u|\leq M} c_1 M^{2s} |\Phi(u)-\Phi_0(u)|^2 du + \int_{|u|>M} c_2 |u|^{2s} |\Phi(u)-\Phi_0(u)|^2 du \\ &\leq c_3\|f-f_0\|_2^2 + \frac{c_2}{M^{2(\beta-s)}} \int_{|u|>M} |u|^{2\beta} |\Phi(u)|^2 du \leq C\|f-f_0\|_2^2, \end{aligned} \quad (26)$$

where, C is a constant depending only on β, L and of the fixed noise probability density g. We also use the fact that for $h = h_*$,

$$\frac{1}{nh^{2(s-\beta)}} / \left(\frac{1}{n^2 h^{4s+1}} \right) = n^{-\frac{1}{4\beta+4s+1}} = o(1).$$

So, the second type error can be bounded as follows

$$\begin{aligned} P_f[T_n^{*2} < \mathcal{C}^* t_n^2] &= P_f[T_n^{*2} - E_f[T_n^*] < \mathcal{C}^* t_n^2 - E_f[T_n^*] + \|f-f_0\|_2^2 - \|f-f_0\|_2^2] \\ &\leq P_f[T_n^{*2} - E_f[T_n^*] < -\|f-f_0\|_2^2 + \mathcal{C}^* t_n^2 + B[T_n^{*2}]] \\ &\leq P_f\left[\frac{T_n^{*2} - E_f[T_n^*]}{\sqrt{V_f[T_n^{*2}]}} \leq \frac{-\|f-f_0\|_2^2 + \mathcal{C}^* t_n^2 + B[T_n^{*2}]}{v_n} \right], \end{aligned} \quad (27)$$

for n large enough. From Butucea [5] we deduce easily the asymptotic normality of center and reduced T_n^{*2}. The proof is based on a result by Hall [17] for U-statistics of order 2 adapted for the case of noisy observations.

So the last probability in (27) is smaller than $\alpha/2 + o(1)$ if

$$\frac{-\|f-f_0\|_2^2 + \mathcal{C}^* t_n^2 + B[T_n^{*2}]}{\left(\frac{C_1}{n^2 h^{4s+1}} + \frac{C_2\|f-f_0\|_2^2}{n} I(\beta \geq s) \right)^{1/2}} \leq -z_{1-\alpha/2}, \quad (28)$$

where z_p denotes the p-quantile of the standard gaussian law $N(0,1)$. So, either $\beta \leq s$, then

$$\frac{-\|f-f_0\|_2^2 + \mathcal{C}^* t_n^2 + B[T_n^{*2}]}{\left(\frac{C_1}{n^2 h^{4s+1}} + \frac{C_2\|f-f_0\|_2^2}{n} I(\beta > s) \right)^{1/2}} \leq \frac{-\mathcal{C}^* \psi_n^2 + \mathcal{C}^* t_n^2 + B[T_n^{*2}]}{n^{-1} h^{-(2s+1/2)}},$$

which is smaller than $-z_{1-\alpha/2}$ for $\psi_n = t_n$ verifying (25). Or, $\beta > s$, then we can solve

$$\frac{-\|f-f_0\|_2^2 + \mathcal{C}^* t_n^2 + B[T_n^{*2}]}{C_4\|f-f_0\|_2/\sqrt{n}} \leq -z_{1-\alpha/2}$$

and (use also (25)) find

$$\begin{aligned}\|f-f_0\|_2 &\geq \frac{C_4 z_{1-\alpha/2}}{2\sqrt{n}} + \frac{1}{2}\sqrt{\frac{C_4^2 z_{1-\alpha/2}^2}{n} + 4(\mathcal{C}^* t_n^2 + B[T_n^{*2}])} \\ &\geq \max\left\{\frac{C_5}{\sqrt{n}}, C_6\left(\frac{C_7}{n^2 h^{4s+1}} + Lh^{2\beta}\right)^{1/2}\right\} \geq \mathcal{C}^{1/2}\psi_n,\end{aligned}$$

for C^* large enough, since $h = h_*$ minimizes the risk $C_7 n^{-2} h^{-(4s+1)} + Lh^{2\beta}$. The upper bounds in (6) are proven. For the lower bounds in (7) see Section 3. ∎

3.2 Supersmooth densities and polynomial noise

Very surprisingly, in this setup, we can estimate d^2 at parametric rate, moreover we can also provide minimax testing at parametric $n^{-1/2}$ rate.

Theorem 7 *The test procedure Δ_n^* defined in (24) for the threshold t_n attains the rate ψ_n and ψ_n is a minimax rate of testing over the class $S(\alpha, r, L)$, where*

$$h = h_* \ll \left(\frac{\log n}{4\alpha}\right)^{-1/r}, \; t_n = \psi_n = n^{-\frac{1}{2}}.$$

Proof. We follow the lines of proof in Theorem 6, using computation from the proof of Theorem 3 in this setup. For the first type error see that

$$E_{f_0}[T_n^{*2}] = o(1)Le^{-2\alpha/h^r} \text{ and } V_{f_0}[T_n^{*2}] = \frac{S\|p_0\|_2^2}{n^2 h^{4s+1}}(1+o(1)),$$

where $p_0 = f_0 * g$. Using Markov's inequality

$$P_{f_0}[T_n^{*2} \geq \mathcal{C}^* t_n^2] \leq \frac{Cn^{-2}h^{-(4s+1)}}{(C^* t_n^2 - o(1)\exp(-2\alpha/h^r))^2} = o(1),$$

since $\max\{n^{-2}h^{-(4s+1)}, \exp(-2\alpha/h^r)\} = o(t_n^2)$.

For an arbitrary density $f \in H_1(\mathcal{C}, \psi_n)$,

$$B[T_n^{*2}] = |\|K_h \star f - f_0\|_2^2 - \|f - f_0\|_2^2| \leq L\exp(-2\alpha/h^r)(1+o(1))$$

and $V_f[T_n^{*2}] \leq C_1\|f-f_0\|_2^2/n =^{Def} v_n$. Indeed, similarly to the proof in (26), we have for some $M > 0$ large enough

$$\Omega_g^2(f-f_0) \leq c_1\|f-f_0\|_2^2 + c_2 e^{-2\alpha M^r}\int_{|u|>M} e^{2\alpha|u|^r}|\Phi(u)|^2 du \leq C_1\|f-f_0\|_2^2,$$

where $C_1 > 0$ depends only on α, r, L and the noise fixed probability density g. Using the asymptotic normality of this U-statistic of order 2, we get

$$P_f[T_n^{*2} < \mathcal{C}^* t_n^2] \leq P_f\left[\frac{T_n^{*2} - E_f[T_n^{*2}]}{\sqrt{V_f[T_n^{*2}]}} \leq \frac{-\|f - f_0\|_2^2 + \mathcal{C}^* t_n^2 + B[T_n^{*2}]}{C_1^{1/2}\|f - f_0\|_2/\sqrt{n}}\right] \leq \frac{\alpha}{2} + o(1),$$

if

$$-C_1^{-1/2}\|f - f_0\|_2\sqrt{n} + C_1^{-1/2}\sqrt{n}(\mathcal{C}^* t_n^2 + B[T_n^{*2}]) \leq -z_{1-\alpha/2}.$$

We actually have

$$-C_1^{-1/2}\|f - f_0\|_2\sqrt{n} + C_1^{-1/2}\sqrt{n}(\mathcal{C}^* t_n^2 + B[T_n^{*2}]) \leq -(\mathcal{C}/C_1)^{1/2} + o(1)$$

which is less than the needed quantile for $\mathcal{C} > \mathcal{C}^*$ large enough. ■

3.3 Sobolev densities and exponential noise

Theorem 8 *The test procedure Δ_n^* defined in (24), for the threshold t_n and for the constant $\mathcal{C}^* = 1$, attains the rate ψ_n and, under Assumption (E), ψ_n is an exact minimax rate of testing over the class $W(\beta, L)$, where*

$$h = h_* = \left(\frac{\log n}{2\gamma} - \frac{2\beta + 1}{2\gamma s}\log\frac{\log n}{2\gamma}\right)^{-1/s},\ t_n = \psi_n = \sqrt{L}\left(\frac{\log n}{2\gamma}\right)^{-\beta/s}.$$

Proof. Again, under the null hypothesis and $h = h_*$

$$E_{f_0}[T_n^{*2}] = Lh^{2\beta}o(1), \quad V_{f_0}[T_n^{*2}] = E_{f_0}[|S_1|^2] \leq c_1\frac{h^{s-1}}{n^2}\exp\left(\frac{4\gamma}{h^s}\right).$$

The first type error can be bounded then:

$$P_{f_0}[T_n^{*2} \geq \mathcal{C}^* t_n^2] \leq \frac{c_1 h^{s-1} n^{-2}\exp(4\gamma/h^s)}{(\mathcal{C}^* t_n^2 - Lh^{2\beta})^2} \leq c_2 h_*^{s+1} = o(1),$$

where we used the facts that $\mathcal{C}^* = 1$, that $t_n^2 \leq Lh^{2\beta}$ and that $h = h_*$ is solution of (23), i.e. $n^{-2}\exp(4\gamma/h^s) = c^{-2}h^{-4\beta-2}$. Under the alternative, if $f \in H_1(\mathcal{C}, \psi_n)$:

$$B_f[T_n^{*2}] \leq Lh^{2\beta}(1 + o(1)), \quad V_f[T_n^{*2}] \leq c_3\frac{h^{2\beta+s-1}}{n}\exp\left(\frac{2\gamma}{h^s}\right) = c_4 h^{4\beta+s},$$

where we used again (23). Then for $\mathcal{C} = \mathcal{C}^*(1 + \delta) > \mathcal{C}^*$, $\delta > 0$, we have

$$\begin{aligned} P_f[T_n^* < \mathcal{C}^* t_n^2] &\leq P_f\left[\frac{E_f[T_n^{*2}] - T_n^{*2}}{\sqrt{V_f[T_n^{*2}]}} \geq \frac{-\mathcal{C}^* t_n^2 - B_f[T_n^{*2}] + \|f - f_0\|_2^2}{\sqrt{c_4}h_*^{2\beta+s/2}}\right] \\ &\leq P_f\left[\frac{E_f[T_n^{*2}] - T_n^{*2}}{\sqrt{V_f[T_n^{*2}]}} \geq \frac{\mathcal{C}^*\delta\psi_n^2 - Lh_*^{2\beta}(1 + o(1))}{\sqrt{c_4}h_*^{2\beta+s/2}}\right] \\ &\leq c_5 h_*^s = o(1), \end{aligned}$$

for $c_5 > 0$ depending on δ and where we used Markov's inequality. ■

3.4 Supersmooth densities and exponential noise

Unknown densities and noise densities are both supersmooth. Nevertheless, there is an essential difference with the case of Sobolev densities and polynomial noise from Subsection 3.1. Nonparametric minimax rates of testing are faster when $r > s$ than in the case $r < s$.

3.4.1 Case $r < s$

Theorem 9 *The test procedure Δ_n^* defined in (24), for the threshold t_n and for the constant $\mathcal{C}^* = 1$, attains the rate ψ_n and, under Assumption (E), ψ_n is an exact minimax rate of testing over the class $S(\alpha, r, L)$, where*

$$h \;= h_* = \; \text{is a solution of } h_*^{r-1-(r-1)_+/2} \exp\left(\frac{2\alpha}{h_*^r} + \frac{2\gamma}{h_*^s}\right) = n(1+o(1)), \quad (29)$$

$$t_n \;= \psi_n \;= \sqrt{L} \exp\left(-\frac{\alpha}{h_*^r}\right).$$

Proof. Under the null hypothesis and for $h = h_*$

$$E_{f_0}[T_n^{*2}] = o(1) L e^{-2\alpha/h^r}, \quad V_{f_0}[T_n^{*2}] \le c_1 \frac{h^{s-1}}{n^2} \exp\left(\frac{4\gamma}{h^s}\right). \quad (30)$$

Then the first type error is bounded by

$$P_{f_0}[T_n^{*2} \ge \mathcal{C}^* t_n^2] \le \frac{c_1 h^{s-1} n^{-2} \exp(4\gamma/h^s)}{(\mathcal{C}^* t_n^2 - L\exp(-2\alpha/h^r))^2} \le c_3 h^{2-2r-(r-1)_+ + s-1} = o(1),$$

where we used the facts that $h = h_*$ is defined by (29) and that $r < s$. Under the alternative, use Theorem 7

$$B_f[T_n^{*2}] \le L e^{-2\alpha/h^r}(1+o(1)), \quad V_f[T_n^{*2}] \le c_1 \frac{h^{s-1}}{n} \exp\left(\frac{2\gamma}{h^s} - \frac{2\alpha}{h^r}\right) + c_2 \frac{h^{s-1}}{n^2} \exp\left(\frac{4\gamma}{h^s}\right).$$

For $\mathcal{C} = \mathcal{C}^*(1+\delta) > \mathcal{C}^*$, $\delta > 0$, use Theorem 5 saying that, for $h = h_*$ defined in (29), $\psi_n^2 = t_n^2$ are of the same order as $\exp(-4\alpha/h^r)$ which is infinitely larger than $\sqrt{V_f[T_n^{*2}]}$ to get

$$\begin{aligned} P_f[T_n^{*2} < \mathcal{C}^* t_n^2] &\le P_f\left[\frac{E_f[T_n^{*2}] - T_n^{*2}}{\sqrt{V_f[T_n^{*2}]}} \ge \frac{-\mathcal{C}^* t_n^2 - B_f[T_n^{*2}] + \|f - f_0\|_2^2}{\sqrt{V_f[T_n^{*2}]}}\right] \\ &\le P_f\left[\frac{E_f[T_n^{*2}] - T_n^{*2}}{\sqrt{V_f[T_n^{*2}]}} \ge \frac{\mathcal{C}^* \delta \psi_n^2 - L\exp(-4\alpha/h^r)(1+o(1))}{\sqrt{V_f[T_n^{*2}]}}\right] \\ &\le \frac{V_f[T_n^{*2}]}{c_4 \exp(-4\alpha/h^r)} = o(1), \end{aligned}$$

by Markov's inequality. ∎

3.4.2 Case $r > s$

Note that no lower bounds are provided for this setup. Nevertheless, we expect the rates to be optimal in the minimax sense.

Theorem 10 *The test procedure Δ_n^* defined in* (24) *for the threshold t_n is a test procedure attaining the rate ψ_n over the class $S(\alpha, r, L)$, (see* (6)*), where*

$$\begin{aligned} h &= h_* = \text{ is a solution of } h_*^{(r-1)/2} \exp\left(\frac{2\alpha}{h_*^r} + \frac{2\gamma}{h_*^s}\right) = n(1+o(1)), \qquad (31) \\ t_n &= \psi_n = \frac{h_*^{(s-1)/2}}{n} \exp\left(\frac{2\gamma}{h_*^s}\right). \end{aligned}$$

Proof. Under the null hypothesis, (30) still holds. For $h = h_*$ defined in (31), t_n is of the order of $h^{(s-1)/2} n^{-1} \exp(2\gamma/h^s)$ and the bias $\exp(-2\alpha/h^r) = c_2 h^{r-s} t_n = o(t_n)$. Thus, the first type error is smaller than $\alpha/2$ for some $\mathcal{C}^*$ large enough.

Under the alternative,

$$B_f[T_n^{*2}] \le L e^{-2\alpha/h^r}(1+o(1)), \quad V_f[T_n^{*2}] \le \frac{4\Omega_g(f-f_0)}{n} + c_2 \frac{h^{s-1}}{n^2} \exp\left(\frac{4\gamma}{h^s}\right)$$

and we can prove as in Theorems 6 and 7 that $4\Omega_g(f-f_0) \le C_1 \|f-f_0\|_2^2$. Then the second type error is bounded by $\alpha/2 + o(1)$ as soon as

$$\frac{-\|f-f_0\|_2^2 + \mathcal{C}^* t_n^2 + L\exp(-2\alpha/h^r)}{(C_1\|f-f_0\|_2^2/n + c_2 h^{s-1} n^{-2} \exp(4\gamma/h^s))^{1/2}} \le -z_{1-\alpha/2}.$$

This is equivalent to

$$\|f-f_0\|_2 \ge \max\left\{\frac{1}{\sqrt{n}}, \frac{h^{(s-1)/2}}{n} e^{2\gamma/h^s} + L e^{-2\alpha/h^r}\right\},$$

since $h = h_*$ defined by (31) minimizes the sum on the right-hand side of the previous inequality. As a result

$$\psi_n = c_3 \frac{h^{(s-1)/2}}{n} e^{2\gamma/h^s} = h^{(s-r)/2} e^{-2\alpha/h^r}$$

which is infinitely smaller than the bias. Note that, the rate is indeed slower than any polynomial n^{-a}, $a > 0$ but faster than any logarithmic rate. ∎

4 Lower bounds

We show in a first part that proofs for minimax lower bounds for the estimation problem of d^2 and for the testing problem in L_2 come down to the same choice of hypotheses and checking similar conditions.

Density \ Noise	Polynomial: $\|u\|^{-s}$	Exponential: $\exp(-\gamma\|u\|^s)$
$f \in W(\beta, L)$, $\beta > 1/2$	$O(1) n^{-\frac{2\beta}{4\beta+4s+1}}$	$\sqrt{L}\,(\log n/(2\gamma))^{-\frac{\beta}{s}}$
$f \in S(\alpha, r, L)$	$O(1) n^{-\frac{1}{2}}$	$r < s: \quad \sqrt{L}\exp\left(-\alpha/h_*^r\right)$ $r > s: \quad O(1)\frac{h_*^{(s-1)/2}}{n}\exp\left(\frac{2\gamma}{h_*^s}\right)$

where h_* is defined in (29) or (31) if $r < s$ or $r > s$, respectively.

Table 2: Testing rates in L_2-norm from noisy data

Let us define

$$\begin{aligned} Rest &:= \inf_{\hat{d}_n^2} \sup_{f \in W(\beta,L)} \varphi_n^{-1} E_f[|\hat{d}_n^2 - d^2|] \\ Rtest &:= \inf_{\Delta_n} \sup_{f \in W(\beta,L)} \left(P_{H_0}(\Delta_n = 1) + P_{H_1(\mathcal{C},\psi_n)}(\Delta_n = 0) \right). \end{aligned}$$

Lemma 1 *Let f_0 and f_1 be two probability densities in the class $W(\beta, L)$, depending on n, and denote by P_0^Y, E_0 and P_1^Y, E_1 the probability measures of our data and the expected value when the true underlying parameters are f_0 and f_1, respectively. If*

a) estimation problem *densities are such that $|\|f_1\|_2^2 - \|f_0\|_2^2| \geq 2\varphi_n$, for some $\varphi_n > 0$,*

a′) test problem *densities are such that $\|f_1 - f_0\|_2 \geq \mathcal{C}\psi_n$, for some $\psi_n > 0$,*

b) *$P_1 \ll P_0$ and there exists $0 < \gamma < 1$ such that*

$$\chi^2(P_0, P_1) :=^{Def} \int \left(\frac{dP_1}{dP_0} - 1 \right)^2 dP_0 \leq \gamma^2$$

then

$$\begin{aligned} Rest &\geq (1-\gamma)(1-\sqrt{\gamma}) && (32) \\ Rtest &\geq (1-\gamma)(1-\sqrt{\gamma}). && (33) \end{aligned}$$

Proof. For the estimation problem we reduce the risk to two hypothesis:

$$Rest \geq \inf_{\hat{d}_n^2} \max_{i=0,1} \varphi_n^{-1} E_{f_i}[|\hat{d}_n^2 - d_i^2|],$$

and then use directly Lemma 4 from Butucea and Tsybakov [6], adapted from Tsybakov [37].

For the testing problem, we choose two hypotheses f_0 the density under H_0 and another density f_1 under H_1 (which implies that $\|f_1 - f_0\|_2 \geq \mathcal{C}\psi_n$, for some $\psi_n > 0$). Then the risk for the test problem becomes

$$
\begin{aligned}
Rtest &\geq \inf_{\Delta_n} \left(P_0(\Delta_n = 1) + (1-\sqrt{\gamma}) P_0 \left(\Delta_n = 0, \frac{dP_1^Y}{dP_0^Y} \geq 1 - \sqrt{\gamma} \right) \right) \\
&\geq (1-\sqrt{\gamma}) P_0 \left(\frac{dP_1^Y}{dP_0^Y} \geq 1 - \sqrt{\gamma} \right) \\
&\geq (1-\sqrt{\gamma}) \left(1 - \frac{1}{\gamma} E_0 \left[\left(\frac{dP_1^Y}{dP_0^Y} - 1 \right)^2 \right] \right) \geq (1-\gamma)(1-\sqrt{\gamma}),
\end{aligned}
$$

if Assumption **b)** holds. ■

We shall use in the proofs the following construction and Lemma 2. Let $0 < \delta < 1$ be small through the remaining proofs of lower bounds. Let f_0 be a density function in the Sobolev class $W(\beta, a(\delta)L)$, respectively, $S(\alpha, r, a(\delta)L)$, where $0 < a(\delta) < 1$ is a constant depending on δ defined for each setup, such that

$$
f_0(x) \geq \frac{c_0}{1 + |x|^2}, \quad \forall x \in \mathbb{R}. \tag{34}
$$

Moreover we want the Fourier transform Φ_0 to have compact support included in $(-2\delta, 2\delta)$.

Let us note immediately that we have a similar property for $f_0^Y = f_0 * g$. Indeed, let $A > 1$ large enough be such that $\int_{-A}^{A} g(x)dx > 1/2$, then

$$
f_0^Y(x) \geq \int_{-A}^{A} f_0(x-y)g(y)dy \geq c_0^Y \min\left\{ \frac{1}{A^2}, \frac{1}{|x|^2} \right\}, \forall x \in \mathbb{R}, \tag{35}
$$

where $c_0^Y > 0$.

Lemma 2 (Lemma 1 in Butucea and Tsybakov [6]) *For any $\delta > 0$ and any $D > 4\delta$ there exists a function $\Phi^G : R \to [0,1]$ such that*

(i) Φ^G is 3 times continuously differentiable on R and its first 3 derivatives are uniformly bounded on R,

(ii) Φ^G is compactly supported on $(\delta, D-\delta)$ and

$$
I_{(2\delta, D-2\delta)}(u) \leq \Phi^G(u) \leq I_{(\delta, D-\delta)}(u),
$$

for all $u \in R$, where $I_A(u)$ denotes the indicator function of the interval A.

Proof of the lower bounds in Theorems 4 and 8. Here we check the assumptions in Lemma 1 and then (32) and (33) imply the needed results. Let us consider the density function f_0 in the class $W(\beta, a(\delta)L)$ for some small $0 < \delta < 1$ such that (34) holds.

Let Φ^G be defined by Lemma 2 with $D = 1$ and the perturbation function H be defined via its Fourier transform

$$\Phi^H(u, h) = \sqrt{\pi L} h^{-\beta} \frac{\Phi^G(|u| - 1/h)}{1 + |u|^{2\beta}},$$

where $\beta > 1/2$ and $h \to 0$ as $n \to \infty$.

Then the second hypothesis function f_1 is defined as follows

$$f_1(x) = f_0(x) + H(x, h), \quad \text{for } h = \left(\frac{\log n}{2\gamma} - \frac{B}{2\gamma s} \log \frac{\log n}{2\gamma} \right)^{-1/s},$$

for some constant $B \in R$ fixed later on. Note that the characteristic functions verify $\Phi_1(u) = \Phi_0(u) + \Phi^H(u, h)$.

Let us see first that f_1 is a probability density function. Indeed, since $\Phi^H(\cdot, h)$ is 3 times continuously differentiable, then via integration by parts we get

$$|H(x, h)| = \left| -\frac{1}{2\pi i x^3} \int e^{-iux} (\Phi^H)'''(u) du \right| \leq \frac{C_H}{1 + |x|^3}, \tag{36}$$

for all $x \in R$, for some constant $C_H > 0$. Note also that

$$\begin{aligned} \|H(\cdot, h)\|_\infty &\leq \frac{1}{2\pi} \int |\Phi^H(u, h)| du \leq \sqrt{\frac{L}{\pi}} \frac{h^{-\beta}}{2} \int_{\delta \leq |u| - 1/h \leq 1 - \delta} \frac{du}{1 + |u|^{2\beta}} \\ &\leq \sqrt{\frac{L}{\pi}} h^{-\beta} \int_{1/h}^{1/h+1} \frac{du}{1 + u^{2\beta}} \leq c h^\beta, \end{aligned} \tag{37}$$

which is $o(1)$ for $\beta > 1/2$ and $h \to 0$. Since f_0 is such that (34) holds, (36) and (37) show that f_1 is a non negative function.

Moreover, $\int f_1(x) dx = \Phi_1(0) = 1$ and f_1 is a probability density function.

Let us check now that it belongs to the class $W(\beta, L)$. Indeed,

$$\begin{aligned} \frac{1}{2\pi} \int |\Phi^H(u, h)|^2 |u|^{2\beta} du &\leq \frac{L}{2} h^{-2\beta} \int_{\delta \leq |u| - 1/h \leq 1 - \delta} \frac{|u|^{2\beta} du}{(1 + |u|^{2\beta})^2} \\ &\leq L h^{-2\beta} \int_{1/h+\delta}^{1/h+1-\delta} \frac{du}{u^{2\beta}} \\ &\leq \frac{L}{2\beta - 1} h^{-2\beta} \left[\left(\frac{1}{h} + \delta \right)^{-2\beta+1} - \left(\frac{1}{h} + 1 - \delta \right)^{-2\beta+1} \right] \\ &\leq L(1 - 2\delta)(1 + o(1)) \leq L(1 - \delta), \end{aligned}$$

for n large enough. So, H belongs to $W(\beta, L(1 - \delta))$, implying that

$$\left\| \frac{1}{2\pi} \int |\Phi_1|^2 |\cdot|^{2\beta} \right\|_2 \leq \left\| \frac{1}{2\pi} \int |\Phi_0|^2 |\cdot|^{2\beta} \right\|_2 + \left\| \frac{1}{2\pi} \int |\Phi^H(\cdot, h)|^2 |\cdot|^{2\beta} \right\|_2 \leq \sqrt{L},$$

for $a(\delta) = (1-\sqrt{1-\delta})^2$. Then $f_1 \in W(\beta, L)$.

Let us check **a)**, respectively **a')** in Lemma 1. Note that Φ_0 and $\Phi^H(\cdot, h)$ have disjoint supports for all $h > 0$ and then by Plancherel formula

$$\left| \|f_1\|_2^2 - \|f_0\|_2^2 \right| = \|f_1 - f_0\|_2^2 = \frac{1}{2\pi} \int |\Phi^H(u,h)|^2 du.$$

Thus, it is enough to deal with

$$\begin{aligned} \|f_1 - f_0\|_2^2 &= \frac{1}{2\pi} \pi L h^{-2\beta} \int \frac{|\Phi^G(|u| - 1/h)|^2}{(1+|u|^{2\beta})^2} du \\ &\geq \frac{L}{2} h^{-2\beta} \int_{1/h+2\delta \leq |u| \leq 1/h+1-2\delta} \frac{du}{(1+|u|^{2\beta})^2} \\ &\geq L h^{-2\beta} \int_{1/h+2\delta}^{1/h+1-2\delta} \frac{du}{(1+u^{2\beta})^2} \\ &\geq L h^{-2\beta} \frac{(1-4\delta)(1+o(1))}{(1+(1/h)^{2\beta})^2} \geq L h^{2\beta}, \end{aligned}$$

for small $0 < \delta < 1$ and large enough n. Note that this construction provides the right constant for testing, but not for the estimation of d^2.

Let us check **b)** of Lemma 1. Note first that $\chi^2(P_0, P_1) \leq C n \chi^2(f_0^Y, f_1^Y)$, for some constant $C > 0$, if $n\chi^2(f_0^Y, f_1^Y)$ is small. We use again the property (35) of f_0^Y:

$$\begin{aligned} n\chi^2(f_0^Y, f_1^Y) &= n \int \frac{(f_1^Y - f_0^Y)^2(y)}{f_0^Y(y)} dy \\ &\leq \frac{n}{c_0^Y} \left(A^2 \int_{|y| \leq A} (H(\cdot, h) \star g)^2(y) dy + \int_{|y|>A} y^2 (H(\cdot, h) \star g)^2(y) dy \right) \\ &\leq \frac{n}{c_0^Y} A^2 \|H(\cdot, h) \star g\|_2^2 + \frac{n}{c_0^Y} \int \left| (\Phi^H(u,h)\Phi^g(u))' \right|^2 du. \end{aligned} \tag{38}$$

On the one hand, let

$$\begin{aligned} T_1 &:= n\|H(\cdot, h) \star g\|_2^2 = \frac{n}{2\pi} \int |\Phi^H(u,h)\Phi^g(u)|^2 du \\ &\leq O(1) n h^{-2\beta} \int_{1/h+\delta \leq |u| \leq 1/h+1-\delta} \frac{\exp(-2\gamma|u|^s)}{(1+|u|^{2\beta})^2} du \\ &\leq O(1) n h^{-2\beta} \int_{1/h}^{\infty} \frac{\exp(-2\gamma u^s)}{(1+u^{2\beta})^2} du \\ &\leq O(1) n h^{2\beta+s-1} \exp\left(-\frac{2\gamma}{h^s}\right), \end{aligned} \tag{39}$$

for $h > 0$ small enough. On the other hand, under the additional Assumption (E)

$$\begin{aligned} T_2 &:= n \int \left|(\Phi^H(u,h)\Phi^g(u))'\right|^2 du \\ &\leq O(1) n h^{-2\beta} \int_{1/h}^{\infty} |u|^{2A} e^{-2\gamma|u|^s} du \\ &\leq O(1) n h^{-2\beta-2A+s-1} \exp\left(-\frac{2\gamma}{h^s}\right), \end{aligned} \tag{40}$$

for some fixed $A \in R$. If we choose some $B < \min\{2\beta + s - 1, -2\beta - 2A + s - 1\}$, we conclude from (38), (39) and (40) that

$$n\chi^2(f_0^Y, f_1^Y) \leq o(1) n h^B \exp\left(-\frac{2\gamma}{h^s}\right) = o(1),$$

by the choice of h. ■

Proof of the lower bounds in Theorems 5 and 9. The proof in this case is very similar to the previous one and it can be easily adapted from Butucea and Tsybakov [6] (L_2 case). Let us choose f_0 such that (34) holds and such that the support of Φ_0 be included in $(-2\delta, 2\delta)$. For Φ^G defined in Lemma 2, let H be defined via its Fourier transform

$$\Phi^H(u,h) = \sqrt{2\pi\alpha r L(d-1)}\, h^{(1-r)/2} e^{(d-1)\alpha/h^r} \exp(-\alpha d|u|^r)\Phi^G(|u|^r - 1/h^r),$$

where $d = \delta^{-1/2}$ and $D = D(\delta) \to \infty$ as $\delta \to 0$, such that $D\delta \to 0$. We know then, that $f_1 = f_0 + H(\cdot, h)$ belongs to $S(\alpha, r, L)$ as soon as $a(\delta) < (1 - e^{-\alpha(d-1)\delta})^2$.

We can actually consider h solution of the equation

$$n \exp\left(-\frac{2\alpha}{h^r} - \frac{2\gamma}{h^s}\right) = \exp(-(\log\log n)^2).$$

Then,

$$\left|\|f_1\|_2^2 - \|f_0\|_2^2\right| = \|f_1 - f_0\|_2^2 \geq L \exp\left(-\frac{2\alpha}{h^r}\right)(1-\sqrt{\delta})[e^{-4\alpha\sqrt{\delta}} - e^{-2\alpha(D-2\delta)/\sqrt{\delta}}](1+o(1)),$$

as $n \to \infty$. It is easy to prove that $L\exp(-2\alpha/h^r) = L\exp(-2\alpha/h_*^r)(1+o(1))$, where h_* is defined in (29). This means we checked **a)** and **a')** in Lemma 1. Following the same ideas as in the previous proof, under the additional Assumption (E), there exists some $B \in R$ fixed such that **b)** of Lemma 1:

$$n\chi^2(f_0^Y, f_1^Y) \leq o(1) n h^B \exp\left(-\frac{2\alpha}{h^r} - \frac{2\gamma}{h^s}\right) \leq o(1)$$

by the choice of h. Again, this proof gives exact minimax testing rates. ■

Proof of the lower bounds in Theorems 2 and 6. This proof is based on a large family of hypotheses. A similar reasoning proves that the same construction is valid for proving lower bounds for both quadratic functional estimation and nonparametric testing in L_2.

Note that this setup includes Theorem 2 for $\beta < s + 1/4$. This is not a contradiction, since the lower bounds here are much slower than the parametric $n^{-1/2}$ rate that the estimator attained, see Theorem 1.

Let θ_j, $j = 1, \ldots, M$, be independent Bernoulli random variables and let Π be the probability measure associated to them. For $h > 0$ small as $n \to \infty$ and for a function H to be defined later, let

$$f_\theta(x) = f_0(x) + \sum_{j=1}^{M} \theta_j h^{\beta+s+1} H_h(x - x_j), \tag{41}$$

where $H_h(\cdot) = 1/hH(\cdot/h)$, $x_j = jh$ and M is an integer such that $M/h = 1 - o(1)$, as $n \to \infty$ and h small. Note that observations Y_i, $i = 1, \ldots, n$, when the underlying density is f_θ, have density

$$f_\theta^Y(x) = f_0^Y(x) + \sum_{j=1}^{M} \theta_j h^{\beta+s+1} G_h(x - x_j), \tag{42}$$

where the function G is defined in Lemma 3 and H is such that

$$\Phi^G(u) = \Phi^H(u)\Phi^g\left(\frac{u}{h}\right). \tag{43}$$

Indeed,

$$(H_h(\cdot - x_j) * g)(x) \quad = \quad H_h * g(x - x_j) = G_h(x - x_j).$$

Using Lemmas 4 and 5, we see that the hypotheses fit into the model, i.e. f_θ are density functions for all θ, belonging to the Sobolev class $W(\beta, L)$ and such that

$$\Pi\left[\|f_\theta - f_0\|_2^2 \geq \mathcal{C} n^{-4\beta/(4\beta+4s+1)}\right] \to 1,$$

as $n \to \infty$, for fixed $\mathcal{C} > 0$.

Lemma 3 *Let the function $G : [-1, 0] \to \mathbb{R}$ be defined by*

$$G(x) = \exp\left(-\frac{1}{1-(4x+3)^2}\right) I(-1 \leq x \leq -1/2) - \exp\left(-\frac{1}{1-(4x+1)^2}\right) I(-1/2 < x \leq 0).$$

Then G is an infinitely differentiable function, such that $\int G(x)dx = 0$ and having all polynomial moments finite. Its Fourier transform is such that

$$|\Phi^G(u)| \leq C_G \exp(-a\sqrt{|u|}), \text{ as } |u| \to \infty,$$

for some positive constants C_G, $a > 0$. Moreover Φ^G is an infinitely differentiable, bounded function.

This construction is based on the function f_{a} in Lepski and Levit [24], p. 133 and the asymptotic behaviour of its Fourier transform follows from the reference therein. All other statements have classical proofs for Fourier and inverse Fourier transforms of functions in L_1 and L_2.

We stress the fact that in this setup, hypotheses functions f_θ belong to $H_1(\mathcal{C}, \psi_n)$ with probability which tends to 1 when $n \to \infty$. In order to bound from below the risk, very small modification is needed in the proof of Lemma 1 that we do not discuss in detail here. The last thing to check is that the distance between resulting models is finite:

$$\begin{aligned} \Delta^2 &:= E_{f_0}\left[\left(\frac{\int \prod_{i=1}^n f_\theta^Y(Y_i)\pi(d\theta) - \prod_{i=1}^n f_0^Y(Y_i)}{\prod_{i=1}^n f_0^Y(Y_i)}\right)^2\right] \\ &= E_{f_0}\left[\left(\int \prod_{i=1}^n \frac{f_\theta^Y}{f_0^Y}(Y_i)\pi(d\theta)\right)^2\right] - 1 \\ &= E_{f_0}\left[\left(\int \prod_{i=1}^n \left(1 + \sum_{j=1}^M \theta_j h^{\beta+s+1} \frac{G_h(Y_i - x_j)}{f_0^Y(Y_i)}\right) \pi(d\theta_j)\right)^2\right] - 1. \end{aligned}$$

Now, call $Y_{i,j}$ those observations Y_i belonging to the support of $G_h(\cdot - x_j)$ and since those intervals are disjoint we write

$$\begin{aligned} \Delta^2 &= E_{f_0}\left[\left(\int \prod_{i=1}^n \prod_{j=1}^M \left(1 + \theta_j h^{\beta+s+1} \frac{G_h(Y_{i,j} - x_j)}{f_0^Y(Y_{i,j})}\right) \pi(d\theta_j)\right)^2\right] - 1 \\ &= \prod_{j=1}^M E_{f_0}\left[\left(\int \prod_{i=1}^n \left(1 + \theta_j h^{\beta+s+1} \frac{G_h(Y_{i,j} - x_j)}{f_0^Y(Y_{i,j})}\right) \pi(d\theta_j)\right)^2\right] - 1 \\ &\leq \prod_{j=1}^M \left\{\frac{1}{2}\left(1 + h^{2\beta+2s+2} E\left[\left(\frac{G_h(Y_{1,j} - x_j)}{f_0^Y(Y_{1,j})}\right)^2\right]\right)^n \right. \\ &\quad \left. + \frac{1}{2}\left(1 - h^{2\beta+2s+2} E\left[\left(\frac{G_h(Y_{1,j} - x_j)}{f_0^Y(Y_{1,j})}\right)^2\right]\right)^n\right\} - 1, \end{aligned}$$

where we used the facts that $(a+b)^2 \leq 2a^2 + 2b^2$ and that $\int G = 0$ giving

$$E\left[\frac{G_h(Y_{1,j} - x_j)}{f_0^Y(Y_{1,j})}\right] = 0.$$

Use Lemma 5 and expressions of M and h in n to get

$$\begin{aligned} \Delta^2 &\leq \left(1 + c_3 n^2 \left(h^{2\beta+2s+2} E\left[\left(\frac{G_h(Y_{1,j} - x_j)}{f_0^Y(Y_{1,j})}\right)^2\right]\right)^2\right)^M - 1 \\ &\leq c_4 M n^2 h^{4\beta+4s+2} \leq c_5. \end{aligned}$$

■

Lemma 4 *For all $\theta \in \{-1,1\}^M$ and for $h>0$ which tends to 0 as defined Theorems 2 and 6, then*

1. *the functions f_θ^Y given by (42), with G defined in Lemma 3 are probability density functions, i.e. non-negative functions of integral equal to 1,*
2. *the functions f_θ given by (41), with H defined by (43) and Lemma 3 are probability density functions, given Assumption (P) and that $\Phi^g(u) \neq 0$ for all $u \in \mathbb{R}$.*

Proof. *1.* It is easy to see that $\int f_\theta^Y(x)dx = 1$, since $\int G(x)dx = 0$ and f_0^Y is a probability density function, positive on R. We have to check that f_θ^Y is non-negative on $[0,1]$. For all $j = 1,\dots,M$ and for x in the support of the function $G_h(\cdot - x_j)$ we have $f_\theta^Y(x) = f_0^Y(x) + \theta_j h^{\beta+s+1} G_h(x-x_j)$. Then

$$f_\theta^Y(x) \geq \inf_{0\leq x\leq 1} f_0^Y(x) - h^{\beta+s}\sup_x |G((x-x_j)/h)| \geq c^Y - o(1) > 0,$$

for n large enough.

2. Let us note first that $\Phi^H(0) = \Phi^G(0)/\Phi^g(0) = 0$, implying that $\int H(x)dx = 0$. Moreover, Φ^H is in L_1 and L_2, uniformly continuous function. Then $\int f_\theta(x)dx = 1$. In order to study its positivity, we use two methods. First, for x small enough we use

$$\begin{aligned} h^{\beta+s+1}|H_h(x)| &\leq h^{\beta+s+1}\frac{1}{2\pi}\int \left|\frac{\Phi^G(hu)}{\Phi^g(u)}\right| du \\ &\leq h^{\beta+s+1}\frac{1}{2\pi}\left(\int_{|u|\leq A} u_0^{-1}|\Phi^G(hu)|du + \int_{|u|>A} c_3|u|^s|\Phi^G(hu)|du\right) \\ &\leq c_4 h^{\beta+s}\int_{|v|\leq Ah}|\Phi^G(v)|dv + c_5 h^\beta \int_{|v|>Ah}|v|^s|\Phi^G(v)|du \leq c_6 h^\beta, \end{aligned} \tag{44}$$

for $A > \max\{A_1, A_2\}$ large enough (see Assumption (P)). For x large, we need a sharper bound that we get using derivability and boundedness properties of Φ^G

$$\begin{aligned} h^{\beta+s+1}|H_h(x)| &= h^{\beta+s+1}\frac{1}{2\pi}\left|\int e^{-ixu}\frac{\Phi^G(hu)}{\Phi^g(u)}du\right| \\ &= h^{\beta+s+1}\left|\left[\frac{\exp(-ixu)}{-2\pi i x}\frac{\Phi^G(hu)}{\Phi^g(u)}\right]_{-\infty}^{\infty} + \frac{1}{2\pi i x}\int e^{-ixu}\frac{\partial}{\partial u}\frac{\Phi^G(hu)}{\Phi^g(u)}du\right| \\ &\leq \frac{h^{\beta+s+1}}{2\pi|x|}\left(\int h\left|\frac{(\Phi^G)'(hu)}{\Phi^g(u)}\right|du + \int\left|\frac{\Phi^G(hu)(\Phi^g(u))'}{(\Phi^g(u))^2}\right|du\right). \end{aligned}$$

We split both integrals as above and obtain

$$\begin{aligned} h\int\left|\frac{(\Phi^G)'(hu)}{\Phi^g(u)}\right|du &\leq h\int_{|u|\leq A} u_0^{-1}|(\Phi^G)'(hu)|du + h\int_{|u|>A} c_7|u|^s|\Phi^G(hu)|du \\ &\leq c_8 + c_9 h^{-s} \leq c_{10}h^{-s}, \end{aligned}$$

respectively, under Assumption (P),

$$\int \left| \frac{\Phi^G(hu)(\Phi^g(u))'}{(\Phi^g(u))^2} \right| du \leq \int_{|u|\leq A} c_{11}|\Phi^G(hu)|du + \int_{|u|>A} c_{12}|u|^{s-1}|\Phi^G(hu)|du$$
$$\leq c_{13}h^{-1} + c_{14}h^{-s} \leq c_{15}h^{-s}.$$

So, for x not equal to 0 we have $h^{\beta+s+1}|H_h(x)| \leq (2\pi|x|)^{-1}h^{\beta+1}(c_{10}+c_{15})$.

This bound is not sufficient, so we repeat integration by parts and, under Assumption (P), we get

$$h^{\beta+s+1}|H_h(x)| \leq c_{16}\frac{h^{\beta+2}}{|x|^2}. \tag{45}$$

Let us go back to f_θ. Whenever x is in $[(j-1)/h, j/h]$ for some $j = 1, \ldots, M$, we apply (44) on the interval and on small neighbouring intervals, respectively (45) for x far enough from x_j. Then

$$\begin{aligned} |f_\theta(x)| &\geq |f_0(x)| - \sum_{k\in\{j,j\pm1\}} h^{\beta+s+1}|H_h(x-x_k)| - \sum_{k=1,|k-j|>1}^{M} h^{\beta+s+1}|H_h(x-x_k)| \\ &\geq c_0^Y - 3c_6h^\beta - \sum_{k=1,|k-j|>1}^{M} \frac{c_{16}h^{\beta+2}}{|k-j|^2h^2} \\ &\geq c_0^Y - c_6h^\beta - c_{17}h^\beta \sum_{k=1}^{M} \frac{1}{k^2} > 0 \end{aligned}$$

for n large enough, since the last sum is finite. For $x<0$ we use $|x-x_j| \geq |x|$ for all $j = 1, \ldots, M$ and

$$|f_\theta(x)| \geq |f_0(x)| - c_{16}h^{\beta+2}\frac{M}{|x|^2} \geq |f_0(x)| - \frac{c_{18}h^{\beta+1}}{|x|^2} > 0,$$

for x in a compact set. For large $|x|$, we apply (35) and integration by parts up to 3rd derivatives of Φ^H using Assumption (P), then

$$|f_\theta(x)| \geq \frac{c_0}{1+|x|^2} - \frac{c_{19}}{|x|^3} > 0.$$

For $x > 1$ we use $|x-x_j| \geq |x-1|$ and a similar reasoning. ∎

Lemma 5 *1. The density functions f_θ given by (41), with H defined by (43) and Lemma 3 are in the Sobolev class for any n large enough;*

2. The density functions f_θ are such that

$$\Pi\left(\|f_\theta - f_0\|_2^2 \geq \mathcal{C}n^{-\frac{4\beta}{4\beta+4s+1}}\right) \to 1, \text{ as } n \to \infty.$$

3. The function G defined in Lemma 4 is such that for all $j = 1, \dots, M$

$$h^{2\beta+2s+2}E\left[\left(\frac{G_h(Y_1 - x_j)}{f_0^Y(Y_1)}\right)^2\right] \le O(1)h^{2\beta+2s+1} = o(1), \text{ as } n \to \infty.$$

Proof. *1.* Let us see that

$$\begin{aligned}
\|f_\theta^{(\beta)} - f_0^{(\beta)}\|^2 &= \sum_{j,k=1}^{M} \theta_j\theta_k h^{2\beta+2s+2} \int H_h^{(\beta)}(x - x_j)H_h^{(\beta)}(x - x_k)dx \\
&= \frac{1}{2\pi}\sum_{j=1}^{M} h^{2\beta+2s+2} \int \Phi^H(hu)\overline{\Phi}^H(hu)|u|^{2\beta}du \\
&\quad + \frac{1}{2\pi}\sum_{j\neq k}^{M} \theta_j\theta_k h^{2\beta+2s+2} \int \Phi^H(hu)\overline{\Phi}^H(hu)|u|^{2\beta}e^{iu(x_j - x_k)}du. \quad (46)
\end{aligned}$$

We prove that the terms in the last sum are bounded by a constant. Indeed,

$$\begin{aligned}
&\frac{1}{2\pi}\sum_{j=1}^{M} h^{2\beta+2s+2}\int \Phi^H(hu)\overline{\Phi}^H(hu)|u|^{2\beta}du \\
\le\ & c_1 h^{2\beta+2s+1}\int \frac{|u|^{2\beta}|\Phi^G(hu)|^2}{|\Phi^g(u)|^2}du \\
\le\ & c_2 h^{2\beta+2s+1}\int_{|u|\le A} |u|^{2\beta}|\Phi^G(hu)|^2 du + c_3 h^{2\beta+2s+1}\int_{|u|>A} |u|^{2\beta+2s}|\Phi^G(hu)|^2 du \\
\le\ & c_4 \int |u|^{2\beta+2s}|\Phi^G(u)|^2 du(1 + c_5 h^{2s}) \le C
\end{aligned}$$

where $A > A_{1,2}$ in Assumption (P) and this is strictly smaller than L if we multiply Φ^G with a constant. Note that Φ^G multiplied by any polynomial is still integrable, which is equivalent to saying that G is an infinitely differentiable function. Moreover, Φ^G and therefore $|\Phi^G|^2 = \Phi^G\overline{\Phi^G}$ are infinitely differentiable functions.

In the second term of (46), we use derivability of order 2 of Φ^g and Assumption (P) as in the proof of Lemma 4, *2.*,

$$\begin{aligned}
&\left|\sum_{j\neq k}^{M}\theta_j\theta_k h^{2\beta+2s+2}\int \Phi^H(hu)\overline{\Phi}^H(hu)|u|^{2\beta}e^{iu(x_j-x_k)}du\right| \\
\leq & \sum_{j,k,j\neq k}^{M} h^{2\beta+2s+2}\left|\left[\frac{e^{iu(x_j-x_k)}}{i(x_j-x_k)}|\Phi^H(hu)|^2|u|^{2\beta}\right]_{-\infty}^{\infty}\right. \\
& \left. +\frac{1}{i(x_j-x_k)}\int e^{iu(x_j-x_k)}\frac{\partial}{\partial u}\frac{|u|^{2\beta}|\Phi^G(hu)|^2}{|\Phi^g(u)|^2}du\right| \\
\leq & \sum_{j,k,j\neq k}^{M}\frac{h^{2\beta+2s+2}}{|x_j-x_k|^2}\int\frac{\partial^2}{\partial u^2}\frac{|u|^{2\beta}|\Phi^G(hu)|^2}{|\Phi^g(u)|^2}du \\
\leq & O(1)h\sum_{j,k,j\neq k}^{M}\frac{1}{|j-k|^2}=c_6h\cdot 2\sum_{k=1}^{M}\frac{M-k}{k^2}\leq C
\end{aligned}$$

and we conclude that f_θ belongs to the class $W(\beta, L)$.

2. Similarly to the above calculations, we write

$$\begin{aligned}
\|f_\theta-f_0\|_2^2 &= \frac{1}{2\pi}\sum_{j=1}^{M}h^{2\beta+2s+2}\int|\Phi^H(hu)|^2du \\
&\quad +\frac{1}{2\pi}\sum_{j,k,j\neq k}^{M}h^{2\beta+2s+2}\theta_j\theta_k\int|\Phi^H(hu)|^2e^{iu(x_j-x_k)}du.
\end{aligned}$$

Note that $E_\Pi[\|f_\theta-f_0\|_2^2]=(2\pi)^{-1}\sum_{j=1}^{M}h^{2\beta+2s+2}\int|\Phi^H(hu)|^2du$.

Using the same splitting technique, (4) and Assumption (P)

$$\begin{aligned}
&\frac{1}{2\pi}\sum_{j=1}^{M}h^{2\beta+2s+2}\int|\Phi^H(hu)|^2du \\
\geq & c_1Mh^{2\beta+2s+2}\left(\int_{|u|>A}|u|^{2s}|\Phi^G(hu)|^2du+\int_{|u|\leq A}|\Phi^G(u)|^2du\right) \\
\geq & c_2Mh^{2\beta+1}\left(\int|u|^{2s}|\Phi^G(u)|^2du+\int|\Phi^G(u)|^2du\right)\geq cn^{-\frac{4\beta}{4\beta+4s+1}}.
\end{aligned}$$

Now, by Chebychev inequality, for $\mathcal{C}=c(1-e_n)$ and for $e_n=(\log n)^{-1}=o(1)$, we have

$$\begin{aligned}
&\Pi\left(\|f_\theta-f_0\|_2^2\geq\mathcal{C}\psi_n^2\right) \\
\geq & \Pi\left(\left|\sum_{j,k,j\neq k}^{M}\theta_j\theta_k h^{2\beta+2s+2}\int|\Phi^H(hu)|^2e^{iu(x_j-x_k)}du\right|\leq ce_n\psi_n^2\right) \\
\geq & 1-\frac{1}{c^2e_n^2\psi_n^4}E_\Pi\left[\left(\left|\sum_{j,k,j\neq k}^{M}\theta_j\theta_k h^{2\beta+2s+2}\int|\Phi^H(hu)|^2e^{iu(x_j-x_k)}du\right|\right)^2\right] \quad (47)
\end{aligned}$$

At last, using again derivatives of order 2 and Assumption (P)

$$
\begin{aligned}
& E_\Pi\left[\left(\left|\sum_{j,k,j\neq k}^{M} \theta_j\theta_k h^{2\beta+2s+2}\int |\Phi^H(hu)|^2 e^{iu(x_j-x_k)}du\right|\right)^2\right] \\
\leq & \sum_{j,k,j\neq k}^{M} h^{4\beta+4s+4}\frac{c_3^2}{|x_j-x_k|^4}\left(\int \frac{\partial^2}{\partial u^2}\left|\Phi^H(hu)\right|^2 du\right)^2 \\
\leq & \sum_{j,k,j\neq k}^{M} h^{4\beta+4s}\frac{c_3^2}{|j-k|^4}\left(\int \frac{\partial^2}{\partial u^2}\frac{\left|\Phi^G(hu)\right|^2}{|\Phi^g(u)|^2}du\right)^2 \\
\leq & c_4 h^{4\beta+2}\sum_{k=1}^{M}\frac{M-k}{k^4} \leq c_5 h^{4\beta+1}.
\end{aligned}
$$

This last term is an $o(e_n^2\psi_n^4)$ and together with (47) it finishes the proof.

3. Use positivity and continuity of f_0^Y to get $\inf_{-1\leq x\leq 2} f_0^Y(x) = c_1^Y > 0$ and obtain

$$
\begin{aligned}
h^{2\beta+2s+2}E\left[\left(\frac{G_h(Y_1-x_j)}{f_0^Y(Y_1)}\right)^2\right] &= h^{2\beta+2s+1}\int_{-1}^{0}\frac{G^2(z)dz}{f_0^Y(x_j+hz)}dx \\
&\leq (c_1^Y)^{-1}h^{2\beta+2s+1}\int G^2(z)dz.
\end{aligned}
$$

■

5 Appendix

Lemma 6 *For all $f\in W(\beta,L)$, $\beta>1/2$, $L>0$, respectively $f\in S(\alpha,r,L)$, $\alpha,r,L>0$, there exists a constant $M^f>0$ depending only on the parameters of the smoothness class such that $\|f\|_2^2\leq M^f$. Moreover, if g is the density of either polynomial or exponential noise, there exists a constant $M^p>0$ depending only on the parameters of the class and on s such that*

$$
\|p\|_2^2 \leq M^p,
$$

and p is at least $(\beta+s-1/2)$-Lipschitz continuous.

Proof. It is easy to use $|\Phi|\leq 1$ and write, e.g. for $f\in W(\beta,L)$ and some $M>0$ large enough

$$
\|f\|_2^2 = \frac{1}{2\pi}\int_{\{|u|\leq M\}\cup\{|u|>M\}}|\Phi(u)|^2du \leq \frac{2M}{2\pi}+\frac{1}{2\pi M^{2\beta}}\int_{|u|>M}|\Phi(u)|^2|u|^{2\beta}du \leq M^f,
$$

where $M^f = M/\pi + LM^{-2\beta}$. Similar reasoning works for all other setups. It is proven in Butucea [5] that p is at least $(\beta+s-1/2)$-Lipschitz. ■

Lemma 7 *If Φ, Φ^g are the characteristic functions of random variables $X_1, \dots, X_n$ and of the noise, respectively, such that $\Phi^g(u) \neq 0$, $\forall u \in R$, then $\Phi/\overline{\Phi^g}$ is a continuous function. Moreover, $\Phi/\overline{\Phi^g}$ is in L_1 and its L_2 norm is uniformly bounded in f, in the following setups:*

1. *if f is in $W(\beta, L)$, see* (2) *and the noise is s-polynomial, as defined in* (4)*, with $\beta \geq s$;*

2. *if f is in $S(\alpha, r, L)$, see* (3) *and the noise is s-polynomial, as defined in* (4)*;*

3. *if f is in $S(\alpha, r, L)$, see* (3) *and the noise is exponential as defined in* (5)*, with $r > s$.*

Proof. Since characteristic functions are uniformly continuous (see Lukacs [28]) and Φ^g is non nul, then $\Phi/\overline{\Phi^g}$ is continuous and integrable on any finite compact set $[-M, M]$, M large. On the other hand, e.g. in the first setup:

$$\int_{|u| \geq M} \left| \frac{\Phi(u)}{\overline{\Phi^g}(u)} \right| du \leq c \int_{|u|>M} |u|^s |\Phi(u)| du$$

and this is finite as soon as f belongs to $W(\beta, L)$ and $\beta \geq s$.

Moreover, $\Phi/\overline{\Phi^g}$ is in L_2 and its L_2 norm is uniformly bounded:

$$\begin{aligned} \|\Phi/\overline{\Phi^g}\|_2^2 &\leq C_1 \int_{|u| \leq M} |\Phi(u)|^2 du + C_2 \int_{|u|>M} |u|^{2s} |\Phi(u)|^2 du \\ &\leq C_1(\beta, L, g) + C_2 M^{2(s-\beta)} \int_{|u|>M} |u|^{2\beta} |\Phi(u)|^2 du \\ &\leq C_1(\beta, L, g) + \frac{C_2}{M^{2(\beta - s)}} 2\pi L = C(\beta, L, g) \end{aligned}$$

and this constant depends only on β, L and the fixed noise density g.

The same can be deduced in a similar way for the remaining setups. ∎

References

[1] Baraud, Y. (2002) Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, **8**, 577–606.

[2] Bickel, P. J. and Ritov, Y. (1988) Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhya Ser. A*, **50**, 381–393.

[3] Birgé, L. and Massart, P. (1995) Estimation of integral functionals of a density. *Ann. Statist.*, **23**, 11–29.

[4] Butucea, C. (2004) Deconvolution of supersmooth densities with smooth noise. *Canad. J. Statist.*, **32**, 181–192.

[5] Butucea, C. (2004) Asymptotic normality of the integrated square error of a density estimator in the convolution model. *SORT*, **28**, 9–26.

[6] Butucea, C. and Tsybakov, A. B. (2004) Sharp optimality for density deconvolution with dominating bias. *Prpublication PMA-898, Universit Paris 6*

[7] Carroll, R. J. and Hall, P. (1988) Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, **83**, 1184–1186.

[8] Comte, F. and Taupin, M.-L. (2003) Penalized contrast estimator for density deconvolution with mixing variables. *Prépublication: 2003-30, Université Paris-Sud*.

[9] Ermakov, M. S. (1994) Minimax nonparametric testing of hypotheses on a distribution density. *Theory Probab. Appl.*, **39**, 396–416.

[10] Ermakov, M. S. (1997) Asymptotic minimaxity of chi-square tests. *Theory Probab. Appl.*, **42**, 589-610.

[11] Fan, J. (1991a) On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, **19**, 1257–1272.

[12] Fan, J. (1991b) Global behavior of deconvolution kernel estimates. *Statist. Sinica*, **1**, 541–551.

[13] Fan, J. (1992) Deconvolution with supersmooth distributions. *Canad. J. Statist.*, **20**, 155–169.

[14] Fromont, M. and Laurent, B. (2004) Adaptive goodness-of-fit tests in a density model. *Prépublication: 2003-74, Université de Paris-Sud*.

[15] Gayraud, G. and Pouet, C. (2001) Minimax testing composite null hypotheses in the discrete regression scheme. *Math. Methods Statist.*, **4**, 375–394.

[16] Gayraud, G. and Pouet, C. (2003) Adaptive minimax testing in the discrete regression scheme. *INSEE, Documents de Travail du CREST, 2003-44*.

[17] Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.*, **14**, 1–16.

[18] Ibragimov, I. A. and Khas'minskii (1991) Asymptotically normal families of distributions and efficient estimation. *Ann. Statist.*, **19**, 1681–1724.

[19] Ingster, Yu. I (1993) Asymptotically minimax hypothesis testing for nonparametric alternatives, I, II, III. *Math. Methods Statist.*, **2**, 85–114, 171–189, 249–268.

[20] Ingster and Suslina (2003) *Nonparametric goodness-of-fit testing under Gaussian models.* Lecture Notes in Statistics, 169, Springer-Verlag, New York.

[21] Kerkyacharian, G. and Picard, D. (1996) Estimating nonquadratic functionals of a density using Haar wavelets. *Ann. Statist.*, **24**, 485–507.

[22] Khoshevnik, Yu. A. and Levit, B. Ya. (1976) On a non parametric ana logue of the information matrix. *Theory Probab. Appl.*, **21**, 738–753.

[23] Laurent, B. (1996) Efficient estimation of integral functionals of a density. *Ann. Statist.*, **24**, 659–681.

[24] Lepski, O. V. and Levit, B. Y. (1998) Adaptive minimax estimation of infinitely differentiable functions. *Mathem. Methods Statist.*, **7**, 123–156.

[25] Lepski, O. V., Nemirovski, A. and Spokoiny, V. G. (1999) On estimation of L_p norm of a regression function. *Probab. Theory Related Fields*, **113**, 221–253.

[26] Lepski, O. V. and Tsybakov, A. B. (2000) Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields*, **117**, 17–48.

[27] Levit, B. Ya. (1978) Asymptotically efficient estimation of nonlinear functionals. *Problems Inform. Transmission*, **14**, 204–209.

[28] Lukacs, E. (1970) *Characteristic Functions.* Hafner Publishing Co., New York.

[29] Matias, C. and Taupin, M.-L. (2004) Minimax estimation of linear functionals in convolution model. *Math. Methods Statist.*, to appear.

[30] Nemirovski, A. (2000) Topics in non-parametric statistics. Lectures on probability theory and statistics, Saint Flour 1998, 1738, Springer-Verlag Berlin Heidelberg.

[31] Pensky, M. and Vidakovic, B. (1999) Adaptive wavelet estimator for nonparametric density deconvolution. it Ann.Statist., **27**, 2033–2053.

[32] Pouet, C. (1999) On testing nonparametric hypotheses for analytic regression functions in Gaussian noise. *Math. Methods Statist.*, **8**, 536–549.

[33] Pouet, C. (2001) An asymptotically optimal test for a parametric set of regression functions against a non-parametric alternative. *J. Statist. Plann. Inference*, **98**, 177–189.

[34] Spokoiny, V.G. (1996) Adaptive hypothesis testing using wavelets. *Ann. Statist.*, **24**, 2477–2498.

[35] Spokoiny, V. G. (1998) Adaptive and spatially adaptive testing of nonparametric hypothesis. *Math. Methods Statist.*, **3**, 245–273.

[36] Tribouley, K. (2000) Adaptive estimation of integrated functionals. *Math. Methods Statist.*, **9**, 19–38.

[37] Tsybakov, A. B. (2004) *Introduction à l'estimation non-paramétrique.* Springer, Berlin-Heidelberg.