IRTG 1792 Discussion Paper 2018-061

PLUG-IN L2-UPPER ERROR BOUNDS IN DECONVOLUTION, FOR A MIXING DENSITY ESTIMATE IN R^d AND FOR ITS DERIVATIVES

Yannis G. Yatracos*

* Cyprus University of Technology, Cyprus

This research was supported by the Deutsche Forschungsgemeinschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

PLUG-IN L_2-UPPER ERROR BOUNDS IN DECONVOLUTION, FOR A MIXING DENSITY ESTIMATE IN R^d AND FOR ITS DERIVATIVES

Yannis G. Yatracos

Faculty of Communication and Media Studies

Cyprus University of Technology

August 28, 2018

Summary

In deconvolution in R^d, $d \geq 1$, with mixing density $p(\in \mathcal{P})$ and kernel h, the mixture density $f_p(\in \mathcal{F}_p)$ can always be estimated with $f_{\hat{p}_n}$, $\hat{p}_n \in \mathcal{P}$, via Minimum Distance Estimation approaches proposed herein, with calculation of $f_{\hat{p}_n}$'s upper L_1-error rate, a_n, in probability or in risk; h is either known or unknown, a_n decreases to zero with n. In applications, a_n is obtained when $\mathcal{P}$ consists either of products of d densities defined on a compact, or L_1 separable densities in R with their differences changing sign at most J times; J is either known or unknown. When h is known and p is $\tilde{q}$-smooth, vanishing outside a compact in R^d, plug-in upper bounds are then also provided for the L_2-error rate of $\hat{p}_n$ and its derivatives, respectively, in probability or in risk; $\tilde{q} \in R^+, d \geq 1$. These L_2-upper bounds depend on h's Fourier transform, $\tilde{h}(\neq 0)$, and have rates $(\log a_n^{-1})^{-N_1}$ and $a_n^{N_2}$, respectively, for h super-smooth and smooth; $N_1 > 0$, $N_2 > 0$. For the typical $a_n \sim (\log n)^{\zeta} \cdot n^{-\delta}$, the former (logarithmic) rate bound is optimal for *any* $\delta > 0$ and the latter misses the optimal rate by the factor $(\log n)^{\xi}$ when $\delta = .5$; $\zeta > 0, \xi > 0$. The exponents N_1 and N_2 appear also in optimal rates and lower error and risk bounds in the deconvolution literature.

1 Introduction

In the deconvolution problem, random vectors Y and X in $R^d,\ d \geq 1$, have densities, respectively, p and f_p and satisfy the equation

$$X = Y + \epsilon; \tag{1}$$

Y is independent of the error ϵ that has density h,

$$f_p(x) = h * p(x) = \int_{R^d} h(x-y)p(y)dy,\ p \in \mathcal{P}, \tag{2}$$

$$\mathcal{F}_{\mathcal{P}} = \mathcal{F}_{\mathcal{P},d} = \{f_p,\ p \in \mathcal{P}\}; \tag{3}$$

$\mathcal{P}$ is any class of densities of interest, " $*$ " denotes convolution. Independent copies $X_1, \ldots, X_n$ of X are observed and the goal is to estimate p, its derivative(s) and calculate the estimation errors. Usually, h is assumed known, with non-vanishing Fourier transform $\tilde{h}$. The classic approach is to estimate p via a kernel estimate of f_p.

Until recently, research has been devoted mainly to the one-dimensional deconvolution problem. However, X-observations in R^d can be used to estimate f_p, e.g., with a kernel estimate; $d > 1$. A Minimum Distance Estimate (*MDE*) $f_{\hat{p}_n}$ with $\hat{p}_n \in \mathcal{P}$ can then be obtained as described in section 3, with calculation of upper L_1-error rates when h is either known or unknown and implementation, as presented in applications. The problem that has not been tackled so far in the literature is to derive "plug-in" upper error and risk bounds for $\hat{p}_n$ and the s-th order mixed partial derivative, $\hat{p}_n^{(s)}$, from the rate of convergence of $f_{\hat{p}_n}$ to f_p.

This problem is addressed herein when $\mathcal{P}$ is a sup-norm compact family of $\tilde{q}$-smooth densities vanishing outside a compact $\mathcal{Y}$ in R^d (see Definition 2.3); $d \geq 1$. The upper bounds in probability for the L_2-errors and their risks are provided for $\hat{p}_n$ and for $\hat{p}_n^{(s)}$ and depend on the L_1-error $||f_{\hat{p}_n} - f_p||_1$, non-vanishing $\tilde{h}$ and the smoothing parameter of a trapezoidal kernel K (Devroye, 1992) that is used as approximation tool.

If $f_{\hat{p}_n}$ is L_1-optimal with respect to some criterion, e.g. minimax, the difference of $\hat{p}_n$'s L_2-error rate from the optimal is not expected to be substantial. For example, if h is super-smooth and $f_{\hat{p}_n}$ converges to f_p in L_1-distance with the typical rate $a_n \sim n^{-\delta} \cdot$

$(\log n)^{\zeta}$ in probability or in risk, it follows from (31) that $\hat{p}_n$ has upper L_2-error rate the optimal, $(\log n)^{-\tilde{q}/k}$, for *any* δ, ζ; k determines the rate of the exponential decay of $\tilde{h}, \delta > 0$, $\zeta \in R$. If h is smooth, for $f_{\hat{p}_n}$'s typical rate and from (32), $\hat{p}_n$ has upper L_2-error rate $[(\log n)^{2\zeta}/n^{2\delta})]^{\frac{\tilde{q}}{2\tilde{q}+2\sum_{i=1}^{d}\beta_i+d}}$, which misses the optimal by the factor $(\log n)^{\xi}$ when $\delta = .5$; $\xi > 0$, $\beta_1, \ldots, \beta_d$ are the exponents determining $\tilde{h}$'s algebraic decay. The exponents in the rates' bounds coincide with those of the lower or optimal rates for the isotropic Hölder and Sobolev classes and for the isotropic and bounded Nikolskii class pointwise, for the mean integrated square error, and in L_u-distance and risk, $2 \leq u \leq \infty$ (Comte and Lacour, 2013, Rebelles, 2015, Lepski and Willer, 2017). The same holds in univariate deconvolution (see, e.g., Fan 1991, 1992, 1993, Hall and Meister, 2007, Meister, 2009, Chapter 2, Lounici and Nickl, 2011).

In multidimensional deconvolution, estimates have been obtained also by Masry (1991) and Youndjé and Wells (2008). For the deconvolution in R, consistent estimates have been provided and, when p is $\tilde{q}$-smooth, optimality of the error rates has been established for smooth and super-smooth h, pointwise and in weighted L_u-distance, among others by Carroll and Hall (1988), Devroye (1989), Stefanski and Carroll (1990), Zhang (1990), Hesse (1995) and Loh and Zhang (1996, 1997), Pensky and Vidakovic (1999); $1 \leq u \leq \infty$. Devroye (1989) showed in particular that one can construct a consistent kernel estimate of p when the set $\{t : \tilde{h}(t) = 0\}$ has Lebesgue measure zero. More recent work includes, among others, Delaigle and Gijbels (2002), Groneboom and Jongbloed (2003), Meister (2006) and Butucea and Tsybakov (2007). Johannes (2009) estimated non-parametrically p when ϵ's distribution in (1) is estimated. Hall and Meister (2007) presented a new estimate for p using *ridging*, "not involving kernels in any way", used also when $\tilde{h}$ has periodic zeros. Meister (2008) proposed also an estimate for p using local polynomials when $\tilde{h}$ has periodic zeros. Under additional assumptions on either p or h, the estimates in Hall and Meister (2007, see page 1542, lines -3, -2) and in Meister (2008, see the Introduction) are optimal but the assumptions and the rates are different.

2 Notation, Definitions and Tools

All the functions used are defined in R^d and are measurable and integrable with real values; $d \geq 1$. The densities are defined with respect to Lebesgue measure. When the domain of integration is R^d, it is omitted. For any function g, its Fourier transform is $\tilde{g}$. The vectors $X,\ Y$ take values, respectively in $\mathcal{X},\ \mathcal{Y}$, which are both sets in R^d. $C,\ c,\ C_1,\ C_2$ denote generic positive constants. For positive a, b, $a \sim b$ means $C_1 b \leq a \leq C_2 b$. Constants $a_n,\ b_n,\ \beta_n,\ \delta_n$ decrease to zero as n increases.

Definition 2.1 (Distances) For densities p_1, p_2 defined in $\mathcal{Y}(\subset R^d)$ their L_u-distance is

$$||p_1 - p_2||_u = [\int_{\mathcal{Y}} |p_1(w) - p_2(w)|^u dw]^{1/u},\ 1 \leq u < \infty.$$

The sup-norm (or L_∞)- distance is

$$||p_1 - p_2||_\infty = \sup_{w \in \mathcal{Y}} |p_1(w) - p_2(w)|.$$

The Hellinger distance is

$$H(p_1, p_2) = [\int_{\mathcal{Y}} (\sqrt{p_1(y)} - \sqrt{p_2(y)})^2 dy]^{1/2}.$$

A well-known inequality between the L_1-distance and Hellinger distance is used:

$$||p_1 - p_2||_1 \leq 2H(p_1, p_2). \tag{4}$$

Notation: If $x = (x_1, \ldots, x_d) \in R^d,\ a \in R$ and $s = (s_1, \ldots, s_d)$ is a d-tuple of non-negative integers,

$$x^s = (x_1^{s_1}, \ldots, x_d^{s_d}),\ xs = x_1 s_1 + \ldots + x_d s_d,\ ax = (ax_1, \ldots, ax_d),\ [s] = s_1 + \ldots + s_d;$$

for $y \in R^d$,

$$|x - y| = \max\{|x_i - y_i|, i = 1, \ldots, d\}.$$

For a real valued function g defined in R^d let $g^{(s)}(x_0)$ denote the $[s]$-th order mixed partial derivative of g at x_0, i.e.

$$g^{(s)}(x_0) = \frac{\partial^{[s]} g(x_0)}{\partial x_1^{s_1} \ldots x_d^{s_d}}.$$

Definition 2.2 The modulus of continuity w of g is a function from R^+ with positive values such that

$$w(\delta) = \sup\{|g(x) - g(y)| : |x - y| \leq \delta\}, \ \delta > 0. \tag{5}$$

If the r-th order mixed partial derivative of g has modulus of continuity w, then

$$|g^{(t)}(x) - g^{(t)}(y)| \leq w(|x - y|), \ [t] = r. \tag{6}$$

Definition 2.3 *Let $\mathcal{P}$ in (2) consist of densities defined on the same compact set $\mathcal{Y}$ ($\subset R^d$), that have all s-mixed order partial derivatives uniformly bounded, $0 \leq [s] \leq q$, with the q-th mixed order derivative having the same and known modulus of continuity w_q.*

When

$$w_q(\delta) = L \cdot \ \delta^\gamma, \ L > 0, \quad 0 < \gamma < 1, \ \tilde{q} = q + \gamma, \tag{7}$$

$\mathcal{P}$ is called $\tilde{q}$-smooth family of densities, ignoring L.

Let $K(x)$ be a symmetric function defined in R^d at least q times continuously differentiable with bounded Fourier transform $\tilde{K}$ having compact support $[-M, M]^d$, $M > 0$, such that for $s \in (R^+)^d$,

$$\int K(x)dx = 1, \qquad \int x^s K(x)dx = 0, \ [s] = 1, \ldots, q, \int (|x|^q + |x|^{q+1})K(x)dx < \infty. \tag{8}$$

Kernel K can be obtained by taking d-fold products of Devroye's trapezoidal kernel (Devroye, 1992) and making smooth enough the linear leg of the trapezoid (Devroye, 2013). For any positive number b_n, let

$$K_n(x) = b_n^{-d} K(x b_n^{-1}), \tag{9}$$

with b_n decreasing to 0 as n increases. If $X_1, X_2, \ldots, X_n$ are independent, identically distributed ($i.i.d.$) vectors in R^d with density f, the kernel estimate of f using K is

$$\hat{f}_n(x) = \frac{1}{n} \sum_{j=1}^{n} K_n(x - X_j). \tag{10}$$

The notion of Total Positivity is introduced from Karlin (1968).

Definition 2.4 *A real function $Q(x, y)$ of two variables ranging over linearly ordered sets $\mathcal{X}$ and $\mathcal{Y}$, respectively, is said to be Totally Positive of order r (abbreviated TP_r) if for all*

$$x_1 < x_2, < \ldots < x_m,\ y_1 < y_2 < \ldots < y_m, \qquad x_i \in \mathcal{X}, y_i \in \mathcal{Y},\ 1 \leq m \leq r, \tag{11}$$

the determinant

$$\begin{vmatrix} Q(x_1, y_1) & Q(x_1, y_2) & \ldots & Q(x_1, y_m) \\ Q(x_2, y_1) & Q(x_2, y_2) & \ldots & Q(x_2, y_m) \\ \ldots & & & \\ Q(x_m, y_1) & Q(x_m, y_2) & \ldots & Q(x_m, y_m) \end{vmatrix} \geq 0. \tag{12}$$

Most often $\mathcal{X}$ and $\mathcal{Y}$ are either intervals of the real line or countable subsets of discrete values along the real line. When r is omitted in TP_r, total positivity holds for any value of r. Many density functions $Q(x, y)$ are totally positive (TP) with respect to a σ-finite measure, with the variable y being a real parameter. Examples include the exponential family, the normal and the non-central t-density (Karlin, 1968, p. 19 and 20).

Proposition 2.1 *(Karlin, 1964, p. 34, 1968, Theorem 3.1 (a), p. 21) Let $Q(x, y)$ be TP_r, let μ denote a σ-finite measure such that $\int_{\mathcal{Y}} Q(x, y)d\mu(y)$ exists for every $x \in \mathcal{X}$ and $\mu(U) > 0$ for each open set U for which $U \cap \mathcal{Y}$ is not empty. Suppose $p(y)$ is bounded, measurable and changes sign $J \leq r - 1$ times. Let*

$$f_p(x) = \int Q(x, y)p(y)d\mu(y),$$

be well defined such that the integral converges absolutely, then $f_p(x)$ changes sign at most J times.

Definition 2.5 *(Vapnik and Cervonenkis, 1971) Given a class $\mathcal{C}$ of subsets of a set V and a finite set U that is subset of V, let $\Delta^{\mathcal{C}}(U)$ be the number of different sets $A \cap U$ for $A \in \mathcal{C}$. Let*

$$m^{\mathcal{C}}(n) = \max\{\Delta^{\mathcal{C}}(U) : U \textit{ has } n \textit{ elements}\},\ n = 1, 2, \ldots,$$

$$v(\mathcal{C}) = \begin{cases} \inf\{n : m^{\mathcal{C}}(n) < 2^n\} \\ \infty, \textit{if } m^{\mathcal{C}}(n) = 2^n \ \textit{ for all } n. \end{cases}$$

The class $\mathcal{C}$ will be called a Vapnik-Cervonenkis (VC) class of exponent $v(\mathcal{C})$ if $v(\mathcal{C}) < \infty$.

3 Estimates $f_{\hat{p}_n}$ with $\hat{p}_n \in \mathcal{P}$ and convergence rates

Let $X_1, \ldots, X_n$ be a sample of d-dimensional vectors from unknown density $g \in \mathcal{G},\ d \geq 1; \mathcal{G}$ is a known family of densities, ρ is a distance for densities.

Definition 3.1 *Let S_n be an estimate of $g(\in \mathcal{G})$.*

a) S_n is uniformly consistent estimate for g in probability, with upper rate of convergece δ_n, if for every $\epsilon > 0$ there is $C(\epsilon)(> 1$ w.l.o.g.) such that

$$\sup_{g \in \mathcal{G}} P_g[\rho(S_n, g) > C(\epsilon)\delta_n] \leq \epsilon,\ \forall n \geq 1; \tag{13}$$

(13) is briefly denoted "S_n has upper ρ-error rate, δ_n, in probability, $\rho(S_n, g) \leq C\delta_n$".

b) The uniform upper risk rate of S_n is δ_n when there is constant $C_U(> 0)$ such that

$$\sup_{g \in \mathcal{G}} E_g \rho(S_n, g) \leq C_U \delta_n,\ \forall n \geq 1; \tag{14}$$

(14) is briefly denoted "S_n has upper ρ-risk rate, δ_n, $E_g\rho(S_n, g) \leq C\delta_n$".

P_g and E_g in (13), (14) denote, respectively, probability and expected value under g which is ommited in the sequel.

For *any* estimate T_n of g in $\mathcal{G}$ with $T_n \notin \mathcal{G}$, a Minimum Distance Estimate (*MDE*) $\hat{g}_n \in \mathcal{G}$ is obtained with the same upper convergence rate as T_n.

Lemma 3.1 *Let $X_1, \ldots, X_n$ be a sample of d-dimensional vectors from unknown density g, element of a known family of densities $\mathcal{G}, \rho$ is a distance; $d \geq 1$. Let T_n be an estimate of g, $T_n \notin \mathcal{G}$, such that the upper ρ-error rate of T_n is δ_n, either in probability or in risk.*

Define MDE $\hat{g}_n \in \mathcal{G}$:

$$\rho(T_n, \hat{g}_n) \leq inf\{\rho(T_n, g^*); g^* \in \mathcal{G}\} + \delta_n. \tag{15}$$

Then, the upper ρ-error rate of $\hat{g}_n$ is $3\delta_n$, either in probability or in risk.

Remark 3.1 *The MDE $\hat{g}_n(\in \mathcal{G})$ always exists, whether achieving the value of the infimum, I_n, in (15) or another value in $(I_n, I_n + a_n]$; $\hat{g}_n$ is not necessarily unique. Lemma 3.1 holds also for any parameter space $\mathcal{G}$ with distance ρ and estimate $T_n \notin \mathcal{G}$.*

In the deconvolution problem (1)-(3), with $\mathcal{F}_{\mathcal{P}}$ replacing $\mathcal{G}$, kernel estimate $\hat{f}_n$ in (10) can be used for T_n in Lemma 3.1 to obtain *MDE* $f_{\hat{p}_n}$. A direct approach to obtain *MDE* $f_{\hat{p}_n}$ is used when $\mathcal{F}_{\mathcal{P}}$ is L_1-totally bounded, first when h is known. The distance ρ in (15) is replaced by a sequence of pseudo-distances ρ_n approximating the L_1 distance as described in Yatracos (1985, 1988).

Proposition 3.1 *In the deconvolution problem (1)-(3), let $X_1, \ldots, X_n$ be a sample from unknown density f_p, $p \in \mathcal{P}$. Assume that $\mathcal{F}_{\mathcal{P},d}$ is L_1 totally bounded and let $N_{\mathcal{F}_{\mathcal{P},d}}(a_n)$ be the smallest number of L_1-balls of radius a_n needed to cover $\mathcal{F}_{\mathcal{P},d}$. Then, a* MDE $f_{\hat{p}_n}$ *can be constructed with upper L_1 error bound in probability*

$$C_1 a_n + C_2 (\frac{\log N_{\mathcal{F}_{\mathcal{P},d}}(a_n)}{n})^{1/2}, \; C_1 > 0, \; C_2 > 0, \tag{16}$$

and upper-L_1-rate of convergence, a_n, to f_p in probability

$$a_n \sim (\frac{\log N_{\mathcal{F}_{\mathcal{P},d}}(a_n)}{n})^{1/2}. \tag{17}$$

The centers of the balls covering $\mathcal{F}_{\mathcal{P},d}$ in Proposition 3.1 are $\mathcal{F}_{\mathcal{P},d}$'s elements and constitute an a_n-L_1-sieve that depends on h and is used to construct $f_{\hat{p}_n}$. The Minimum Distance Estimation method to obtain $f_{\hat{p}_n}$ can be used also when model parameters, like h or the smoothness $\tilde{q}$ are not known. The unknown parameters are included in the *MDE* criterion and the size of the sieve used is increased. An example is provided in the next proposition that is the only one where h is assumed unknown, element of a family of densities $\mathcal{H}$ that is L_1 totally bounded.

Proposition 3.2 *In the deconvolution problem (1)-(3), let $X_1, \ldots, X_n$ be a sample from unknown density f_p, $p \in \mathcal{P}$. Assume that h is unknown, element of a space of densities $\mathcal{H}$ and that $\mathcal{P}$ and $\mathcal{H}$ are both L_1 totally bounded. Let $N_{\mathcal{P}}(a_n)$ and $N_{\mathcal{H}}(\xi_n)$ be, respectively the smallest numbers of L_1-balls of radius a_n and ξ_n needed to cover $\mathcal{P}$ and $\mathcal{H}$. Then, a MDE $\hat{h}_n * \hat{p}_n$ can be constructed with upper-L_1-rate of convergence* $\max\{a_n, \xi_n\}$ *to f_p in probability, with*

$$a_n \sim (\frac{\log N_{\mathcal{P}}(a_n)}{n})^{1/2}, \; \xi_n \sim (\frac{\log N_{\mathcal{H}}(\xi_n)}{n})^{1/2}. \tag{18}$$

Remark 3.2 *Upper error rates as those in Propositions 3.1 and 3.2 can be obtained also under weak dependence, with the mixing coefficient $\phi(p_n)$ appearing under the square-root in (16); with the proper choice of p_n the upper rate is the same with that in the i.i.d. case (Roussas and Yatracos, 1996, 1997).*

Upper rates are now obtained for particular examples of deconvolution problems (1)-(3).

Proposition 3.3 *For the d-dimensional deconvolution problem (1)-(3), assume that Y consists of d* independent *random variables, h is standard multivariate normal $\mathcal{N}(0, I_d)$ and $\mathcal{P}$ is the family of d-products of $\tilde{q}$-smooth densities, each having known support* $[-a, a], a \in R^+$; *I_d is unit matrix in R^d,* $d \geq 1$.

A MDE $f_{\hat{p}_n}$ can be obtained with upper-L_1-rate of convergence a_n in probability,

$$a_n \sim \{\frac{[\log(1/a_n)]^2}{n}\}^{1/2} \sim \frac{\log n}{\sqrt{n}}. \tag{19}$$

A family $\mathcal{P}$ is considered that has not been used often in the literature and a_n is obtained via Yatracos (1988), without using the metric entropy, $\log N_{\mathcal{F}_{\mathcal{P},d}}(a_n)$, that appears in (17).

Proposition 3.4 *For the deconvolution problem (1)-(3) in R, let h be Totally Positive (TP) and let $\mathcal{P}_J$ be a family of bounded and measurable densities that is L_1-separable (to avoid measurability problems), such that for every $p_1,\ p_2$ in $\mathcal{P}_J$ their difference $(p_1 - p_2)$ changes sign at most J times; $0 < J < \infty$. Assume also that the σ-finite measure determined via h and Lebesgue measure satisfy the conditions in Proposition 2.1. Then, a MDE $f_{\hat{p}_n}$ can be obtained with upper-L_1-rate of convergence in probability,*

a) *when J is known,*

$$a_n \sim \frac{(\log n)^{.5}}{n^{.5}}, \tag{20}$$

b) when J is unknown,

$$a_n \sim \frac{m_n^{.5}(\log n)^{.5}}{n^{.5}}, \tag{21}$$

with m_n increasing to infinity as slow as it is wished.

The next Lemma is used to obtain $\hat{p}_n$'s upper rate of convergence from that of $f_{\hat{p}_n}$.

Lemma 3.2 *For the deconvolution problem (1)-(3), if the upper ρ_1-error rate of $f_{\hat{p}_n}$ is a_n, in probability and/or in risk, and for the ρ_2-error of $\hat{p}_n$ holds*

$$\rho_2(\hat{p}_n, p) \leq \lambda_n + \mu_n \rho_1(f_{\hat{p}_n}, f_p), \ \lambda_n > 0, \ \mu_n > 0, \tag{22}$$

then, the upper ρ_2-error rate of $\hat{p}_n$ is $\lambda_n + \mu_n a_n$, respectively, in probability and/or in risk.

4 L_2-upper rates of convergence for $\hat{p}_n, \ \tilde{h} \neq 0$

For the deconvolution problem in R^d, let $X_1, \ldots, X_n$ be *i.i.d.* vectors with values in $\mathcal{X}(\subset R^d)$ and density f_p satisfying (2) with p defined on $\mathcal{Y}(\subset R^d); d \geq 1$

The Assumptions:

($\mathcal{A}1$) h is known, $\tilde{h} \neq 0, \ ||\tilde{h}||_2 < \infty$,

($\mathcal{A}2$) $\mathcal{Y}$ is compact,

($\mathcal{A}3$) $\mathcal{P}$ is the family of $\tilde{q}$-smooth densities (Definition 2.3),

($\mathcal{A}4$) $\mathcal{Y} \subset \mathcal{X} \subset R^d, \ d \geq 1$,

($\mathcal{A}5$) $f_{\hat{p}_n}$ is an estimate of f_p, obtained as described in section 3, with upper L_1-error rate, a_n, in probability and in risk; $\hat{p}_n \in \mathcal{P}$.

Let $\tilde{h}$ and $\tilde{K}_n$ be, respectively, the Fourier transforms of h and K_n (see (9)). Since $\tilde{h} \neq 0$, let ψ_n be the inverse Fourier transform of

$$\tilde{\psi}_n = \frac{\tilde{K}_n}{\tilde{h}}. \tag{23}$$

By the convolution theorem,

$$\psi_n * h = K_n. \tag{24}$$

An upper bound for $||\psi_n||_2$ is obtained. The set $[-M, M]^d$ is the support of $\tilde{K}$.

Lemma 4.1 *Under* ($\mathcal{A}1$),

$$||\psi_n||_2 = C||\tilde{\psi}_n||_2 \leq C \cdot [\int_{[-\frac{M}{b_n}, \frac{M}{b_n}]^d} |\tilde{K}(tb_n)|^2 |\tilde{h}(t)|^{-2} dt \ ]^{1/2} \leq C \cdot \frac{\sup_{t \in [-M/b_n, M/b_n]^d} |\tilde{h}(t)|^{-1}}{b_n^{.5d}}. \tag{25}$$

An upper bound for $||\hat{p}_n - p||_2$ is provided when $\tilde{h} \neq 0$.

Proposition 4.1 *Under assumptions* $(\mathcal{A}1) - (\mathcal{A}5)$,

$$||\hat{p}_n - p||_2 \leq C[b_n^q w_q(b_n) + \frac{\sup_{t\in[-M/b_n, M/b_n]^d} |\tilde{h}(t)|^{-1}}{b_n^{.5d}} ||f_{\hat{p}_n} - f_p||_1]; \tag{26}$$

$[-M, M]^d$ *is* $\tilde{K}$*'s support.*

Careful choice of b_n determines the least upper bound (26). When $\tilde{h}(t)$ varies exponentially as t increases, it determines the upper bound in (26). For algebraic variation of $\tilde{h}(t)$ as t increases, b_n satisfies

$$\frac{b_n^{q+.5d} w_q(b_n)}{\sup_{t\in[-M/b_n, M/b_n]^d} |\tilde{h}(t)|^{-1}} \sim a_n. \tag{27}$$

A small b_n-value satisfying (27) exists and is unique since when b_n decreases to zero, the numerator in the left side of (27) decreases to zero, the denominator increases to infinity and w_q, $\tilde{h}$ are continuous.

Models for h are now presented. Let $0 < C_1 \leq C_2 < \infty$, $|t| = (|t_1|, \ldots, |t_d|)$, $k > 0$, $\alpha_j \geq 0$, $\beta_j > .5, j = 1, \ldots, d$, $\bar{\alpha} = \frac{1}{d}\sum_{j=1}^d \alpha_j$, $\bar{\beta} = \frac{1}{d}\sum_{j=1}^d \beta_j$.

$(\mathcal{M}1)$ h is super-smooth when $\tilde{h} \neq 0$ and for large $|t|$-values, $d\bar{\alpha} > 0$,

$$C_1 e^{-\sum_{j=1}^d \alpha_j |t_j|^k} \Pi_{j=1}^d |t_j|^{\beta_j} \leq |\tilde{h}(t_1, \ldots, t_d)| \leq C_2 e^{-\sum_{j=1}^d \alpha_j |t_j|^k} \Pi_{j=1}^d |t_j|^{\beta_j}. \tag{28}$$

$(\mathcal{M}2)$ h is smooth when $\tilde{h} \neq 0$ and for large $|t|$-values

$$C_1 \Pi_{j=1}^d |t_j|^{-\beta_j} \leq |\tilde{h}(t_1, \ldots, t_d)| \leq C_2 \Pi_{j=1}^d |t_j|^{-\beta_j}, \tag{29}$$

The upper error rates for $||\hat{p}_n - p||_2$ in probability and for $E||\hat{p}_n - p||_2$ are now given explicitly as function of a_n in $(\mathcal{A}5)$ for super-smooth and smooth h, using in (30)-(35) and in Examples 4.1 and 4.2 the brief notations for *a)* and *b)* in Definition 3.1 .

Proposition 4.2 *Assume that* $(\mathcal{A}1) - (\mathcal{A}5)$ *hold.*

i) For super-smooth h *from model* $(\mathcal{M}1)$*, an upper rate in probability on* $\hat{p}_n$*'s* L_2*-error is*

$$||\hat{p}_n - p||_2 \leq C_{\bar{\alpha},d,k,M} \cdot (\log a_n^{-1})^{-q/k} w_q[C(\log a_n^{-1})^{-1/k}]. \tag{30}$$

When $w_q(b_n) = b_n^\gamma, 0 < \gamma < 1, \tilde{q} = q + \gamma,$

$$||\hat{p}_n - p||_2 \le C_{\bar{\alpha},d,k,M} \cdot (\log a_n^{-1})^{-\tilde{q}/k}. \tag{31}$$

The dimension d *affects only constant* $C_{\bar{\alpha},d,k,M}$.

ii) For smooth h *from model* $(\mathcal{M}2)$, *an upper rate on* $||\hat{p}_n - p||_2$ *is obtained when* b_n *satisfies*

$$b_n^q w_q(b_n) \sim \frac{a_n}{b_n^{d\bar{\beta}+.5d}}.$$

When $w_q(b_n) = b_n^\gamma,\ 0 < \gamma < 1,$ *an upper rate in probability on* $\hat{p}_n$*'s* L_2*-error is*

$$||\hat{p}_n - p||_2 \le c_M a_n^{\tilde{q}/(\tilde{q}+d\bar{\beta}+.5d)},\ \tilde{q} = q + \gamma. \tag{32}$$

iii) When $E||f_{\hat{p}_n} - f_p||_1 \le a_n$ *and* $w_q(b_n) = b_n^\gamma$, *the upper rates in (31) and (32) hold also for* $E||\hat{p}_n - p||_2$.

Remark 4.1 *Model* $(\mathcal{M}1)$ *can be enlarged, with* k *in (28) replaced by positive* $k_j, j = 1, \ldots, d$. *Then, upper bounds (30), (31) remain valid with* $\max\{k_1, \ldots, k_d\}$ *replacing* k. *In the proof,* k *of the upper bound in (42) will be replaced by* $\max\{k_1, \ldots, k_d\}$.

Upper rates on the L_2-error and risk of the derivative of $\hat{p}_n$ follow.

Corollary 4.1 *Assume* $(\mathcal{A}1) - (\mathcal{A}5)$ *hold,* δ_n *is the upper bound obtained in Proposition 4.2,* $w_q(b) = b^\gamma, 0 < \gamma < 1,\ \tilde{q} = q + \gamma,\ s = (s_1, \ldots, s_d)$ *is a d-tuple of non-negative integers,* $[s] = s_1 + \ldots + s_d \le q$.

i) If $||\hat{p}_n - p||_2 \le \delta_n$ *in probability, then in probability*

$$||\hat{p}_n^{(s)} - p^{(s)}||_2 \le C \cdot \delta_n^{\frac{\tilde{q}-[s]}{\tilde{q}}}. \tag{33}$$

ii) If the upper rate of $E||\hat{p}_n - p||_2$ *is* δ_n, *then*

$$E||\hat{p}_n^{(s)} - p^{(s)}||_2 \le C \cdot \delta_n^{\frac{\tilde{q}-[s]}{\tilde{q}}}. \tag{34}$$

The next result indicates the reason that, when h is super-smooth, estimates of p and $p^{(s)}$ are frequently minimax optimal.

Corollary 4.2 *Under the assumptions in Proposition 4.2 a) i) and Corollary 4.1 and if* $||f_{\hat{p}_n} - f_p||_1 \sim n^{-\delta}$ *in probability,* $0 < \delta < 1$,

$$||\hat{p}_n^{(s)} - p^{(s)}||_2 \le C_{\bar{\alpha},d,k,M}(\delta \log n)^{-(\tilde{q}-[s])/k},\ [s] \ge 0. \tag{35}$$

If $E||\hat{f}_n - f_p||_1 \sim n^{-\delta}$, *the upper bound in (35) is valid for the risk* $E||\hat{p}_n^{(s)} - p^{(s)}||_2$.

Remark 4.2 *When* $d = 1$, $\hat{p}_n^{(s)}$ *is risk minimax optimal for any* $\delta > 0$ *for the weighted* L_2*-distance (Fan, 1992) and the* L_2*-distance (see, e.g., Meister, 2009). The same holds for* $d > 1$, *(see, e.g. Comte and Lacour, 2013, Theorem 3, Case B).*

We searched the literature for density estimates of f_p. For p defined on a compact subset in R, estimates for location-scale Gaussian mixtures have Hellinger error rates in probability $(\log n)^\zeta/n^\delta,\ 0 < \delta \le .5,\ \zeta > 0$ (Genovese and Wasserman, 2000, Ghosal and van der Vaart, 2001, and Zhang, 2009). From (4) these bounds hold also for L_1-distance and estimates with form $f_{\hat{p}_n}$ can be obtained via Lemma 3.1, with the same upper L_1-error rates. These rates and additional results in the literature, e.g., Ibragimov (2001), as well as (19)-(21) suggest to use $a_n \sim n^{-1/2}(\log n)^\zeta,\ 0 < \zeta$.

Example 4.1 Assume $(\mathcal{A}1) - (\mathcal{A}5)$ with $a_n \sim n^{-1/2}(\log n)^\zeta$ in probability, $d = 1, w_q(b) = b^\gamma, \gamma > 0, \tilde{q} = q + \gamma$. Then:

a) for h the standard normal, $\tilde{h}(t) \sim e^{-t^2}$ for large $|t|$, and from (31), (33) in probability

$$||\hat{p}_n^{(s)} - p^{(s)}||_2 \le C(\log n)^{(\tilde{q}-[s])/2},\ [s] \ge 0.$$

b) for h the Cauchy, $\tilde{h}(t) \sim e^{-|t|}$ for large $|t|$, and from (31), (33) in probability

$$||\hat{p}_n^{(s)} - p^{(s)}||_2 \le C(\log n)^{\tilde{q}-[s]},\ [s] \ge 0.$$

c) for h the exponential, $\tilde{h}(t) \sim |t|^{-\beta}$ for large $|t|$, and from (32) in probability

$$||\hat{p}_n^{(s)} - p^{(s)}||_2 \le C\frac{(\log n)^\xi}{n^{(\tilde{q}-[s])/(2\tilde{q}+2\beta+1)}},\ \xi = \zeta(\tilde{q} - [s])/(\tilde{q} + \beta + .5),\ [s] \ge 0.$$

The bound in *c)* misses by the factor $(\log n)^\xi$ the weighted L_2-minimax rate (Fan, 1992) and the L_2-minimax rate (see, e.g., Meister, 2009).

The bounds in *a)-c)* remain valid when a_n is the risk rate.

Example 4.2 For the d-dimensional deconvolution in Proposition 3.3, the upper rate of convergence in probability of $f_{\hat{p}_n}$ to f_p is $\frac{\log n}{n^{.5}}$. Thus, the upper L_2-rates of convergence in probability to $p^{(s)}$ for super-smooth and smooth h are, respectively, $(\log n)^{(q-[s])/k}$ and $(\log n)^{\frac{2(\tilde{q}-[s])}{2\tilde{q}+2\sum_{i=1}^d \beta_i+d}} \cdot n^{-\frac{q-[s]}{2\tilde{q}+2\sum_{i=1}^d \beta_i+d}}$, $[s] \geq 0$.

Remark 4.3 *When h is smooth, we compare the upper L_2-risk rate herein with that of the lower L_u-risk bound provided by Lepski and Willer (2017, p. 892-895) for the isotropic and bounded Nikolskii class, $\mathcal{N}_{r,d}(\tilde{q}, L)$, in the generalized deconvolution with density of the X's in R^d having form $(1-\alpha)p + \alpha(h * p); 0 \leq \alpha \leq 1,\ d \geq 1,\ 2 \leq u < \infty$, r is d-vector $(u, u, \ldots, u)$, $\tilde{q}$ and L as defined in (7). The rate of the lower bound is $\delta_n^{-\rho(\alpha)}$; $\rho(\alpha)$ depends on parameters $\beta(\alpha)$ and $\omega(\alpha)$ and on whether a parameter $\kappa_\alpha(u)$ is larger than $u \cdot \omega(\alpha)$ or not. With our notation and for $\alpha = 1$ that corresponds to the problem herein,*

$$\beta(1) = \frac{\tilde{q}}{2\sum_{j=1}^d \beta_j + d}, \qquad \omega(1) = \frac{u\tilde{q}}{2\sum_{j=1}^d \beta_j + d}$$

and

$$\kappa_1(u) = \omega(1)[2 + \frac{1}{\beta(1)}] - u = \frac{2u\tilde{q}}{2\sum_{j=1}^d \beta_j + d} = 2\omega(1).$$

Since $\kappa_1(u)$ is positive and $\kappa_1(u) \leq u\omega(1)$, for $u \geq 2$,

$$\rho(1) = \frac{\beta(1)}{2\beta(1)+1} = \frac{\tilde{q}}{2\sum_{j=1}^d \beta_j + d}/(2\frac{\tilde{q}}{2\sum_{j=1}^d \beta_j + d} + 1) = \frac{\tilde{q}}{2\tilde{q} + 2\sum_{j=1}^d \beta_j + d}.$$

The rate of the L_u-lower bound is $n^{-\frac{\tilde{q}}{2\tilde{q}+2\sum_{j=1}^d \beta_j+d}}$, $2 \leq u < \infty$. When $a_n \sim (\log n)^\zeta n^{-.5}$, the rate of the plug-in upper L_2-error bound herein is $[\frac{(\log n)^{2\zeta}}{n}]^{-\frac{\tilde{q}}{2\tilde{q}+2\sum_{j=1}^d \beta_j+d}}$, missing the lower bound by a power of $\log n$. However, the exponents in both bounds coincide.

5 Appendix

Proof of Lemma 3.1: From (15), for the estimate $\hat{g}_n \in \mathcal{G}$ it holds

$$\rho(\hat{g}_n, g) \leq \rho(\hat{g}_n, T_n) + \rho(T_n, g) \leq inf\{\rho(T_n, g^*); g^* \in \mathcal{G}\} + \delta_n + \rho(T_n, g) \leq 2\rho(T_n, g) + \delta_n. \quad (36)$$

When T_n's risk is bounded by δ_n, it follows from (15) that

$$\sup_{g \in \mathcal{G}} E\rho(\hat{g}_n, g) \leq 3\delta_n. \quad .$$

For the upper rate in probability, for $\epsilon > 0$, let $C(\epsilon)(> 1,)$ be the constant in (13) such that

$$\sup_{g \in G} P[\rho(T_n, g) > C(\epsilon)\delta_n] \leq \epsilon.$$

Then, from (36),

$$\sup_{g \in G} P[\rho(\hat{g}_n, g) > C(\epsilon)3\delta_n] \leq \sup_{g \in G} P[2\rho(T_n, g) + \delta_n > C(\epsilon)3\delta_n] \leq \sup_{g \in G} P[\rho(T_n, g) > C(\epsilon)\delta_n]. \square$$

Proof of Proposition 3.1: Follows from Yatracos (1985, Theorem 1). $\square$

Proof of Proposition 3.2: Let $h_1^*, \ldots, h_{N_{\mathcal{H}}(\xi_n)}^*$ be a ξ_n-L_1-sieve for $\mathcal{H}$ and $p_1^*, \ldots, p_{N_{\mathcal{P}}(a_n)}^*$ be a a_n-L_1-sieve for $\mathcal{P}$. For $h \in \mathcal{H},\ p \in \mathcal{P}$, let $h_i^*,\ p_k^*$ be such that

$$||h - h_i^*||_1 \leq \xi_n,\ ||p_1 - p_k^*||_1 \leq a_n. \tag{37}$$

Then, using (37) and Young's inequality it follows that,

$$||h * p - h_i^* * p_k^*||_1 \leq ||h * p - h_i^* * p||_1 + ||h_i^* * p - h_i^* * p_k^*||_1 \leq \xi_n + a_n.$$

Thus,

$$\{h_i^* * p_k^*, 1 \leq i \leq N_{\mathcal{H}}(\xi_n), 1 \leq k \leq N_{\mathcal{P}}(a_n)\}$$

is a $(a_n + \xi_n)$-L_1-sieve with cardinality $N_{\mathcal{P}}(a_n) \cdot N_{\mathcal{H}}(\xi_n)$ in the space

$$\{h * p,\ h \in \mathcal{H}, p \in \mathcal{P}\}.$$

Thus, from (16) the upper bound in probability of the *MDE* $\hat{h}_n * \hat{p}_n$ is

$$C_1(a_n+\xi_n)+C_2(\frac{\log[N_{\mathcal{P}}(a_n) \cdot N_{\mathcal{H}}(\xi_n)]}{n})^{1/2} \leq C[a_n+\xi_n+(\frac{\log N_{\mathcal{P}}(a_n)}{n})^{1/2}+(\frac{\log N_{\mathcal{H}}(\xi_n)}{n})^{1/2}]. \quad \square$$

Proof of Proposition 3.3: Since $\mathcal{P}$ is sup-norm compact (Lorentz, 1986, p. 153), it is also L_1-totally bounded and by Young's inequality, $\mathcal{F}_{\mathcal{P},d}$ is also L_1-totally bounded. Thus, for every $a_n > 0$ there is a Ca_n-L_1-sieve of densities in $\mathcal{F}_{\mathcal{P},d}$. An upper bound for the $\log N_{\mathcal{F}_{\mathcal{P},d}}(a_n)$ is obtained using the a_n-L_1-sieve for p continuous on $[-a, a]$, with the logarithm of the sieve's cardinality bounded by $C_1[\log(1/a_n)]^2$ (Ghosal and van der Vaart, 2001, Theorem 3.1, p. 1240, with known a, $\sigma = 1$ and $\gamma = .5$). In every a_n-L_1 ball with center in this sieve, replace the center by a density in $\mathcal{F}_{\mathcal{P},1}$ from the same ball, if it exists. The so-obtained densities are a $(2a_n)$-L_1 sieve for $\mathcal{F}_{\mathcal{P},1}$ with cardinality bounded by $C_1[\log(1/a_n)]^2$. Thus, d-products of these densities are a C_2a_n-L_1 sieve of $\mathcal{F}_{\mathcal{P},d}$ and

$$\log N_{\mathcal{F}_{\mathcal{P},d}}(C_2a_n) \leq c[\log(1/a_n)]^2. \tag{38}$$

The rate (19) follows from Proposition 3.1. $\square$.

Proof of Proposition 3.4: *a)* Consider the L_1-separable subset $\mathcal{P}_J^* = \{p_1^*, p_2^*, \ldots, p_n^*, \ldots, \}$ of $\mathcal{P}_J$. For every $p \in \mathcal{P}_J$ denote by F_p the probability measure with density f_p. For $\beta_n = \frac{(\log n)^{.5}}{n^{.5}}$, there is $p^* \in \mathcal{P}_J^*$ such that $||p - p^*||_1 \leq \beta_n$ and $||f_p - f_{p^*}||_1 \leq \beta_n$. If

$$\mathcal{S}_J = \{\ \{x : h * p_i^*(x) > h * p_j^*(x)\},\ i \neq j\}, \tag{39}$$

then, for $p_1,\ p_2 \in \mathcal{P}_J$,

$$||f_{p_1} - f_{p_2}||_1 \leq 2\beta_n + ||f_{p_1^*} - f_{p_2^*}||_1 \leq 2\beta_n + 2\sup_{A \in \mathcal{S}_J} |F_{p_1^*}(A) - F_{p_2^*}(A)| \leq 6\beta_n + 2\sup_{A \in \mathcal{S}_J} |F_{p_1}(A) - F_{p_2}(A)|. \tag{40}$$

By Total Positivity of h for any r and from Proposition 2.1 for every $i \neq j, h * (p_i - p_j)$ has at most J changes of sign and the sets in $\mathcal{S}_J$ are unions of at most J disjoint intervals, thus $\mathcal{S}_J$ is VC class with exponent $2J + 1$. From Yatracos (1988, Theorem 2, p. 287, with $a_n = \beta_n,\ l_k = 2,\ v_k = 2J + 1,\ F_{a_k} = \mathcal{S}_J$) it follows that the upper L_1 rate of convergence of $f_{\hat{p}_n}$ is

$$a_n \sim \frac{(\log n)^{.5}}{n^{.5}}.$$

b) When J is unknown, consider L_1-separable families of densities $\mathcal{P}_I$ with the same properties as $\mathcal{P}_{\mathcal{J}}$ and I the maximum number of sign changes for the densities' differences; $I \geq 1$. Observe that $\mathcal{P}_I \subset \mathcal{P}_L,\ I \leq L$. Let I_n increase to infinity with n and assume *w.l.o.g* that it takes integer values. For $n \geq n_0,\ \mathcal{P}_J \subset \mathcal{P}_{I_n}$ and for $p_1, p_2 \in \mathcal{P}_{I_n}$, (40) holds with $\beta_n = \frac{(2I_n+1)^{.25}(\log n)^{.5}}{n^{.5}}$ and S_J replaced by S_{I_n} with VC-exponent $2I_n + 1$. From Yatracos (1988, Theorem 2) it follows that the upper L_1 rate of convergence of $f_{\hat{p}_n}$ is

$$a_n \sim \frac{(2I_n + 1)^{.5}(\log n)^{.5}}{n^{.5}}. \qquad \square$$

Proof of Lemma 3.2: By taking expected values in both sides of (22) and the supremum over all $p \in \mathcal{P}$ the upper bound follows. For the upper bound in probability, let $\epsilon > 0$, and let $C(\epsilon)(>1)$ be such that

$$\sup_{p \in \mathcal{P}} P[\rho_1(f_{\hat{p}_n}, f_p) > C(\epsilon)\delta_n] \leq \epsilon.$$

Then,

$$P[\rho_2(\hat{p}_n, p) > C(\epsilon)(\lambda_n + \mu_n\delta_n)] \leq P[\lambda_n + \mu_n\rho_1(f_{\hat{p}_n}, f_p) > C(\epsilon)(\lambda_n + \mu_n\delta_n)] \leq P[\rho_1(f_{\hat{p}_n}, f_p) > C(\epsilon)\delta_n]$$

and the result follows by taking the supremum for $p \in \mathcal{P}$. $\square$

Lemma 5.1 *Let g be a function defined on a set $\mathcal{Y}$ in R^d that has all s-mixed order partial derivatives uniformly bounded for $0 \leq [s] \leq q$, with the q-th derivative having modulus of continuity w_q. Then, for the kernel K satisfying (8), K_n defined in (9) and $\mathcal{Y}$ compact,*

$$||g - K_n * g||_u \leq c b_n^q w_q(b_n), \ c > 0, \ u \geq 1. \tag{41}$$

Proof of Lemma 5.1: The result follows from Yatracos (1989, p. 173, Proposition 1). $\square$

Proof of Lemma 4.1: For the Fourier transform $\tilde{K}_n(x)$ it holds,

$$\tilde{K}_n(x) = C\int e^{-ixy} b_n^{-d} K(y/b_n)dy = C\int e^{i(xb_n)yb_n^{-1}} K(yb_n^{-1})d(yb_n^{-1}) = C\tilde{K}(xb_n).$$

Boundedness of $\tilde{K}$ and Parseval's identity imply that

$$||\psi_n||_2 = C||\tilde{\psi}_n||_2 = C[\int_{[-M/b_n, M/b_n]^d} \frac{|\tilde{K}(b_n t)|^2}{|\tilde{h}(t)|^2} dt]^{.5} \leq C\frac{\sup_{t \in [-M/b_n, M/b_n]^d} |\tilde{h}(t)|^{-1}}{b_n^{.5d}}. \quad \square$$

Proof of Proposition 4.1:

$$[\int_{\mathcal{Y}} |\hat{p}_n(y) - p(y)|^2 dy]^{1/2} \leq [\int_{\mathcal{Y}} |\hat{p}_n(y) - K_n * \hat{p}_n(y)|^2 dy]^{1/2}$$

$$+[\int_{\mathcal{X}} |K_n * \hat{p}_n(x) - K_n * p(x)|^2 dx]^{1/2} + [\int_{\mathcal{Y}} |K_n * p(y) - p(y)|^2 dy]^{1/2}$$

$$\leq Cb_n^q w_q(b_n) + ||\psi_n * h * (\hat{p}_n - p)||_2 \leq Cb_n^q w_q(b_n) + ||\psi_n||_2 \cdot ||f_{\hat{p}_n} - f_p||_1.$$

The first inequality is due to the triangular property of the $||\cdot||_2$-distance and to $\mathcal{Y} \subset \mathcal{X}$. The second inequality is due to Lemma 5.1 and (24). The third inequality follows from

Young's inequality for convolutions. The result follows from Lemma 4.1. □

Proof of Proposition 4.2: i) When h follows the super-smooth model (28), the second term in the upper bound (26) has an exponential rate but the first term decreases at algebraic rate. Since

$$\sup_{t\in[-M/b_n,M/b_n]^d} |\tilde{h}(t)|^{-1} \leq C \cdot e^{\sum_{j=1}^d \alpha_j M^k b_n^{-k}} \leq C \cdot e^{d\bar{\alpha}M^k b_n^{-k}}, \tag{42}$$

the second term in upper bound (26) converges to zero as n increases if

$$\lim_{n\to\infty} \frac{\exp\{d\bar{\alpha}M^k b_n^{-k}\}}{b_n^{.5d}} a_n = 0 \iff \lim_{n\to\infty} d\bar{\alpha}M^k b_n^{-k} - .5d\log b_n - \log a_n^{-1} = -\infty. \tag{43}$$

Choosing

$$b_n^k = \frac{4d\bar{\alpha}M^k}{\log a_n^{-1}} \text{ or } b_n = \frac{(4d\bar{\alpha})^{1/k}M}{(\log a_n^{-1})^{1/k}}$$

(43) holds and from Lemma 3.2 the terms in upper bound (26) are

$$b_n^q w_q(b_n) \sim (\log a_n^{-1})^{-q/k} w_q[C(\log a_n^{-1})^{-1/k}], \tag{44}$$

$$\frac{\sup_{t\in[-M/b_n,M/b_n]^d} |\tilde{h}(t)|^{-1}}{b_n^{.5d}} a_n \leq a_n^{3/4}(\log a_n^{-1})^{5d/k}, \tag{45}$$

with (45) converging faster to 0 as n increases than (44).

When $w_q(b_n) \sim b_n^\gamma$, (44) determines the upper convergence rate $(\log a_n^{-1})^{-(q+\gamma)/k}$.

ii) When h follows the smooth model (29), both terms in upper bound (26) have algebraic rate. Since

$$\sup_{t\in[-M/b_n,M/b_n]^d} |\tilde{h}(t)|^{-1} \leq C \cdot (\frac{M}{b_n})^{d\bar{\beta}}$$

and from Lemma 3.2 we choose b_n such that

$$b_n^q w_q(b_n) \sim \frac{a_n}{b_n^{d\bar{\beta}+.5d}}.$$

When $w_q(b_n) \sim b_n^\gamma,\ \tilde{q} = q + \gamma$,

$$b_n^{\tilde{q}} \sim \frac{1}{b_n^{d\bar{\beta}} \cdot b_n^{.5d}} a_n \text{ or } b_n \sim a_n^{1/(\tilde{q}+d\bar{\beta}+.5d)} \tag{46}$$

and

$$||\hat{p}_n - p||_2 \leq c_M a_n^{\tilde{q}/(\tilde{q}+d\bar{\beta}+.5d)}.$$

iii) Follows using the approach in *i)* and *ii)*. $\square$

Proof of Corollary 4.1: Follows along the lines in Yatracos (1989), Proposition 2, p. 174 and Remarks (i) and (ii) pages 174, 175, since p and $p^{(s)}$ have compact support. $\square$

Proof of Corollary 4.2: The bounds are obtained by plugging $a_n \sim n^{-\delta}$ in the bounds in Proposition 4.2 *a) i)* and in (33) and (34). For densities in R, optimality for any $\delta > 0$ follows from the optimal rates in Fan (1991,1992, 1993). $\square$

References

[1] Butucea, C. and Tsybakov, A. B. (2007) Sharp optimality in density deconvolution with dominationg bias . I, II *Theory. Probab. Appl.* **52**, 24-39, 237-249.

[2] Carrol, R. J. and Hall, P. (1988) Optimal rates of convergence for deconvolving a density. *JASA* **83**, 1184-1186.

[3] Comte, F. and Lacour, C. (2013) Anisotropic adaptive kernel deconvolution. *Ann. Inst. H. Poincaré Probab. Statist.* **49**, 569-609.

[4] Delaigle, A. and Gijbels, I. (2002). Estimation of integrated squared density derivatives from a contaminated sample. *JRSS B* **64**, 869-886.

[5] Devroye, L. (2013) Personal Communication

[6] Devroye, L. (1992) A note on the usefulness of superkernels in density estimation. *Ann. Statist.* **20**, 2037-2056.

[7] Devroye, L. (1989) Consistent deconvolution in density estimation.*Canad. J. Statist* **17**, 235-239.

[8] Fan, J. (1993) Adaptive local one-dimensional sub-problems with application to a deconvolution problem. *Ann. Statist.* **21**, 600-610.

[9] Fan, J. (1992) Deconvolution with supersmooth distributions. *Can. J. Stat.* **20**, **2**, 155-169.

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

[11] Genovese, C. R. and Wasserman, L. (2000) Rates of convergence for the Gaussian mixture sieve. *Ann. Statist.* **28**, 1105-1127.

[12] Ghosal, S. and van der Vaart, A. W. (2001) Entropy and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29**, 1233-1263.

[13] Groeneboom, P. and Jongbloed, G. (2003) Density estimation in the uniform deconvolution model. *Statist. Nederlandica* **57**, 136-157.

[14] Hall, P. and Meister, A. (2007) A ridge-parameter approach to deconvolution. *Ann. Stat.* **35**, 1535-1558.

[15] Hesse, C. H. (1995) Deconvolving a density from partially contaminated observations. *J. Multivariate Anal.* **55**, 246-260.

[16] Ibragimov, I. A. (2001) Estimation of analytic functions. *State of the Art in Probability and Statistics. Festschrift for W. R. van Zwet.* IMS, Hayward.

[17] Johannes, J. (2009) Deconvolution with unknown error distribution. *Ann. Stat.* **37**, 2301-2323.

[18] Karlin, S. (1968) *Total Positivity Vol. I* Stanford University Press, Stanford California.

[19] Karlin, S. (1964) Total Positivity, Absorption Probabilities and Applications. *Trans. Amer. Math. Soc.* **111**, 33-107.

[20] Lepski, O. V. and Willer, T. (2017) Lower bounds in the convolution structure density model. *Bernoulli* **23**, 884-926.

[21] Loh, W- L. and Zhang, C.-H. (1996) Global properties of kernel estimators for mixing densities in exponential family models for discrete variables. *Statist. Sinica* **6**, 561-678.

[22] Loh, W- L. and Zhang, C.-H. (1997) Estimating mixing densities in exponential family models for discrete variables. *Scand. J. Statist.* **24**, 15-32.

[23] Lorentz, G. G. (1986) *Approximation of Functions.* Chelsea Publishing Company. New York.

[24] Lounici, K. and Nickl, R. (2011) Global uniform risk bound for wavelet deconvolution estimators. *Ann. Statist.* **39**, 201-231.

[25] Masry, E. (1991) Multivariate probability density deconvolution for stationary random processes . *IEEE Trans. Inform. Theory*, **37**, 1105-1115.

[26] Meister, A. (2009) Deconvolution problems in nonparametric statistics. *Lecture Notes in Statistics*, **193**, Springer-Verlag, Berlin.

[27] Meister, A. (2008). Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Problems* **24**, **1**, 015003.

[28] Meister, A. (2006). Density estimation with normal measurement error with unknown variance. *Statist. Sinica* **16**, 195-211.

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

[30] Rebelles, G. (2015) Structural adaptive deconvolution under L_p-losses. *arXiv:1504.06246v2*

[31] Roussas, G. G. and Yatracos, Y.G. (1997) Minimum distance estimates with rates under ϕ-mixing. *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics,* p. 337-345. Editors: D. Pollard, E. Torgersen, G. L. Yang. Springer, New York.

[32] Roussas, G. G. and Yatracos, Y.G. (1996) Minimum distance regression-type estimates with rates under weak dependence. *Ann. Inst. Stat. Math.* **48, 2,** 267-281.

[33] Shapiro (1969) *Smoothing and approximation of functions.* Van Nostrand Reinhold Company, New York.

[34] Stefanski, L. A. and Carroll, R. J. (1990) Deconvoluting kernel density estimators. *Statistics* **21**, 169-184.

[35] Vapnik, V. N. and Cervonenkis, A. Y. (1971) On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.*, **16**, 264-280.

[36] Yatracos, Y. G. (1989) On the estimation of the derivatives of a function with the derivatives of an estimate. *J. of Multivariate Analysis* **28**, 172-175.

[37] Yatracos, Y. G. (1988) A note on L_1 consistent estimation. *Can. J. Stat.* **16**, 283-292.

[38] Yatracos, Y. G. (1985) Rates of convergence of minimum distance estimators and Kolmogorov's entropy. *Ann. Statist.* **13**, 768-774.

[39] Youndjé, E. and Wells, M. T. (2008) Optimal bandwidth selection for multivariate kernel deconvolution density estimation. *TEST* **17**, 138-162.

[40] Zhang, C.-H. (2009) Generalized maximum likelihood estimation of normal mixture densities. *Statist. Sinica* **19**, 1297-1318.

[41] Zhang, C.-H. (1990) Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806-831.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

001 "Data Driven Value-at-Risk Forecasting using a SVR-GARCH-KDE Hybrid" by Marius Lux, Wolfgang Karl Härdle and Stefan Lessmann, January 2018.
002 "Nonparametric Variable Selection and Its Application to Additive Models" by Zheng-Hui Feng, Lu Lin, Ruo-Qing Zhu asnd Li-Xing Zhu, January 2018.
003 "Systemic Risk in Global Volatility Spillover Networks: Evidence from Option-implied Volatility Indices " by Zihui Yang and Yinggang Zhou, January 2018.
004 "Pricing Cryptocurrency options: the case of CRIX and Bitcoin" by Cathy YH Chen, Wolfgang Karl Härdle, Ai Jun Hou and Weining Wang, January 2018.
005 "Testing for bubbles in cryptocurrencies with time-varying volatility" by Christian M. Hafner, January 2018.
006 "A Note on Cryptocurrencies and Currency Competition" by Anna Almosova, January 2018.
007 "Knowing me, knowing you: inventor mobility and the formation of technology-oriented alliances" by Stefan Wagner and Martin C. Goossen, February 2018.
008 "A Monetary Model of Blockchain" by Anna Almosova, February 2018.
009 "Deregulated day-ahead electricity markets in Southeast Europe: Price forecasting and comparative structural analysis" by Antanina Hryshchuk, Stefan Lessmann, February 2018.
010 "How Sensitive are Tail-related Risk Measures in a Contamination Neighbourhood?" by Wolfgang Karl Härdle, Chengxiu Ling, February 2018.
011 "How to Measure a Performance of a Collaborative Research Centre" by Alona Zharova, Janine Tellinger-Rice, Wolfgang Karl Härdle, February 2018.
012 "Targeting customers for profit: An ensemble learning framework to support marketing decision making" by Stefan Lessmann, Kristof Coussement, Koen W. De Bock, Johannes Haupt, February 2018.
013 "Improving Crime Count Forecasts Using Twitter and Taxi Data" by Lara Vomfell, Wolfgang Karl Härdle, Stefan Lessmann, February 2018.
014 "Price Discovery on Bitcoin Markets" by Paolo Pagnottoni, Dirk G. Baur, Thomas Dimpfl, March 2018.
015 "Bitcoin is not the New Gold - A Comparison of Volatility, Correlation, and Portfolio Performance" by Tony Klein, Hien Pham Thu, Thomas Walther, March 2018.
016 "Time-varying Limit Order Book Networks" by Wolfgang Karl Härdle, Shi Chen, Chong Liang, Melanie Schienle, April 2018.
017 "Regularization Approach for NetworkModeling of German EnergyMarket" by Shi Chen, Wolfgang Karl Härdle, Brenda López Cabrera, May 2018.
018 "Adaptive Nonparametric Clustering" by Kirill Efimov, Larisa Adamyan, Vladimir Spokoiny, May 2018.
019 "Lasso, knockoff and Gaussian covariates: a comparison" by Laurie Davies, May 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

020 "A Regime Shift Model with Nonparametric Switching Mechanism" by Haiqiang Chen, Yingxing Li, Ming Lin and Yanli Zhu, May 2018.
021 "LASSO-Driven Inference in Time and Space" by Victor Chernozhukov, Wolfgang K. Härdle, Chen Huang, Weining Wang, June 2018.
022 " Learning from Errors: The case of monetary and fiscal policy regimes" by Andreas Tryphonides, June 2018.
023 "Textual Sentiment, Option Characteristics, and Stock Return Predictability" by Cathy Yi-Hsuan Chen, Matthias R. Fengler, Wolfgang Karl Härdle, Yanchu Liu, June 2018.
024 "Bootstrap Confidence Sets For Spectral Projectors Of Sample Covariance" by A. Naumov, V. Spokoiny, V. Ulyanov, June 2018.
025 "Construction of Non-asymptotic Confidence Sets in 2 -Wasserstein Space" by Johannes Ebert, Vladimir Spokoiny, Alexandra Suvorikova, June 2018.
026 "Large ball probabilities, Gaussian comparison and anti-concentration" by Friedrich Götze, Alexey Naumov, Vladimir Spokoiny, Vladimir Ulyanov, June 2018.
027 "Bayesian inference for spectral projectors of covariance matrix" by Igor Silin, Vladimir Spokoiny, June 2018.
028 "Toolbox: Gaussian comparison on Eucledian balls" by Andzhey Koziuk, Vladimir Spokoiny, June 2018.
029 "Pointwise adaptation via stagewise aggregation of local estimates for multiclass classification" by Nikita Puchkin, Vladimir Spokoiny, June 2018.
030 "Gaussian Process Forecast with multidimensional distributional entries" by Francois Bachoc, Alexandra Suvorikova, Jean-Michel Loubes, Vladimir Spokoiny, June 2018.
031 "Instrumental variables regression" by Andzhey Koziuk, Vladimir Spokoiny, June 2018.
032 "Understanding Latent Group Structure of Cryptocurrencies Market: A Dynamic Network Perspective" by Li Guo, Yubo Tao and Wolfgang Karl Härdle, July 2018.
033 "Optimal contracts under competition when uncertainty from adverse selection and moral hazard are present" by Natalie Packham, August 2018.
034 "A factor-model approach for correlation scenarios and correlation stress-testing" by Natalie Packham and Fabian Woebbeking, August 2018.
035 "Correlation Under Stress In Normal Variance Mixture Models" by Michael Kalkbrener and Natalie Packham, August 2018.
036 "Model risk of contingent claims" by Nils Detering and Natalie Packham, August 2018.
037 "Default probabilities and default correlations under stress" by Natalie Packham, Michael Kalkbrener and Ludger Overbeck, August 2018.
038 "Tail-Risk Protection Trading Strategies" by Natalie Packham, Jochen Papenbrock, Peter Schwendner and Fabian Woebbeking, August 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

039 "Penalized Adaptive Forecasting with Large Information Sets and Structural Changes" by Lenka Zbonakova, Xinjue Li and Wolfgang Karl Härdle, August 2018.
040 "Complete Convergence and Complete Moment Convergence for Maximal Weighted Sums of Extended Negatively Dependent Random Variables" by Ji Gao YAN, August 2018.
041 "On complete convergence in Marcinkiewicz-Zygmund type SLLN for random variables" by Anna Kuczmaszewska and Ji Gao YAN, August 2018.
042 "On Complete Convergence in Marcinkiewicz-Zygmund Type SLLN for END Random Variables and its Applications" by Ji Gao YAN, August 2018.
043 "Textual Sentiment and Sector specific reaction" by Elisabeth Bommes, Cathy Yi-Hsuan Chen and Wolfgang Karl Härdle, September 2018.
044 "Understanding Cryptocurrencies" by Wolfgang Karl Härdle, Campbell R. Harvey, Raphael C. G. Reule, September 2018.
045 "Predicative Ability of Similarity-based Futures Trading Strategies" by Hsin-Yu Chiu, Mi-Hsiu Chiang, Wei-Yu Kuo, September 2018.
046 "Forecasting the Term Structure of Option Implied Volatility: The Power of an Adaptive Method" by Ying Chen, Qian Han, Linlin Niu, September 2018.
047 "Inferences for a Partially Varying Coefficient Model With Endogenous Regressors" by Zongwu Cai, Ying Fang, Ming Lin, Jia Su, October 2018.
048 "A Regime Shift Model with Nonparametric Switching Mechanism" by Haiqiang Chen, Yingxing Li, Ming Lin, Yanli Zhu, October 2018.
049 "Strict Stationarity Testing and GLAD Estimation of Double Autoregressive Models" by Shaojun Guo, Dong Li, Muyi Li, October 2018.
050 "Variable selection and direction estimation for single-index models via DC-TGDR method" by Wei Zhong, Xi Liu, Shuangge Ma, October 2018.
051 "Property Investment and Rental Rate under Housing Price Uncertainty: A Real Options Approach" by Honglin Wang, Fan Yu, Yinggang Zhou, October 2018.
052 "Nonparametric Additive Instrumental Variable Estimator: A Group Shrinkage Estimation Perspective" by Qingliang Fan, Wei Zhong, October 2018.
053 "The impact of temperature on gaming productivity: evidence from online games" by Xiaojia Bao, Qingliang Fan, October 2018.
054 "Topic Modeling for Analyzing Open-Ended Survey Responses" by Andra-Selina Pietsch, Stefan Lessmann, October 2018.
055 "Estimation of the discontinuous leverage effect: Evidence from the NASDAQ order book" by Markus Bibinger, Christopher Neely, Lars Winkelmann, October 2018.
056 "Cryptocurrencies, Metcalfe's law and LPPL models" by Daniel Traian Pele, Miruna Mazurencu-Marinescu-Pele, October 2018.
057 "Trending Mixture Copula Models with Copula Selection" by Bingduo Yang, Zongwu Cai, Christian M. Hafner, Guannan Liu, October 2018.

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

058 "Investing with cryptocurrencies – evaluating the potential of portfolio allocation strategies" by Alla Petukhina, Simon Trimborn, Wolfgang Karl Härdle, Hermann Elendner, October 2018.

059 "Towards the interpretation of time-varying regularization parameters in streaming penalized regression models" by Lenka Zbonakova, Ricardo Pio Monti, Wolfgang Karl Härdle, October 2018.

060 "Residual's Influence Index (Rinfin), Bad Leverage And Unmasking In High Dimensional L2-Regression" by Yannis G. Yatracos, October 2018.

060 "Plug-In L2-Upper Error Bounds In Deconvolution, For A Mixing Density Estimate In R^d And For Its Derivatives" by Yannis G. Yatracos, October 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.