ON DECONVOLUTION OF DISTRIBUTION FUNCTIONS

BY I. DATTNER[1], A. GOLDENSHLUGER[1] AND A. JUDITSKY

University of Haifa, University of Haifa and Université Grenoble I

The subject of this paper is the problem of nonparametric estimation of a continuous distribution function from observations with measurement errors. We study minimax complexity of this problem when unknown distribution has a density belonging to the Sobolev class, and the error density is ordinary smooth. We develop rate optimal estimators based on direct inversion of empirical characteristic function. We also derive minimax affine estimators of the distribution function which are given by an explicit convex optimization problem. Adaptive versions of these estimators are proposed, and some numerical results demonstrating good practical behavior of the developed procedures are presented.

1. Introduction. In this paper we study the problem of estimating a distribution function in the presence of measurement errors.

Let $X_1, \ldots, X_n$ be a sequence of independent, identically distributed random variables with common distribution F. Suppose that we observe random variables $Y_1, \ldots, Y_n$ given by

$$(1) \qquad Y_j = X_j + \zeta_j, \qquad j = 1, \ldots, n,$$

where ζ_j are i.i.d. random variables, independent of X_j's with the density f_ζ w.r.t. the Lebesgue measure on the real line. The objective is to estimate the value $F(t_0)$ of the distribution function F of X at a given point $t_0 \in \mathbb{R}$ from the observations $Y^n = (Y_1, \ldots, Y_n)$.

By an estimator we mean any measurable function $\widetilde{F} = \widetilde{F}(Y^n)$ of the observations Y^n. We adopt the minimax approach for measuring estimation accuracy. Let $\mathcal{F}$ be a given family of probability distributions on $\mathbb{R}$. Given an estimator $\widetilde{F}$ of $F(t_0)$, we consider two types of maximal over $\mathcal{F}$ risks:

- quadratic risk,

$$\mathrm{Risk}_2[\widetilde{F}; \mathcal{F}] := \sup_{F \in \mathcal{F}} \{\mathrm{E}|\widetilde{F} - F(t_0)|^2\}^{1/2}.$$

- ϵ-risk: given a tolerance level $\epsilon \in (0, 1/2)$ we define

$$\mathrm{Risk}_\epsilon[\widetilde{F}; \mathcal{F}] := \min\Big\{\delta : \sup_{F \in \mathcal{F}} \mathrm{P}[|\widetilde{F} - F(t_0)| > \delta] \le \epsilon\Big\}.$$

An estimator $\widetilde{F}^*$ is said to be *rate optimal* or *optimal in order* with respect to Risk if

$$\mathrm{Risk}[\widetilde{F}^*; \mathcal{F}] \le C \inf_{\widetilde{F}} \mathrm{Risk}[\widetilde{F}; \mathcal{F}],$$

where inf is taken over all possible estimators of $F(t_0)$, and $C < \infty$ is independent of n. We will be particularly interested in the classes of distributions having density with respect to the Lebesgue measure on the real line.

The outlined problem is closely related to the density deconvolution problem that has been extensively studied in the literature; see, for example, [4, 5, 13, 18, 24, 27, 28] and references therein. In these works the minimax rates of convergence have been derived under different assumptions on the error density and on the smoothness of the density to be estimated. Depending on the tail behavior of the characteristic function $\widehat{f}_\zeta$ of ζ the following two cases are usually distinguished:

(i) *ordinary smooth* errors, when the tails of $\widehat{f}_\zeta$ are polynomial, that is,

$$|\widehat{f}_\zeta(\omega)| \asymp |\omega|^{-\beta}, \qquad |\omega| \to \infty,$$

for some $\beta > 0$;

(ii) *supersmooth* errors, when the tails are exponential, that is,

$$|\widehat{f}_\zeta(\omega)| \asymp \exp\{-c|\omega|^\beta\}, \qquad |\omega| \to \infty,$$

for some $c > 0$ and $\beta > 0$.

The afore cited papers derive minimax rates of convergence for different functional classes under ordinary smooth and supersmooth errors.

In contrast to existence of the voluminous literature on density deconvolution, the problem of deconvolution of the distribution function F has attracted much less attention and has been studied in very few papers (see [24], Section 2.7.2, for a recent review of corresponding contributions). A consistent estimator of a distribution function from observations with additive Gaussian measurement errors was developed by [14]. A "plug-in" estimator based on integration of the density estimator in the density deconvolution problem has been studied under moment conditions on F in [28]. The paper [13] also considered the estimator based on integration of the density deconvolution estimator. It was shown there that under a tail condition on F the estimator achieves optimal rates of convergence provided that the errors are supersmooth. For the case of ordinary smooth errors there is a gap between the upper and lower bounds reported in [13] which leaves open the question of constructing optimal estimators. More recently, some minimax rates of estimation of distribution functions in models with measurement errors were

reported in [17]. Note also that [3] considered a general problem of optimal and adaptive estimation of linear functionals $\ell(f) = \int_{-\infty}^{\infty} \phi(t) f(t) \, dt$ in the model (1). However, their results hold only for representative $\phi \in \mathbb{L}_1(\mathbb{R})$ which is clearly not the case in the problem of recovery of distribution function.

The objective of this paper is to develop optimal methods of minimax deconvolution of distribution functions and to answer several questions raised by known results on this problem: Is a smoothness assumption alone on F sufficient in order to secure minimax rates of estimation of the sort $O(n^{-\gamma})$ for $\gamma > 0$ in the case of ordinary smooth errors? Do we need tail or moment conditions on F?

Our contribution is two-fold. First, we characterize the minimax rates of convergence in the case when the unknown distribution belongs to a Sobolev ball, and the observation errors are ordinary smooth. The rates of convergence depend crucially on the relation between the *smoothness index* α of the Sobolev ball and the parameter β [the rate at which the characteristic function of errors tends to zero; see (i) above]. In contrast to the density deconvolution problem, it turns out that there are different regions in the (α, β)-plane where different rates of convergence are attained. We show that in some regions of the (α, β)-plane the minimax rates of convergence are attained by a linear estimator, which is based on direct inversion of the distribution function from the corresponding characteristic function; cf. [17]. It is worth noting that we do not require any additional tail or moment conditions on the unknown distribution. In the case when the parameters of the regularity class of the distribution F are unknown, we also construct an adaptive estimator based on Lepski's adaptation scheme [23]. The ϵ-risk of this estimator is within a $\ln \ln n$-factor of the minimax ϵ-risk.

Second, using recent results on estimating linear functionals developed in [19], we propose minimax and adaptive affine estimators of the cumulative distribution function for a discrete distribution deconvolution problem; see also [6, 9–12] for the general theory of affine estimation. These estimators can be applied to the original deconvolution problem provided that it can be efficiently discretized. By efficient discretization we mean that:

(1) the support of the distributions of X (Y) can be "compactified" [one can point out a compact subset of $\mathbb{R}$ such that the probability of X (Y) being outside this set is "small"] and binned into small intervals;

(2) the class $\mathcal{X}$ of discrete distributions, obtained by the corresponding finite-dimensional cross-section of the class $\mathcal{F}$ of continuous distributions is a computationally tractable convex closed set.[2]

Under these conditions one can efficiently implement the minimax affine estimator for F based on the approach proposed in [19]. This estimator is rate minimax

---

[2]Roughly speaking, a computationally tractable set can be interpreted as a set given by a finite system of inequalities $p_i(x) \leq 0$, $i = 1, \ldots, m$, where p_i are convex polynomials; see, for example, [2], Chapter 4.

with respect to Risk$_\epsilon$ (within a factor ≈ 2 for small ϵ) *whatever are the noise distribution and a convex and closed class $\mathcal{X}$*.

We describe construction of the minimax affine estimator of F when the class $\mathcal{X}$ is known and provide an adaptive version of the estimation procedure when the available information allows us to construct an embedded family of classes.

The rest of the paper is structured as follows. We present our results on estimation over the Sobolev classes in Section 2. Section 3 deals with minimax and adaptive affine estimation. Section 4 presents a numerical study of proposed adaptive estimators and discusses their relative merits. Proofs of all results are given in the supplementary article [8].

2. Estimation over Sobolev classes.

2.1. *Notation*. We denote by f_Y and f_ζ the densities of random variables Y and ζ; with certain abuse of notation we simply denote by f the density of unknown distribution of X.

Let g be a function on $\mathbb{R}$; we denote by $\widehat{g}$ the Fourier transform of g,

$$\widehat{g}(\omega) = \int_{-\infty}^{\infty} g(x)e^{i\omega x}\,\mathrm{d}x, \qquad \omega \in \mathbb{R}.$$

We consider the classes of absolutely continuous distributions.

DEFINITION 2.1. Let $\alpha > -1/2$, $L > 0$. We say that F belongs to the class $\mathcal{F}_\alpha(L)$ if it has a density f with respect to the Lebesgue measure on $\mathbb{R}$, and

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |\widehat{f}(\omega)|^2 (1+\omega^2)^\alpha\,\mathrm{d}\omega \leq L^2.$$

The set $\mathcal{F}_\alpha(L)$ with $\alpha > -1/2$ contains absolutely continuous distributions. If $\alpha > 1/2$, then the distributions F from $\mathcal{F}_\alpha(L)$ have bounded continuous densities. Usually $\mathcal{F}_\alpha(L)$ is referred to as the Sobolev class.

We use extensively the following inversion formula: for a continuous distribution F one has

$$(2) \qquad F(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \omega^{-1}\Im\{e^{-i\omega x}\widehat{f}(\omega)\}\,\mathrm{d}\omega, \qquad x \in \mathbb{R},$$

where $\Im\{\cdot\}$ stands for the imaginary part, and the above integral is interpreted as an improper Riemann integral $\lim_{T\to\infty} \int_{1/T}^{T} \omega^{-1}\Im\{e^{-i\omega x}\widehat{f}(\omega)\}\,\mathrm{d}\omega$. For the proof of (2) see [15, 16] and [20], Section 4.3.

Throughout this section we assume that the error characteristic function does not vanish:

$$|\widehat{f}_\zeta(\omega)| \neq 0 \qquad \forall \omega \in \mathbb{R}.$$

This is a standard assumption in deconvolution problems.

2.2. *Minimax rates of estimation.* In model (1) we have $\widehat{f}(\omega) = \widehat{f}_Y(\omega) / \widehat{f}_\zeta(\omega)$, and $\widehat{f}_Y(\omega)$ can be easily estimated by the empirical characteristic function of the observations Y. This motivates the following construction: for $\lambda > 0$ we define the estimator $\widetilde{F}_\lambda$ of $F(t_0)$ by

$$(3) \qquad \widetilde{F}_\lambda = \frac{1}{2} - \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\pi} \int_0^\lambda \frac{1}{\omega} \Im \left\{ \frac{e^{i\omega(Y_j - t_0)}}{\widehat{f}_\zeta(\omega)} \right\} d\omega.$$

Here λ is the design parameter to be specified. Note that if the density f_ζ is symmetric around the origin, then $\widehat{f}_\zeta$ is real, and the estimator $\widetilde{F}_\lambda(t_0)$ takes the form (cf. [17])

$$\widetilde{F}_\lambda = \frac{1}{2} - \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\pi} \int_0^\lambda \frac{\sin\{\omega(Y_j - t_0)\}}{\widehat{f}_\zeta(\omega)\omega} d\omega.$$

Note that $\widetilde{F}_\lambda$ may be truncated to the interval $[0, 1]$; obviously, the risk of such a "projected" estimator is smaller than that of $\widetilde{F}_\lambda$.

Our current goal is to establish an upper bound on the risk of the estimator $\widetilde{F}_\lambda$ over the classes $\mathcal{F}_\alpha(L)$. We need the following assumptions on the distribution of the measurement errors ζ_i:

(E1) There exist real numbers $\beta > 0$, $c_\zeta > 0$ and $C_\zeta > 0$ such that

$$c_\zeta (1 + \omega^2)^{-\beta/2} \leq |\widehat{f}_\zeta(\omega)| \leq C_\zeta (1 + \omega^2)^{-\beta/2} \qquad \forall \omega \in \mathbb{R}.$$

(E2) There exist positive real numbers ω_0, b_ζ and τ such that

$$|\widehat{f}_\zeta(\omega)| \geq 1 - b_\zeta |\omega|^\tau \qquad \forall |\omega| \leq \omega_0.$$

Assumption (E1) characterizes the case of the *ordinary smooth errors*. Assumption (E2) describes the local behavior of $\widehat{f}_\zeta$ near the origin. It is well known that for any distribution of a nondegenerate random variable there exist positive constants b and δ such that $|\widehat{f}(\omega)| \leq 1 - b|\omega|^2$ for all $|\omega| \leq \delta$ (see, e.g., [25], Lemma 1.5). Thus in (E2) we have $\tau \in (0, 2]$. Typical examples of distributions satisfying (E1) and (E2) are the Laplace and Gamma distributions. For example, for the Laplace distribution (E1) holds with $\beta = 2$, and (E2) holds with $\tau = 2$. The Gamma distribution provides an example of the distribution satisfying (E1) with $\beta > 0$ being the shape parameter of the distribution.

As we will see in the sequel, the rates of convergence of the risks $\mathrm{Risk}_2[\widetilde{F}_\lambda; \mathcal{F}_\alpha(L)]$ and $\mathrm{Risk}_\epsilon[\widetilde{F}_\lambda; \mathcal{F}_\alpha(L)]$ are mainly determined by the relationship between parameters α and β. Consider the following two subsets of the parameter set $\Theta := \{(\alpha, \beta) : \alpha > -1/2, \beta > 0\}$ for the pair (α, β):

$$\Theta_r := \{(\alpha, \beta) \in \Theta : \alpha + \beta > 1/2\}, \qquad \Theta_s := \{(\alpha, \beta) \in \Theta : \alpha + \beta < 1/2\}.$$

If $(\alpha, \beta) \in \Theta_s$, then necessarily $\widehat{f}_\zeta \notin \mathbb{L}_1(\mathbb{R})$; in addition, because $\alpha < 1/2$, the density f can be discontinuous. That is why we will refer to Θ_s as the *singular*
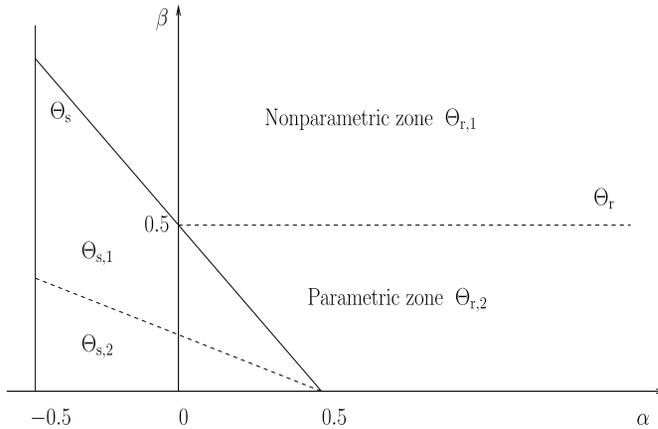
FIG. 1. *Division of the parameter set for* (α, β).

zone, while the subset Θ_r will be called the *regular zone*. We denote by Θ_b the *border zone* between Θ_r and Θ_s:

$$\Theta_b := \{(\alpha, \beta) \in \Theta : \alpha + \beta = 1/2\}.$$

Division of the parameter set Θ into zones Θ_r, Θ_s and Θ_b is displayed in Figure 1. The figure also shows the sub-regions $\Theta_{r,i}$ and $\Theta_{s,i}$, $i = 1, 2$, that are defined by the following formulas:

$$\Theta_{r,1} := \{(\alpha, \beta) \in \Theta_r : \beta > 1/2\}, \qquad \Theta_{r,2} := \{(\alpha, \beta) \in \Theta_r : \beta < 1/2\},$$

$$\Theta_{s,1} := \{(\alpha, \beta) \in \Theta_s : \alpha + 3\beta \geq 1/2\}, \qquad \Theta_{s,2} := \{(\alpha, \beta) \in \Theta_s : \alpha + 3\beta < 1/2\}.$$

The next two theorems present bounds on the risks in the regular zone: Theorem 2.1 states upper bounds on the risks of $\widetilde{F}_\lambda$, while Theorem 2.2 contains the corresponding lower bounds on the minimax risks.

For $z \geq 1$ define

$$\lambda(z) = z^{1/[2\alpha + (2\beta \vee 1)]}, \qquad \psi(z) = \begin{cases} z^{-(2\alpha+1)/(4\alpha+4\beta)}, & \beta > 1/2, \\ \sqrt{\ln z / z}, & \beta = 1/2, \\ 1/\sqrt{z}, & \beta \in (0, 1/2). \end{cases}$$

THEOREM 2.1. *Let assumptions* (E1) *and* (E2) *hold, and suppose that* $(\alpha, \beta) \in \Theta_r$. *If* $\widetilde{F}_{\lambda_\star}$ *is estimator* (3) *associated with* $\lambda_\star = C_1(\alpha, L)\lambda(n)$, *then for all* $t_0 \in \mathbb{R}$ *and large enough* n,

$$\mathrm{Risk}_2[\widetilde{F}_{\lambda_\star}; \mathcal{F}_\alpha(L)] \leq \psi_n(\alpha, L) := C_2(\alpha, L)\psi(n).$$

In addition, if $\lambda_\star = C_1(\alpha, L)\lambda(n/\ln[2\epsilon^{-1}])$, *then for all* $t_0 \in \mathbb{R}$ *and large enough* n,

$$\mathrm{Risk}_\epsilon[\widetilde{F}_{\lambda_\star}; \mathcal{F}_\alpha(L)] \leq \psi_{n,\epsilon}(\alpha, L) := C_3(\alpha, L)\psi(n/\ln[2\epsilon^{-1}]),$$

provided that $\epsilon \geq 2 \exp\{-C_4(\alpha, L)n\}$. *The constants* C_i, $i = 1, \ldots, 4$, *are specified in the proof of the theorem* (*see* (A.15)–(A.22) *in* [8]).

Theorem 2.1 shows that if (α, β) is in the regular zone Θ_r and $\beta \in (0, 1/2)$, then the estimator $\widetilde{F}_{\lambda_\star}$ attains the parametric rate of convergence. In the case $\beta = 1/2$ this rate is within a logarithmic factor of the parametric rate. The natural question is if the estimator $\widetilde{F}_{\lambda_\star}$ is rate optimal whenever $\beta > 1/2$, and $(\alpha, \beta) \in \Theta_r$. The answer is provided by Theorem 2.2.

We need the following assumption.

(E3) The characteristic function $\widehat{f}_\zeta$ is twice differentiable, and there exist real numbers $\beta > 1/2$, $C_\zeta > 0$ and $\omega_* > 0$ such that

$$(1 + \omega^2)^{\beta/2} \max_{j=0,1,2}\{|\widehat{f}_\zeta^{(j)}(\omega)|\} \leq C_\zeta \qquad \forall |\omega| \geq \omega_*.$$

Assumption (E3) is rather standard in derivations of lower bounds for deconvolution problems. This assumption should be compared to condition (G3) in [13]; it is assumed there that for $j = 0, 1, 2$ one has $|\widehat{f}_\zeta^{(j)}(\omega)||\omega|^{\beta+j} \leq C_\zeta$ as $|\omega| \to \infty$. Note that (E3) is a weaker assumption.

THEOREM 2.2. *Let assumption* (E3) *hold. Suppose that the class* $\mathcal{F}_\alpha(L)$ *is such that* $L^2 \geq \pi^{-1}2^{1+(\alpha-1)_+}\Gamma(2\alpha + 1)$ *and* $\alpha > 1/2$. *Then there exist constants* c_1 *and* c_2 *depending on* α, β *and* f_ζ *only such that, for all* n *large enough,*

$$\inf_{\widetilde{F}} \mathrm{Risk}_2[\widetilde{F}; \mathcal{F}_\alpha(L)] \geq c_1 L^{(2\beta-1)/(2\alpha+2\beta)}\phi_n,$$

$$\inf_{\widetilde{F}} \mathrm{Risk}_\epsilon[\widetilde{F}; \mathcal{F}_\alpha(L)] \geq c_2 L^{(2\beta-1)/(2\alpha+2\beta)}\phi_{n,\epsilon},$$

where $\phi_n := \phi(n)$, $\phi_{n,\epsilon} := \phi(n/\ln \epsilon^{-1})$, $\phi(z) := z^{-(2\alpha+1)/(4\alpha+4\beta)}$, *and* $\inf$ *is taken over all possible estimators of* $F(t_0)$.

The results of Theorems 2.1 and 2.2 deal with the regular zone. While we do not present the lower bound for the case of $\alpha \leq 1/2$ we expect that the bounds of Theorem 2.2 hold for the whole regular zone.

It is important to realize that the risks of $\widetilde{F}_\lambda$ converge to zero for all $(\alpha, \beta) \in \Theta$, and, in particular, for $(\alpha, \beta) \in \Theta_s$ and $(\alpha, \beta) \in \Theta_b$. The next statement establishes upper bounds on $\mathrm{Risk}_2[\widetilde{F}_\lambda; \mathcal{F}_\alpha(L)]$ in the singular and border zones, Θ_s and Θ_b.

THEOREM 2.3. *Let assumptions* (E1) *and* (E2) *hold. If* $\widetilde{F}_{\lambda_\star}$ *is the estimator* (3) *associated with* $\lambda_\star = C_1(\alpha, L)\lambda(n)$, *then for all* $t_0 \in \mathbb{R}$ *and large enough* n

$$\mathrm{Risk}_2[\widetilde{F}_{\lambda_\star}; \mathcal{F}_\alpha(L)] \leq C_2(\alpha, L)\varphi(n),$$

where the sequences $\lambda(n)$ *and* $\varphi(n)$ *are given in Table* 1, *and constants* C_1 *and* C_2 *are specified in the proof* (*see* (A.15)–(A.22) *in* [8]). *In addition, if*

TABLE 1
The bandwidth order $\lambda(n)$ and the convergence rate of the maximal risk $\varphi(n)$ in the singular and border zones

	Border zone Θ_b: $\alpha + \beta = 1/2$			Singular zone Θ_s: $\alpha + \beta < 1/2$	
	$\beta > 1/2$	$\beta = 1/2$	$\beta < 1/2$	$\alpha + 3\beta \geq 1/2$	$\alpha + 3\beta < 1/2$
$\lambda(n)$	$\frac{n}{\sqrt{\ln n}}$	$\frac{n}{(\ln n)^{3/2}}$	$(\frac{n}{\sqrt{\ln n}})^{1/(2\alpha+1)}$	$n^{2/(2\alpha+3-2\beta)}$	$n^{1/(2\alpha+2\beta+1)}$
$\varphi(n)$	$(\frac{\sqrt{\ln n}}{n})^{\alpha+1/2}$	$\frac{(\ln n)^{3/4}}{\sqrt{n}}$	$\frac{(\ln n)^{1/4}}{\sqrt{n}}$	$n^{-(2\alpha+1)/(2\alpha+3-2\beta)}$	$n^{-(2\alpha+1)/(4\alpha+4\beta+2)}$

$\lambda_\star = C_3(\alpha, L)\lambda(n/\ln[2\epsilon^{-1}])$, *then for large enough n*

$$\text{Risk}_\epsilon[\widetilde{F}_{\lambda_\star}; \mathcal{F}_\alpha(L)] \leq C_4(\alpha, L)\varphi(n/\ln[2\epsilon^{-1}]).$$

Several remarks on the results of Theorems 2.1–2.3 are in order.

Remarks. (1) Theorem 2.1 shows that the regular zone Θ_r is decomposed into three disjoint regions with respect to the upper bounds on the risks of $\widetilde{F}_{\lambda_\star}$. In the zone $\Theta_{r,2}$ where $\beta < 1/2$, the rates of convergence are parametric; because of roughness of the error density, here the estimation problem is essentially a parametric one. The region $\Theta_{r,1}$ is characterized by nonparametric rates, while in the border zone between $\Theta_{r,1}$ and $\Theta_{r,2}$ ($\beta = 1/2$) the rate of convergence differs from the parametric one by a $\ln n$-factor.

(2) The condition on L stated in Theorem 2.2 is purely technical; it requires that the family $\mathcal{F}_\alpha(L)$ is rich enough. It follows from Theorems 2.1 and 2.2 that the estimator $\widetilde{F}_{\lambda_\star}$ is optimal in order in the regular zone if $\alpha > 1/2$.

(3) The subdivision of the singular zone Θ_s into two zones $\Theta_{s,1} = \{(\alpha, \beta) \in \Theta_s : 3\beta + \alpha \geq \frac{1}{2}\}$ and $\Theta_{s,2} = \{(\alpha, \beta) \in \Theta_s : 3\beta + \alpha < \frac{1}{2}\}$ is a consequence of two types of upper bounds that we have on the variance term; see (14) in [8]. In the border zone Θ_b the upper bounds on the risk differ from those in the regular zone only by logarithmic in n factors. We do not know if the estimator $\widetilde{F}_{\lambda_\star}$ is rate optimal in the singular and border zones.

(4) Note that the results of Theorems 2.1 and 2.3, when put together, allow us to establish risk bounds for any pair (α, β) from the parameter set $\Theta = \{(\alpha, \beta) : \alpha > -1/2, \beta > 0\}$. In particular, for any fixed $\alpha > -1/2$, the rate of convergence of the maximal risk approaches the parametric rate when β approaches zero. We would like to stress the fact that no tails or moment conditions on F are required to obtain these results; such conditions were systematically imposed in the previous work on deconvolution of distribution functions.

2.3. *Adaptive estimation.* The choice of the smoothing parameter λ in (3) is crucial in order to achieve the optimal estimation accuracy. As Theorems 2.1 and 2.2 show, if parameters α and L of the class $\mathcal{F}_\alpha(L)$ are known, then one

can choose λ in such a way that the resulting estimator is optimal in order. In practice the functional class $\mathcal{F}_\alpha(L)$ is hardly known; in these situations the estimator of Section 2 cannot be implemented. Note, however, that this does not pose a serious problem in the regular zone when $\beta \in (0, 1/2)$. Indeed, here if we choose $\lambda = \sqrt{n}$, then the resulting estimator will be optimal in order for any functional class $\mathcal{F}_\alpha(L)$ satisfying $\lambda_\star = \lambda_\star(\alpha, L) \leq \sqrt{n}$, where $\lambda_\star$ is defined in Theorem 2.1.

The situation is completely different in the case $\beta > 1/2$. In this section we develop an estimator that is nearly optimal for the ϵ-risk over a scale of classes $\mathcal{F}_\alpha(L)$. The construction of our adaptive estimator is based on the general scheme by [23].

2.3.1. *Estimator construction.* Consider the family of estimators $\{\widetilde{F}_\lambda, \lambda \in \Lambda\}$, where $\widetilde{F}_\lambda$ is defined in (3), $\Lambda := \{\lambda_j, j = 1, \ldots, N\}$ with $\lambda_{\min} := \lambda_1$, $\lambda_{\max} := \lambda_N$, and $\lambda_j = 2^j \lambda_{\min}$, $j = 2, \ldots, N$. The adaptive estimator $\widetilde{F}$ is obtained by selection from the family $\{\widetilde{F}_\lambda, \lambda \in \Lambda\}$ according to the following rule.

Let

$$(4) \qquad \omega_1 := \min\{\omega_0, (4b_\zeta)^{-1/\tau}\}, \qquad c_* := 2\pi^{-2}[2 + (1/\tau)]^2,$$

where constants ω_0, b_ζ and τ appear in assumption (E2). For any $\lambda \in \Lambda$ we define

$$(5) \qquad \widetilde{\sigma}_\lambda^2 := c_* + \frac{2}{\pi^2 n} \sum_{j=1}^n \int_{\omega_1}^\lambda \int_{\omega_1}^\lambda \frac{1}{\omega\mu} \Im\left\{\frac{e^{i\omega(Y_j - t_0)}}{\widehat{f}_\zeta(\omega)}\right\} \Im\left\{\frac{e^{i\mu(Y_j - t_0)}}{\widehat{f}_\zeta(\mu)}\right\} d\omega\, d\mu,$$

$$\widetilde{\Sigma}_\lambda^2 := \max_{\mu \in \Lambda : \mu \leq \lambda} \widetilde{\sigma}_\mu^2.$$

Note that $\widetilde{\sigma}_\lambda^2$ can be computed from the data (the parameters τ and ω_1 are determined completely by $\widehat{f}_\zeta$; hence they are known). In fact, $\widetilde{\sigma}_\lambda^2 n^{-1}$ is a plug-in estimator of an upper bound on the variance of $\widetilde{F}_\lambda$, while $\widetilde{\Sigma}_\lambda^2$ is a "monotonization" of $\widetilde{\sigma}_\lambda^2$ with respect to λ.

Define

$$\widetilde{v}_\lambda^2 := \widetilde{\Sigma}_\lambda^2 + 11\bar{m}^2 \lambda^{2\beta} n^{-1} \ln(4N^2\epsilon^{-1}), \qquad \lambda \in \Lambda,$$

where

$$(6) \qquad \bar{m} := \sqrt{2c_*} + (\pi c_\zeta \beta)^{-1} 2^{1+(\beta/2-1)_+}[2 + \beta \ln_+(1/\omega_1)],$$

and constant c_ζ appears in assumption (E1).

Let $\vartheta := \sqrt{2}(\sqrt{2} - 1)^{-1}[1 + \sqrt{3\ln(4N\epsilon^{-1})}]$; then with every estimator $\widetilde{F}_\lambda$, $\lambda \in \Lambda$ we associate the interval

$$(7) \qquad Q_\lambda := [\widetilde{F}_\lambda - \vartheta\widetilde{v}_\lambda n^{-1/2}, \widetilde{F}_\lambda + \vartheta\widetilde{v}_\lambda n^{-1/2}].$$

Define

$$(8) \qquad \widetilde{\lambda} := \min\left\{\lambda \in \Lambda : \bigcap_{\mu \geq \lambda, \mu \in \Lambda} Q_\mu \neq \varnothing\right\},$$

and set finally

$$(9) \qquad \widetilde{F} := \widetilde{F}_{\widetilde{\lambda}}.$$

Note that $\widetilde{\lambda}$ is well defined: the intersection in (8) is nonempty for $\lambda = \lambda_{\max}$.

2.3.2. *Oracle inequality.* We will show that the estimator $\widetilde{F}$ mimics the oracle estimator $\widetilde{F}_o$ which is defined as follows:

Let

$$\sigma_\lambda^2 := c_* + \frac{2}{\pi^2} \mathrm{E}\left[\int_{\omega_1}^{\lambda} \frac{1}{\omega} \Im\left\{ \frac{e^{i\omega(Y_j - t_0)}}{\widehat{f}_\zeta(\omega)} \right\} d\omega \right]^2,$$

$$\Sigma_\lambda^2 := \max_{\mu \in \Lambda : \mu \leq \lambda} \sigma_\mu^2, \qquad \lambda \in \Lambda.$$

It is shown in the proof of Lemma 5.2 (see Section A.1.2 in [8]) that $\sigma_\lambda^2 n^{-1}$ is an upper bound on the variance of the estimator $\widetilde{F}_\lambda$ associated with parameter λ. Note that $\widetilde{\sigma}_\lambda^2$ defined in (5) is the empirical counterpart of the quantity σ_λ^2. Define

$$v_\lambda^2 := \Sigma_\lambda^2 + 11\bar{m}^2 \lambda^{2\beta} n^{-1} \ln(4N^2 \epsilon^{-1}).$$

Given $\alpha > 0$ and $L > 0$ let

$$\lambda_o = \lambda_o(\alpha, L) := \min\{\lambda \in \Lambda : v_\lambda n^{-1/2} \geq 2\sqrt{2}\pi^{-1/2} L \lambda^{-\alpha - 1/2}\}$$

and define $\widetilde{F}_o := \widetilde{F}_{\lambda_o}$.

The oracle estimator $\widetilde{F}_o$ has attractive minimax properties over classes $\mathcal{F}_\alpha(L)$. In particular, it is easily verified that for any class $\mathcal{F}_\alpha(L)$ such that $\lambda_o \leq [11\bar{m}^2 \ln(4N\epsilon^{-1})]^{-1} n$ one has

$$\mathrm{Risk}_\epsilon[\widetilde{F}_o; \mathcal{F}_\alpha(L)] \leq 2v_{\lambda_o} n^{-1/2} \leq \varkappa_1 \psi_{n,\epsilon}(\alpha, L) + \varkappa_2 \phi_{n,\epsilon}.$$

Here $\psi_{n,\epsilon}$ is the upper bound of Theorem 2.1 on the risk of the estimator $\widetilde{F}_{\lambda_\star}$ that "knows" α and L, $\phi_{n,\epsilon}$ is defined in Theorem 2.2, and $\varkappa_1$ and $\varkappa_2$ are constants independent of α and L. Thus, the risk of the oracle estimator admits the same upper bound as the risk of the estimator $\widetilde{F}_{\lambda_\star}$ that is based on the knowledge of the class parameters α and L.

Now we are in a position to state a bound on the risk of the estimator $\widetilde{F}_{\widetilde{\lambda}}$.

THEOREM 2.4. *Suppose that assumptions* (E1), (E2) *hold,* $\beta > 1/2$ *and let*

$$\lambda_{\max} = [11\bar{m}^2 \ln(4N\epsilon^{-1})]^{-1} n.$$

If $\widetilde{F}_{\widetilde{\lambda}}$ *is the estimator defined in* (7)–(9) *then for any class* $\mathcal{F}_\alpha(L)$ *with* $\alpha > 0$ *such that* $\lambda_{\min} \leq \lambda_o(\alpha, L) \leq \lambda_{\max}$, *one has*

$$\mathrm{Risk}_\epsilon[\widetilde{F}_{\widetilde{\lambda}}; \mathcal{F}_\alpha(L)] \leq (3 - 1/\sqrt{2})\vartheta v_{\lambda_o} n^{-1/2}.$$

Estimator (7)–(9) attains the optimal rates of convergence with respect to ϵ-risk within a $\ln(N\epsilon^{-1})$-factor over the collection of functional classes $\mathcal{F}_\alpha(L)$.

In particular, if $\lambda_{\min}$ is chosen to be a constant, and $\lambda_{\max} \asymp n^l$ for some $l \geq 1$, then $N = \ln(\lambda_{\max}/\lambda_{\min})/\ln 2 \asymp \ln n$, and the ϵ-risk of the adaptive estimator $\widetilde{F}_{\widetilde{\lambda}}$ is within a $\ln\ln n$-factor of the minimax ϵ-risk for a scale of Sobolev classes. It can be shown that this $\ln\ln n$-factor is unavoidable price for adaptation when the accuracy is measured by the ϵ-risk; see, for example, [26].

3. Minimax and adaptive affine estimation in discrete deconvolution model. The results of Section 2 imply that in the regular zone the minimax rates of convergence on the Sobolev classes are attained by linear estimator (3). It seems interesting to compare the performance of estimator (3) and its adaptive version in Section 2.3 with that of the minimax linear estimator.

Consider the estimation problem as follows; cf. [19], Problem 2.2:

PROBLEM D. We observe n independent realizations $\eta_1, \ldots, \eta_n$ of a random variable η, taking values in $\mathbb{S} = \{1, \ldots, m\}$. The distribution of η is identified with a vector p from the m-dimensional simplex $\mathcal{P}_m = \{y \in \mathbb{R}^m : y \geq 0, \sum_i y_i = 1\}$ by setting $p_k = \mathrm{P}\{\eta = k\}$, $1 \leq k \leq m$. Suppose that vector p is affinely parameterized by an M-dimensional "signal"-vector of unknown "parameters" $x \in \mathcal{X} \subset \mathcal{P}_M : p = Ax = [[Ax]_1; \ldots; [Ax]_m]$. Here Ax is the linear mapping with $A\mathcal{X} \subset \mathcal{P}_m$, and $[a]_j$ stands for the jth element of a. Our goal is to estimate a given linear form $g(x) = g^T x$ at the point x underlying the observation η^n.

It is obvious that if distributions of X and ζ are compactly supported, or can be "compactified" (i.e., for any $\varepsilon > 0$ one can point out bounded intervals of probability $1 - \varepsilon$ for X and ζ), then under very minor regularity conditions on f_ζ and F, the Problem D approximates the initial distribution deconvolution problem with "arbitrary accuracy." The latter means that given $\varepsilon > 0$ we can compile the discretized problem such that its δ-solution is the solution to the initial continuous problem with the accuracy $\delta + \varepsilon$ with probability $1 - \varepsilon$.

We consider the following discretization of the deconvolution problem:

(1) Let $J = [a_0, a_m]$ be the (finite) observation domain, and let $a_0 < a_1 < a_2 < \cdots < a_{m-1} < a_m$. We split J into m intervals $J_1 = [a_0, a_1]$, $J_2 = (a_1, a_2], \ldots$, $J_m = (a_{m-1}, a_m]$. We denote $p_k = \mathrm{P}\{Y \in J_k\}$, $k = 1, \ldots, m$.

(2) Suppose that the (finite) interval $I = [b_0, b_M]$ contains the support of all $F \in \mathcal{F}$. Let $b_0 < b_1 < b_2 < \cdots < b_M$, we partition I into M intervals $I_1 = [b_0, b_1]$, $I_2 = (b_1, b_2], \ldots, I_M = (b_{M-1}, b_M]$. We denote $x_k = \mathrm{P}\{X \in I_k\}$, $k = 1, \ldots, M$.

(3) Denote $\bar{b}_k = (b_{k-1} + b_k)/2$. Define the $m \times M$ matrix $A = (A_{jk})$ with elements

$$A_{jk} = \mathrm{P}\{\bar{b}_k + \zeta \in J_j\}$$

$$= \begin{cases} \mathrm{P}\{a_0 - \bar{b}_k \leq \zeta \leq a_1 - \bar{b}_k\}, & k = 1, \ldots, M, j = 1, \\ \mathrm{P}\{a_{j-1} - \bar{b}_k < \zeta \leq a_j - \bar{b}_k\}, & k = 1, \ldots, M, j = 2, \ldots, m, \end{cases}$$

and the vector $g = g(t_0) \in \mathbb{R}^M$, with $g_k = \mathbb{1}(\bar{b}_k \leq t_0)$, $k = 1, \ldots, M$. The elements A_{jk} of A are the approximations of conditional probabilities $\mathrm{P}\{Y \in J_j | X \in I_k\}$, and $g^T x$ is an approximation of $F(t_0)$.

(4) Consider discrete observations $\eta_i \in \{1, \ldots, m\}$ as follows:

$$\eta_i = \mathbb{1}(a_0 \leq Y_i \leq a_1) + \sum_{j=2}^{m} j \cdot \mathbb{1}(a_{j-1} < Y_i \leq a_j), \qquad i = 1, \ldots, n.$$

If the sets I and J are selected so that $\mathrm{P}\{X \in I\} \geq 1 - \varepsilon$, $\mathrm{P}\{Y \in J\} \geq 1 - \varepsilon$ for any $F \in \mathcal{F}$, if $\mathcal{F}$ is the class of "regular distributions" and the noise distribution possesses some regularity, and if the partitions of I and J are "fine enough," then solving Problem D with $\mathcal{X}$ being the corresponding M-dimensional cross-section of $\mathcal{F}$ will provide us with an estimation $\widetilde{g}$ of $F(t_0)$ in the continuous deconvolution problem.

We now concentrate on solving the deconvolution problem in the discrete model.

3.1. *Minimax estimation in the discrete model.* An estimate of $g(x)$—a candidate solution to our problem—is a measurable function $\widetilde{g} = \widetilde{g}(\eta^n) : \mathbb{S}^n \to \mathbb{R}$. Given tolerance $\epsilon \in (0, 1)$, we define the ϵ-*risk* of such an estimate on $\mathcal{X}$ as

$$\mathrm{Risk}_\epsilon(\widetilde{g}; \mathcal{X}) = \inf\left\{\delta : \sup_{x \in \mathcal{X}} \mathrm{P}_x\{|\widetilde{g}(\eta^n) - g^T x| > \delta\} < \epsilon\right\},$$

where P_x stands for the distribution of observations η^n associated with the "signal" x. The minimax optimal ϵ-risk is

$$\mathrm{Risk}_\epsilon^*(\mathcal{X}) = \inf_{\widetilde{g}(\cdot)} \mathrm{Risk}_\epsilon(\widetilde{g}; \mathcal{X}).$$

We are particularly interested in the family of estimators of the following structure:

$$\widetilde{g}_{\varphi,c}(\eta^n) = \frac{1}{n} \sum_{i=1}^{n} \varphi(\eta_i) + c = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{m} \varphi_k \mathbb{1}(\eta_i = k) + c.$$

We refer to such estimators $\widetilde{g}_\varphi$ as *affine*. In other words, $\widetilde{g}_\varphi$ is an affine function of empirical distribution: for some $\varphi \in \mathbb{R}^m$ and $c \in \mathbb{R}$,

$$\widetilde{g}_{\varphi,c}(\eta^n) = \sum_{k=1}^{m} \varphi_k \widetilde{P}_n(k) + c,$$

where $\widetilde{P}_n$ is the empirical distribution of the observation sample $\widetilde{P}_n(k) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(\eta_i = k)$. An important property of the class of affine estimators, when applied to Problem D with convex set $\mathcal{X}$, is that one can choose an estimator from the class such that its ϵ-risk attains (up to a moderate constant ≈ 2; see Theorem 3.1 below) the minimax ϵ-risk $\mathrm{Risk}_\epsilon^*(\mathcal{X})$.

From now on let us assume that $\mathcal{X} \subset \mathbb{R}^M$ is a convex closed (and, being a subset of an M-dimensional simplex, compact) set.

Let us consider the affine estimator $\widetilde{g}_\epsilon$ of $g^T x$

$$\widetilde{g}_\epsilon(\eta^n) \equiv \widetilde{g}_{\bar{\varphi}, \bar{c}}(\eta^n) = \sum_{k=1}^{m} \bar{\varphi}_k \widetilde{P}_n(k) + \bar{c},$$

in which the parameters $\bar{\varphi}$ and $\bar{c}$ of $\widetilde{g}_\epsilon$ are defined as follows.

Consider the optimization problem

$$\overline{S}(\epsilon) = \max_{x, y \in \mathcal{X}} \left\{ \frac{1}{2} g^T (y - x), \right.$$

(10)

$$\left. h(x, y; \epsilon) \equiv n \ln \left(\sum_{j=1}^{m} \sqrt{[Ax]_j [Ay]_j} \right) + \ln(2/\epsilon) \geq 0 \right\}.$$

Let $(\bar{x}, \bar{y})$ be an optimal solution to (10), and let $\nu \geq 0$ be the Lagrange multiplier of the constraint $h(x, y; \epsilon) \geq 0$. We set

$$\bar{c} = \frac{1}{2} g^T [\bar{y} + \bar{x}] \quad \text{and} \quad \bar{\varphi}_j = \nu n \ln \left[\sqrt{\frac{[A\bar{y}]_j}{[A\bar{x}]_j}} \right], \qquad j = 1, \ldots, m.$$

We have the following result.

THEOREM 3.1. *Let $\epsilon \in (0, 1/4]$. Then the ϵ-risk of the estimator $\widetilde{g}_\epsilon$ satisfies*

$$(11) \qquad \mathrm{Risk}_\epsilon(\widetilde{g}_\epsilon; \mathcal{X}) \leq \overline{S}(\epsilon) \leq \vartheta(\epsilon) \, \mathrm{Risk}_\epsilon^*(\mathcal{X}), \qquad \vartheta(\epsilon) = \frac{2 \ln(2/\epsilon)}{\ln[1/(4\epsilon)]}.$$

Note that $\vartheta(\epsilon) \to 2$ as $\epsilon \to 0$; thus for small tolerance levels the ϵ-risk of the estimator $\widetilde{g}_\epsilon$ is within factor ≈ 2 of the minimax ϵ-risk. It is important to emphasize that $\widetilde{g}_\epsilon$ is readily given by a solution to the explicit convex program (10), and as such, it can be found in a computationally efficient fashion, provided that $\mathcal{X}$ is computationally tractable.

In the "historical perspective" the affine estimator $\widetilde{g}_\epsilon$ represents an alternative to the binary search estimator $\widetilde{g}_B$, proposed in [10] for the case of "direct" observations. It can be shown that the ϵ-risk $\mathrm{Risk}_\epsilon(\widetilde{g}_B; \mathcal{X})$ of that estimator satisfies $\mathrm{Risk}_\epsilon(\widetilde{g}_B; \mathcal{X}) \leq C \, \mathrm{Risk}_\epsilon^*(\mathcal{X})$ for small ϵ (e.g., one can prove that $C \leq 26$ whenever $\epsilon \leq 0.01$). To the best of our knowledge, risk bound (11) in Theorem 3.1 for the estimator $\widetilde{g}_\epsilon$ is much better than those available for the binary search estimator.

Note that the constraint $h(x, y; \epsilon) \geq 0$ of the problem (10) can be rewritten as follows:

$$\rho(x, y) \geq (\epsilon/2)^{1/n},$$

where

$$\rho(x, y) = \sum_{k=1}^{m} \sqrt{[Ax]_k [Ay]_k}$$

is the *Hellinger affinity* of distributions $A(x)$ and $A(y)$; cf. [21] and [22], Chapter 4. Thus the optimal value $\overline{S}(\epsilon)$ of the optimization problem (10) can be seen as modulus of continuity of the linear functional $g(\cdot)$ over the class $\mathcal{X}$ of distributions "with respect to Hellinger affinity." If $\frac{1}{n} \ln[1/\epsilon] = o(1)$ we have $\rho(x, y) \approx 1$ and

$$H^2(x, y) = 1 - \rho(x, y) \approx -\ln \rho(x, y),$$

where $H(x, y)$ is the *Hellinger distance* between x and y. In this limit we have

$$\overline{S}(\epsilon) \approx \frac{1}{2} \omega\left(\sqrt{\frac{\ln[2/\epsilon]}{2N}}\right) \equiv \max_{x, y \in \mathcal{X}} \left\{ \frac{1}{2} g^T(y - x), H(x, y) \le \sqrt{\frac{\ln[2/\epsilon]}{2n}} \right\}.$$

Here $\omega(\cdot)$ is the "modulus of continuity of g over $\mathcal{X}$ with respect to Hellinger distance," introduced in [10]. Therefore, bound (11) can be seen as a finite-dimensional nonasymptotic counterpart of [10], Theorem 3.1.

3.2. *Adaptive version of the estimate.* Consider a modification of our estimation problem where the set $\mathcal{X}$, instead of being given in advance, is known to be one of the sets from the collection of nonempty convex compact sets $\mathcal{X}^1, \mathcal{X}^2, \ldots, \mathcal{X}^N$ in $\mathbb{R}^M$. We aim to construct an *adaptive estimator* of the linear form $g^T x$, given that x is an element of some $\mathcal{X}_i$ in the collection. Here we consider the simple case where the sets are nested. $\mathcal{X}^1 \subset \mathcal{X}^2 \subset \cdots \subset \mathcal{X}^N$. Note that in the case of nonnested sets an adaptive estimator can be constructed following the ideas of [7].

Given a linear form $g^T z$ on $\mathbb{R}^M$, let $\mathrm{Risk}^k(\widetilde{g})$ and Risk^k_* be, respectively, the ϵ-risk of an estimate $\widetilde{g}$ on $\mathcal{X}^k$, and the minimax optimal ϵ-risk of recovering $g^T x$ on $\mathcal{X}^k$. Let also $S_k(\cdot)$ be the function $\overline{S}(\cdot)$ in (10) associated with $\mathcal{X} = \mathcal{X}^k$. As it is immediately seen, the functions $S_k(\cdot)$ grow with k. Our goal is to modify the estimate $\widetilde{g}$ we have built in such a way that the ϵ-risk of the modified estimate on $\mathcal{X}^k$ will be "nearly" Risk^k_* *for every* $k \le N$. This goal can be achieved by a straightforward application of Lepski's adaptation scheme as follows.

Given $\epsilon > 0$, let $\widetilde{g}^k(\cdot)$ be the affine estimate with the (ϵ/N)-risk on $\mathcal{X}^k$ not exceeding $S_k(\epsilon/N)$ as provided by Theorem 3.1 which is applied with ϵ/N substituted for ϵ and $\mathcal{X}^k$ substituted for $\mathcal{X}$. Then

$$\sup_{x \in \mathcal{X}^k} \mathrm{P}_x\{|\widetilde{g}^k(\eta^n) - g^T x| > S_k(\epsilon/N)\} \le \epsilon/N \qquad \forall k \le N.$$

Given observation η^n, let us say that the index $k \le N$ is η^n-*good*, if for all k' satisfying $k \le k' \le N$ one has

$$|\widetilde{g}^{k'}(\eta^n) - \widetilde{g}^k(\eta^n)| \le S_k(\epsilon/N) + S_{k'}(\epsilon/N).$$

Note that η^n-good indices do exist (e.g., $k = N$). Given η^n, we can find the smallest η^n-good index $k = k(\eta^n)$; our estimate is nothing but $\widetilde{g}(\eta^n) = \widetilde{g}^{k(\eta^n)}(\eta^n)$.

PROPOSITION 3.1. *Assume that $\epsilon \in (0, 1/4)$, and let*

$$\vartheta = 3\frac{\ln(2N/\epsilon)}{\ln(2/\epsilon)}.$$

Then

$$\sup_{x \in \mathcal{X}^k} P_x\{|\widetilde{g}(\eta^n) - g^T x| > \vartheta S_k(\epsilon)\} < \epsilon \qquad \forall(k, 1 \le k \le N);$$

whence also

$$\mathrm{Risk}^k(\widetilde{g}) \le \frac{6\ln(2N/\epsilon)}{\ln[1/(4\epsilon)]}\mathrm{Risk}^k_* \qquad \forall(k, 1 \le k \le N).$$

The proof of the proposition follows exactly same steps as that of Proposition 5.1 of [19], and it is omitted.

4. Numerical examples. To illustrate our results we present here examples of implementation of the adaptive estimation procedures of Sections 2.3 and 3.2.

We consider three measurement error distributions scenarios:

(i) Gamma distribution $\Gamma(0, 2, 1/(2\sqrt{2}))$ with the shape parameter 2 and the scale $\frac{1}{2\sqrt{2}}$ (the standard deviation of the error is equal to 0.5). Here $\Gamma(\mu, \alpha, \theta)$ stands for the Gamma distribution with location μ, shape parameter α and scale θ, such that its density is $[\Gamma(\alpha)\theta^\alpha]^{-1}(x - \mu)^{\alpha-1}\exp\{-(x - \mu)/\theta\}\mathbb{1}(x \ge \mu)$.

(ii) Mixture of Laplace distributions $\frac{1}{2}\mathcal{L}(-1, \frac{1}{2}) + \frac{1}{2}\mathcal{L}(1, \frac{1}{2})$; here $\mathcal{L}(\mu, a)$ stands for the Laplace distribution with the density $(2a)^{-1}e^{-|x-\mu|/a}$.

(iii) Normal mixture $\frac{1}{2}\mathcal{N}(0, \frac{1}{4}) + \frac{1}{2}\mathcal{N}(2, \frac{1}{4})$.

We consider three distributions of X:

(1) mixture of "shifted" Gamma distributions: $0.3\Gamma(0, 0.5, 2) + 0.7\Gamma(5, 0.5, 2)$;
(2) mixture of Laplace distributions $0.3\mathcal{L}(-1.5, 0.5) + 0.7\mathcal{L}(1.7, 0.25)$;
(3) normal mixture $0.6\mathcal{N}(0.15827, 1) + 0.4\mathcal{N}(1, 0.0150)$.

Note that in the case (i) of $\Gamma(0, 2, \theta)$ error distribution the estimator (3) can be computed explicitly: we have $\widetilde{F}_\lambda = \frac{1}{2} - \frac{1}{\pi n}\sum_{i=1}^n I_\lambda(Y_i - t_0)$, where

$$I_\lambda(y) = \mathrm{Si}(\lambda y) + y^{-1}[\theta^2\lambda\cos(\lambda y) - 2\theta\sin(\lambda y)] - y^{-2}\theta^2\sin(\lambda y),$$

and $\mathrm{Si}(x) = \int_0^x \omega^{-1}\sin\omega\,d\omega$ is the sine integral function. Then the adaptive estimation algorithm of Section 2.3 is implemented for the grid $\Lambda = \{\lambda \in [0.01 : 0.05 : 10]\}$.

Estimation procedures, described in Section 3.1, were implemented using `Mosek` optimization software [1]. The observation space and the signal space were split into $m = M = 200$ bins. The adaptation procedure was implemented over 17 linear estimators corresponding to the classes $\mathcal{X}^1, \ldots, \mathcal{X}^{17}$ of "Lipschitz-continuous" discrete distributions with Lipschitz constants on the geometric grid,

scaled from 0.001 to 1 [if reduced to continuous densities, it corresponds to the approximate range of Lipschitz constant from $O(0.1)$ to $O(100)$].

The simulation has been repeated for 100 observation samples of size $n = 2{,}000$. On Figure 2 we present simulation results for the scenario (i) when the
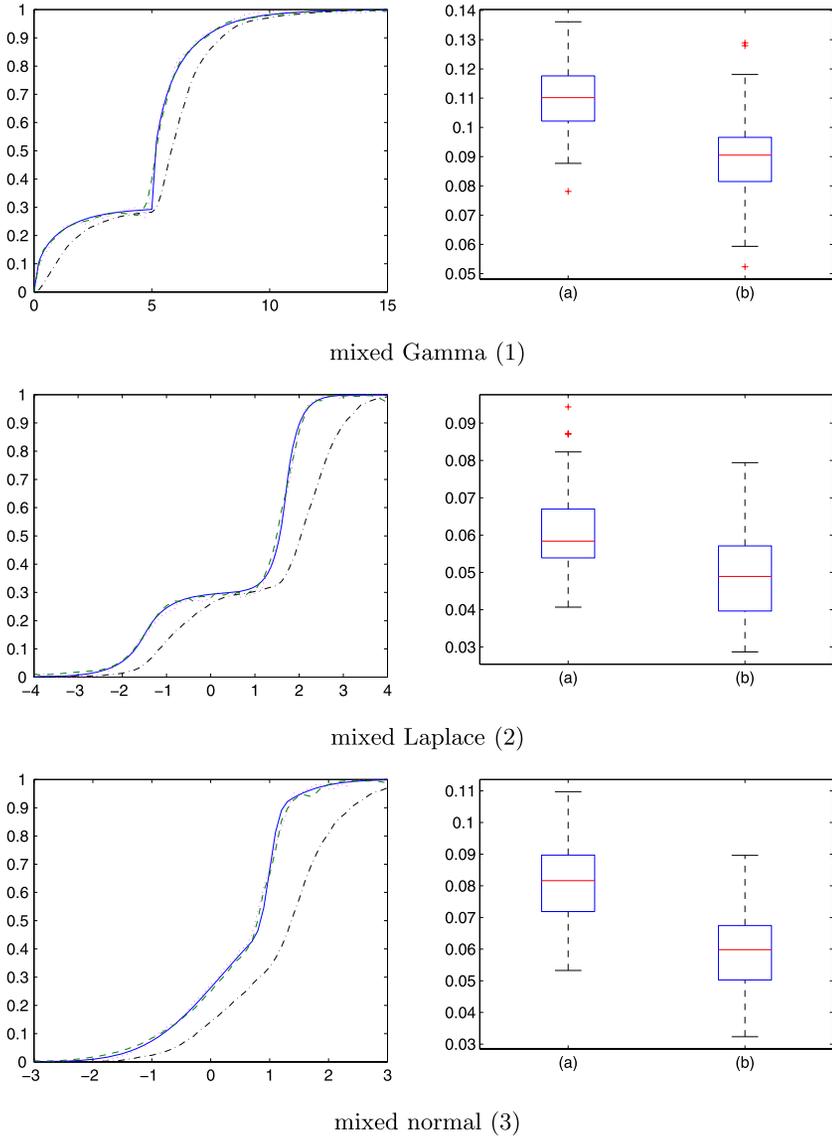

FIG. 2.   *Simulation results for the Gamma error scenario. On the left*: *true cdf* (*solid line*), *adaptive estimator* $\widetilde{g}(\eta^n)$ *of Section* 3.2 (*dashed line*), *adaptive estimator* $\widetilde{F}_{\widetilde{\lambda}}$ *of Section* 2.3 (*dotted line*) *and the edf of the observations* (*dash–dot line*). *On the right*: *the boxplots of the maximal estimation error of* $\widetilde{g}(\eta^n)$ (a) *and* $\widetilde{F}_{\widetilde{\lambda}}$ (b).

error distribution follows the $\Gamma(0, 2, 1/(2\sqrt{2}))$ law. The left column displays "typical" results of estimation corresponding to three signal distributions. We present the true distribution (solid line), the estimate $\widetilde{F}_{\widetilde{\lambda}}$ of Section 2.3 (dotted line), the estimate $\widetilde{g}(\eta^n)$ of Section 3.2 (dashed line) and the empirical distribution of the observations (dash–dot line). The boxplots on the right display resume the corresponding empirical distributions of the maximal estimation error over 50 points of the regular grid on the support of f for two estimators: (a) for $\widetilde{g}(\eta^n)$ of Section 3.2 and (b) for the $\widetilde{F}_{\widetilde{\lambda}}$ of Section 2.3. On Figure 3 we present "typical" results for adaptive estimator $\widetilde{g}(\eta^n)$ of Section 3.2 under the error scenarios (ii) (on the left) and (iii) (on the right). Similarly to Figure 2 we plot true cdf (solid line), adaptive estimator $\widetilde{g}(\eta^n)$ of Section 3.2 (dashed line) and the observation edf (dash–dot line). The results of this simulation are summarized on Figure 4. The first boxplot (the left column plots) represents the distribution of the maximal estimation error over 50 points of the regular grid on the support of f. Next, for each point in the grid we compute the maximal estimation error over 100 simulations, the distribution of maximal errors "over the points of the grid" is represented on the second boxplot (plots on the right column).

Remarks. The numerical examples in this section illustrate strong and weak points of the proposed estimators related to practical implementation. They can be summarized as follows.

The adaptive estimator of Section 2.3 is based on the choice of the unique smoothing parameter λ. This imposes a "natural" family of nested classes and facilitates implementation of the adaptation scheme. Yet, this estimator should be "explicitly tuned" for a specific distribution of the errors. In particular, the integral computation in (3) for a given distribution of ζ may become very tedious. Even though our theoretical results are proved under the condition that $|\widehat{f}_\zeta(\omega)| \neq 0$ for all $\omega \in \mathbb{R}$, in practical implementation the estimator (3) could be modified in order to allow characteristic functions $\widehat{f}_\zeta$ vanishing at finite number of points on $\mathbb{R}$. In this case the integration domain in (3) should exclude some properly specified vicinities of the points where $\widehat{f}_\zeta$ vanishes.

In contrast to this, the adaptive estimator in Section 3.2 can be easily tuned to *any noise distribution* and convex target distribution class. For instance, the characteristic function of noise in the Laplace scenario (ii) vanishes at some points, what precludes the possibility of utilizing the estimator of Section 2.3 without proper modifications. Note that one can easily incorporate any additional available information on the unknown distribution that can be expressed as a convex constraint in the corresponding optimization problem. The typical examples of such constraints are unimodality, symmetry, monotonicity and moment bounds. However, this freedom comes at a price: the family $\mathcal{X}^1 \subset \cdots \subset \mathcal{X}^N$ of the embedded classes for the adaptive estimator in Section 3.2 should be constructed "by hand." The computation of the adaptive affine estimator of Section 3.2 is also a heavy
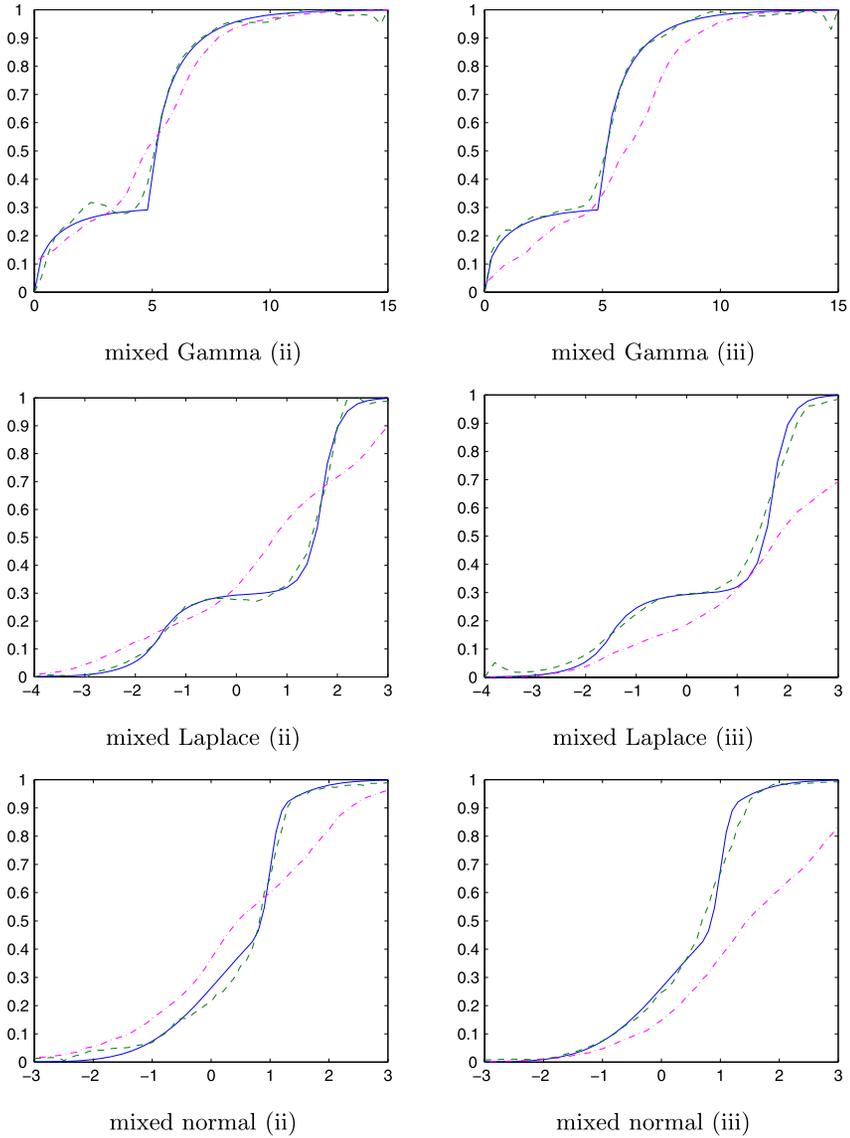
FIG. 3.   *Simulation results*: *true cdf* (*solid line*), *adaptive estimator* (*dashed line*) *and empirical distribution function of the observation* (*dash–dot line*). *On the left*, (ii) *are the results for mixed Laplace noise*; *on the right*, (iii) *are the results for the mixed normal noise*.

numerical task. In particular, in our setting it involves solving 17 conic quadratic optimization problems with 1,006 variables, 809 linear and 202 conic constraints.

It is well known that the normal noise in the deconvolution problem results in a very poor quality of estimation [13]. In particular, the minimax rate of convergence in this case is $O((\ln n)^{-\gamma})$ with $\gamma > 0$ depending on the exponent α
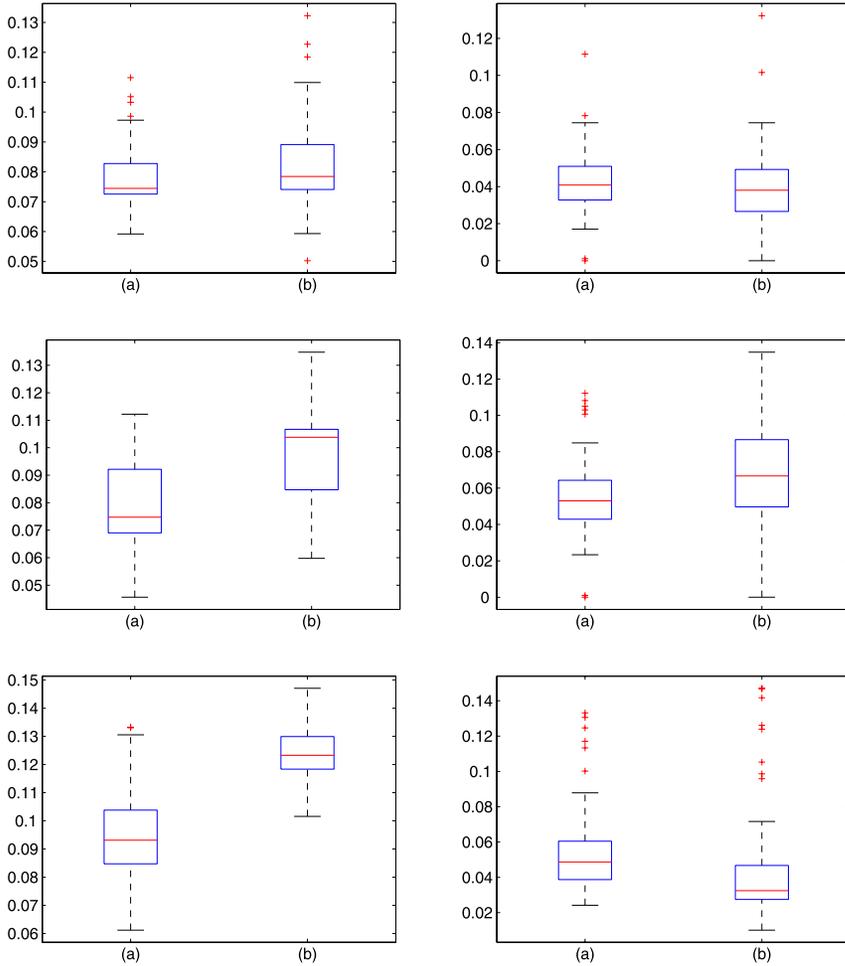
FIG. 4. *Estimation error distribution. Left column: empirical distribution of the maximal error of estimation over a regular grid; right column: distribution of the maximal over* 100 *simulations estimation error over the points of the grid. On each plot the left boxplot* (a) *corresponds to the mixed Laplace noise, while the right boxplot* (b) *corresponds to the mixed normal noise.*

of the regularity class $\mathcal{F}_\alpha(L)$. Fortunately, these pessimistic results are concerned with *the asymptotic* as $n \to \infty$ behavior of the estimators. We observed that the estimation procedures exhibit much better performance for small or moderately sized observation samples. On the other hand, this performance does not improve when the sample size grows up: in our experiments, for instance, the estimation accuracy, measured by ℓ_∞-error over a regular grid in the distribution domain, improved only by the factor ≈ 2 when we increased the sample size from $n = 2,000$ to $n = 100,000$.

5. Proofs. This section is organized as follows. In Section 5.1 we state main results that are used in the proof of Theorems 2.1 and 2.3 and briefly discuss the proof outline. Then in Section 5.2 we prove Theorem 2.4. Full proofs of all auxiliary results and additional technical details are given in the supplementary paper [8].

5.1. *Proofs of Theorems 2.1 and 2.3.* Proofs of Theorems 2.1 and 2.3 go along the same lines and exploits three basic statements presented here. Lemmas 5.1 and 5.2 given below establish upper bounds on the bias and variance of the estimator $\widetilde{F}_\lambda$. Then we present Lemma 5.3 that states an exponential inequality on the stochastic error of $\widetilde{F}_\lambda$. This result is used for derivation bounds on the ϵ-risk. Finally we briefly explain how the stated results are combined in order to complete the proof of Theorems 2.1 and 2.3.

We start with the standard decomposition of the error of estimator (3).

$$|\widetilde{F}_\lambda - F(t_0)| \leq |\mathrm{E}\widetilde{F}_\lambda - F(t_0)| + |\widetilde{F}_\lambda - \mathrm{E}\widetilde{F}_\lambda| = B_\lambda(t_0; F) + |V_\lambda|,$$
$$\mathrm{E}|\widetilde{F}_\lambda - F(t_0)|^2 = B_\lambda^2(t_0; F) + \mathrm{E}|V_\lambda|^2,$$

where we have denoted

$$B_\lambda(t_0, F) := \left| \frac{1}{\pi} \int_\lambda^\infty \frac{1}{\omega} \Im(e^{-i\omega t_0} \widehat{f}(\omega)) \, d\omega \right|, \qquad V_\lambda := \frac{1}{n} \sum_{j=1}^n [\xi_j(\lambda) - \mathrm{E}\xi_j(\lambda)]$$

and

$$\xi_j(\lambda) := \frac{1}{\pi} \int_0^\lambda \frac{1}{\omega} \Im \left\{ \frac{e^{i\omega(Y_j - t_0)}}{\widehat{f}_\zeta(\omega)} \right\} \, d\omega, \qquad j = 1, \ldots, n.$$

5.1.1. *Bounds on bias and variance.* First we bound the bias of $\widetilde{F}_\lambda$.

LEMMA 5.1. *Let $\widetilde{F}_\lambda$ be the estimator defined in* (3); *then for every class $\mathcal{F}_\alpha(L)$ with $\alpha > -\frac{1}{2}$, $L > 0$ and for any $\lambda \geq 1$ one has*

(12) $$\sup_{F \in \mathcal{F}_\alpha(L)} B_\lambda(F; t_0) \leq K_0 L \lambda^{-\alpha - 1/2}, \qquad K_0 := \sqrt{2/\pi}[1 + (2\alpha + 1)^{-1/2}].$$

Now we establish an upper bound on the variance of $\widetilde{F}_\lambda$. Recall that ω_1 and c_* are given in (4) and depend on the constants ω_0, b_ζ and τ appearing in assumption (E2). Define

(13) $$w(\lambda) := \begin{cases} \lambda^{2\beta - 1}, & \beta > 1/2, \\ 1 \vee \ln(\lambda/\omega_1), & \beta = 1/2, \\ 1, & \beta \in (0, 1/2). \end{cases}$$

LEMMA 5.2.   *Let assumptions* (E1), (E2) *hold and* $\widetilde{F}_\lambda$ *be the estimator defined in* (3). *Then there exist constants* $K_1 = K_1(\alpha, \beta, \omega_1)$ *and* $K_2 = K_2(\beta, \omega_1)$ *such that for every* $\lambda \geq 1 \vee \omega_1$ *the following statements hold*:

(i) *If* $\alpha + \beta > 1/2$, *then*

$$\mathrm{var}\{\widetilde{F}_\lambda\} \leq K_1 L C_\zeta c_\zeta^{-2} w(\lambda) n^{-1} + c_* n^{-1}.$$

If $\beta > 1$, *then the upper bound can be made independent of* α *and* L

$$\mathrm{var}\{\widetilde{F}_\lambda\} \leq K_2 C_\zeta c_\zeta^{-2} \lambda^{2\beta-1} n^{-1} + c_* n^{-1}.$$

(ii) *If* $\alpha + \beta = 1/2$, *then*

$$\mathrm{var}\{\widetilde{F}_\lambda\} \leq K_1 L C_\zeta c_\zeta^{-2} w(\lambda) \sqrt{\ln(\lambda/\omega_1)} n^{-1} + c_* n^{-1}.$$

(iii) *If* $\alpha + \beta < 1/2$, *then*

(14) $\mathrm{var}\{\widetilde{F}_\lambda\} \leq K_1 c_\zeta^{-2} \min[L C_\zeta \lambda^{1/2-\beta-\alpha}, \ln^2(\lambda/\omega_1) + \lambda^{2\beta}] n^{-1} + c_* n^{-1}.$

Explicit expressions for K_1 *and* K_2 *are given in the proof*; *see* (A.12) *in* [8].

It is worth noting that if $\beta > 1$, then the upper bound on the variance of $\widetilde{F}_\lambda$ stated in part (i) does not depend on paramaters α and L. This is particularly important when the problem of adaptive estimation of $F(t_0)$ is considered.

5.1.2. *An exponential inequality.*   First we recall some notation.

$$\sigma_\lambda^2 := c_* + \frac{2}{\pi^2} \mathrm{E}\left(\int_{\omega_1}^{\lambda} \frac{1}{\omega} \Im\left\{\frac{e^{i\omega(Y_j - t_0)}}{\widehat{f_\zeta}(\omega)}\right\} d\omega\right)^2,$$

where $\omega_1 = \min\{\omega_0, (2b_\zeta)^{-1/\tau}\}$, $c_* = 2\pi^{-2}[2 + (1/\tau)]^2$ and constants ω_0, b_ζ and τ appear in assumption (E2). Define

$$m_\lambda := \sqrt{2c_*} + 2^{1+(\beta/2-1)_+} (\pi c_\zeta)^{-1}[\ln(\lambda/\omega_1) + \beta^{-1}\lambda^\beta].$$

It is easily seen that $m_\lambda \leq \bar{m}\lambda^\beta$, $\forall \lambda \geq 1$, where $\bar{m}$ is defined in (6). We also put

$$\bar{\sigma}^2 := c_* + C_\zeta c_\zeta^{-2}\{K_1 L \mathbb{1}(\beta \leq 1) + [(K_1 L) \vee K_2]\mathbb{1}(\beta > 1)\},$$

where constants K_1 and K_2 are given in (A.12) in [8].

LEMMA 5.3.   *Suppose that assumptions* (E1) *and* (E2) *hold*; *then for any* $\lambda > 0$ *and* $z > 0$ *one has*

(15) $$\mathrm{P}\{|V_\lambda| \geq z\} \leq 2\exp\left\{-\frac{nz^2}{2\sigma_\lambda^2 + (2/3)m_\lambda z}\right\}.$$

In particular, if $\alpha + \beta > 1/2$, *then for any* $\lambda \geq 1 \vee \omega_1$ *and* $z > 0$ *one has*

(16) $$\mathrm{P}\{|V_\lambda| \geq z\} \leq 2\exp\left\{-\frac{nz^2}{2\bar{\sigma}^2 w(\lambda) + (2/3)\bar{m}\lambda^\beta z}\right\},$$

where $w(\lambda)$ *is given in* (13).

5.1.3. *Outline of the proofs of Theorems* 2.1 *and* 2.3.   The upper bounds on the quadratic risk stated in Theorems 2.1 and 2.3 are immediate consequence of Lemmas 5.1 and 5.2. Balancing the upper bounds on the bias and variance with respect to the smoothing parameter λ, we come to the announced results. Lemma 5.3 along with Lemma 5.1 are used in order to derive upper bounds on the ϵ-risk. Full technical details are provided in the supplementary paper [8].

5.2. *Proof of Theorem* 2.4.   The next preparatory lemma establishes an exponential probability inequality on deviation of $\widetilde{\Sigma}_\lambda^2$ from Σ_λ^2.

LEMMA 5.4.   *Suppose that assumptions* (E1) *and* (E2) *hold.*

(i) *For every* $\lambda \in \Lambda$

$$\mathrm{P}\{|\widetilde{\Sigma}_\lambda^2 - \Sigma_\lambda^2| \geq v_\lambda^2/2\} \leq \frac{\epsilon}{2N}.$$

(ii) *Let* $q(\epsilon) := \sqrt{3\ln(4N\epsilon^{-1})}$; *then for every* $\lambda \in \Lambda$

$$\mathrm{P}\{|V_\lambda| \geq q(\epsilon)v_\lambda n^{-1/2}\} \leq \frac{\epsilon}{2N}.$$

Proof of Lemma 5.4 is given in [8].

5.2.1. *Proof of Theorem* 2.4.   Define the following events:

$$A(\lambda) := \{|V_\lambda| \leq q(\epsilon)v_\lambda n^{-1/2}\} \cap \{|\widetilde{\Sigma}_\lambda^2 - \Sigma_\lambda^2| \leq v_\lambda^2/2\},$$
$$A(\Lambda) := \bigcap_{\lambda \in \Lambda} A(\lambda).$$

It follows from Lemma 5.4 and $\#(\Lambda) = N$ that $\mathrm{P}\{A(\Lambda)\} \geq 1 - \epsilon$. By the triangle inequality,

$$(17) \qquad |\widetilde{F}_{\widetilde{\lambda}} - F(t_0)| \leq |\widetilde{F}_{\lambda_o} - F(t_0)| + |\widetilde{F}_{\widetilde{\lambda}} - \widetilde{F}_{\lambda_o}|.$$

By definition of λ_o and by the fact that v_λ is monotone increasing with λ we have that $v_\lambda n^{-1/2} \geq \bar{B}_\lambda$ for all $\lambda \geq \lambda_o$, where we have denoted $\bar{B}_\lambda := 2(2/\pi)^{1/2}L\lambda^{-\alpha-1/2}$. Therefore, on the event $A(\Lambda)$

$$(18) \qquad |\widetilde{F}_{\lambda_o} - F(t_0)| \leq \bar{B}_{\lambda_o} + |V_{\lambda_o}| \leq [1 + q(\epsilon)]v_{\lambda_o} n^{-1/2}.$$

Furthermore, if $A(\Lambda)$ holds, then for any pair $\lambda, \mu \in \Lambda$ satisfying $\lambda \geq \lambda_o$ and $\mu \geq \lambda_o$ one has $Q_\lambda \cap Q_\mu \neq \varnothing$. Indeed, by definition of λ_o for any $\lambda \geq \lambda_o$ one has $\bar{B}_\lambda \leq v_\lambda/\sqrt{n}$; therefore

$$|\widetilde{F}_\lambda - F(t_0)| \leq \bar{B}_\lambda + q(\epsilon)v_\lambda n^{-1/2} \leq [1 + q(\epsilon)]v_\lambda n^{-1/2}.$$

In addition, on the set $A(\Lambda)$ we have

$$|\widetilde{v}_\lambda - v_\lambda| \leq |\widetilde{v}_\lambda^2 - v_\lambda^2|^{1/2} = |\widetilde{\Sigma}_\lambda^2 - \Sigma_\lambda^2|^{1/2} \leq v_\lambda/\sqrt{2}.$$

This yields

$$|\widetilde{F}_\lambda - F(t_0)| \le \frac{\sqrt{2}}{\sqrt{2}-1}[1 + q(\epsilon)]\widetilde{v}_\lambda n^{-1/2}$$

$$= \vartheta \widetilde{v}_\lambda n^{-1/2} \qquad \forall \lambda \ge \lambda_o.$$

Thus one has $F(t_0) \in Q_\lambda$ and $F(t_0) \in Q_\mu$ for all $\lambda \ge \lambda_o$ and $\mu \ge \lambda_o$; hence $Q_\mu \cap Q_\lambda \ne \varnothing$. Then by the procedure definition, $\widetilde{\lambda} \le \lambda_o$ and $Q_{\widetilde{\lambda}} \cap Q_{\lambda_o} \ne \varnothing$ on the event $A(\Lambda)$. Therefore

$$\begin{aligned}|\widetilde{F}_{\widetilde{\lambda}} - \widetilde{F}_{\lambda_o}| &\le \vartheta n^{-1/2}[\widetilde{v}_{\widetilde{\lambda}} + \widetilde{v}_{\lambda_o}] \\ &\le 2\vartheta n^{-1/2}\widetilde{v}_{\lambda_o} \\ &\le \sqrt{2}(1 + \sqrt{2})\vartheta n^{-1/2}v_{\lambda_o}.\end{aligned}$$

(19)

Here the second line follows from $\widetilde{v}_{\widetilde{\lambda}} \le \widetilde{v}_{\lambda_o}$, and the fact that $\widetilde{v}_{\lambda_o} \le (1 + 2^{-1/2})v_{\lambda_o}$ on the event $A(\Lambda)$. Combining (19), (18) and (17) we obtain that on the set $A(\Lambda)$

$$|\widetilde{F}_{\widetilde{\lambda}} - F(t_0)| \le \left(\frac{3\sqrt{2}-1}{\sqrt{2}-1}\right)[1 + q(\epsilon)]v_{\lambda_o}n^{-1/2}.$$

This completes the proof.

Acknowledgments. The authors are grateful to the Associate Editor and two anonymous referees for careful reading and useful remarks that led to improvements in the presentation.

SUPPLEMENTARY MATERIAL

Supplement to "On deconvolution of distribution functions" (DOI: 10. 1214/11-AOS907SUPP; .pdf). In the supplementary paper [8] we prove results stated here and provide additional details for the proofs appearing in Section 5. In particular, [8] is partitioned in two Appendices, A and B. Appendix A contains proofs for Section 2: full technical details for Theorems 2.1, 2.3 and 2.4 are presented, and the proof of Theorem 2.2 is given. In Appendix B we prove Theorem 3.1 from Section 3.

REFERENCES

[1] ANDERSEN, E. D. and ANDERSEN, K. D. The MOSEK optimization tools manual. Version 5.0. Available at http://www.mosek.com/.

[2] BEN-TAL, A. and NEMIROVSKI, A. (2001). *Lectures on Modern Convex Optimization*: *Analysis*, *Algorithms*, *and Engineering Applications*. SIAM, Philadelphia, PA. MR1857264

[3] BUTUCEA, C. and COMTE, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli* **15** 69–98. MR2546799

[4] BUTUCEA, C. and TSYBAKOV, A. B. (2008). Sharp optimality in density deconvolution with dominating bias. I. *Theory Probab. Appl.* **52** 24–39.

[5] BUTUCEA, C. and TSYBAKOV, A. B. (2008). Sharp optimality in density deconvolution with dominating bias. II. *Theory Probab. Appl.* **52** 237–249.

[6] CAI, T. T. and LOW, M. G. (2003). A note on nonparametric estimation of linear functionals. *Ann. Statist.* **31** 1140–1153. MR2001645

[7] CAI, T. T. and LOW, M. G. (2004). Minimax estimation of linear functionals over nonconvex parameter spaces. *Ann. Statist.* **32** 552–576. MR2060169

[8] DATTNER, I., GOLDENSHLUGER, A. and JUDITSKY, A. (2011). Supplement to "On deconvolution of distribution functions." DOI:10.1214/11-AOS907SUPP.

[9] DONOHO, D. and LIU, R. (1987). Geometrizing rates of convergence. I. Technical Report 137a, Dept. Statistics, Univ. California, Berkeley.

[10] DONOHO, D. and LIU, R. (1991). Geometrizing rates of convergence. II. *Ann. Statist.* **19** 633–667.

[11] DONOHO, D. and LIU, R. (1991). Geometrizing rates of convergence. III. *Ann. Statist.* **19** 668–701.

[12] DONOHO, D. L. (1994). Statistical estimation and optimal recovery. *Ann. Statist.* **22** 238–270. MR1272082

[13] FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324

[14] GAFFEY, W. R. (1959). A consistent estimator of a component of a convolution. *Ann. Math. Statist.* **30** 198–205. MR0102154

[15] GIL-PELAEZ, J. (1951). Note on the inversion theorem. *Biometrika* **38** 481–482. MR0045992

[16] GURLAND, J. (1948). Inversion formulae for the distribution of ratios. *Ann. Math. Statist.* **19** 228–237. MR0025005

[17] HALL, P. and LAHIRI, S. N. (2008). Estimation of distributions, moments and quantiles in deconvolution problems. *Ann. Statist.* **36** 2110–2134. MR2458181

[18] JOHANNES, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist.* **37** 2301–2323. MR2543693

[19] JUDITSKY, A. B. and NEMIROVSKI, A. S. (2009). Nonparametric estimation by convex programming. *Ann. Statist.* **37** 2278–2300. MR2543692

[20] KENDALL, M., STUART, A. and ORD, J. K. (1987). *Kendall's Advanced Theory of Statistics. Vol. 1, Distribution Theory*, 5th ed. Oxford Univ. Press, New York. MR0902361

[21] LE CAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53. MR0334381

[22] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York. MR0856411

[23] LEPSKIĬ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466.

[24] MEISTER, A. (2009). *Deconvolution Problems in Nonparametric Statistics. Lecture Notes in Statistics* **193**. Springer, Berlin. MR2768576

[25] PETROV, V. V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables. Oxford Studies in Probability* **4**. Oxford Univ. Press, New York. MR1353441

[26] SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498. MR1425962

[27] STEFANSKI, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.* **9** 229–235. MR1045189

[28] ZHANG, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–831. MR1056338

I. DATTNER
A. GOLDENSHLUGER
DEPARTMENT OF STATISTICS
UNIVERSITY OF HAIFA
31905 HAIFA
ISRAEL
E-MAIL: idattner@stat.haifa.ac.il
          goldensh@stat.haifa.ac.il

A. JUDITSKY
LMC, B. P. 53
UNIVERSITÉ GRENOBLE I
38041 GRENOBLE CEDEX 9
FRANCE
E-MAIL: anatoli.iouditski@imag.fr