EURANDOM PREPRINT SERIES
2010-034

DECONVOLUTION FOR AN ATOMIC DISTRIBUTION: RATES OF CONVERGENCE

(revised 24-02-2011)

Shota Gugushvili
ISSN 1389-2355

DECONVOLUTION FOR AN ATOMIC DISTRIBUTION: RATES OF CONVERGENCE

SHOTA GUGUSHVILI, BERT VAN ES, AND PETER SPREIJ

Abstract. Let $X_1, \ldots, X_n$ be i.i.d. copies of a random variable $X = Y + Z$, where $X_i = Y_i + Z_i$, and Y_i and Z_i are independent and have the same distribution as Y and Z, respectively. Assume that the random variables Y_i's are unobservable and that $Y = AV$, where A and V are independent, A has a Bernoulli distribution with probability of success equal to $1 - p$ and V has a distribution function F with density f. Let the random variable Z have a known distribution with density k. Based on a sample $X_1, \ldots, X_n$, we consider the problem of nonparametric estimation of the density f and the probability p. Our estimators of f and p are constructed via Fourier inversion and kernel smoothing. We derive their convergence rates over suitable functional classes. By establishing in a number of cases the lower bounds for estimation of f and p we show that our estimators are rate-optimal in these cases.

1. Introduction

Let $X_1, \ldots, X_n$ be i.i.d. copies of a random variable $X = Y + Z$, where $X_i = Y_i + Z_i$, and Y_i and Z_i are independent and have the same distribution as Y and Z, respectively. Assume that the random variables Y_i's are unobservable and that $Y = AV$, where A and V are independent, A has a Bernoulli distribution with probability of success equal to $1 - p$ and V has a distribution function F with density f. Furthermore, let the random variable Z have a known distribution with density k. Based on a sample $X_1, \ldots, X_n$, we consider the problem of nonparametric estimation of the density f and the probability p. This problem has been recently introduced in van Es et al. (2008) for the case when Z is normally distributed and Lee et al. (2010) for the class of more general error distributions. It is referred to as deconvolution for an atomic distribution, which reflects the fact that the distribution of Y has an atom of size p at zero and that we have to reconstruct ('deconvolve') p and f from the observations from the convolution structure $X = Y + Z$. When p is known to be equal to zero, i.e. when Y has a density, the problem reduces to the classical and much studied deconvolution problem, see e.g. Meister (2009) for an introduction to the latter and many recent references.

The above problem arises in a number of practical situations. For instance, suppose that a measurement device is used to measure some quantity of interest. Let it have a probability of failure to detect this quantity equal to p, in which case it renders zero. Repetitive measurements of the quantity of interest can be modelled by random variables Y_i defined as above. Assume that our goal is to estimate the

Date: February 25, 2011.
2000 *Mathematics Subject Classification.* Primary: 62G07, Secondary: 62G20.
Key words and phrases. Atomic distribution, deconvolution, Fourier inversion, kernel smoothing, mean square error, mean integrated square error, optimal convergence rate.

density f and the probability of failure p. If we could use the measurements Y_i directly, then when estimating f, zero measurements could be discarded and we could use the nonzero observations to base our estimator of f on. The probability p could be estimated by the proportion of zero observations. However, in practice it is often the case that some measurement error is present. This can be modelled by random variables Z_i and assuming the additive measurement error structure, in such a case the observations are $X_i = Y_i + Z_i$. Now notice that due to the measurement error, the zero Y_i's cannot be distinguished from the nonzero Y_i's. If we do not want to impose parametric assumptions on f, the use of nonparametric deconvolution techniques will be unavoidable when estimating f.

Another example comes from evolutionary biology, see Section 4 in Lee et al. (2010): suppose that a virus lineage is grown in a lab for a number of days in a manner that promotes accumulation of mutations. Plaque size can be used as a measure of viral fitness. Assume that it is measured every day and let the mutation effect on viral fitness be defined as a change in plaque size. If a high fitness virus is used, during any time interval in terms of mutations there are only two possibilities: either 1) no mutation, or only silent mutation occurs, or 2) a deleterious mutation occurs. Due to the fact that a silent mutation does not affect fitness, theoretically it will not change the plaque size and hence the mutation effect is zero for the first case. Deleterious mutations on the other hand will affect the plaque size. Since the distribution of deleterious mutation effects is usually considered to be continuous, the distribution of mutation effects can be expressed as a mixture of a point mass at zero, which corresponds to scenario 1), and a continuous distribution, which corresponds to scenario 2). Presence of measurement errors (which can be assumed to be additive) when measuring the plaque size leads precisely to the deconvolution problem for an atomic distribution.

Deconvolution for an atomic distribution is also closely related to empirical Bayes estimation of a mean of a high-dimensional normally distributed vector, see e.g. Jiang and Zhang (2009) for the description of the problem and many references. In more detail, let $X_i \sim N(\theta_i, 1), i = 1, \ldots, n$ be i.i.d., where $N(\theta_i, 1)$ denotes the normal distribution with mean θ_i and variance 1, and suppose that based on $X_1, \ldots, X_n$ the goal is to estimate the mean vector $\theta = (\theta_1, \ldots, \theta_n)$. This has applications e.g. in denoising a noisy signal or image. It is often the case that the vector θ is sparse in some sense in that many of θ_i's are zero or close to zero. The notion of sparsity can be naturally modelled in a Bayesian way by putting independent priors $\Pi_i(dx) = p1_{[x=0]}dx + (1-p)F(dx)$ on each component θ_i of θ, where $0 \leq p < 1$ and F is a continuous distribution function. Notice that excess of zeros among θ_i's is matched by choosing the prior Π_i that has a point mass at zero. In the empirical Bayes approach to estimation of θ the hyperparameters p and F of the priors Π_i are estimated from the data $X_1, \ldots, X_n$. This leads precisely to the deconvolution problem for an atomic distribution.

A related problem is estimation of the proportion of non-null effects in large-scale multiple testing framework, see e.g. Cai and Jin (2010). In large-scale multiple testing one is interested in simultaneous testing of a large number of hypotheses $H_1, \ldots, H_n$. Suppose that with every hypothesis H_i there is associated a corresponding test statistic X_i. A popular framework for large-scale multiple testing is the two-group random mixture model, where one assumes that each hypothesis H_i has a certain unknown probability π of being true (the approach is empirical Bayes in its essence) and the test statistics X_i are independent and are generated from a mixture of two densities, $X_i \sim (1-\pi) f_{\text{null}} + \pi f_{\text{alt}}$. Here π (the same for all i) is called the probability of null effects, f_{null} is the null density and f_{alt} is the non-null density. Often f_{null} is modelled as a density of a normal distribution $N(\mu_0, \sigma_0)$, while the density f_{alt} is modelled as a Gaussian location-scale mixture

$$f_{\text{alt}}(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) dG(\mu, \sigma),$$

where φ is the standard normal density and G is the mixing distribution which is assumed to be unknown. Observe that π in this case plays a role similar to $1-p$ in the deconvolution problem for an atomic distribution. Estimation of the probability π and the mixing distribution G based on $X_1, \ldots, X_n$ leads to a problem strongly related to the deconvolution problem for an atomic distribution.

After these motivating examples we return to the deconvolution problem for an atomic distribution and move to the construction of estimators of p and f (our notation is as in the first paragraph of this section). Because of a great similarity of our problem to the classical deconvolution problem, one natural approach to estimation of p and f is based on the use of Fourier inversion and kernel smoothing, cf. Section 2.2.1 in Meister (2009). In the sequel ϕ_ξ will denote the characteristic function of a random variable ξ. The Fourier transform of a function g will be denoted by ϕ_g. Suppose that $\phi_Z(t) \neq 0$ for all $t \in \mathbb{R}$. Following van Es et al. (2008), we define an estimator p_{ng_n} of p as

$$p_{ng_n} = \frac{g_n}{2} \int_{-1/g_n}^{1/g_n} \frac{\phi_{emp}(t)\phi_u(g_n t)}{\phi_Z(t)} dt, \tag{1}$$

where a number $g_n > 0$ denotes a bandwidth, ϕ_u is the Fourier transform of some fixed function (a kernel) u chosen beforehand and $\phi_{emp}(t) = n^{-1} \sum_{j=1}^n e^{itX_j}$ is the empirical characteristic function. To make the definition of p_{ng_n} meaningful, we assume that ϕ_u has support on $[-1,1]$. This guarantees integrability of the integrand in (1). We also assume that ϕ_u is real-valued, bounded, symmetric and integrates to two. Other conditions on u will be stated in the next section. Notice that p_{ng_n} is real-valued, because for its complex conjugate we have $\overline{p_{ng_n}} = p_{ng_n}$. The heuristics behind the definition of p_{ng_n} are the same as in van Es et al. (2008): using $\phi_X(t) = \phi_Y(t)\phi_Z(t)$ and $\phi_Y(t) = p + (1-p)\phi_f(t)$, we have

$$\begin{aligned}
\lim_{g_n \to 0} \frac{g_n}{2} \int_{-1/g_n}^{1/g_n} \frac{\phi_X(t)\phi_u(g_n t)}{\phi_Z(t)} dt &= \lim_{g_n \to 0} \frac{g_n}{2} \int_{-1/g_n}^{1/g_n} \phi_Y(t)\phi_u(g_n t) dt \\
&= \lim_{g_n \to 0} \frac{g_n}{2} \int_{-1/g_n}^{1/g_n} p\phi_u(g_n t) dt \\
&+ \lim_{g_n \to 0} \frac{g_n}{2} \int_{-1/g_n}^{1/g_n} (1-p)\phi_f(t)\phi_u(g_n t) dt \\
&= p,
\end{aligned}$$

provided $\phi_f(t)$ is integrable. The last equality follows from the dominated convergence theorem and the fact that ϕ_u integrates to two. Notice that this estimator coincides with the one in Lee et al. (2010) when u is the sinc kernel, i.e. $u(x) = \sin(x)/(\pi x)$. The Fourier transform of this kernel is given by $\phi_u(t) = 1_{[-1,1]}(t)$. In general p_{ng_n} might take on negative values, even though for large n the probability of this event will be small. In any case this is of minor importance, because we can always truncate p_{ng_n} from below at zero, i.e. we can define an estimator of p as $p_{ng_n}^{+} = \max(0, p_{ng_n})$. This new estimator of p has risk (quantified by the mean square error) not larger than that of p_{ng_n}:

$$\mathbb{E}_{p,f}[(p_{ng_n}^{+} - p)^2] \leq \mathbb{E}_{p,f}[(p_{ng_n} - p)^2].$$

Remark 1. In order to keep our notation compact, in the sequel instead of writing the expectation under the parameter pair (p, f) as $\mathbb{E}_{p,f}[\cdot]$, we will simply write $\mathbb{E}[\cdot]$. □

Next we turn to the construction of an estimator of f. Let

$$\hat{p}_{ng_n} = \max(-1 + \epsilon_n, \min(p_{ng_n}, 1 - \epsilon_n)), \tag{2}$$

where $0 < \epsilon_n < 1$ and $\epsilon_n \downarrow 0$ at a suitable rate to be specified later on. Notice that $|\hat{p}_{ng_n}| \leq 1 - \epsilon_n$. Truncating p_{ng_n} from below at $-1 + \epsilon_n$ and not at zero will make proofs of the asymptotic results for an estimator of f somewhat shorter, although truncation at zero is still a valid option. As in van Es et al. (2008), we propose the following estimator of f,

$$f_{nh_n g_n}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{\phi_{emp}(t) - \hat{p}_{ng_n} \phi_Z(t)}{(1 - \hat{p}_{ng_n}) \phi_Z(t)} \phi_w(h_n t) dt, \tag{3}$$

where w is a kernel function with a real-valued and symmetric Fourier transform ϕ_w supported on $[-1, 1]$ and $h_n > 0$ is a bandwidth. Notice that $\overline{f_{nh_n g_n}(x)} = f_{nh_n g_n}(x)$ and hence $f_{nh_n g_n}(x)$ is real-valued. It is clear that p_{ng_n} is truncated to $\hat{p}_{ng_n}$ in order to control the factor $(1 - \hat{p}_{ng_n})^{-1}$ in (3). The definition of $f_{nh_n g_n}$ is motivated by the fact that

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{\phi_X(t) - p\phi_Z(t)}{(1 - p)\phi_Z(t)} dt,$$

cf. equation (1.2) in van Es et al. (2008). Thus $f_{nh_n g_n}$ is obtained by replacing ϕ_X and p by their estimators and application of appropriate regularisation determined by the kernel w and bandwidth h. The estimator $f_{nh_n g_n}$ essentially coincides with the one in Lee et al. (2010) when both u and w are taken to be the sinc kernels. Again, notice that with positive probability $f_{nh_n g_n}(x)$ might become negative for some $x \in \mathbb{R}$, a little drawback often shared by kernel-type density estimators. Some correction method can be used to remedy this drawback, for instance one can define $f_{nh_n g_n}^{+}(x) = \max(0, f_{nh_n g_n}(x))$, as this does not increase the pointwise risk of the estimator. Note that this possible negativity of $f_{nh_n g_n}$ cannot be remedied only by truncating p_{ng_n} from below at zero and then using this new estimator instead of $\hat{p}_{ng_n}$ in (3). Observe also that $f_{nh_n g_n}^{+}$ can be rescaled to integrate to one and thus can be turned into a probability density. An alternative correction method to turn a possibly negative density estimator into a probability density is described in Glad et al. (2003). We do not pursue these questions any further.

In the present work we assume that the distribution of Z is known. In practice this is not always the case. If the distribution of Z is totally unknown, then next to the sample $X_1, \ldots, X_n$ one typically will need some additional data in order to construct consistent estimators of f and p. For instance, the case when additional measurements on Z, say $Z_1, \ldots, Z_m$, are available in the classical deconvolution problem with a priori known $p = 0$ is dealt with in Johannes (2009). Furthermore, one can also consider the case when the distribution of Z is known up to a scale parameter. The relevant papers in the classical deconvolution context are

Butucea and Matias (2005) and Meister (2006). Although conceivable in principle, extension of our results to these cases is beyond the scope of the present work.

In the rest of the paper we concentrate on asymptotics of the estimators p_{ng_n} and $f_{nh_ng_n}$. In particular, we derive upper bounds on the supremum of the mean square error of the estimator p_{ng_n} and the supremum of the mean integrated square error of the estimator $f_{nh_ng_n}$ taken over an appropriate class of the densities f and an appropriate interval for the probability p. Our results complement those in van Es et al. (2008), where the asymptotic normality of the estimators p_{ng_n} and $f_{nh_ng_n}$ is established. However, the present results are also more general, as we consider more general error distributions, and not necessarily the normal distribution as in van Es et al. (2008). Weak consistency of the estimators (1) and (3) based on the sinc kernel has been established under wide conditions in Lee et al. (2010). Here, however, we also derive convergence rates, much in the spirit of the classical deconvolution problems. Notice also that the fixed parameter asymptotics of the estimators of p and f were studied in Lee et al. (2010), in particular the rate of convergence of their estimator of f (but not of p) was derived. On the other hand, we prefer to study asymptotics uniformly in p and f, since fixed parameter statements are difficult to interpret from the asymptotic optimality point of view in nonparametric curve estimation, see e.g. Low et al. (1997) for a discussion. Furthermore, in case of estimation of f we quantify the risk globally in terms of the mean integrated squared error and not pointwise by the mean squared error as done in Lee et al. (2010). We also derive a lower risk bound for estimation of f, which shows that our estimator is rate-optimal over an appropriate functional class. Our final results are lower bounds for estimation of p. These lower bounds entail rate-optimality of p_{ng_n} in a large class of examples.

The structure of the paper can be outlined as follows: in Section 2 we state the main results of the paper. The proofs of these results are given in Section 3, while the Appendix contains several technical lemmas used in Section 3.

2. Results

The classical deconvolution problems are usually divided into two groups, ordinary smooth deconvolution problems and supersmooth deconvolution problems, see e.g. Fan (1991) or p. 35 in Meister (2009). In the former case it is assumed that the characteristic function ϕ_Z of a random variable Z decays to zero algebraically at plus and minus infinity (an example of such a Z is a random variable with Laplace distribution), while in the latter case the decay is essentially exponential (for instance, Z can be a normally distributed random variable). The rate of decay of ϕ_Z at infinity determines smoothness of the density of Z and hence the names ordinary smooth and supersmooth. Here too we will adopt the distinction between ordinary smooth and supersmooth deconvolution problems. The ordinary smooth deconvolution problems for an atomic distribution will be defined by the following condition on ϕ_Z.

Condition 1. *Let $\phi_Z(t) \neq 0$ for all $t \in \mathbb{R}$ and let*

$$d_0|t|^{-\beta} \leq |\phi_Z(t)| \text{ as } |t| \to \infty, \tag{4}$$

where d_0 and β are some strictly positive constants. Furthermore, let ϕ_Z be integrable.

Remark 2. Note that the assumption of integrability of ϕ_Z puts certain restriction on the tail behaviour of ϕ_Z and therefore implicitly on β too. In particular, in order that Condition 1 does not lead to an empty assumption, we must have $\beta > 1$. Notice that a lower bound on the rate of decay of ϕ_Z as in (4) is needed in order to derive upper risk bounds for the estimators p_{ng_n} and $f_{nh_ng_n}$, cf. p. 1260 in Fan (1991) and p. 35 in Meister (2009). When deriving lower bounds for estimation of p and f, (4) has to be further refined by adding an explicit upper bound on the rate of decay of ϕ_Z, see below. □

For the supersmooth deconvolution problems for an atomic distribution we will need the following condition on ϕ_Z.

Condition 2. *Let $\phi_Z(t) \neq 0$ for all $t \in \mathbb{R}$ and let*

$$d_0|t|^{\beta_0}e^{-|t|^\beta/\gamma} \leq |\phi_Z(t)| \text{ as } |t| \to \infty, \tag{5}$$

where β_0 is some real constant and d_0, β and γ are some strictly positive constants. Furthermore, let ϕ_Z be integrable.

Next we need to impose conditions on the class of target densities f.

Condition 3. *Define the class of target densities f as*

$$\Sigma(\alpha, K_\Sigma) = \left\{ f : \int_{-\infty}^{\infty} |\phi_f(t)|^2(1 + |t|^{2\alpha})dt \leq K_\Sigma \right\}, \tag{6}$$

Here α and K_Σ are some strictly positive numbers.

Smoothness conditions of this type are typical in nonparametric curve estimation problems, cf. p. 25 in Tsybakov (2009) or p. 34 in Meister (2009). Some smoothness assumptions have to be imposed on the class of target densities, because e.g. the class of all continuous densities is usually too large to be handled when dealing with uniform asymptotics. A possibility, different from Condition 3, is to assume that f belongs to the class of supersmooth densities

$$\Sigma(\alpha, \gamma, K_\Sigma) = \left\{ f : \int_{-\infty}^{\infty} |\phi_f(t)|^2 \exp(2\gamma|t|^\alpha)dt \leq K_\Sigma \right\},$$

for some strictly positive α, γ and K_Σ. The class $\Sigma(\alpha, \gamma, K_\Sigma)$ is much smaller than the class $\Sigma(\alpha, K_\Sigma)$ and the estimators p_{ng_n} and $f_{ng_nh_n}$ will enjoy better convergence rates in this case than in the case when the class of target densities is $\Sigma(\alpha, K_\Sigma)$, cf. Butucea and Tsybakov (2008a) and Butucea and Tsybakov (2008b) for a similar result in the classical deconvolution problem. In order not to overstretch the length of the paper, we decided however not to cover this case in the present work.

Remark 3. In the sequel we will use the symbols $\lesssim$ and $\gtrsim$ to compare two sequences a_n and b_n indexed by n, meaning respectively that a_n is less or equal than b_n for all n, or greater or equal, up to a universal constant that does not depend on n. □

The following theorem deals with asymptotics of the estimator p_{ng_n}. Its proof, as well as the proofs of all other results of the paper, is given in Section 3.

Theorem 1. *Let a function u be such that its Fourier transform ϕ_u is symmetric, real-valued, continuous in some neighbourhood of zero and is supported on $[-1, 1]$. Furthermore, let*

$$\int_{-1}^{1} \phi_u(t)dt = 2, \quad \left| \frac{\phi_u(t)}{t^\alpha} \right| \leq U \text{ for all } t \in \mathbb{R}, \tag{7}$$

where the constant α *is the same as in Condition 3,* U *is a strictly positive constant and for* $t=0$ *the ratio* $\phi_u(t)t^{-\alpha}$ *is defined by continuity at zero as the limit* $\lim_{t\to 0}\phi_u(t)t^{-\alpha}$*, which we assume to exist. Then*

(i) under Condition 1, by selecting $g_n = dn^{-1/(2\alpha+2\beta)}$ *for some constant* $d>0$*, we have*

$$\sup_{f\in\Sigma(\alpha,K_\Sigma),p\in[0,1)} \mathbb{E}\left[(p_{ng_n}-p)^2\right] \lesssim n^{-(2\alpha+1)/(2\alpha+2\beta)}; \tag{8}$$

(ii) under Condition 2, by selecting $g_n=(4/\gamma)^{1/\beta}(\log n)^{-1/\beta}$*, we have*

$$\sup_{f\in\Sigma(\alpha,K_\Sigma),p\in[0,1)} \mathbb{E}\left[(p_{ng_n}-p)^2\right] \lesssim (\log n)^{-(2\alpha+1)/\beta}. \tag{9}$$

Thus the rate of convergence of the estimator p_{ng_n} is slower than the root-n rate for estimation of a finite-dimensional parameter in regular parametric models. For Theorem 1 (ii) this is evident, while for Theorem 1 (i) this follows from Remark 2, which entails the fact that $2\alpha+1<2\alpha+2\beta$. However, see Theorems 4 and 5 below, where for a practically important case of a normally distributed Z, as well as Z with ordinary smooth distribution, by establishing the lower bounds for estimation of p we show that the slow convergence rate is intrinsic to the deconvolution problem and is not a quirk of our particular estimator.

Remark 4. The function u in the statement of Theorem 1 will not be a probability density, not even a function that integrates to one, and hence by calling it a kernel we somewhat abuse the established terminology in kernel estimation. Notice that condition (7) and the assumption $\alpha>0$ in Condition 3 preclude the kernel u from being the sinc kernel. We refer to van Es et al. (2008) for one particular example of u that produced good results in simulations. Its Fourier transform is given by

$$\phi_u(t)=\frac{693}{8}t^6(1-t^2)^2 1_{[-1,1]}(t).$$

Here $\alpha=6$ and $U=693/8$. An explicit, but rather complicated expression for u can be found in van Es et al. (2008). □

Next we will study the asymptotic behaviour of the estimator $f_{nh_ng_n}$ of f. We select the mean integrated square error as a criterion of its performance.

Due to technical reasons, see the proof of Theorem 2, in the ordinary smooth case it is convenient to split the sample $X_1,\ldots,X_n$ into two parts and next to base the estimator $\hat{p}_{ng_n}$ on the first part of the sample only, i.e. on $X_1,\ldots,X_{\lfloor n/2\rfloor}$, and to redefine $f_{nh_ng_n}$ as

$$f_{nh_ng_n}(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{\widetilde{\phi}_{emp}(t)-\hat{p}_{ng_n}}{(1-\hat{p}_{ng_n})\phi_Z(t)}\phi_w(h_nt)dt, \tag{10}$$

where

$$\widetilde{\phi}_{emp}(t)=\frac{1}{n-\lfloor n/2\rfloor}\sum_{j=\lfloor n/2\rfloor+1}^{n} e^{itX_j}.$$

Thus $\widetilde{\phi}_{emp}$ is based on the second half of the sample $X_1,\ldots,X_n$ only. Note that $\mathbb{E}\left[\phi_{emp}(t)\right]=\mathbb{E}[\widetilde{\phi}_{emp}(t)]=\phi_X(t)$. From now on we will assume that p_{ng_n} and $f_{nh_ng_n}$ are defined in this way in the ordinary smooth case, but will retain the old definition in the supersmooth case. Splitting the sample does not affect the convergence rate of $f_{nh_ng_n}$ in the ordinary smooth case, but only the constant factor in the upper bound on its mean integrated squared error. The general case without sample splitting in principle can also be handled, but we anticipate longer and more technical proofs, cf. the remarks at the end of the proof of Theorem 2. Since in the present work we are only concerned with convergence rates, sample splitting does not lead to a significant loss of generality.

The following theorem holds.

Theorem 2. *Let a kernel u satisfy the assumptions in Theorem 1. Furthermore, let a kernel w be such that its Fourier transform is symmetric, real-valued, is supported on $[-1,1]$ and*

$$\phi_w(0)=1, \quad |\phi_w(t)-1|\leq W|t|^{\alpha} \text{ for all } t\in\mathbb{R}, \quad \int_{-1}^{1}|\phi_w(t)|^2dt<\infty, \tag{11}$$

where W is some strictly positive constant. Moreover, let $p\in[0,p^*]$, where $p^*<1$. Then

(i) under Condition 1, by selecting $h_n=d(n-\lfloor n/2\rfloor)^{-1/(2\alpha+2\beta+1)}$ for some $d>0$, $g_n=d\lfloor n/2\rfloor^{-1/(2\alpha+2\beta)}$ and $\epsilon_n=(\log 3n)^{-1}$, we have

$$\sup_{f\in\Sigma(\alpha,K_\Sigma),p\in[0,p^*]}\mathbb{E}\left[\int_{-\infty}^{\infty}(f_{nh_ng_n}(x)-f(x))^2dx\right]\lesssim n^{-2\alpha/(2\alpha+2\beta+1)}, \tag{12}$$

where $f_{nh_ng_n}$ is defined by (10).

(ii) under Condition 2, by selecting $h_n=g_n=(4/\gamma)^{1/\beta}(\log n)^{-1/\beta}$ and $\epsilon_n=(\log 3n)^{-1}$, we have

$$\sup_{f\in\Sigma(\alpha,K_\Sigma),p\in[0,p^*]}\mathbb{E}\left[\int_{-\infty}^{\infty}(f_{nh_ng_n}(x)-f(x))^2dx\right]\lesssim (\log n)^{-2\alpha/\beta}, \tag{13}$$

where $f_{nh_ng_n}$ is defined by (3).

Remark 5. As it will become clear from the proof of this theorem, without the assumption $p^*<1$ one cannot study the asymptotics of $f_{nh_ng_n}$ uniformly in (p,f) for $p\in[0,p^*]$ and $f\in\Sigma(\alpha,K_\Sigma)$. Since p^* is allowed to be arbitrarily close to 1, from a practical point of view $p^*<1$ is not an important restriction. Observe that one can also study the case when $p^*=p_n^*$ depends on the sample size n and $p_n^*\to 1$ at a suitable rate. □

Remark 6. The condition $h_n=g_n$ in Theorem 2 (ii) is imposed for simplicity of the proofs only. In practice the two bandwidths need not be the same, cf. van Es et al. (2008), where unequal h_n and g_n are used in simulation examples. Also notice that our conditions on h_n and g_n in Theorems 1 and 2 are of asymptotic nature. For practical suggestions on bandwidth selection for the case when both u and w are sinc kernels, see Lee et al. (2010), where also a number of simulation examples is given. □

Remark 7. We refer to van Es et al. (2008) for one particular example of a kernel w. Any kernel that is known to produce good results in the classical deconvolution problem can be used as a kernel w. A relevant paper on the choice of a kernel in the context of the classical deconvolution problems is Delaigle and Hall (2006), to which we refer for a discussion and more examples. □

The upper risk bounds derived in Theorem 2 coincide with the upper risk bounds for kernel-type estimators in the classical deconvolution problems, i.e. in the case when p is a priori known to be zero, see Theorem 2.9 in Meister (2009). Naturally, a discussion on the optimality of convergence rates of the estimators $f_{nh_ng_n}$ and p_{ng_n} is in order. Let $\widetilde{f}_n$ denote an arbitrary estimator of f based on a sample $X_1, \ldots, X_n$. Consider

$$\mathcal{R}_n^* \equiv \inf_{\widetilde{f}_n} \sup_{f \in \Sigma, p \in [0,p^*]} \mathbb{E}\left[\int_{-\infty}^{\infty} (\tilde{f}_n(x) - f(x))^2 dx\right],$$

i.e. the minimax risk for estimation of f over some functional class Σ and the interval $[0, p^*]$ for p that is associated with our statistical model, cf. p. 78 in Tsybakov (2009). Notice that

$$\mathcal{R}_n^* \geq \inf_{\widetilde{f}_n} \sup_{f \in \Sigma, p=0} \mathbb{E}\left[\int_{-\infty}^{\infty} (\tilde{f}_n(x) - f(x))^2 dx\right].$$

The quantity on the right-hand side coincides with the minimax risk for estimation of a density f in the classical deconvolution problem, i.e. when $p = 0$ and the random variable Y has a density f. Using this fact, by Theorem 2.14 of Meister (2009) it is easy to obtain lower bounds for $\mathcal{R}_n^*$, but first we need to formulate two addition conditions on the rate of decay of ϕ_Z at plus and minus infinity. These two conditions correspond to the ordinary smooth and supersmooth deconvolution problems, cf. Conditions 1 and 2.

Condition 4. *Let ϕ_Z be such that*

$$|\phi_Z(t)| \leq \frac{d_1}{1+|t|^\beta}, \quad |\phi_Z'(t)| \leq \frac{d_1}{1+|t|^\beta} \quad \text{for all } t \in \mathbb{R}$$

for some strictly positive constants d_1 and β.

Condition 5. *Let ϕ_Z be such that*

$$|\phi_Z(t)| \leq d_1 e^{-|t|^\beta/\gamma}, \quad |\phi_Z'(t)| \leq d_1 e^{-|t|^\beta/\gamma} \quad \text{for all } t \in \mathbb{R}$$

for some strictly positive constants d_1, β and γ.

The following result holds.

Theorem 3. *Let $\widetilde{f}_n$ denote any estimator of f based on a sample $X_1, \ldots, X_n$ and let $\alpha \geq 1/2$. Suppose that K_Σ is large enough. Then*

(i) under Condition 4 we have

$$\inf_{\widetilde{f}_n} \sup_{f \in \Sigma(\alpha, K_\Sigma), p \in [0,p^*]} \mathbb{E}\left[\int_{-\infty}^{\infty} (\hat{f}(x) - f(x))^2 dx\right] \gtrsim n^{-2\alpha/(2\alpha+2\beta+1)}; \tag{14}$$

(ii) under Condition 5 the inequality

$$\inf_{\widetilde{f}_n} \sup_{f \in \Sigma(\alpha, K_\Sigma), p \in [0,p^*]} \mathbb{E}\left[\int_{-\infty}^{\infty} (\hat{f}(x) - f(x))^2 dx\right] \gtrsim (\log n)^{-2\alpha/\beta} \tag{15}$$

holds.

These lower bounds are of the same order as upper bounds in Theorem 2. It then follows that our estimator of f is rate-optimal under the combined conditions in Theorems 2 and 3. For a discussion on the conditions in Theorem 3 see p. 35 in Meister (2009).

Derivation of the lower risk bounds for estimation of probability p appears to be more involved. We will establish the lower bound for the case when Z follows the standard normal distribution. This is an important case, as the assumption of normality of measurement errors is frequently imposed in practice. The following result holds true.

Theorem 4. *Let Z have the standard normal distribution and let $\widetilde{p}_n$ denote any estimator of p based on a sample $X_1,\ldots,X_n$. Then*

$$\inf_{\widetilde{p}_n} \sup_{f\in\Sigma(\alpha,K_\Sigma),p\in[0,1)} \mathbb{E}\left[(\widetilde{p}_n-p)^2\right] \gtrsim (\log n)^{-(\alpha+1/2)} \tag{16}$$

holds.

A consequence of this theorem and (9) is that our estimator p_{ng_n} is rate-optimal in the case when Z follows the normal distribution.

The arguments used in the proof of Theorem 4 can be easily extended to the case when the distribution of Z is ordinary smooth. Below we provide the corresponding statement in the ordinary smooth case.

Theorem 5. *Let the characteristic function of Z satisfy Condition 4 for $\beta>1/2$. Let $\widetilde{p}_n$ denote any estimator of p based on the sample $X_1,\ldots,X_n$. Then*

$$\inf_{\widetilde{p}_n} \sup_{f\in\Sigma(\alpha,K_\Sigma),p\in[0,1)} \mathbb{E}\left[(\widetilde{p}_n-p)^2\right] \gtrsim n^{-(2\alpha+1)/(2\alpha+2\beta)}$$

holds.

This theorem and Theorem 1 (i) imply that under the combined conditions in Theorems 1 (i) and 5 the estimator p_{ng_n} is rate-optimal.

3. Proofs

Proof of Theorem 1. The proof uses some arguments from Fan (1991). To make the notation less cumbersome, let $\sup_{f,p}\equiv\sup_{f\in\Sigma(\alpha,K_\Sigma),p\in[0,1)}$. We first prove (i). We have

$$\sup_{f,p}\mathbb{E}\left[(p_{ng_n}-p)^2\right] \leq \sup_{f,p}(\mathbb{E}\left[p_{ng_n}\right]-p)^2+\sup_{f,p}\mathbb{V}\mathrm{ar}\left[p_{ng_n}\right]. \tag{17}$$

Observe that

$$\begin{aligned}
|\mathbb{E}\left[p_{ng_n}\right]-p| &= \frac{1-p}{2}\left|\int_{-1}^{1}\phi_f\left(\frac{t}{g_n}\right)\phi_u(t)dt\right| \\
&\leq \frac{1}{2}\int_{-1}^{1}\left|\phi_f\left(\frac{t}{g_n}\right)\left(\frac{t}{g_n}\right)^{\alpha}\right|\left|\left(\frac{g_n}{t}\right)^{\alpha}\phi_u(t)\right|1_{[t\neq0]}dt \\
&\leq \frac{1}{2}\sqrt{\int_{-1}^{1}\left|\phi_f\left(\frac{t}{g_n}\right)\left(\frac{t}{g_n}\right)^{\alpha}\right|^2dt}\sqrt{\int_{-1}^{1}\left|\frac{\phi_u(t)}{t^\alpha}g_n^\alpha\right|^2 1_{[t\neq0]}dt} \\
&\leq \frac{1}{\sqrt{2}}\sqrt{K_\Sigma}Ug_n^{\alpha+1/2},
\end{aligned} \tag{18}$$

where we used (7), (6) and the Cauchy-Schwarz inequality. Therefore

$$\sup_{f,p}(\mathbb{E}\left[p_{ng_n}\right]-p)^2 \lesssim g_n^{2\alpha+1} \tag{19}$$

holds. Furthermore, using independence of the random variables X_i's,

$$\mathbb{V}\mathrm{ar}\left[p_{ng_n}\right]=\frac{1}{4}\frac{g_n^2}{n}\mathbb{V}\mathrm{ar}\left[\int_{-1/g_n}^{1/g_n}e^{itX_1}\frac{\phi_u(g_nt)}{\phi_Z(t)}dt\right]$$

$$
\begin{aligned}
&\le \frac{1}{4}\frac{g_n^2}{n}\mathbb{E}\left[\left(\int_{-1/g_n}^{1/g_n} e^{itX_1}\frac{\phi_u(g_nt)}{\phi_Z(t)}dt\right)^2\right]\\
&= \frac{1}{4}\frac{g_n^2}{n}\int_{-\infty}^{\infty}\left(\int_{-1/g_n}^{1/g_n} e^{itx}\frac{\phi_u(g_nt)}{\phi_Z(t)}dt\right)^2 q(x)dx,
\end{aligned}
$$

where q is the density of X_1. Notice that

$$
q(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi_Y(t)\phi_Z(t)dt \le \frac{1}{2\pi}\int_{-\infty}^{\infty}|\phi_Z(t)|dt < \infty,
$$

where we used integrability of ϕ_Z. Therefore

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}\,[p_{ng_n}] &\lesssim \frac{g_n^2}{n}\int_{-\infty}^{\infty}\left(\int_{-1/g_n}^{1/g_n} e^{itx}\frac{\phi_u(g_nt)}{\phi_Z(t)}dt\right)^2 dx\\
&= \frac{1}{2\pi}\frac{g_n^2}{n}\int_{-\infty}^{\infty}\left|\frac{\phi_u(g_nt)}{\phi_Z(-t)}\right|^2 dt\\
&= \frac{1}{2\pi}\frac{g_n}{n}\int_{-1}^{1}\left|\frac{\phi_u(t)}{\phi_Z(-t/g_n)}\right|^2 dt
\end{aligned}
$$

by Parseval's identity. This inequality and an argument as on p. 1266 of Fan (1991) entail that

$$
\sup_{f,p}\mathbb{V}\mathrm{ar}\,[p_{ng_n}] \lesssim \frac{1}{ng_n^{2\beta-1}}. \tag{20}
$$

Formula (8) is then a consequence of (17), (19), (20) and our specific choice of g_n in (i).

Now we prove (ii). Since the first term on the right-hand side of (17) can be treated as in the ordinary smooth case (in particular (19) holds), we concentrate on the second term. Using independence of the random variables X_i's,

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}\,[p_{ng_n}] &= \frac{1}{4}\frac{1}{n}\mathbb{V}\mathrm{ar}\left[\int_{-1}^{1} e^{itX_1/g_n}\frac{\phi_u(t)}{\phi_Z(t/g_n)}dt\right]\\
&\le \frac{1}{4}\frac{1}{n}\left(\int_{-1}^{1}\left|\frac{\phi_u(t)}{\phi_Z(t/g_n)}\right|dt\right)^2. \qquad (21)
\end{aligned}
$$

By the same arguments as on pp. 1265–1266 of Fan (1991), one can show that

$$
\int_{-1}^{1}\left|\frac{\phi_u(t)}{\phi_Z(t/g_n)}\right|dt \le \begin{cases} C' e^{1/(\gamma g_n^{\beta})}, & \text{if } \beta_0 \ge 0\\ C' g_n^{\beta_0} e^{1/(\gamma g_n^{\beta})}, & \text{if } \beta_0 < 0, \end{cases} \tag{22}
$$

where the constant C' does not depend on n. In either case, because of our choice of g_n, the righthand side of (22) is of order $o(n^{1/3})$. This and (21) imply that

$$
\sup_{f,p}\mathbb{V}\mathrm{ar}\,[p_{ng_n}] = o(n^{-1/3}).
$$

The latter together with (17), (19) and our choice of g_n in (ii) proves (9). $\square$

Proof of Theorem 2. We use the shorthand notation $\sup_{f,p} \equiv \sup_{f\in\Sigma(\alpha,K_\Sigma),p\in[0,p^*]}$. By Fubini's theorem and the standard squared bias plus variance decomposition we have

$$\begin{aligned}\sup_{f,p}\mathbb{E}\left[\int_{-\infty}^{\infty}(f_{nh_ng_n}(x)-f(x))^2dx\right] &\leq \sup_{f,p}\int_{-\infty}^{\infty}(\mathbb{E}\left[f_{nh_ng_n}(x)\right]-f(x))^2dx\\ &+\sup_{f,p}\int_{-\infty}^{\infty}\mathbb{V}\mathrm{ar}\left[f_{nh_ng_n}(x)\right]dx\\ &=T_1+T_2.\end{aligned}$$

Keeping in mind the remarks surrounding (10), let

$$\hat{f}_{nh_n}(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\frac{\widetilde{\phi}_{emp}(t)\phi_w(h_nt)}{\phi_Z(t)}dt$$

in the ordinary smooth case, while

$$\hat{f}_{nh_n}(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\frac{\phi_{emp}(t)\phi_w(h_nt)}{\phi_Z(t)}dt$$

in the supersmooth case. Introduce

$$f_{nh_n}(x)=\frac{\hat{f}_{nh_n}(x)}{1-p}-\frac{p}{1-p}w_{h_n}(x), \tag{23}$$

where $w_{h_n}(x)=(1/h_n)w(x/h_n)$. We first study T_1, i.e. the supremum of the integrated squared bias. By the c_2-inequality it can be bounded as

$$\begin{aligned}T_1 &\lesssim \sup_{f,p}\int_{-\infty}^{\infty}(\mathbb{E}\left[f_{nh_n}(x)\right]-f(x))^2dx\\ &+\sup_{f,p}\int_{-\infty}^{\infty}(\mathbb{E}\left[f_{nh_ng_n}(x)-f_{nh_n}(x)\right])^2dx\\ &=T_3+T_4.\end{aligned}$$

By Parseval's identity and the dominated convergence theorem

$$\begin{aligned}\int_{-\infty}^{\infty}(\mathbb{E}\left[f_{nh_n}(x)\right]-f(x))^2dx &= \frac{1}{2\pi}\int_{-\infty}^{\infty}|\phi_f(t)|^2|\phi_w(h_nt)-1|^2dt\\ &=h_n^{2\alpha}\frac{1}{2\pi}\int_{-\infty}^{\infty}|t|^{2\alpha}|\phi_f(t)|^2\frac{|\phi_w(h_nt)-1|^2}{|h_nt|^{2\alpha}}1_{[t\neq 0]}dt\\ &\lesssim h_n^{2\alpha}.\end{aligned}$$

Here in the second equality we used the fact that $\phi_w(0)=1$. The dominated convergence theorem is applicable because of Condition 3 and (11). Hence $T_3\lesssim h_n^{2\alpha}$ in view of the fact that $f\in\Sigma(\alpha,K_\Sigma)$. It is also straightforward to see that in fact $\sup_{f,p}T_3\lesssim h_n^{2\alpha}$. We deal with T_4. By the c_2-inequality

$$\begin{aligned}\int_{-\infty}^{\infty}(\mathbb{E}\left[f_{nh_ng_n}(x)-f_{nh_n}(x)\right])^2dx &\lesssim \left(\mathbb{E}\left[\frac{\hat{p}_{ng_n}-p}{(1-\hat{p}_{ng_n})(1-p)}\right]\right)^2\int_{-\infty}^{\infty}(w_{h_n}(x))^2dx\\ &+\int_{-\infty}^{\infty}\left(\mathbb{E}\left[\hat{f}_{nh_n}(x)\frac{(\hat{p}_{ng_n}-p)}{(1-\hat{p}_{ng_n})(1-p)}\right]\right)^2dx\\ &=T_5+T_6.\end{aligned}$$

Notice that

$$\int_{-\infty}^{\infty}(w_{h_n}(x))^2dx=\frac{1}{h_n}\int_{-\infty}^{\infty}(w(x))^2dx<\infty,$$

because by our assumptions and Parseval's identity w is square integrable. We first consider T_5. By the Cauchy-Schwarz inequality we have

$$T_5 \leq \frac{1}{h_n}\int_{-\infty}^{\infty}(w(u))^2 du \mathbb{E}\left[\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}\right].$$

With our choice of the smoothing parameters h_n and g_n it follows from Lemma 2 of the Appendix that $\sup_{p,f} T_5 \lesssim g_n^{2\alpha}$. Now let us turn to T_6. By the Cauchy-Schwarz inequality

$$T_6 \leq \mathbb{E}\left[\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}\right]\int_{-\infty}^{\infty}\mathbb{E}\left[(\hat{f}_{nh_n}(x))^2\right]dx.$$

By Lemma 2 of the Appendix the first term in the product in the above display is of order $g_n^{2\alpha+1}$. The same holds true for its supremum over f and p. Hence it remains to study the second factor in the above upper bound on T_6. We have

$$\begin{aligned}\int_{-\infty}^{\infty}\mathbb{E}\left[(\hat{f}_{nh_n}(x))^2\right]dx &= \int_{-\infty}^{\infty}\mathbb{V}\mathrm{ar}\left[\hat{f}_{nh_n}(x)\right]dx + \int_{-\infty}^{\infty}\left(\mathbb{E}\left[\hat{f}_{nh_n}(x)\right]\right)^2 dx\\ &= T_7 + T_8.\end{aligned}$$

Let the function W_n is defined by

$$W_n(x) = \frac{1}{2\pi}\int_{-1}^{1} e^{-itx}\frac{\phi_w(t)}{\phi_Z(t/h_n)}dt.$$

Notice that by independence of X_i's

$$T_7 = \frac{1}{nh_n^2}\int_{-\infty}^{\infty}\mathbb{V}\mathrm{ar}\left[W_n\left(\frac{x-X_1}{h_n}\right)\right]dx \leq \frac{1}{nh_n^2}\int_{-\infty}^{\infty}\mathbb{E}\left[\left(W_n\left(\frac{x-X_1}{h_n}\right)\right)^2\right]dx$$

in the supersmooth case, and

$$T_7 \leq \frac{1}{(n-\lfloor n/2\rfloor)h_n^2}\int_{-\infty}^{\infty}\mathbb{E}\left[\left(W_n\left(\frac{x-X_1}{h_n}\right)\right)^2\right]dx$$

in the ordinary smooth case. Then by Fubini's theorem

$$\begin{aligned}T_7 &\leq \frac{1}{nh_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(W_n\left(\frac{x-s}{h_n}\right)\right)^2 q(s)dsdx\\ &= \frac{1}{nh_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(W_n\left(\frac{x-s}{h_n}\right)\right)^2 dxq(s)ds\\ &= \frac{1}{nh_n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(W_n(x))^2 dxq(s)ds\\ &= \frac{1}{nh_n}\int_{-1}^{1}\frac{|\phi_w(t)|^2}{|\phi_Z(t/h_n)|^2}dt\end{aligned}$$

in the supersmooth case, and

$$T_7 \leq \frac{1}{(n-\lfloor n/2\rfloor)h_n}\int_{-1}^{1}\frac{|\phi_w(t)|^2}{|\phi_Z(t/h_n)|^2}dt$$

in the ordinary smooth case. Here we used the fact that q, being a probability density, integrates to one, as well as Parseval's identity. The integrals in the last equalities of the above two displayed formulae can be analysed by exactly the same arguments as on pp. 1265-1266 in Fan (1991). Thus

$$T_7 \lesssim \begin{cases} \frac{1}{nh_n^{2\beta+1}}, & \text{if } Z \text{ is ordinary smooth,} \\ \frac{1}{nh_n} e^{2/(\gamma h_n^\beta)}, & \text{if } Z \text{ is supersmooth and } \beta_0 \geq 0, \\ \frac{h_n^{2\beta_0-1}}{n} e^{2/(\gamma h_n^\beta)}, & \text{if } Z \text{ is supersmooth and } \beta_0 < 0. \end{cases} \tag{24}$$

The same order bounds hold for $\sup_{f,p} T_7$ as well. As a consequence, $\sup_{f,p} T_7 \to 0$. Let us now study T_8. By Parseval's identity and the fact that $|\phi_Y(t)| \leq 1$, we have

$$\begin{aligned} T_8 &= \int_{-\infty}^{\infty} \left(\frac{1}{2\pi} \int_{-1/h_n}^{1/h_n} e^{-itx} \phi_Y(t) \phi_w(h_n t) dt \right)^2 dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |\phi_Y(t)\phi_w(h_n t)|^2 1_{[-h^{-1},h^{-1}]}(t) dt \\ &\leq \frac{1}{h_n} \frac{1}{2\pi} \int_{-1}^{1} |\phi_w(t)|^2 dt \\ &\lesssim \frac{1}{h_n}, \end{aligned}$$

where the last line follows from our assumptions on w. It follows that $\sup_{p,f} T_8 \lesssim 1/h_n$. Combination of the above bounds on $\sup_{p,f} T_7$ and $\sup_{p,f} T_8$ entails that $\sup_{f,p} T_6 \lesssim g_n^{2\alpha}$, where we also used the fact that $g_n \lesssim h_n$. Therefore T_4, as well as T_1, i.e. the supremum of the integrated squared bias, is of order $h_n^{2\alpha}$. For the ordinary smooth case this gives an upper bound of order $n^{-2\alpha/(2\alpha+2\beta+1)}$ on T_1, while for the supersmooth case an upper bound of order $(\log n)^{-2\alpha/\beta}$.

Now we turn to T_2, i.e. the supremum of the integrated variance. We have

$$\begin{aligned} \int_{-\infty}^{\infty} \mathbb{V}\mathrm{ar}\,[f_{nh_ng_n}(x)] dx &= \int_{-\infty}^{\infty} \mathbb{V}\mathrm{ar}\,[f_{nh_ng_n}(x) - f_{nh_n}(x) + f_{nh_n}(x)] dx \\ &\lesssim \int_{-\infty}^{\infty} \mathbb{V}\mathrm{ar}\,[f_{nh_n}(x)] dx + \int_{-\infty}^{\infty} \mathbb{V}\mathrm{ar}\,[f_{nh_ng_n}(x) - f_{nh_n}(x)] dx \\ &= T_9 + T_{10}, \end{aligned}$$

where we used the fact that for random variables ξ and η

$$\mathbb{V}\mathrm{ar}\,[\xi + \eta] \leq 2(\mathbb{V}\mathrm{ar}\,[\xi] + \mathbb{V}\mathrm{ar}\,[\eta]).$$

Since T_9 up to a constant is the same as T_7, cf. (23), the term $\sup_{f,p} T_9$ can be bounded as before, see (24). We consider T_{10}. Let ψ_n be as in (34) in the proof of Lemma 2 of the Appendix. Then

$$\begin{aligned} T_{10} &\leq \int_{-\infty}^{\infty} \mathbb{E}\,[(f_{nh_ng_n}(x) - f_{nh_n}(x))^2 1_{[|\hat{p}_{ng_n} - p| > \psi_n]}] dx \\ &+ \int_{-\infty}^{\infty} \mathbb{E}\,[(f_{nh_ng_n}(x) - f_{nh_n}(x))^2 1_{[|\hat{p}_{ng_n} - p| \leq \psi_n]}] dx \\ &= T_{11} + T_{12}. \end{aligned}$$

By the c_2-inequality

$$T_{11} \lesssim \frac{1}{h_n} \int_{-\infty}^{\infty} (w(x))^2 dx \mathbb{E} \left[\frac{(\hat{p}_{ng_n} - p)^2}{(1-\hat{p}_{ng_n})^2 (1-p)^2} 1_{[|\hat{p}_{ng_n} - p| > \psi_n]} \right]$$

$$+\int_{-\infty}^{\infty}\mathbb{E}\left[(\hat{f}_{nh_n}(x))^2\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}1_{[|\hat{p}_{ng_n}-p|>\psi_n]}\right]dx$$
$$=T_{13}+T_{14}.$$

Since $T_{13}\lesssim h_n^{-1}\epsilon_n^{-2}\sup_{p,f}P(|\hat{p}_{ng_n}-p|>\psi_n)$, which follows from the fact that

$$\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}\leq\frac{2(1-\epsilon_n)^2+2p^{*2}}{\epsilon_n^2(1-p^*)^2},$$

by Lemma 3 of the Appendix with our conditions on h_n and ϵ_n it certainly holds true that $\sup_{p,f}T_{13}\lesssim h_n^{2\alpha}$. As far as T_{14} is concerned, by Fubini's theorem and Parseval's identity

$$T_{14}=\mathbb{E}\left[\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}1_{[|\hat{p}_{ng_n}-p|>\psi_n]}\int_{-\infty}^{\infty}(\hat{f}_{nh_n}(x))^2dx\right]$$
$$=\mathbb{E}\left[\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}1_{[|\hat{p}_{ng_n}-p|>\psi_n]}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{|\phi_{emp}(t)\phi_w(h_nt)|^2}{|\phi_Z(t)|^2}dt\right]$$
$$\lesssim\frac{1}{\epsilon_n^2}\frac{1}{h_n}\int_{-\infty}^{\infty}\frac{|\phi_w(t)|^2}{|\phi_Z(t/h_n)|^2}dt\,\mathrm{P}(|\hat{p}_{ng_n}-p|>\psi_n).$$

Hence

$$T_{14}\lesssim\frac{1}{\epsilon_n^2}\frac{1}{h_n^{2\beta+1}}\mathrm{P}(|\hat{p}_{ng_n}-p|>\psi_n)$$

in the ordinary smooth case, and

$$T_{14}\lesssim\begin{cases}\frac{1}{\epsilon_n^2}\frac{1}{h_n}e^{2/(\gamma h_n^{\beta})}\mathrm{P}(|\hat{p}_{ng_n}-p|>\psi_n), & \text{if } \beta_0\geq 0,\\ \frac{1}{\epsilon_n^2}h_n^{2\beta_0-1}e^{2/(\gamma h_n^{\beta})}\mathrm{P}(|\hat{p}_{ng_n}-p|>\psi_n), & \text{if } \beta_0<0\end{cases}$$

in the supersmooth case, cf. pp. 1265-1266 of Fan (1991). Similar order bounds are true for $\sup_{p,f}T_{14}$. Again by Lemma 3 and our conditions on h_n and ϵ_n, we have $\sup_{p,f}T_{14}\lesssim h_n^{2\alpha}$.

To complete establishing an upper bound on T_{10}, it remains to study T_{12}. As in the case of T_{11}, by the c_2-inequality

$$T_{12}\lesssim\frac{1}{h_n}\int_{-\infty}^{\infty}(w(x))^2dx\mathbb{E}\left[\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}1_{[|\hat{p}_{ng_n}-p|\leq\psi_n]}\right]$$
$$+\int_{-\infty}^{\infty}\mathbb{E}\left[(\hat{f}_{nh_n}(x))^2\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}1_{[|\hat{p}_{ng_n}-p|\leq\psi_n]}\right]dx$$

holds. By Lemma 1 of the Appendix the first term on the righthand side is up to a constant bounded by $(1/h_n)g_n^{2\alpha+1}$ and hence is of order $g_n^{2\alpha}$. The same is true for its supremum over p and f. As far as the second term is concerned, in the supersmooth case it is bounded by $\psi_n^2\int_{-\infty}^{\infty}\mathbb{E}\left[(\hat{f}_{nh_n}(x))^2\right]dx$. It follows from the upper bounds on $\sup_{p,f}T_7$ and $\sup_{p,f}T_8$ that in the supersmooth case we have $\sup_{p,f}T_{12}\lesssim h_n^{2\alpha}$. As far as the ordinary smooth case is concerned,

$$\int_{-\infty}^{\infty}\mathbb{E}\left[(\hat{f}_{nh_n}(x))^2\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}1_{[|\hat{p}_{ng_n}-p|\leq\psi_n]}\right]dx$$
$$=\int_{-\infty}^{\infty}\mathbb{E}\left[(\hat{f}_{nh_n}(x))^2\right]dx\,\mathbb{E}\left[\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}1_{[|\hat{p}_{ng_n}-p|\leq\psi_n]}\right]$$

holds. This is precisely the place where we use independence between $\hat{f}_{nh}(x)$ and $\hat{p}_{ng_n}$ implied by sample splitting, cf. the remarks around (10). Then in this case too $\sup_{p,f} T_{12} \lesssim h_n^{2\alpha}$. Had not we used the sample splitting trick, in the above display we would have to apply the Cauchy-Schwarz inequality apparently leading to rather lengthy computations.

Combination of the bounds on $\sup_{p,f} T_{11}$ and $\sup_{p,f} T_{12}$ implies that $\sup_{f,p} T_{10} \lesssim h_n^{2\alpha}$. The bounds on and $\sup_{f,p} T_9$ and $\sup_{f,p} T_{10}$ induce the bound on T_2. The statement of the theorem then follows from the bounds on T_1 and T_2. $\square$

Proof of Theorem 3. The result is a straightforward consequence of Theorem 2.14 of Meister (2009). $\square$

Proof of Theorem 4. A general idea of the proof can be outlined as follows: we will consider two pairs (p_1, f_1) and (p_2, f_2) (depending on n) of the parameter (p, f) that parametrises the density of X, such that the probabilities p_1 and p_2 are separated as much as possible, while at the same time the corresponding product densities $q_1^{\otimes n}$ and $q_2^{\otimes n}$ of observations $X_1, \ldots, X_n$ are close in the χ^2-divergence and hence cannot be distinguished well using the observations $X_1, \ldots, X_n$. By Lemma 8 of Butucea and Tsybakov (2008b) the squared distance between p_1 and p_2 will then give (up to a constant that does not depend on n) the desired lower bound (16) for estimation of p.

Our construction of the two alternatives (p_1, f_1) and (p_2, f_2) is partially motivated by the construction used in the proof of Theorem 3.5 of Chen et al. (2010). Let $\lambda_1 = \lambda + \delta^{\alpha+1/2}$, where $\lambda > 0$ is a fixed constant and $\delta \downarrow 0$ as $n \to \infty$. Define $p_1 = e^{-\lambda_1}$ and notice that $p_1 \in [0, 1)$. Next set $\phi_{g_1}(t) = e^{-|t|}$ and observe that this is the characteristic function corresponding to the Cauchy density $g_1(x) = 1/(\pi(1 + x^2))$. Finally, define

$$\phi_{f_1}(t) = \frac{1}{e^{\lambda_1} - 1} \left(e^{\lambda_1 \phi_{g_1}(t)} - 1 \right).$$

Denote by W_j the i.i.d. random variables that have the common density g_1 and by N_{λ_1} the random variable that has Poisson distribution with parameter λ_1. Then the function ϕ_{f_1} will be the characteristic function corresponding to the density f_1 of the Poisson sum $Y = \sum_{j=1}^{N_{\lambda_1}} W_j$ of i.i.d. W_j's conditional on the fact that the number of its summands $N_{\lambda_1} > 0$, see pp. 14–15 of Gugushvili (2008). Notice that we have an inequality

$$|\phi_{f_1}(t)| \leq \frac{\lambda_1 e^{\lambda_1}}{e^{\lambda_1} - 1} |\phi_{g_1}(t)|,$$

cf. inequality (2.10) on p. 22 of Gugushvili (2008). Keeping this inequality in mind, without loss of generality we can assume that K_Σ is already such that $\phi_{f_1} \in \Sigma(\alpha, K_\Sigma/4)$. Otherwise we can always consider $\phi_{g_1}(t) = e^{-\alpha'|t|}$ with a fixed and large enough constant $\alpha' > 0$, so that $\phi_{f_1} \in \Sigma(\alpha, K_\Sigma/4)$. It is not difficult to see that the fact that $\alpha' \neq 1$ will not affect seriously our subsequent argumentation in this proof. Next define the density q_1 corresponding to the pair (p_1, f_1) via its characteristic function

$$\phi_{q_1}(t) = (p_1 + (1 - p_1)\phi_{g_1}(t))e^{-t^2/2}$$

and remark that it has the convolution structure required for our problem.

Now we proceed to the definition of the second alternative (p_2, f_2). Set $\lambda_2 = \lambda$ and $p_2 = e^{-\lambda_2}$. The fact that $p_2 \in [0, 1)$ follows from the fact that $\lambda > 0$. Let

H be a function, such that its Fourier transform ϕ_H is symmetric and real-valued with support on $[-2,2]$, $\phi_H(t)=1$ for $t\in[-1,1]$ and ϕ_H is two times continuously differentiable. Such a function can be constructed e.g. in the same way as a flat-top kernel in Section 3 of McMurry and Politis (2004). Define

$$\phi_{g_2}(t)=\phi_{g_1}(t)+\tau(t),$$

where the perturbation function τ is given by

$$\tau(t)=\frac{\delta^{\alpha+1/2}}{\lambda_2}(\phi_{f_1}(t)-1)\phi_H(\delta t).$$

We claim that for all n large enough ϕ_{g_2} is a characteristic function, i.e. its inverse Fourier transform g_2 is a probability density. This involves showing that g_2 integrates to one and is nonnegative. The former easily follows from the fact that

$$\int_{-\infty}^{\infty} g_2(x)dx=\phi_{g_2}(0)=\phi_{g_1}(0)=1, \tag{25}$$

since $\tau(0)=0$ by construction and ϕ_{g_1} is a characteristic function. As far as the latter is concerned, we argue as follows: observe that g_2 is real-valued, because ϕ_{g_2} is symmetric and real-valued. By the Fourier inversion argument

$$\sup_x|g_2(x)-g_1(x)|\leq\frac{1}{2\pi}\int_{-\infty}^{\infty}|\tau(t)|dt\to 0$$

as $n\to\infty$, by definition of τ and because $\delta\to 0$. Since g_1, being the Cauchy density, is strictly positive on the whole real line, provided n is large enough it follows that

$$g_2(x)\geq 0,\quad x\in B, \tag{26}$$

where B is a certain neighbourhood around zero. Next, we need to consider those x's, that lie outside this certain fixed neighbourhood of zero. We have

$$\begin{aligned}
g_2(x)&=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\left(\phi_{g_1}(t)+\frac{\delta^{\alpha+1/2}}{\lambda_2}(\phi_{g_1}(t)-1)\phi_H(\delta t)\right)dt\\
&=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\left(\left(1+\frac{\delta^{\alpha+1/2}}{\lambda_2}\right)\phi_{g_1}(t)-\frac{\delta^{\alpha+1/2}}{\lambda_2}\phi_{g_1}(t)+\frac{\delta^{\alpha+1/2}}{\lambda_2}(\phi_{g_1}(t)-1)\phi_H(\delta t)\right)dt\\
&=\left(1+\frac{\delta^{\alpha+1/2}}{\lambda_2}\right)g_1(x)+\frac{\delta^{\alpha+1/2}}{\lambda_2}\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\phi_{g_1}(t)(\phi_H(\delta t)-1)dt\\
&-\frac{\delta^{\alpha+1/2}}{\lambda_2}\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\phi_H(\delta t)dt\\
&=T_1(x)+T_2(x)+T_3(x).
\end{aligned}$$

Both $T_2(x)$ and $T_3(x)$ are real-valued by symmetry of ϕ_{g_1} and ϕ_H and the fact that these Fourier transforms are real-valued. Consequently, g_2 itself is also real-valued. Since g_1 is the Cauchy density and $\delta>0$, the inequality

$$T_1(x)\geq\frac{1}{\pi}\frac{1}{1+x^2} \tag{27}$$

holds for all $x\in\mathbb{R}$. Assuming that $x\neq 0$ and integrating by parts, we get

$$T_2(x)=-\frac{1}{ix}\frac{\delta^{\alpha+1/2}}{\lambda_2}\frac{1}{2\pi}\int_{\mathbb{R}\setminus[-\delta^{-1},\delta^{-1}]}\phi_{g_1}(t)(\phi_H(\delta t)-1)de^{-itx}$$

$$= \frac{1}{ix} \frac{\delta^{\alpha+1/2}}{\lambda_2} \frac{1}{2\pi} \int_{\mathbb{R}\setminus[-\delta^{-1},\delta^{-1}]} e^{-itx} [\phi_{g_1}(t)(\phi_H(\delta t) - 1)]' dt.$$

Applying integration by parts to the last equality one more time, we obtain that

$$T_2(x) = \frac{1}{x^2} \frac{\delta^{\alpha+1/2}}{\lambda_2} \frac{1}{2\pi} \int_{\mathbb{R}\setminus[-\delta^{-1},\delta^{-1}]} e^{-itx} [\phi_{g_1}(t)(\phi_H(\delta t) - 1)]'' dt,$$

which implies that

$$|T_2(x)| \leq \frac{1}{x^2} C \delta^{\alpha+1/2} \int_{\mathbb{R}\setminus[-\delta^{-1},\delta^{-1}]} |[\phi_{g_1}(t)(\phi_H(\delta t) - 1)]''| dt,$$

where the constant C does not depend on x and n. Since $\delta \to 0$ and the first and the second derivatives of ϕ_H are bounded on $\mathbb{R}$, it follows that

$$|T_2(x)| \leq \frac{1}{x^2} C' \delta^{\alpha+1/2} \int_{t>\delta^{-1}} e^{-t} dt,$$

where the constant C' is independent of n and x. In particular,

$$|T_2(x)| \leq C' \delta^{\alpha+1/2} \frac{1}{x^2} \tag{28}$$

for all n large enough. Finally, using integration by parts twice, one can also show that for $x \neq 0$

$$T_3(x) = \frac{1}{x^2} \frac{\delta^{\alpha+5/2}}{\lambda_2} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_H''(\delta t) dt$$

and hence

$$|T_3(x)| \leq C'' \delta^{\alpha+3/2} \frac{1}{x^2}, \tag{29}$$

where the constant C'' does not depend on n and x. Therefore, by gathering (27)–(29), we conclude for all n large enough and all $x \in \mathbb{R}$ the inequality

$$g_2(x) = T_1(x) + T_2(x) + T_3(x) \geq 0$$

is valid. Combining this with (25), we obtain that g_2 is a probability density.

Now we turn to the model defined by the pair (p_2, f_2). Again by the argument on pp. 22–23 of Gugushvili (2008),

$$|\phi_{f_2}(t)| \leq \frac{\lambda_2 e^{\lambda_2}}{e^{\lambda_2} - 1} |\phi_{g_2}(t)|.$$

Notice that by selecting α' in the definition of $\phi_{g_1}(t) = e^{-\alpha'|t|}$ large enough and λ large enough, one can arrange that $f_2 \in \Sigma(\alpha, K_\Sigma)$, at least for all n large enough. Without loss of generality we take $\alpha' = 1$. Set

$$\phi_{q_2}(t) = (p_2 + (1 - p_2)\phi_{g_2}(t)) e^{-t^2/2}.$$

This has the convolution structure as needed in our problem. Hence both pairs (p_1, f_1) and (p_2, f_2) belong to the class required in the statement of the theorem and generate the required models.

It is easy to see that

$$|p_2 - p_1| \asymp \delta^{\alpha+1/2} \tag{30}$$

as $\delta \to 0$, where $\asymp$ means that two sequences are asymptotically of the same order. Consequently, by Lemma 8 of Butucea and Tsybakov (2008b) the lower bound in (16) will be of order $\delta^{2\alpha+1}$, provided we can prove that $n\chi^2(q_2, q_1) \to 0$ as $n \to \infty$ for an appropriate $\delta \to 0$. Here $\chi^2(q_2, q_1)$ is the χ^2 divergence between the probability measures with densities q_2 and q_1, i.e.

$$\chi^2(q_2, q_1) = \int_{-\infty}^{\infty} \frac{(q_2(x) - q_1(x))^2}{q_1(x)} dx,$$

see p. 86 in Tsybakov (2009).

Notice that we have

$$q_1(x) = e^{-\lambda_1}\varphi(x) + (1 - e^{-\lambda_1}) f_1 * \varphi(x),$$

where φ denotes the standard normal density. Let δ_1 denote the first element of the sequence $\delta = \delta_n \downarrow 0$. Then

$$\begin{aligned} f_1(x) &= \sum_{n=1}^{\infty} g_1^{*n}(x) P(N_{\lambda_1} = n | N_{\lambda_1} > 0) \\ &\geq g_1(x) P(N_{\lambda_1} = 1 | N_{\lambda_1} > 0) \\ &= g_1(x) \frac{P(N_{\lambda_1} = 1)}{1 - P(N_{\lambda_1} = 0)} \\ &\geq \frac{\lambda e^{-\lambda - \delta_1^{\alpha+1/2}}}{1 - e^{-\lambda_1}} g_1(x), \end{aligned}$$

cf. p. 23 in Gugushvili (2008). It follows that for all x

$$q_1(x) \geq (1 - e^{-\lambda_1}) f_1 * \varphi(x) \geq \kappa_A \lambda e^{-\lambda - \delta_1^{\alpha+1/2}} g_1(|x| + A) = c_\lambda g_1(|x| + A) \tag{31}$$

for some large enough (but fixed) constant $A > 0$. Here the constant $\kappa_A = \int_{-A}^{A} k(t) dt$. The inequalities in (31) hold, because

$$\begin{aligned} (1 - e^{-\lambda_1}) f_1 * \varphi(x) &= (1 - e^{-\lambda_1}) \int_{-\infty}^{\infty} f_1(x - t) \varphi(t) dt \\ &\geq \lambda e^{-\lambda - \delta_1^{\alpha+1/2}} \int_{-\infty}^{\infty} g_1(x - t) \varphi(t) dt \\ &\geq \lambda e^{-\lambda - \delta_1^{\alpha+1/2}} \int_{-A}^{A} g_1(x - t) \varphi(t) dt \\ &\geq g_1(|x| + A) \lambda e^{-\lambda - \delta_1^{\alpha+1/2}} \kappa_A \end{aligned}$$

by positivity of g_1 and k and the fact that the Cauchy density is symmetric at zero and is decreasing on $[0, \infty)$.

Now we will use (31) to bound the χ^2-divergence between the densities q_2 and q_1. Write

$$\begin{aligned} \chi^2(q_2, q_1) &= \int_{-\infty}^{\infty} \frac{(q_2(x) - q_1(x))^2}{q_1(x)} dx \\ &= \int_{-A}^{A} \frac{(q_2(x) - q_1(x))^2}{q_1(x)} dx + \int_{\mathbb{R} \setminus [-A, A]} \frac{(q_2(x) - q_1(x))^2}{q_1(x)} dx \\ &= S_1 + S_2. \end{aligned}$$

Using (31), for S_1 we have

$$S_1 \leq \frac{1}{c_\lambda \inf_{|x| \leq A} g_1(x)} \int_{-\infty}^{\infty} (q_2(x) - q_1(x))^2 dx = c_{\lambda, g_1} \int_{-\infty}^{\infty} (q_2(x) - q_1(x))^2 dx,$$

where $c_{\lambda,g_1} > 0$ is a constant. By Parseval's identity the asymptotic behaviour of the integral on the righthand side of the last equality can be studied as follows,

$$\begin{aligned}\int_{-\infty}^{\infty}(q_2(x)-q_1(x))^2dx &= \frac{1}{2\pi}\int_{-\infty}^{\infty}|\phi_{q_2}(t)-\phi_{q_1}(t)|^2dt\\ &= \frac{1}{2\pi}\int_{\mathbb{R}\setminus[-\delta^{-1},\delta^{-1}]} e^{-t^2}\left|e^{\lambda_2(\phi_{g_2}(t)-1)}-e^{\lambda_1(\phi_{g_1}(t)-1)}\right|^2dt\\ &\asymp \frac{1}{2\pi}\int_{\mathbb{R}\setminus[-\delta^{-1},\delta^{-1}]} e^{-t^2}|\delta^{\alpha+1/2}(\phi_{g_1}(t)-1)|^2|1-\phi_H(\delta t)|^2dt.\end{aligned}$$

Using this fact and boundedness of ϕ_H on the whole real line, we get that

$$\int_{-\infty}^{\infty}(q_2(x)-q_1(x))^2dx \lesssim \delta^{2\alpha+1}\int_{1/\delta}^{\infty}e^{-t^2}dt \lesssim \delta^{2\alpha+2}e^{-1/\delta^2}.$$

Thus by taking $\delta = c_\delta(\log n)^{-1/2}$ with a constant $0 < c_\delta < 1$ we can ensure that the righthand side of the above display is $o(n^{-1})$ and consequently also that $S_1 = o(n^{-1})$.

Next we deal with S_2. By (31) and Parseval's identity we have that

$$q_1(x) \geq \frac{c_\lambda}{\pi}\frac{1}{1+(|x|+A)^2}.$$

Therefore by Parseval's identity

$$S_2 \lesssim \int_{\mathbb{R}\setminus[-\delta^{-1},\delta^{-1}]}|[\phi_{q_2}(t)-\phi_{q_1}(t)]'|^2dt + \int_{\mathbb{R}\setminus[-\delta^{-1},\delta^{-1}]}|\phi_{q_2}(t)-\phi_{q_1}(t)|^2dt.$$

Exactly by the same type of an argument as for S_1, after some laborious but easy computations, one can show that $S_2 = o(n^{-1})$, provided $\delta \asymp (\log n)^{-1/2}$ with a small enough constant. Consequently, with such a choice of δ, we have $n\chi^2(q_2,q_1) \to 0$ as $n \to \infty$ and the theorem follows from Lemma 8 of Butucea and Tsybakov (2008b) and (30). □

Proof of Theorem 5 . We use the same alternatives (p_1,f_1) and (p_2,f_2) as in the proof of Theorem 4. One needs to show that the χ^2-divergence between the corresponding probability densities q_1 and q_2 is of order $O(n^{-1})$. The arguments used in the proof of Theorem 4 go through and for that end it suffices to show that

$$\int_{-\infty}^{\infty}|\phi_{q_1}(t)-\phi_{q_2}(t)|^2dt = O(n^{-1}) \tag{32}$$

and that

$$\int_{-\infty}^{\infty}|(\phi_{q_1}(t)-\phi_{q_2}(t))'|^2dt = O(n^{-1}). \tag{33}$$

Observe that for these two integrals to be finite, we need that $\beta > 1/2$, cf. the argument below. We have

$$\begin{aligned}\int_{-\infty}^{\infty}|\phi_{q_2}(t)-\phi_{q_1}(t)|^2dt &= \int_{\mathbb{R}\setminus[-\delta^{-1},\delta^{-1}]}|\phi_Z(t)|^2\left|e^{\lambda_2(\phi_{g_2}(t)-1)}-e^{\lambda_1(\phi_{g_1}(t)-1)}\right|^2dt\\ &\asymp \int_{\mathbb{R}\setminus[-\delta^{-1},\delta^{-1}]}|\phi_Z(t)|^2|\delta^{\alpha+1/2}(\phi_{g_1}(t)-1)|^2|1-\phi_H(\delta t)|^2dt.\end{aligned}$$

Now change the integration variable in the last equality from t to $s = \delta_n t$ and use the fact that for all $s \geq 1$ and for δ_n small enough by assumption on ϕ_Z it holds that $|\phi_Z(s/\delta_n)||s/\delta_n|^{\beta} \leq d_1$, to conclude that the lefthand side of (32) is of order $\delta_n^{2\alpha+2\beta}$. Selecting $\delta_n \asymp n^{-1/(2\alpha+2\beta)}$ then yields (32). A similar argument works in case of (33). We also remark that the condition on ϕ_Z' given in the statement of the theorem is needed to treat (33). Application of Lemma 8 of Butucea and Tsybakov (2008b) as in Theorem 4 concludes the proof. $\square$

Appendix A

Lemma 1. *Let $p^* < 1$ and let $\hat{p}_{ng_n}$ be defined by* (2) *(with p_{ng_n} defined by* (1)*). Under the same conditions as in Theorem 1 (i), we have*

$$\sup_{f\in\Sigma(\alpha,K_\Sigma),p\in[0,p^*]} \mathbb{E}\left[(\hat{p}_{ng_n}-p)^2\right] \lesssim n^{-(2\alpha+1)/(2\alpha+2\beta)},$$

while under conditions of Theorem 1 (ii) the inequality

$$\sup_{f\in\Sigma(\alpha,K_\Sigma),p\in[0,p^*]} \mathbb{E}\left[(\hat{p}_{ng_n}-p)^2\right] \lesssim (\log n)^{-(2\alpha+1)/\beta}$$

holds.

Proof of Lemma 1. Introduce the notation $\sup_{f,p} \equiv \sup_{f\in\Sigma(\alpha,K_\Sigma),p\in[0,p^*]}$. Let n be so large that $p^* < 1-\epsilon_n$, which is possible, because $p^*<1$ and $\epsilon_n \downarrow 0$. Then

$$\mathbb{E}\left[(\hat{p}_{ng_n}-p)^2\right] \leq \mathbb{E}\left[(p_{ng_n}-p)^2\right].$$

This and Theorem 1 entail the desired result. $\square$

Lemma 2. *Under the same conditions as in Theorem 1 and provided $\epsilon_n = (\log 3n)^{-1}$, the inequality*

$$\sup_{f\in\Sigma(\alpha,K_\Sigma),p\in[0,p^*]} \mathbb{E}\left[\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}\right] \lesssim g_n^{2\alpha+1}$$

holds.

Proof. Introduce the sequence

$$\psi_n = 100\sqrt{K_\Sigma}U\left\{\left(\frac{4}{\gamma}\right)^{1/\beta}(\log n)^{-1/\beta}\right\}^{\alpha+1/2} \tag{34}$$

and notice that $\psi_n = 100\sqrt{K_\Sigma}Uh_n^{\alpha+1/2}$ in the supersmooth case, i.e in the setting of Theorem 1 (ii). The constants in the definition of ψ_n are rather arbitrary, but they suffice for our purposes. Notice that on the set $\{|\hat{p}_{ng_n}-p| \leq \psi_n\}$ for all n large enough the inequality

$$|1-\hat{p}_{ng_n}| \geq 1-p^*-\psi_n$$

holds, because $\psi_n \to 0$. We have

$$\begin{aligned}\mathbb{E}\left[\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}\right] &= \mathbb{E}\left[\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}1_{[|\hat{p}_{ng_n}-p|\leq\psi_n]}\right]\\ &\quad+\mathbb{E}\left[\frac{(\hat{p}_{ng_n}-p)^2}{(1-\hat{p}_{ng_n})^2(1-p)^2}1_{[|\hat{p}_{ng_n}-p|>\psi_n]}\right]\\ &\lesssim \mathbb{E}\left[(\hat{p}_{ng_n}-p)^2\right]\\ &\quad+\frac{1}{\epsilon_n^2}P(|\hat{p}_{ng_n}-p|>\psi_n)\end{aligned}$$

$$\lesssim g_n^{2\alpha+1} + \frac{1}{\epsilon_n^2} P(|\hat{p}_{ng_n} - p| > \psi_n),$$

where in the last inequality we used Lemma 1 and Theorem 1. It is easy to see that for all $f \in \Sigma(\alpha, K_\Sigma)$ and $p \in [0, p^*]$ the constants in this chain of inequalities can be made independent of a particular f and a particular p. Then applying Lemma 3 and taking supremum over $f \in \Sigma(\alpha, K_\Sigma)$ and $p \in [0, p^*]$ on the righthand side of the last equality establishes the desired result, because

$$\sup_{p,f} \left(\frac{1}{\epsilon_n^2} P(|\hat{p}_{ng_n} - p| > \psi_n) \right) = o(g_n^{2\alpha+1})$$

holds under our conditions on ϵ_n and g_n. □

Lemma 3. *Define the sequence ψ_n by (34) and let $\epsilon_n = (\log 3n)^{-1}$. Let $\hat{p}_{ng_n}$ be defined by (2) (with p_{ng_n} defined by (1)). Under the same conditions as in Theorem 1 (i) we have*

$$\sup_{f \in \Sigma(\alpha, K_\Sigma), p \in [0,p^*]} P(|\hat{p}_{ng_n} - p| > \psi_n) \lesssim \frac{1}{\psi_n g_n^\beta} \exp\left(-const \times n g_n^{2\beta}\right) + \exp\left(-const' \times \psi_n^2 g_n^{2\beta} n\right),$$

while under those in Theorem 1 (ii) it holds that

$$\sup_{f \in \Sigma(\alpha, K_\Sigma), p \in [0,p^*]} P(|\hat{p}_{ng_n} - p| > \psi_n) \lesssim \frac{e^{1/(\gamma g_n^\beta)}}{\psi_n} \exp\left(-const \times n e^{-2/(\gamma g_n^\beta)}\right) + \exp\left(-const' \times \psi_n^2 e^{-2/(\gamma g_n^\beta)} n\right)$$

for the case when $\beta_0 \geq 0$, and

$$\sup_{f \in \Sigma(\alpha, K_\Sigma), p \in [0,p^*]} P(|\hat{p}_{ng_n} - p| > \psi_n) \lesssim \frac{g_n^{\beta_0} e^{1/(\gamma g_n^\beta)}}{\psi_n} \exp\left(-const \times n g_n^{-2\beta_0} e^{-2/(\gamma g_n^\beta)}\right) + \exp\left(-const \times \psi_n^2 g_n^{-2\beta_0} e^{-2/(\gamma g_n^\beta)} n\right)$$

for the case when $\beta_0 < 0$. Here const and const′ are some universal constants (not necessarily the same in all three cases) independent of particular n, $p \in [0, p^*]$ and $f \in \Sigma(\alpha, K_\Sigma)$.

Proof. In this proof we continue numbering of the terms from the proof of Theorem 2, because it is the proof of Theorem 2 where this lemma finds its primary use. Observe that

$$\begin{aligned} \mathrm{P}(|\hat{p}_{ng_n} - p| > \psi_n) &\leq \mathrm{P}(|\mathbb{E}[\hat{p}_{ng_n}] - p| > \psi_n/2) + \mathrm{P}(|\hat{p}_{ng_n} - \mathbb{E}[\hat{p}_{ng_n}]| > \psi_n/2) \\ &= T_{15} + T_{16}. \end{aligned}$$

We have

$$\begin{aligned} |\mathbb{E}[\hat{p}_{ng_n}] - p| &\leq |\mathbb{E}[p_{ng_n}] - p| + |\mathbb{E}[\hat{p}_{ng_n} - p_{ng_n}]| \\ &\leq |\mathbb{E}[p_{ng_n}] - p| + |\mathbb{E}[(1 - \epsilon_n - p_{ng_n}) 1_{[p_{ng_n} > 1 - \epsilon_n]}]| \end{aligned}$$

$$
\begin{aligned}
&+ |\mathbb{E}\,[(-1+\epsilon_n - p_{ng_n})1_{[p_{ng_n}<-1+\epsilon_n]}]| \\
&\leq \frac{1}{\sqrt{2}}\sqrt{K_\Sigma} U g_n^{\alpha+1/2} \\
&+ \mathbb{E}\,[|1-\epsilon_n - p_{ng_n}|1_{[p_{ng_n}>1-\epsilon_n]}] \\
&+ \mathbb{E}\,[|-1+\epsilon_n - p_{ng_n}|1_{[p_{ng_n}<-1+\epsilon_n]}] \\
&= T_{17} + T_{18} + T_{19}.
\end{aligned}
$$

We put the study of T_{17} aside for a while and consider the other two terms. Since T_{18} and T_{19} can be studied in the similar manner, we consider only T_{18}. Our goal is to show that T_{18} (and by extension T_{19}) is negligible in comparison to T_{17}. We have

$$
T_{18} \leq \left(1+\epsilon_n + \frac{1}{2}\int_{-1}^{1} \frac{|\phi_u(t)|}{|\phi_Z(t/g_n)|}dt\right) \mathrm{P}(p_{ng_n} > 1-\epsilon_n).
$$

The righthand side in both cases of the ordinary smooth or supersmooth Z is of smaller order than T_{17}, which can be seen by employing the arguments on pp. 1265-1266 from Fan (1991) used to bound the integral on the righthand side of the above display and by the exponential bounds on $\mathrm{P}(p_{ng_n} > 1-\epsilon_n)$, which we formulate separately in Lemma 4. With our conditions on g_n these bounds imply that $\sup_{p,f} T_{18}$ is of lower order than T_{17}. The same is true for $\sup_{p,f} T_{19}$. As a consequence, $\sup_{p,f}(T_{18}+T_{19}) < T_{17}$ for all n large enough. Thus $T_{15}=0$, provided n is large enough, because $T_{17} < \psi_n/4$ for all n large enough, and in fact $\sup_{p,f} T_{15} = 0$ for all n large enough.

It remains to study T_{16}. We have

$$
\begin{aligned}
T_{16} &\leq \mathrm{P}(|\hat{p}_{ng_n} - p_{ng_n}| > \psi_n/4) + \mathrm{P}(|p_{ng_n} - \mathbb{E}\,[\hat{p}_{ng_n}]| > \psi_n/4) \\
&\leq \mathrm{P}(|\hat{p}_{ng_n} - p_{ng_n}| > \psi_n/4) + \mathrm{P}(|p_{ng_n} - \mathbb{E}\,[p_{ng_n}]| > \psi_n/8) \\
&+ \mathrm{P}(|\mathbb{E}\,[p_{ng_n}] - \mathbb{E}\,[\hat{p}_{ng_n}]| > \psi_n/8) \\
&= T_{20} + T_{21} + T_{22}.
\end{aligned}
$$

Notice that

$$
\begin{aligned}
T_{20} \leq\ & \mathrm{P}(|1-\epsilon_n - p_{ng_n}|1_{[p_{ng_n}>1-\epsilon_n]} > \psi_n/8) \\
&+ \mathrm{P}(|-1+\epsilon_n - p_{ng_n}|1_{[p_{ng_n}<-1+\epsilon_n]} > \psi_n/8).
\end{aligned}
$$

We consider e.g. the first term on the righthand side. It is bounded by

$$
\frac{8}{\psi_n}\left(1-\epsilon_n + \frac{1}{2}\int_{-1}^{1} \frac{|\phi_u(t)|}{|\phi_Z(t/g_n)|}dt\right) \mathrm{P}(p_{ng_n} > 1-\epsilon_n).
$$

Next, as we did above, we use the order bound on the integral on the righthand side, cf. pp. 1265-1266 in Fan (1991), and the exponential bounds on $P(p_{ng_n} > 1-\epsilon_n)$ from (35) and (36) from Lemma 4 to bound the first term in the upper bound on T_{20}. Similar reasoning applies to the second term in the upper bound on T_{20}. There we use Lemma 5. These bounds give the first term on the righthand side of the three different formulae in the statement of the lemma.

To bound T_{21}, we apply the exponential inequalities from Lemma 6. The terms on the righthand side will then give the second terms in the three formulae on the righthand side in the statement of the lemma.

Finally, we turn to T_{22}. Our goal is to show that there exists n' independent of p and f, such that for all $n \geq n'$ we have $T_{22} = 0$. It holds that

$$\begin{aligned}|\mathbb{E}[p_{ng_n}] - \mathbb{E}[\hat{p}_{ng_n}]| \leq \mathbb{E}[|p_{ng_n} - 1 + \epsilon_n| 1_{[p_{ng_n} > 1-\epsilon_n]}] \\ + \mathbb{E}[|p_{ng_n} + 1 - \epsilon_n| 1_{[p_{ng_n} < 1-\epsilon_n]}].\end{aligned}$$

As the arguments for both terms on the righthand side are similar, we consider only the first term. We have

$$\mathbb{E}[|p_{ng_n} - 1 + \epsilon_n| 1_{[p_{ng_n} > 1-\epsilon_n]}] \leq \left(1 + \epsilon_n + \frac{1}{2}\int_{-1}^{1} \frac{|\phi_u(t)|}{|\phi_Z(t/g_n)|} dt\right) \mathrm{P}(p_{ng_n} > 1 - \epsilon_n).$$

By Lemmas 4 and 5 and the argument as on pp. 1265-1266 of Fan (1991) the righthand side is negligible compared to ψ_n and it follows that T_{22} is zero for all large enough n. In fact n' can be found, such that this holds true uniformly in p and f for all $n \geq n'$. Gathering all the above bounds entails the statement of the lemma. $\square$

Lemma 4. *Let p_{ng_n} be defined by* (1)*. Under the conditions of Theorem 1 (i) we have*

$$\sup_{p \in [0,p^*], f \in \Sigma(\alpha, K_\Sigma)} \mathrm{P}(p_{ng_n} > 1 - \epsilon_n) \lesssim \exp\left(-const \times n g_n^{2\beta}\right), \tag{35}$$

while under conditions of Theorem 1 (ii) we have

$$\sup_{p \in [0,p^*], f \in \Sigma(\alpha, K_\Sigma)} \mathrm{P}(p_{ng_n} > 1 - \epsilon_n) \lesssim \begin{cases} \exp\left(-const \times n e^{-2/(\gamma g_n^{\beta})}\right), & \text{if } \beta_0 \geq 0, \\ \exp\left(-const \times n g_n^{-2\beta_0} e^{-2/(\gamma g_n^{\beta})}\right), & \text{if } \beta_0 < 0. \end{cases} \tag{36}$$

Here $const$ is a universal constant independent of particular $n, p \in [0, p^*]$ and $f \in \Sigma(\alpha, K_\Sigma)$.

Proof. We have

$$\begin{aligned}\mathrm{P}(p_{ng_n} > 1 - \epsilon_n) &= \mathrm{P}(p_{ng_n} - \mathbb{E}[p_{ng_n}] > 1 - \epsilon_n - \mathbb{E}[p_{ng_n}]) \\ &\leq \mathrm{P}(|p_{ng_n} - \mathbb{E}[p_{ng_n}]| > 1 - \epsilon_n - \mathbb{E}[p_{ng_n}]) \\ &= \mathrm{P}\left(\left|\sum_{j=1}^{n} U_n\left(\frac{-X_j}{g_n}\right) - \mathbb{E}\left[\sum_{j=1}^{n} U_n\left(\frac{-X_j}{g_n}\right)\right]\right| > n\frac{(1 - \epsilon_n - \mathbb{E}[p_{ng_n}])}{\pi}\right),\end{aligned}$$

where

$$U_n(x) = \frac{1}{2\pi}\int_{-1}^{1} e^{-itx} \frac{\phi_u(t)}{\phi_Z(t/g_n)} dt.$$

Under the conditions of Theorem 1 (i) we have

$$|U_n(x)| \leq \frac{C}{2\pi}\frac{1}{g_n^{\beta}},$$

while under those of Theorem 1 (ii) the inequality

$$|U_n(x)| \leq \begin{cases} \frac{C'}{2\pi} e^{1/(\gamma g_n^{\beta})}, & \text{if } \beta_0 \geq 0, \\ \frac{C''}{2\pi} g_n^{\beta_0} e^{1/(\gamma g_n^{\beta})}, & \text{if } \beta_0 < 0 \end{cases}$$

holds. Here $C, C^{'}$ and C'' are some constants independent of n. By (18) we have

$$|\mathbb{E}[p_{ng_n}]| \leq |\mathbb{E}[p_{ng_n}] - p| + p \leq p^* + \frac{1}{\sqrt{2}}\sqrt{K_\Sigma} U g_n^{\alpha+1/2}. \tag{37}$$

By taking n_0 so large that for all $n \geq n_0$

$$p^* + \frac{1}{\sqrt{2}}\sqrt{K_\Sigma} U g_n^{\alpha+1/2} < 1 - \epsilon_n \tag{38}$$

holds, one can ensure that uniformly in f and p, $1 - \epsilon_n - \mathbb{E}[p_{ng_n}] > 0$ for $n \geq n_0$. Then by Hoeffding's inequality, see Lemma A.4 on p. 198 of Tsybakov (2009), we obtain

$$\mathrm{P}(p_{ng_n} > 1 - \epsilon_n) \leq 2\exp\left(-2\frac{(1 - \epsilon_n - \mathbb{E}[p_{ng_n}])^2}{C^2} n g_n^{2\beta}\right)$$

for the setting of Theorem 1 (i), and

$$\mathrm{P}(p_{ng_n} > 1 - \epsilon_n) \leq \begin{cases} 2\exp\left(-2\frac{(1-\epsilon_n - \mathbb{E}[p_{ng_n}])^2}{(C')^2} n e^{-2/(\gamma g_n^\beta)}\right), & \text{if } \beta_0 \geq 0, \\ 2\exp\left(-2\frac{(1-\epsilon_n - \mathbb{E}[p_{ng_n}])^2}{(C'')^2} n g_n^{-2\beta_0} e^{-2/(\gamma g_n^\beta)}\right), & \text{if } \beta_0 < 0 \end{cases}$$

for the setting of Theorem 1 (ii). Since

$$1 - \epsilon_n - \mathbb{E}[p_{ng_n}] \geq 1 - \epsilon_n - p^* - \frac{1}{\sqrt{2}}\sqrt{K_\Sigma} U g_n^{\alpha+1/2} > 0 \tag{39}$$

for all n large enough and uniformly in f and p, see (37), there exists a constant *const* independent of $n, p \in [0, p^*]$ and $f \in \Sigma(0, K_\Sigma)$, such that

$$\sup_{p,f} \mathrm{P}(p_{ng_n} > 1 - \epsilon_n) \lesssim \exp\left(-\mathit{const} \times n g_n^{2\beta}\right)$$

for the setting of Theorem 1 (i), and

$$\sup_{p,f} \mathrm{P}(p_{ng_n} > 1 - \epsilon_n) \lesssim \begin{cases} \exp\left(-\mathit{const} \times n e^{-2/(\gamma g_n^\beta)}\right), & \text{if } \beta_0 \geq 0, \\ 2\exp\left(-\mathit{const} \times n g_n^{-2\beta_0} e^{-2/(\gamma g_n^\beta)}\right), & \text{if } \beta_0 < 0 \end{cases}$$

for the setting of Theorem 1 (ii). This concludes the proof. □

Lemma 5. *Let p_{ng_n} be defined by* (1). *Under the conditions of Theorem 1 (i) we have*

$$\sup_{p \in [0,p^*], f \in \Sigma(\alpha, K_\Sigma)} \mathrm{P}(p_{ng_n} < -1 + \epsilon_n) \lesssim \exp\left(-\mathit{const} \times n g_n^{2\beta}\right),$$

while under conditions of Theorem 1 (ii) we have

$$\sup_{p \in [0,p^*], f \in \Sigma(\alpha, K_\Sigma)} \mathrm{P}(p_{ng_n} < -1 + \epsilon_n) \lesssim \begin{cases} \exp\left(-\mathit{const} \times n e^{-2/(\gamma g_n^\beta)}\right), & \text{if } \beta_0 \geq 0, \\ \exp\left(-\mathit{const} \times n g_n^{-2\beta_0} e^{-2/(\gamma g_n^\beta)}\right), & \text{if } \beta_0 < 0. \end{cases}$$

Here const is a universal constant independent of particular $n, p \in [0, p^*]$ and $f \in \Sigma(\alpha, K_\Sigma)$.

Proof. The proof is analogous to the proof of Lemma 4 and is therefore omitted. □

Lemma 6. *Let p_{ng_n} be defined by* (1). *Under the conditions of Theorem 1 (i) we have*

$$\sup_{p \in [0,p^*, f \in \Sigma(\alpha, K_\Sigma)]} P(|p_{ng_n} - \mathbb{E}[p_{ng_n}]| > \psi_n/8) \lesssim \exp\left(-\mathit{const}' \times \psi_n^2 n g_n^{2\beta}\right), \tag{40}$$

while under conditions of Theorem 1 (ii)

$$\sup_{p\in[0,p^*,f\in\Sigma(\alpha,K_\Sigma)]} P(|p_{ng_n} - \mathbb{E}\,[p_{ng_n}]| > \psi_n/8) \lesssim \exp\left(-const' \times \psi_n^2 n e^{2/(\gamma g_n^\beta)}\right) \tag{41}$$

holds. Here $const'$ is a universal constant independent of particular $n, p \in [0,p^*]$ and $f \in \Sigma(\alpha, K_\Sigma)$.

Proof. These inequalities can be established by using Hoeffding's inequality in the same way as the exponential bounds on $P(p_{ng_n} > 1 - \epsilon_n)$ from Lemma 4. □

References

L.D. Brown, M.G. Low and L.H. Zhao. Superefficiency in nonparametric function estimation. *Ann. Statist.*, 25:2607-2625, 1997.

C. Butucea and C. Matias. Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, 11:309–340, 2005.

C. Butucea and A.B. Tsybakov. Sharp optimality for density deconvolution with dominating bias, I. *Theory Probab. Appl.*, 52:111-128, 2008.

C. Butucea and A.B. Tsybakov. Sharp optimality for density deconvolution with dominating bias, II. *Theory Probab. Appl.*, 52:237–249, 2008.

T.T. Cai and J. Jin. Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. *Ann. Statist.*, 38:100–145, 2010.

S.X. Chen, A. Delaigle and P. Hall. Nonparametric estimation for a class of Lévy processes. *J. Econometrics*, doi:10.1016/j.jeconom.2009.12.005, 2010.

A. Delaigle and P. Hall. On the optimal kernel choice for deconvolution. *Statist. Probab. Lett.*, 76:15941602, 2006.

B. van Es, S. Gugushvili and P. Spreij. Deconvolution for an atomic distribution. *Electron. J. Stat.*, 2:265–297, 2008.

J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19:1257–1272, 1991.

I.K. Glad, N.L. Hjort and N.G. Ushakov. Correction of density estimators that are not densities. *Scand. J. Statist.*, 30:415–427, 2003.

S. Gugushvili. *Nonparametric Inference for Partially Observed Lévy Processes*. PhD thesis, Universiteit van Amsterdam, 2008.

W. Jiang and C-H. Zhang. General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.*, 37:1647-1684, 2009.

J. Johannes. Deconvolution with unknown error distribution. *Ann. Statist.* 37:2301–2323, 2009.

M. Lee, H. Shen, C. Burch and J.S. Marron. Direct deconvolution density estimation of a mixture distribution motivated by mutation effects distribution. *J. Nonparametr. Stat.*, 22:1–22, 2010.

T.L. McMurry and D.N. Politis. Nonparametric regression with infinite order flat-top kernels. *J. Nonparametr. Stat.*, 16:549–562, 2004.

A. Meister. Density estimation with normal measurement error with unknown variance. *Stat. Sinica* 16:195–211, 2006.

A. Meister. *Deconvolution Problems in Nonparametric Statistics*. Springer, Berlin, 2009.

A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, 2009.

Department of Mathematics, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
E-mail address: `s.gugushvili@vu.nl`

Korteweg-de Vries Institute for Mathematics, Universiteit van Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
E-mail address: `a.j.vanes@uva.nl`

Korteweg-de Vries Institute for Mathematics, Universiteit van Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
E-mail address: `spreij@uva.nl`