

Deconvolution for an atomic distribution : rates of convergence

Citation for published version (APA):

Gugushvili, S., Es, van, B., & Spreij, P. J. C. (2010). *Deconvolution for an atomic distribution : rates of convergence*. (Report Eurandom; Vol. 2010034). Eurandom.

Document status and date:

Published: 01/01/2010

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

EURANDOM PREPRINT SERIES
2010-034

**DECONVOLUTION FOR AN ATOMIC DISTRIBUTION:
RATES OF CONVERGENCE**

S. Gugushvili, B. van Es, P. Spreij
ISSN 1389-2355

DECONVOLUTION FOR AN ATOMIC DISTRIBUTION: RATES OF CONVERGENCE

SHOTA GUGUSHVILI, BERT VAN ES, AND PETER SPREIJ

ABSTRACT. Let X_1, \dots, 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 $Y = UV$, where U and V are independent, U 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, \dots, 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 and show that the estimators are rate-optimal.

1. INTRODUCTION

Let X_1, \dots, 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 $Y = UV$, where U and V are independent, U 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, \dots, 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 might arise in a number of practical situations, some of which are mentioned in van Es et al. (2008) and Lee et al. (2010). For instance, suppose that a measurement device is used to measure some quantity of interest. Let it have a probability p of failure to detect this quantity, 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 density f

Date: July 13, 2010.

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.

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 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.

Another example comes from the evolutionary biology, see Section 4 in Lee et al. (2010) for background and additional details, and deals with estimation of the distribution of mutation effects on fitness of virus lineage. Here we mention only that the fact that silent mutations can occur leads in this case 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.

We move to the construction of estimators of p and f . 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). 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

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

where a number $g_n > 0$ denotes a bandwidth, ϕ_u is the Fourier transform of a kernel function u 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

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

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. $u(x) = \sin(x)/(\pi x)$. In general p_{ng_n} might take on negative values, even though for large n the probability of this event will be small. At any rate this is of minor

importance, because we can always truncate p_{ng_n} from below at zero, i.e. 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} :

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

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

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

where $0 < \epsilon_n < 1$ and $\epsilon_n \downarrow 0$ at a suitable rate. Notice that $|\hat{p}_{ng_n}| \leq 1 - \epsilon_n$. As in van Es et al. (2008), we propose the following estimator of f ,

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

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 $\overline{f_{nh_{ng_n}}(x)} = f_{nh_{ng_n}}(x)$ and hence $f_{nh_{ng_n}}(x)$ 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 $f_{nh_{ng_n}}$ 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 $f_{nh_{ng_n}}$ 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 $f_{nh_{ng_n}}$ 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 $f_{nh_{ng_n}}(x)$ might become negative for some $x \in \mathbb{R}$, a little drawback often shared by kernel-type estimators in deconvolution problems. If this is the case, then some correction method can be used, for instance one can define $f_{nh_{ng_n}}^+(x) = \max(0, f_{nh_{ng_n}}(x))$, as this does not increase the pointwise risk of the estimator. Furthermore, $f_{nh_{ng_n}}^+$ can be rescaled to integrate to one and thus can be turned into a probability density. We do not pursue these questions any further.

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 from van Es et al. (2008), where the asymptotic normality of the estimators p_{ng_n} and $f_{nh_{ng_n}}$ is established. However, our 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. See the next section for details. 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 the lower risk bound for estimation of f , which shows that our estimator is rate-optimal over an appropriate functional class. Our final result is a lower bound for estimation of p for the case when Z is normally distributed. This lower bound entails rate-optimality of p_{ng_n} .

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

$$(4) \quad d_0|t|^{-\beta} \leq |\phi_Z(t)| \leq d_1|t|^{-\beta}, \text{ as } |t| \rightarrow \infty$$

where d_0, d_1 and β are some strictly positive constants.

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

$$(5) \quad d_0|t|^{\beta_0}e^{-|t|^{\beta}/\gamma} \leq |\phi_Z(t)| \leq d_1|t|^{\beta_1}e^{-|t|^{\beta}/\gamma} \text{ as } |t| \rightarrow \infty,$$

where β_0 and β_1 are some real constants and d_0, d_1, β and γ are some strictly positive constants.

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

Condition 3. Define two classes of target densities f as

$$(6) \quad \Sigma_1(\alpha, K_\Sigma) = \left\{ f : \int_{-\infty}^{\infty} |\phi_f(t)|(1 + |t|^\alpha)dt \leq K_\Sigma \right\}$$

and

$$(7) \quad \Sigma_2(\alpha, L_\Sigma) = \left\{ f : \int_{-\infty}^{\infty} |\phi_f(t)|^2(1 + |t|^{2\alpha})dt \leq L_\Sigma \right\},$$

and let $\Sigma(\alpha, K_\Sigma, L_\Sigma) = \Sigma_1(\alpha, K_\Sigma) \cap \Sigma_2(\alpha, L_\Sigma)$. Here α, K_Σ and L_Σ are some strictly positive numbers.

Conditions of this type are typical in nonparametric curve estimation problems, cf. p. 25 in Tsybakov (2009) or p. 34 in Meister (2009), and for an integer α condition (6) is roughly equivalent to the assumption that f is α times differentiable. At the same time (7) puts some restriction on the L_2 norms of f and $f^{(\alpha)}$. 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.

In the sequel we will use the symbols \lesssim and \gtrsim , meaning respectively less or equal, 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 in the paper, is given in Section 3. In order to keep our notation compact, instead of writing the expectation under the parameter pair (p, f) as $\mathbb{E}_{p,f}[\cdot]$, we will simply write $\mathbb{E}[\cdot]$.

Theorem 1. *Let a kernel 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*

$$(8) \quad \int_{-1}^1 \phi_u(t) dt = 2, \quad \left| \frac{\phi_u(t)}{t^\alpha} \right| \leq U,$$

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 $\lim_{t \rightarrow 0} \phi_u(t)t^{-\alpha}$, which we assume to exist. Then

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

$$(9) \quad \sup_{f \in \Sigma_1(\alpha, K_\Sigma), p \in [0, 1]} \mathbb{E}[(p_{ng_n} - p)^2] \lesssim n^{-(2\alpha+2)/(2\alpha+2\beta+1)},$$

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

$$(10) \quad \sup_{f \in \Sigma_1(\alpha, K_\Sigma), p \in [0, 1]} \mathbb{E}[(p_{ng_n} - p)^2] \lesssim (\log n)^{-(2\alpha+2)/\beta}.$$

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. However, see Theorem 4 below, where for a practically important case of a normally distributed Z by establishing the lower bound for estimation of p we show that the slow convergence rate is intrinsic to the problem and is not a quirk of our particular estimator.

Next we study the asymptotic behaviour of the estimator $f_{nh_n g_n}$ of f . We selected the mean integrated square error as a criterion of its performance. The following theorem holds.

Theorem 2. *Let a kernel u and the bandwidth g_n satisfy the assumptions in Theorem 1. Furthermore, let a kernel w be such that its Fourier transform is symmetric, real-valued and is supported on $[-1, 1]$, $\phi_w(0) = 1$ and*

$$(11) \quad |\phi_w(s) - 1| \leq W|s|^\alpha, \quad \int_{-1}^1 |\phi_w(t)|^2 dt < \infty,$$

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

(i) under Condition 1, by selecting $h_n = g_n = dn^{-1/(2\alpha+2\beta+1)}$ for some $d > 0$ and $\epsilon_n \downarrow 0$ such that $h_n/\epsilon_n^2 \rightarrow 0$, we have

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

(ii) under Condition 2, by selecting $h_n = g_n = (4/\gamma)^{1/\beta}(\log n)^{-1/\beta}$ and $\epsilon_n \downarrow 0$ such that $h_n/\epsilon_n^2 \rightarrow 0$, we have

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

where the sequence $\epsilon_n \downarrow 0$ is the same as in (2).

The condition $h_n = g_n$ 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 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 considered.

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. Naturally, a discussion on the optimality of convergence rates of the estimators $f_{nh_n g_n}$ and p_{ng_n} is in order. Let \tilde{f}_n denote an arbitrary estimator of f based on a sample X_1, \dots, X_n . Consider

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

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

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

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^* . In particular, the following result holds.

Theorem 3. *Let \tilde{f}_n denote any estimator of f based on a sample X_1, \dots, X_n . Then*

(i) *under Condition 1 we have*

$$(14) \quad \inf_{\tilde{f}_n} \sup_{f \in \Sigma_2(\alpha, L_\Sigma), p \in [0, p^*]} \mathbb{E} \left[\int_{-\infty}^{\infty} (\hat{f}(x) - f(x))^2 dx \right] \gtrsim n^{-2\alpha/(2\alpha+2\beta+1)};$$

(ii) *under Condition 2 the inequality*

$$(15) \quad \inf_{\tilde{f}_n} \sup_{f \in \Sigma_2(\alpha, L_\Sigma), p \in [0, p^*]} \mathbb{E} \left[\int_{-\infty}^{\infty} (\hat{f}(x) - f(x))^2 dx \right] \gtrsim (\log n)^{-2\alpha/\beta}$$

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.

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 (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 \tilde{p}_n denote any estimator of p based on a sample X_1, \dots, X_n . Then*

$$(16) \quad \inf_{\tilde{p}_n} \sup_{f \in \Sigma_1(\alpha, L_\Sigma), p \in [0,1]} \mathbb{E} [(\tilde{p}_n - p)^2] \gtrsim (\log n)^{-(\alpha+1)}$$

holds.

A slight modification of the proof shows that the same lower bound also holds for the case when the supremum is taken over $p \in [0, p^*]$ and not only $p \in [0, 1]$. A consequence of this theorem and (10) is that our estimator p_{ng_n} is rate-optimal for the case when Z follows the normal distribution.

3. PROOFS

Proof of Theorem 1. The proof uses some arguments from Fan (1991). To make the notation less cumbersome, let $\sup_{f,p} \equiv \sup_{f \in \Sigma_1(\alpha, K_\Sigma), p \in [0,1]}$. We first prove (i). We have

$$(17) \quad \sup_{f,p} \mathbb{E} [(p_{ng_n} - p)^2] \leq \sup_{f,p} (\mathbb{E} [p_{ng_n}] - p)^2 + \sup_{f,p} \text{Var} [p_{ng_n}].$$

Observe that

$$(18) \quad |\mathbb{E} [p_{ng_n}] - p| = \frac{1-p}{2} \left| \int_{-1}^1 \phi_f \left(\frac{t}{g_n} \right) \phi_u(t) dt \right| \leq \frac{1}{2} K_\Sigma U g_n^{\alpha+1},$$

where we used (8), as well as (6). Therefore

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

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

$$(20) \quad \begin{aligned} \text{Var} [p_{ng_n}] &= \frac{1}{4} \frac{1}{n} \text{Var} \left[\int_{-1}^1 e^{itX_1/g_n} \frac{\phi_u(t)}{\phi_Z(t/g_n)} dt \right] \\ &\leq \frac{1}{4} \frac{1}{n} \left(\int_{-1}^1 \left| \frac{\phi_u(t)}{\phi_Z(t/g_n)} \right| dt \right)^2. \end{aligned}$$

Let M be a large enough (but fixed) constant. Suppose also that $n \geq n_0$ and $Mg_n < 1$ for all $n \geq n_0$. If M is selected appropriately and n_0 is large enough, then we have

$$(21) \quad |\phi_Z(t/g_n)| \geq \frac{d_0}{2} \left| \frac{t}{g_n} \right|^{-\beta}$$

for all $Mg_n \leq |t| \leq 1$, which follows from Condition 1. Moreover, for $|t| \leq Mg_n$

$$(22) \quad |\phi_Z(t/g_n)| \geq \inf_{s \in [-M, M]} |\phi_Z(s)| > 0,$$

because ϕ_Z does not vanish on the whole real line. Now write

$$(23) \quad \int_{-1}^1 \left| \frac{\phi_u(t)}{\phi_Z(t/g_n)} \right| dt = \left(\int_{[-Mg_n, Mg_n]} + \int_{[-1,1] \setminus [-Mg_n, Mg_n]} \right) \left| \frac{\phi_u(t)}{\phi_Z(t/g_n)} \right| dt.$$

Formulae (21)–(23) imply that

$$(24) \quad \int_{-1}^1 \left| \frac{\phi_u(t)}{\phi_Z(t/g_n)} \right| dt \leq C \frac{1}{g_n^\beta},$$

where C does not depend on n . This and (20) entail that

$$(25) \quad \sup_{f,p} \text{Var} [p_{ng_n}] \lesssim \frac{1}{ng_n^{2\beta}}.$$

Formula (9) is then a consequence of (17), (19), (25) 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. Notice that in this case (20) holds true as well. By the same arguments as in (21)–(23), one can show that

$$(26) \quad \int_{-1}^1 \left| \frac{\phi_u(t)}{\phi_Z(t/g_n)} \right| dt \leq \begin{cases} C' e^{1/(\gamma g_n^\beta)}, & \text{if } \beta_0 \geq 0 \\ C' g_n^{\beta_0} e^{1/(\gamma g_n^\beta)}, & \text{if } \beta_0 < 0, \end{cases}$$

where the constant C' does not depend on n . In either case, because of our choice of g_n , the righthand side of (26) is of order $o(n^{1/3})$. Thus

$$\sup_{f,p} \text{Var} [p_{ng_n}] = o(n^{-2/3}).$$

This together with (17) and (19) proves (10). \square

The following lemma will be used in the proof of Theorem 2.

Lemma 1. *Under the same conditions as in Theorem 1 (i), we have*

$$\sup_{f \in \Sigma_1(\alpha, K_\Sigma), p \in [0, p^*]} \mathbb{E} [(\hat{p}_{ng_n} - p)^2] \lesssim n^{-(2\alpha+2)/(2\alpha+2\beta+1)},$$

while under conditions of Theorem 1 (ii) the inequality

$$\sup_{f \in \Sigma_1(\alpha, K_\Sigma), p \in [0, p^*]} \mathbb{E} [(\hat{p}_{ng_n} - p)^2] \lesssim (\log n)^{-(2\alpha+2)/\beta}$$

holds.

Proof of Lemma 1. Introduce the notation $\sup_{f,p} \equiv \sup_{f \in \Sigma_1(\alpha, K_\Sigma), p \in [0, p^*]}$. Let n be so large that $p^* < 1 - \epsilon_n$, which is possible, because $p^* < 1$ and $\epsilon_n \downarrow 0$. Then

$$\begin{aligned} \mathbb{E} [(\hat{p}_{ng_n} - p)^2] &\leq \mathbb{E} [(p_{ng_n} - p)^2] \\ &= T_1. \end{aligned}$$

Observe that because of (19) and our choice of g_n ,

$$\sup_{f,p} T_1 \lesssim n^{-(2\alpha+2)/(2\alpha+2\beta+1)}$$

in the setting of Theorem 1 (i), and

$$\sup_{f,p} T_1 \lesssim (\log n)^{-(2\alpha+2)/\beta}$$

in the setting of Theorem 1 (ii). This entails the desired result. \square

Proof of Theorem 2. We use the notation $\sup_{f,p} \equiv \sup_{f \in \Sigma(\alpha, K_\Sigma, L_\Sigma), p \in [0, p^*]}$. We have

$$\begin{aligned} \sup_{f,p} \mathbb{E} \left[\int_{-\infty}^{\infty} (f_{nh_n g_n}(x) - f(x))^2 dx \right] &\leq \sup_{f,p} \int_{-\infty}^{\infty} (\mathbb{E} [f_{nh_n g_n}(x)] - f(x))^2 dx \\ &\quad + \sup_{f,p} \int_{-\infty}^{\infty} \text{Var} [f_{nh_n g_n}(x)] dx \end{aligned}$$

$$= T_1 + T_2.$$

Let

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

and introduce

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

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

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

By Parseval's identity and the dominated convergence theorem

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

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 $f \in \Sigma(\alpha, K_\Sigma, L_\Sigma)$. With our choice of the bandwidths h_n and g_n , T_3 is the dominating term in the upper risk bound for the estimator $f_{nh_n g_n}$. The rest of the proof is dedicated to showing that the other terms are negligible. We deal with T_4 . By the c_2 -inequality

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

Consider T_5 . By the Cauchy-Schwarz inequality and a change of the integration variable from x into $v = x/h_n$ we have

$$\begin{aligned} T_5 &\leq \frac{1}{h_n} \int_{-\infty}^{\infty} (w(x))^2 dx \mathbb{E} \left[\frac{(\hat{p}_{ng_n} - p)^2}{(1 - \hat{p}_{ng_n})^2 (1 - p)^2} \right] \\ &\leq \int_{-\infty}^{\infty} (w(x))^2 dx \frac{1}{(1 - p^*)^2} \frac{1}{h_n} \frac{1}{\epsilon_n^2} \mathbb{E} [(\hat{p}_{ng_n} - p)^2], \end{aligned}$$

where we used the fact that $(1 - \hat{p}_{ng_n})^2 \geq \epsilon_n^2$ and $p \leq p^* < 1$ to see the last line. Since $g_n = h_n$, it follows from the proof of Lemma 1 that $\sup_{p,f} T_5 \lesssim g_n^{2\alpha+1}/\epsilon_n^2$. Now let us turn to T_6 . By the Cauchy-Schwarz inequality and (19)

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

By the same arguments as we used for T_5 , the first term in the product in the above display is of order $g_n^{2\alpha+2}/\epsilon_n^2$. The same holds true for its supremum over f and p . Hence it remains to study the second factor in T_6 . We have

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

Notice that by the independence of X_i 's

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

where the function W_n is defined by

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

Let q denote the density of X_1 . Then by Fubini's theorem

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

where we also used the fact that q , being a probability density, integrates to one, as well as Parseval's identity. The integral in the last equality of the above display can be analysed by exactly the same arguments as the integral (23). Thus

$$(27) \quad T_7 \lesssim \begin{cases} \frac{1}{nh_n^{2\beta+1}}, & \text{if } Z \text{ is ordinary smooth,} \\ \frac{1}{nh_n^{\beta}} e^{2/(\gamma h_n^{\beta})}, & \text{if } Z \text{ is supersmooth and } \beta_0 \geq 0, \\ \frac{h_n^{2\beta_0-1}}{n} e^{2/(\gamma h_n^{\beta})}, & \text{if } Z \text{ is supersmooth and } \beta_0 < 0. \end{cases}$$

It also follows that the same order bounds hold for $\sup_{f,p} T_7$. Let us now study T_8 . By Parseval's identity and the fact that $|\phi_Y(t)| \leq 1$, we have

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

Notice that because of (11), $\int_{-1}^1 |\phi_w(t)|^2 dt$ is finite. Combination of the above bounds for T_7 and T_8 entails that $\sup_{f,p} T_6$ is of order $g_n^{2\alpha+1}/\epsilon_n^2$. Therefore T_4 , as well as T_1 , i.e. the supremum of the integrated squared bias is of order $g_n^{2\alpha}$. 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

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

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

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

Since T_9 up to a constant is the same as T_7 , $\sup_{f,p} T_9$ can be bounded as before, see (27). We consider T_{10} . Let $\psi_n = 2K_{\Sigma} U g_n^{\alpha+1}$. Then

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

By the c_2 -inequality

$$\begin{aligned} T_{11} &\lesssim \frac{1}{h_n} \int_{-\infty}^{\infty} (w(x))^2 dx \mathbb{E} \left[\frac{(\hat{p}_{ng_n} - p)^2}{(1 - \hat{p}_{ng_n})^2 (1 - p)^2} \mathbf{1}_{\{|\hat{p}_{ng_n} - p| > \psi_n\}} \right] \\ &\quad + \int_{-\infty}^{\infty} \mathbb{E} \left[(\hat{f}_{nh_n}(x))^2 \frac{(\hat{p}_{ng_n} - p)^2}{(1 - \hat{p}_{ng_n})^2 (1 - p)^2} \mathbf{1}_{\{|\hat{p}_{ng_n} - p| > \psi_n\}} \right] dx \\ &= T_{13} + T_{14}. \end{aligned}$$

Since $T_{13} \lesssim h_n^{-1} \epsilon_n^{-2} \mathbb{E} [(\hat{p}_{ng_n} - p)^2]$, we have $\sup_{f,p} T_{13} \lesssim g_n^{2\alpha+1} / \epsilon_n^2$. As far as T_{14} is concerned, by Fubini's theorem and Parseval's identity

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

Here we used the facts that $|\hat{p}_{ng_n}| \leq 1 - \epsilon_n$ and $p \leq p^* < 1$. Hence

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

in the ordinary smooth case, and

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

in the supersmooth case, see the proof of Theorem 1. We thus have to study $\mathbb{P}(|\hat{p}_{ng_n} - p| > \psi_n)$. Observe that

$$\mathbb{P}(|\hat{p}_{ng_n} - p| > \psi_n) \leq \mathbb{P}(|\mathbb{E} [\hat{p}_{ng_n}] - p| > \psi_n/2) + \mathbb{P}(|\hat{p}_{ng_n} - \mathbb{E} [\hat{p}_{ng_n}]| > \psi_n/2)$$

$$= T_{15} + T_{16}.$$

Similar to the proof of Lemma 1,

$$\begin{aligned} |\mathbb{E}[\hat{p}_{ng_n}] - p| &\leq |\mathbb{E}[p_{ng_n}] - p| + |\mathbb{E}[\hat{p}_{ng_n} - p_{ng_n}]| \\ &\leq |\mathbb{E}[p_{ng_n}] - p| + |\mathbb{E}[(1 - \epsilon_n - p_{ng_n})\mathbf{1}_{[p_{ng_n} > 1 - \epsilon_n]}]| \\ &\quad + |\mathbb{E}[(-1 + \epsilon_n - p_{ng_n})\mathbf{1}_{[p_{ng_n} < -1 + \epsilon_n]}]| \\ &\leq \frac{1}{2}K_{\Sigma}Ug_n^{\alpha+1} \\ &\quad + \mathbb{E}[|1 - \epsilon_n - p_{ng_n}|\mathbf{1}_{[p_{ng_n} > 1 - \epsilon_n]}] \\ &\quad + \mathbb{E}[|-1 + \epsilon_n - p_{ng_n}|\mathbf{1}_{[p_{ng_n} < -1 + \epsilon_n]}] \\ &= T_{17} + T_{18} + T_{19}. \end{aligned}$$

Since T_{18} and T_{19} can be studied in the same manner, we consider only T_{18} . By bounding p_{ng_n} , we have

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

The right-hand side in both cases of the ordinary smooth or supersmooth Z is of smaller order than ψ_n , which can be seen by using (24), (26) and the following reasoning used to bound $\mathbb{P}(p_{ng_n} > 1 - \epsilon_n)$:

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

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) by (24) we have

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

while under those of Theorem 1 (ii)

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

By (18), we have

$$|\mathbb{E}[p_{ng_n}]| \leq |\mathbb{E}[p_{ng_n}] - p| + p \leq p^* + \frac{1}{2}K_{\Sigma}Ug_n^{\alpha+1}.$$

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

$$p^* + \frac{1}{2}K_{\Sigma}Ug_n^{\alpha+1} < 1 - \epsilon_n$$

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

$$\mathbb{P}(p_{ng_n} > 1 - \epsilon_n) \leq 2 \exp\left(-8 \frac{(1 - \epsilon_n - \mathbb{E}[p_{ng_n}])^2}{C^2} n g_n^{2\beta}\right)$$

for the setting of Theorem 1 (i) and

$$\mathbb{P}(p_{ng_n} > 1 - \epsilon_n) \leq \begin{cases} 2 \exp\left(-8 \frac{(1 - \epsilon_n - \mathbb{E}[p_{ng_n}])^2}{(C')^2} n e^{-2/(\gamma g_n^\beta)}\right), & \text{if } \beta_0 \geq 0, \\ 2 \exp\left(-8 \frac{(1 - \epsilon_n - \mathbb{E}[p_{ng_n}])^2}{(C')^2} n g_n^{-2\beta_0} e^{-2/(\gamma g_n^\beta)}\right), & \text{if } \beta_0 < 0 \end{cases}$$

for the setting of Theorem 1 (ii). Since

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

for all n large enough and uniformly in f and p , further bounding yields

$$\mathbb{P}(p_{ng_n} > 1 - \epsilon_n) \leq 2 \exp\left(-8 \frac{(1 - \epsilon_n - p^* - (1/2) K_\Sigma U g_n^{\alpha+1})^2}{C^2} n g_n^{2\beta}\right)$$

for the setting of Theorem 1 (i) and

$$\mathbb{P}(p_{ng_n} > 1 - \epsilon_n) \leq \begin{cases} 2 \exp\left(-8 \frac{(1 - \epsilon_n - p^* - (1/2) K_\Sigma U g_n^{\alpha+1})^2}{(C')^2} n e^{-2/(\gamma g_n^\beta)}\right), & \text{if } \beta_0 \geq 0, \\ 2 \exp\left(-8 \frac{(1 - \epsilon_n - p^* - (1/2) K_\Sigma U g_n^{\alpha+1})^2}{(C')^2} n g_n^{-2\beta_0} e^{-2/(\gamma g_n^\beta)}\right), & \text{if } \beta_0 < 0 \end{cases}$$

for the setting of Theorem 1 (ii). Consequently, T_{18} is of lower order than ψ_n . The same is true for T_{19} .

Thus $T_{15} = 0$, provided n is large enough. In fact, this will hold true uniformly in p and f , which follows from (18). It remains to study T_{16} . This can be done in much the same way as in case of T_{15} , but nevertheless, we provide the complete proof. In fact,

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

Notice that

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

We consider e.g. the first term on the right-hand side. By Chebyshev's inequality it is bounded by

$$\frac{8}{\psi_n} T_{18} = \frac{8}{\psi_n} \left(1 - \epsilon_n + \frac{1}{2} \int_{-1}^1 \frac{|\phi_u(t)|}{|\phi_Z(t/g_n)|} dt\right) \mathbb{P}(p_{ng_n} > 1 - \epsilon_n).$$

The order bound on the latter term, which is also uniform in p and f , can be established just as above by using (24), (26) and an exponential bound on $\mathbb{P}(p_{ng_n} > 1 - \epsilon_n)$, which was proved above. This will be of smaller order than $g_n^{2\alpha}$. To bound T_{21} , we apply an exponential bound on $\mathbb{P}(p_{ng_n} > 1 - \epsilon_n)$. Again, this will be negligible in comparison to $g_n^{2\alpha}$. Finally, we turn to T_{22} . Our goal is to show that for all n large enough and uniformly in p and f , $T_{22} = 0$. We have

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

$$+ \mathbb{E} [|p_{ng_n} + 1 - \epsilon_n| \mathbf{1}_{[p_{ng_n} < 1 - \epsilon_n]}].$$

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

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

Since the right-hand side is negligible compared to ψ_n , it follows that T_{22} is zero for all large enough n and in fact this holds true uniformly in p and f . To complete establishing an upper bound on T_{10} , it remains to study T_{12} . By the c_2 -inequality

$$T_{12} \lesssim \frac{1}{\epsilon_n^2} \frac{1}{h_n} \psi_n^2 \int_{-\infty}^{\infty} (w(x))^2 dx + \frac{1}{\epsilon_n^2} \psi_n^2 \int_{-\infty}^{\infty} \mathbb{E} [(\hat{f}_{nh_n}(x))^2] dx.$$

Since

$$\int_{-\infty}^{\infty} \mathbb{E} [(\hat{f}_{nh_n}(x))^2] dx = \int_{-\infty}^{\infty} \text{Var} [(\hat{f}_{nh_n}(x))^2] dx + \int_{-\infty}^{\infty} (\mathbb{E} [\hat{f}_{nh_n}(x)])^2 dx,$$

it follows from upper bounds on T_7 and T_8 that $T_{12} \lesssim g_n^{2\alpha+1}/\epsilon_n^2$. Combination of the above intermediate results and taking suprema over f and p implies that $\sup_{f,p} T_{10} \lesssim g_n^{2\alpha+1}/\epsilon_n^2$. The statement of the theorem is then a consequence of our choice of h_n and g_n . \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, \dots, X_n are close in the χ^2 -divergence and hence cannot be distinguished well using the observations X_1, \dots, X_n . By Lemma 8 of Butucea and Tsybakov (2008) 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 $\lambda_1 = \lambda + \delta^{\alpha+1}$, where $\lambda > 0$ is a fixed constant and $\delta \downarrow 0$ as $n \rightarrow \infty$. Define $p_1 = e^{-\lambda_1}$ and notice that $p_1 \in [0, 1)$ for all n large enough. 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

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

Assume that the i.i.d. random variables W_j 's have the common density g_1 and the random variable N_{λ_1} 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 $\phi_{f_1} \in$

$\Sigma_1(\alpha, K_\Sigma/2)$. Otherwise we can always consider $\phi_{g_1}(t) = e^{-\alpha'|t|}$ with a fixed and large enough constant $\alpha' > 0$, so that $\phi_{f_1} \in \Sigma_1(\alpha, K_\Sigma/2)$. 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

$$\tau(t) = \frac{\delta^{\alpha+1}}{\lambda_2}(\phi_{g_1}(t) - 1)\phi_H(\delta t).$$

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

$$(28) \quad \int_{\mathbb{R}} g_2(x) dx = \phi_{g_2}(0) = \phi_{g_1}(0) = 1,$$

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

$$\sup_x |g_2(x) - g_1(x)| \leq \frac{1}{2\pi} \int_{\mathbb{R}} |\tau(t)| dt \rightarrow 0$$

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

$$(29) \quad g_2(x) \geq 0, \quad x \in B,$$

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

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

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. Since g_1 is the Cauchy density and $\delta > 0$, the inequality

$$(30) \quad T_1(x) \geq \frac{1}{\pi} \frac{1}{1+x^2}$$

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

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

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

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

which implies that

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

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

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

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

$$(31) \quad |T_2(x)| \leq C' \delta^{\alpha+1} \frac{1}{x^2}$$

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

$$T_3(x) = \frac{1}{x^2} \frac{\delta^{\alpha+3}}{\lambda_2} \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi_H''(\delta t) dt$$

and hence

$$(32) \quad |T_3(x)| \leq C'' \delta^{\alpha+2} \frac{1}{x^2},$$

where the constant C'' does not depend on n and x . Therefore, by gathering (30)–(32), 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 (28), 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, one can arrange that $f_2 \in \Sigma_1(\alpha, K_\Sigma)$, 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

$$(33) \quad |p_2 - p_1| \asymp \delta^{\alpha+1}$$

as $\delta \rightarrow 0$, where \asymp means that two sequences are asymptotically of the same order. Consequently, by Lemma 8 of Butucea and Tsybakov (2008) the lower bound in (16) will be of order $\delta^{2\alpha+2}$, provided we can prove that $n\chi^2(q_2, q_1) \rightarrow 0$ as $n \rightarrow \infty$ for an appropriate $\delta \rightarrow 0$. 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.

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

see p. 86 in Tsybakov (2009).

Notice that we have

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

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

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

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

$$(34) \quad q_1(x) \geq (1 - e^{-\lambda_1}) f_1 * k(x) \geq \kappa_A \lambda e^{-\lambda - \delta_1} g_1(|x| + A) = c_\lambda g_1(|x| + A)$$

for some large enough (but fixed) constant $A > 0$. Here the constant $\kappa_A = \int_{-A}^A k(t) dt$. The inequalities in (34) hold, because

$$\begin{aligned} (1 - e^{-\lambda_1}) f_1 * k(x) &= (1 - e^{-\lambda_1}) \int_{\mathbb{R}} f_1(x - t) k(t) dt \\ &\geq \lambda e^{-\lambda - \delta_1} \int_{\mathbb{R}} g_1(x - t) k(t) dt \\ &\geq \lambda e^{-\lambda - \delta_1} \int_{-A}^A g_1(x - t) k(t) dt \\ &\geq g_1(|x| + A) \lambda e^{-\lambda - \delta_1} \kappa_A \end{aligned}$$

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

Now we will use (34) to bound the χ^2 -divergence between the densities q_2 and q_1 . Write

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

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

Using (34), for S_1 we have

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

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

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

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

$$\int_{\mathbb{R}} (q_2(x) - q_1(x))^2 dx \lesssim \delta^{2\alpha+2} \int_{1/\delta}^{\infty} e^{-t^2} dt \lesssim \delta^{2\alpha+3} e^{-1/\delta^2}.$$

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 (34) 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

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

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 $n\chi^2(q_2, q_1) \rightarrow 0$ as $n \rightarrow \infty$ and the theorem follows from Lemma 8 of Butucea and Tsybakov (2008) and (33). \square

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, II. *Theory Probab. Appl.*, 52:237-249, 2008.
- 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.

- 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.
- 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.
- 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. *Deconvolution Problems in Nonparametric Statistics*. Springer, Berlin, 2009.
- A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, 2009.

DEPARTMENT OF MATHEMATICS, VU UNIVERSITY AMSTERDAM, DE BOELELAAN 1081, 1081 HV AMSTERDAM, THE NETHERLANDS

E-mail address: `shota@few.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`