Proceedings of the 7th International Conference on Statistics: Theory and Applications (ICSTA'25)
Paris, France - August, 2025
Paper No. 159
DOI: 10.11159/icsta25.159

Minimax Rates of Convergence for Multivariate Distribution Function L^1-Deconvolution with Known Ordinary Smooth Errors

Catia Scricciolo
Dipartimento di Scienze Economiche, Università di Verona
Via Cantarane 24, 37129 Verona, Italy
catia.scricciolo@univr.it

Abstract **-** In this paper, we study the problem of estimating a distribution function on $\mathbb{R}^d$, $d \geq 1$, from data contaminated by additive noise, using the L^1-distance between distribution functions. We assume that the error distribution is known and belongs to a class of anisotropic ordinary smooth distributions. We derive minimax-optimal convergence rates for L^1-deconvolution in arbitrary dimensions.

Keywords**:** deconvolution, distribution functions, L^1-metric, minimax rates, ordinary smooth distributions.

1. Introduction

We observe random vectors in $\mathbb{R}^d$, $d \geq 1$,

$$\boldsymbol{Y}_i = \boldsymbol{X}_i + \boldsymbol{\varepsilon}_i, \qquad i=1, \ldots, n, \tag{1}$$

where each $\boldsymbol{Y}_i$ consists of a signal $\boldsymbol{X}_i$ additively corrupted by a noise $\boldsymbol{\varepsilon}_i$. We assume that

- the (column) random vectors $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,j}, \ldots, X_{i,d})^{\mathrm{T}}$ are independent and identically distributed (i.i.d.) according to an *unknown* probability measure μ_X;
- the random vectors $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,j}, \ldots, \varepsilon_{i,d})^{\mathrm{T}}$ are i.i.d. according to a *known* probability measure μ_ε;
- the sequences $(\boldsymbol{X}_i)_{i \in \mathbb{N}}$ and $(\boldsymbol{\varepsilon}_i)_{i \in \mathbb{N}}$ are independent.

The distribution of the $\boldsymbol{Y}_i$'s is then the convolution $\mu_X * \mu_\varepsilon$,

$$\boldsymbol{Y}_i \sim \mu_X * \mu_\varepsilon, \qquad i=1, \ldots, n.$$

In this paper, we derive minimax-optimal convergence rates relative to the L^1-distance for estimating the distribution function F_X associated with the probability measure μ_X. For probability measures μ and γ on $\mathbb{R}^d$, $d \geq 1$, let F and G be the associated distribution functions, respectively. Then, the distance in the L^1-metric between them is given by

$$\|F - G\|_1 = \int_{\mathbb{R}^d} |F(\mathbf{x}) - G(\mathbf{x})|\, \mathbf{dx}.$$

The distance in the L^1-metric between distribution functions on $\mathbb{R}^d$, $d \geq 1$, can be viewed as a multivariate extension of the L^1-Wasserstein distance between probability measures in dimension one. Indeed, in the univariate case, the L^1-Wasserstein distance between two probability measures μ and γ, defined as $W_1(\mu, \gamma) := \inf_{\mu \leftarrow \pi \rightarrow \gamma} \int_{\mathbb{R}^2} |x - y| \mathrm{d}\pi(x, y)$, where the infimum is taken over all joint distributions π on $\mathbb{R}^2$ with fixed marginals μ and γ, coincides with the L^1-distance between their distribution functions, that is $W_1(\mu, \gamma) = \|F - G\|_1$, even though, in the multivariate case, the L^1-Wasserstein distance between two probability measures on $\mathbb{R}^d$ is not equal to the L^1-distance between their distribution functions. More generally, for $1 \leq p < \infty$, the distance in the L^p-metric $\|F - G\|_p = (\int_{\mathbb{R}^d} |F(\mathbf{x}) - G(\mathbf{x})|^p\, \mathbf{dx})^{1/p}$ provides a natural way to compare distribution functions. This metric is homogeneous of order d/p, that is $\|F(\cdot / r)\|_p = r^{d/p} \|F\|_p$ for any scalar $r > 0$. Furthermore, L^p-distances between distribution functions on $\mathbb{R}^d$, $d \geq 1$, can be controlled by their L^1-distance by virtue of the inequality $\|F - G\|_p \leq (2^{p-1} \|F - G\|_1)^{1/p} = 2^{1-1/p} (\|F - G\|_1)^{1/p}$.

1.1. Assumptions on the Noise

Let $\varphi_\varepsilon(\mathbf{t})$, $\mathbf{t} \in \mathbb{R}^d$, denote the characteristic function of the noise distribution μ_ε defined as

$$\varphi_\varepsilon(\mathbf{t}) = \int_{\mathbb{R}^d} e^{i\langle \mathbf{t}, \mathbf{x}\rangle} \mu_\varepsilon(\mathbf{dx}), \qquad \mathbf{t} \in \mathbb{R}^d,$$

where, for given vectors $\mathbf{t}, \mathbf{x} \in \mathbb{R}^d$, we denote by $\langle \mathbf{t}, \mathbf{x}\rangle := \sum_{j=1}^d t_j x_j$ the standard scalar product in $\mathbb{R}^d$. We assume that

- the error coordinates $(\varepsilon_{1,j})_{1 \le j \le d}$ are independent random variables. Let $\mu_{\varepsilon,j}$ be the distribution of the jth coordinate of the random vector $\boldsymbol{\varepsilon}_1$ and let $\varphi_{\varepsilon,j}$ be its characteristic function. Then,

$$\boldsymbol{\varepsilon}_1 \sim \mu_{\varepsilon,1} \otimes \dots \otimes \mu_{\varepsilon,d}; \tag{2}$$

- the characteristic function of the noise has a polynomial decrease: there exists $\boldsymbol{\beta} \in \mathbb{R}^d$, with $\beta_j > 0$ for all $j = 1, \dots, d$, and constants $R_1, R_2 > 0$ such that the characteristic function of the noise satisfies

$$\frac{1}{R_1} \prod_{j=1}^d (1 + |t_j|)^{-\beta_j} \le |\varphi_\varepsilon(\mathbf{t})| = \left| \prod_{j=1}^d \varphi_{\varepsilon,j}(t_j) \right| \le R_2 \prod_{j=1}^d (1 + |t_j|)^{-\beta_j}, \qquad \mathbf{t} \in \mathbb{R}^d; \tag{3}$$

- defined the function $r_j(t) := 1 / \varphi_{\varepsilon,j}(t)$, $t \in \mathbb{R}$, there exists a constant $R > 0$ such that $r_j(\cdot)$ is twice continuously differentiable and, for $l = 0, 1$,

$$|r_j^{(l)}(t)| \le R(1 + |t_j|)^{(\beta_j - l)}, \qquad t \in \mathbb{R}. \tag{4}$$

Condition (3) implies that $|\varphi_\varepsilon(\mathbf{t})| \neq 0$ for all $\mathbf{t} \in \mathbb{R}^d$. Random variables with Laplace or gamma distributions (with shape parameter $\beta > 0$ and scale parameter equal to 1) are ordinary smooth. When all error coordinates $\varepsilon_{1,j}$, $j = 1, \dots, d$, are ordinary smooth, potentially with different orders $\beta_j > 0$, this is referred to as the *homogeneous* case.

Many studies have addressed the problem of recovering a signal's distribution from measurements contaminated by additive noise with known density; see, *e.g.*, [3], [6] and the references therein. Most existing work focuses on the one-dimensional setting and the problem of density function recovery. Few papers address the multidimensional deconvolution problem: [3] proposes adaptive anisotropic kernel estimators for the signal density under the MISE criterion, while [5] investigates convergence rates for kernel-based estimators in Wasserstein distances under supersmooth errors, though without achieving optimal rates. In this work, we study the problem of recovering the distribution function in a multidimensional setting, employing a multivariate extension of the integrated classical deconvolution kernel density estimator and we establish minimax-optimal L^1-risk bounds for this estimator, without imposing any regularity assumptions on the mixing distribution beyond the existence of a Lebesgue density.

1.2. Notations

For two functions u and v on $\mathbb{R}^d$, $d \ge 1$, we write $u(\mathbf{x}) \lesssim v(\mathbf{x})$ if there exists a constant $C > 0$ (not depending on $\mathbf{x}$) such that $u(\mathbf{x}) \le Cv(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$. The same convention applies to $u(\mathbf{x}) \gtrsim v(\mathbf{x})$. For a function $g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, let $\mathcal{F}[g(x)](t) := \int_{\mathbb{R}} e^{itx} g(x)\, dx$, $t \in \mathbb{R}$, be its Fourier transform. Let $h(t) = \mathcal{F}[g(x)](t)$, $t \in \mathbb{R}$. Then, the inverse Fourier transform of h is $g(x) = \mathcal{F}^{-1}[h(t)](x) := \int_{\mathbb{R}} e^{-itx} h(t)\, dt$, $x \in \mathbb{R}$.

1.3. Organization of the Paper

The paper is organized as follows. In Section **2**, we derive an upper bound on the convergence rate with respect to the L^1-metric of the integrated classical deconvolution kernel density estimator. In Section **3**, we establish a matching lower bound for the same estimation problem. Finally, in Section **4**, we discuss possible extensions of the results.

2. Upper Bound

In this section, we first define the distribution function estimator in the multivariate case. We then derive an upper bound for its convergence rate under the L^1-risk, thereby extending Theorem 3.1 in [4], p. 243, to the multivariate case. Specifically, we generalize the upper bound for L^1-Wasserstein deconvolution in the univariate case to L^1-deconvolution in the multivariate setting. Since the distribution function estimator is obtained by integrating a kernel density estimator, we begin by specifying the required assumptions on the kernel.

2.1. Assumptions on the Kernel

For simplicity and without loss of generality, we assume that

- $K(\mathbf{x}) = \prod_{j=1}^{d} K_j(x_j)$, $\mathbf{x} \in \mathbb{R}^d$, where, for every $j = 1, \ldots, d$, the kernel function $K_j \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. We denote by $\varphi_{K,j}(t)$, $t \in \mathbb{R}$, the Fourier transform of K_j. In symbols, $\varphi_{K,j}(t) := \mathcal{F}[K_j(x)](t)$, $t \in \mathbb{R}$;
- for every $j = 1, \ldots, d$, the kernel K_j is a symmetric density with $\int_{\mathbb{R}} |z| K_j(z)\, \mathrm{d}z < \infty$ and its Fourier transform $\varphi_{K,j}(t)$, $t \in \mathbb{R}$, is symmetric about zero ($\varphi_{K,j}(-t) = \varphi_{K,j}(t)$, $t \in \mathbb{R}$) and has $\mathrm{supp}(\varphi_{K,j}) = [-1, 1]$;
- for every $j = 1, \ldots, d$, the Fourier transform $\varphi_{K,j}(t)$, $t \in \mathbb{R}$, is continuously differentiable with Lipschitz derivative.

An example of univariate kernel density K_j satisfying these assumptions is provided in (7) of [4], p. 240. From the first kernel assumption stated above, it follows that the Fourier transform of the kernel K is given by

$$\varphi_K(\mathbf{t}) := \mathcal{F}[K(\mathbf{x})](\mathbf{t}) = \prod_{j=1}^{d} \varphi_{K,j}(t_j), \qquad \mathbf{t} \in \mathbb{R}^d.$$

2.2. The Distribution Function Estimator

In this section, we give the definition of the distribution function estimator as the integral of the multivariate version of the standard deconvolution kernel density estimator on $\mathbb{R}$, first introduced by [2]. For $h_j > 0$, $j = 1, \ldots, d$, let

$$K_{j,h_j}(x_j) := \mathcal{F}^{-1}\left[\frac{\varphi_{K,j}(t_j)}{\varphi_{\varepsilon,j}\left(t_j/h_j\right)}\right](x_j) = \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i} t_j x_j} \frac{\varphi_{K,j}(t_j)}{\varphi_{\varepsilon,j}\left(t_j/h_j\right)}\, \mathrm{d}t_j = h_j \mathcal{F}^{-1}\left[\frac{\varphi_{K,j}(h_j t_j)}{\varphi_{\varepsilon,j}(t_j)}\right](h_j x_j), \qquad x_j \in \mathbb{R}.$$

The deconvolution kernel *density* estimator is then defined as

$$\hat{f}_n(\mathbf{x}) = \hat{f}_n(x_1, \ldots, x_d) := \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\frac{1}{h_j}K_{j,h_j}\left(\frac{x_j - Y_{i,j}}{h_j}\right) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\mathcal{F}^{-1}\left[\frac{\varphi_{K,j}(h_j t_j)}{\varphi_{\varepsilon,j}(t_j)}\right](x_j - Y_{i,j}), \qquad \mathbf{x} \in \mathbb{R}^d. \tag{5}$$

The estimator in (5) is the multivariate version of the standard deconvolution kernel density estimator and has been studied in [1], [3] and [5]. It has Fourier transform

$$\mathcal{F}\left[\hat{f}_n(\mathbf{x})\right](\mathbf{t}) = \varphi_{n,Y}(\mathbf{t}) \times \prod_{j=1}^{d}\frac{\varphi_{K,j}(h_j t_j)}{\varphi_{\varepsilon,j}(t_j)}, \qquad \mathbf{t} \in \mathbb{R}^d,$$

where $\varphi_n(\mathbf{t})$ is the empirical characteristic function of $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ defined as

$$\varphi_{n,Y}(\mathbf{t}) := \frac{1}{n}\sum_{i=1}^{n}\mathrm{e}^{\mathrm{i}\langle \mathbf{t}, \boldsymbol{Y}_i\rangle} = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\mathrm{e}^{\mathrm{i} t_j Y_{i,j}}, \qquad \mathbf{t} \in \mathbb{R}^d.$$

As a distribution function estimator, we consider the integral of the kernel density estimator in (5)

$$\mathcal{F}_n(\mathbf{x}) := \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_d} f_n(\mathbf{u})\,\mathrm{d}\mathbf{u} = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_d} f_n(u_1, \ldots, u_d)\,\mathrm{d}u_1 \ldots \mathrm{d}u_d$$

$$= \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_d} \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\frac{1}{h_j}K_{j,h_j}\left(\frac{u_j - Y_{i,j}}{h_j}\right)\mathrm{d}u_1 \ldots \mathrm{d}u_d$$

$$= \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_d} \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\mathcal{F}^{-1}\left[\frac{\varphi_{K,j}(h_j t_j)}{\varphi_{\varepsilon,j}(t_j)}\right](u_j - Y_{i,j})\,\mathrm{d}u_1 \ldots \mathrm{d}u_d$$

$$= \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_d} \mathcal{F}^{-1}\left[\varphi_{n,Y}(\mathbf{t}) \times \prod_{j=1}^{d}\frac{\varphi_{K,j}(h_j t_j)}{\varphi_{\varepsilon,j}(t_j)}\right](u_1, \ldots, u_d)\,\mathrm{d}u_1 \ldots \mathrm{d}u_d, \quad \mathbf{x} \in \mathbb{R}^d. \tag{6}$$

The estimator in (6) is the extension to the multivariate setting of the deconvolution distribution function estimator considered in (8) of [4], p. 240.

2.3. Rates of Convergence

In this section, we derive an upper bound for the L^1-risk $\mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}[\|\mathcal{F}_n - F_X\|_1]$ of the estimator $\mathcal{F}_n$ involving the regularity assumptions on the error coordinates stated in Section 1.1 and a tail condition imposed on the distribution of each coordinate $Y_{1,j}$, $j = 1, \ldots, d$, of $\boldsymbol{Y}_1$.

Theorem 1. *Suppose that we observe a sample $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ from the multivariate convolution model* (1). *Assume that conditions* (2)-(4) *on the error distribution are satisfied and, for every $j = 1, \ldots, d$,*

$$\int_0^{+\infty}\sqrt{\mathrm{P}(|Y_{1,j}| > y)}\,\mathrm{d}y < \infty. \tag{7}$$

Suppose that the mixing distribution μ_X possesses a Lebesgue density on $\mathbb{R}^d$ and that the coordinates $(X_{1,j})_{1\le j\le d}$ of $\boldsymbol{X}_1$ are independent. Then, for $\mathbf{h} = (h_1, \ldots, h_d)^{\mathrm{T}} \in (0, 1)^d$,

$$\mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}[\|\mathcal{F}_n - F_X\|_1] \lesssim \sum_{j=1}^{d} h_j + \frac{1}{\sqrt{n}}\prod_{j=1}^{d}\sqrt{I_j(h_j)},$$

where

$$I_j(h_j) := \int_{|t| \le 1/h_j}\left(\sum_{l=0}^{1}\frac{|r_j^{(l)}(t)|^2}{t^2}1_{[-1,1]^c}(t)\right)\mathrm{d}t, \qquad j = 1, \ldots, d. \tag{8}$$

Taking $h_{1,\mathrm{opt}} = \ldots = h_{d,\mathrm{opt}} = n^{-1/\sum_{j=1}^{d}[(2\beta_j \vee 1) + 1]}$, where $(2\beta_j \vee 1) := \max\{2\beta_j, 1\}$, we obtain

$$\mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}[\|\mathcal{F}_n - F_X\|_1] \lesssim n^{-1/\sum_{j=1}^{d}[(2\beta_j \vee 1) + 1]}(\log n)^{\#\{\beta_j = 1/2,\, 1 \le j \le d\}/2}.$$

Proof. We begin by deriving a bias-variance decomposition of the L^1-risk $\mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}[\|\mathcal{F}_n - F_X\|_1]$. For this purpose, we define

$$K_{\mathbf{h}}(\mathbf{x}) := \frac{1}{h_1 \cdot \ldots \cdot h_d}K\left(\frac{x_1}{h_1}, \cdots, \frac{x_d}{h_d}\right) = \prod_{j=1}^{d}\frac{1}{h_j}K_j\left(\frac{x_j}{h_j}\right) = \prod_{j=1}^{d}K_{j,h_j}(x_j), \qquad \mathbf{x} \in \mathbb{R}^d,$$

where the factorization follows from the first kernel assumption stated in Section 2.1 and $K_{j,h_j}(\cdot) := (1/h_j)K_j(\cdot/h_j)$ represents the re-scaled version of K_j. In what follows, we take the same kernel density $K_j = K$ for all $j = 1, \ldots, d$.

Bounding the bias of $\mathcal{F}_n$

Taking into account that $\mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}[\mathcal{F}_n] = F_X * K_{\mathbf{h}}$, which can be deduced from (6), using the assumption on the existence of the density of μ_X, the bias $\left\| \mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}[\mathcal{F}_n] - F_X \right\|_1$ of the estimator $\mathcal{F}_n$ can be bounded as follows:

$$\left\| \mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}[\mathcal{F}_n] - F_X \right\|_1 = \left\| F_X * K_{\mathbf{h}} - F_X \right\|_1 \le \mathrm{B}(\mathbf{h}),$$

where $\mathrm{B}(\mathbf{h}) := \left(\int_{\mathbb{R}} |z| K(z)\, \mathrm{d}z\right) \sum_{j=1}^{d} h_j$.

Bias-variance decomposition of the L^1-risk of $\mathcal{F}_n$

The expression of $\mathcal{F}_n(\mathbf{x})$ admits the following factorization:

$$\begin{aligned}
\mathcal{F}_n(\mathbf{x}) : &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \frac{1}{h_j} K_{j,h_j}\left(\frac{u_j - Y_{i,j}}{h_j}\right) \mathrm{d}u_1 \ldots \mathrm{d}u_d \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} \prod_{j=1}^{d} \frac{1}{h_j} K_{j,h_j}\left(\frac{u_j - Y_{i,j}}{h_j}\right) \mathrm{d}u_1 \ldots \mathrm{d}u_d \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(\prod_{j=1}^{d} \int_{-\infty}^{x_j} \frac{1}{h_j} K_{j,h_j}\left(\frac{u_j - Y_{i,j}}{h_j}\right) \mathrm{d}u_j \right).
\end{aligned}$$

The following bias-variance decomposition of the L^1-risk of $\mathcal{F}_n$ holds:

$$\begin{aligned}
\mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}\left[\left\| \mathcal{F}_n - F_X \right\|_1\right] &\le \mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}\left[\left\| \mathcal{F}_n - F_X * K_{\mathbf{h}} \right\|_1\right] + \left\| F_X * K_{\mathbf{h}} - F_X \right\|_1 \\
&\le \mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}\left[\left\| \mathcal{F}_n - \mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}[\mathcal{F}_n] \right\|_1\right] + \mathrm{B}(\mathbf{h}) \\
&\le \int_{\mathbb{R}^d} \sqrt{\mathrm{Var}\left(\mathcal{F}_n(\mathbf{x})\right)}\, \mathbf{dx} + \mathrm{B}(\mathbf{h}).
\end{aligned}$$

Bounding the variance of $\mathcal{F}_n$

Let ϕ denote a function symmetric about zero, equal to 1 on $[-1, 1]$ and to 0 on $[-2, 2]^{\mathrm{c}}$, twice continuously differentiable. For every $j = 1, \ldots, d$, we write

$$\begin{aligned}
\int_{-\infty}^{x_j} K_{j,h_j}(u_j)\, \mathrm{d}u_j &= \int_{-\infty}^{x_j} \mathcal{F}^{-1}\left[\frac{\varphi_{K,j}(t_j)\phi(t_j/h_j)}{\varphi_{\varepsilon,j}(t_j/h_j)}\right](u_j)\, \mathrm{d}u_j \\
&+ \int_{-\infty}^{x_j} \mathcal{F}^{-1}\left[\frac{\varphi_{K,j}(t_j)[1 - \phi(t_j/h_j)]}{\varphi_{\varepsilon,j}(t_j/h_j)}\right](u_j)\, \mathrm{d}u_j \\
&=: G_{h_j 1}(x_j) + G_{h_j 2}(x_j), \qquad x_j \in \mathbb{R}.
\end{aligned}$$

By the independence of the coordinates $(Y_{1,j})_{1 \le j \le d}$ of $\mathbf{Y}_1$, using the inequality $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ valid for every pair of real numbers $a, b > 0$, we obtain

$$\int_{\mathbb{R}^d} \sqrt{\mathrm{Var}\left(\mathcal{F}_n(\mathbf{x})\right)}\, \mathbf{dx} \le \frac{1}{\sqrt{n}} \int_{\mathbb{R}^d} \sqrt{\mathrm{E}_{(\mu_X * \mu_\varepsilon)}\left[\left(\prod_{j=1}^{d} \int_{-\infty}^{x_j} \frac{1}{h_j} K_{j,h_j}\left(\frac{u_j - Y_{1,j}}{h_j}\right) \mathrm{d}u_j\right)^2\right]}\, \mathbf{dx}$$

$$= \frac{1}{\sqrt{n}} \int_{\mathbb{R}^d} \sqrt{\mathrm{E}_{(\mu_X * \mu_\varepsilon)}\left[\prod_{j=1}^d \left(G_{h_j,1}\left(\frac{x_j - Y_{1,j}}{h_j}\right) + G_{h_j,2}\left(\frac{x_j - Y_{1,j}}{h_j}\right)\right)^2\right]} \, \mathbf{dx}$$

$$\leq \sqrt{\frac{2^d}{n}} \prod_{j=1}^d \int_{\mathbb{R}} \sqrt{\mathrm{E}_{(\mu_{X,j} * \mu_{\varepsilon,1})}\left[\left(G_{h_j,1}\left(\frac{x_j - Y_{1,j}}{h_j}\right)\right)^2\right] + \mathrm{E}_{(\mu_{X,j} * \mu_{\varepsilon,1})}\left[\left(G_{h_j,2}\left(\frac{x_j - Y_{1,j}}{h_j}\right)\right)^2\right]} \, \mathrm{d}x_j$$

$$\leq \sqrt{\frac{2^d}{n}} \prod_{j=1}^d \left(\underbrace{\int_{\mathbb{R}} \sqrt{\mathrm{E}_{(\mu_{X,j} * \mu_{\varepsilon,1})}\left[\left(G_{h_j,1}\left(\frac{x_j - Y_{1,j}}{h_j}\right)\right)^2\right]} \, \mathrm{d}x_j}_{=: T_{1,j}} + \underbrace{\int_{\mathbb{R}} \sqrt{\mathrm{E}_{(\mu_{X,j} * \mu_{\varepsilon,1})}\left[\left(G_{h_j,2}\left(\frac{x_j - Y_{1,j}}{h_j}\right)\right)^2\right]} \, \mathrm{d}x_j}_{=: T_{2,j}} \right). \tag{9}$$

The integrals $T_{1,j}$ and $T_{2,j}$ in (9) can be bounded as the terms *I* and *J*, respectively, in [4], pp. 248-253. For every $j = 1, \ldots, d$, we have $T_{1,j} = O(1)$, while, recalling the definition in (8),

$$T_{2,j} \lesssim \sqrt{I_j(h_j)} \lesssim h_j^{-[(\beta_j \vee 1/2) - 1/2]} (|\log h_j|)^{1_{\{1/2\}}(\beta_j)/2},$$

so that

$$\int_{\mathbb{R}^d} \sqrt{\mathrm{Var}(\mathcal{F}_n(\mathbf{x}))} \, \mathbf{dx} \lesssim \frac{1}{\sqrt{n}} \prod_{j=1}^d \sqrt{I_j(h_j)} \lesssim \frac{1}{\sqrt{n}} \prod_{j=1}^d h_j^{-[(\beta_j \vee 1/2) - 1/2]} (|\log h_j|)^{1_{\{1/2\}}(\beta_j)/2}.$$

The assertion follows by combining the bounds on the bias and the variance and taking $h_{1,\mathrm{opt}} = \ldots = h_{d,\mathrm{opt}} = n^{-1/\sum_{j=1}^d [(2\beta_j \vee 1) + 1]}$. Note that, in the bias-variance decomposition, the bias is the dominant term.
□

Remark 1. For $d = 1$, Theorem 1 recovers, as a special case, the convergence rate of the deconvolution distribution function estimator $\mathcal{F}_n$ studied in [4], Theorem 3.1, p. 243.

Remark 2. The assumption in (7) has been discussed in Remark 3.4 of [4], pp. 244-245. It is shown to be equivalent to the two conditions $\int_0^{+\infty} \sqrt{\mathrm{P}(|X_{1,j}| > x)} \, \mathrm{d}x < \infty$ and $\int_0^{+\infty} \sqrt{\mathrm{P}(|\varepsilon_{1,j}| > u)} \, \mathrm{d}u < \infty$ jointly considered. Each of these conditions implies that the corresponding distribution has a finite first absolute moment; hence, the corresponding characteristic function is continuously differentiable.

3. Lower Bound

In this section, we derive a lower bound on the L^1-risk over classes of uniformly bounded densities.

Theorem 2. *Suppose that we observe a sample* $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ *from the multivariate convolution model* (1). *Assume that conditions* (2)-(4) *on the error distribution are satisfied for* $\beta_j \geq 1/2$ *for all* $j = 1, \ldots, d$. *For* $M > 0$, *let* $\mathcal{B}(M)$ *be the class of probability measures on* $\mathbb{R}^d$ *with Lebesgue densities uniformly bounded by M. Then, there exists a constant C>0 such that, for any distribution function estimator* $\tilde{\mathcal{F}}_n$, *we have*

$$\liminf_{n \to \infty} n^{1/\sum_{j=1}^d (2\beta_j + 1)} \sup_{\mu_X \in \mathcal{B}(M)} \mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}\left[\|\tilde{\mathcal{F}}_n - F_X\|_1\right] \geq C.$$

Proof. The proof combines key arguments of the proofs of Theorem 3 in [5], pp. 281-284, and Theorem 3 in [3], pp. 599-603. For $h = n^{-1/\sum_{j=1}^d (2\beta_j + 1)}$, let $M = \lfloor h^{-1} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the greatest integer smaller than or equal to h^{-1}. Let

$\mathcal{K} := \{1, \ldots, M\}^d$. For $\mathbf{k} \in \mathcal{K}$, let $\mathbf{x_k}$ be the column vector with jth coordinate $x_{\mathbf{k}j} := a_j + k_j(b_j - a_j)h$, where $b_j - a_j \geq (2\pi)^{-1}$, $j = 1, \ldots, d$. We define a finite class of probability measures $\mu_{\boldsymbol{\theta}}$'s on $\mathbb{R}^d$ with Lebesgue densities

$$f_{\boldsymbol{\theta}}(\mathbf{x}) := f_0(\mathbf{x}) + c\sum_{\mathbf{k} \in \mathcal{K}} \theta_{\mathbf{k}} \prod_{j=1}^{d} H\left(\frac{x_j - x_{\mathbf{k}j}}{2h}\right), \qquad \mathbf{x} \in \mathbb{R}^d, \tag{10}$$

where c is a positive constant, $f_0(\mathbf{x}) := \prod_{j=1}^{d} c_j^{-1} g_{s_j}(x_j / c_j)$ for suitably chosen large enough positive constants c_j's and functions g_{s_j}'s that are densities of symmetric stable laws with characteristic functions $\mathcal{F}[g_{s_j}(x)](t) = \exp(-|t|^{s_j})$, $t \in \mathbb{R}$, and $0 < s_j < 1$. By Lemma 4 in [3], p. 590, the densities g_{s_j}'s are bounded. The real function H is such that $\int H = 0$, $H(0) \neq 0$, $|H(x)| = O(x^{-\delta})$ (as $|x| \to \infty$) with $\delta > 3$, and $\mathcal{F}[H(x)](t) = 0$ when $|t| \notin [1, 2]$. Let $\boldsymbol{\theta} \in \{0, 1\}^{M^d}$. For $i = 0, 1$ and $\mathbf{k} \in \mathcal{K}$, let $\boldsymbol{\theta}_{\mathbf{k}}^i$ be the sequence such that

$$\left(\boldsymbol{\theta}_{\mathbf{k}}^i\right)_{\mathbf{l}} = \begin{cases} i & \mathbf{l} = \mathbf{k}, \\ \theta_{\mathbf{l}} & \mathbf{l} \neq \mathbf{k}, \end{cases} \qquad \mathbf{l} \in \mathcal{K}.$$

It is known from the proof of Theorem 3 in [3], pp. 599-603, that the hypothesis functions $f_{\boldsymbol{\theta}}$ in (10) are uniformly bounded densities. Also, denoted by f_ε the density of $\boldsymbol{\varepsilon}_1$, then, uniformly in $\mathbf{k} \in \mathcal{K}$, we have that

$$\chi^2(f_{\boldsymbol{\theta}_{\mathbf{k}}^1} * f_\varepsilon; f_{\boldsymbol{\theta}_{\mathbf{k}}^0} * f_\varepsilon) := \int \frac{(f_{\boldsymbol{\theta}_{\mathbf{k}}^1} * f_\varepsilon - f_{\boldsymbol{\theta}_{\mathbf{k}}^0} * f_\varepsilon)^2}{f_{\boldsymbol{\theta}_{\mathbf{k}}^0} * f_\varepsilon} = O(n^{-1}). \tag{11}$$

For $\boldsymbol{\theta} \in \{0, 1\}^{M^d}$, let $F_{\boldsymbol{\theta}}$ be the distribution function having density $f_{\boldsymbol{\theta}}$ defined in (10). Then,

$$\sup_{\mu_X \in \mathcal{B}(M)} \mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}\left[\left\|\mathcal{F}_n - F_X\right\|_1\right] \geq \max_{\boldsymbol{\theta} \in \{0, 1\}^{M^d}} \mathrm{E}_{(\mu_{\boldsymbol{\theta}} * \mu_\varepsilon)^{\otimes n}}\left[\left\|\mathcal{F}_n - F_{\boldsymbol{\theta}}\right\|_1\right].$$

Let $\boldsymbol{\theta}$ be a sequence of Bernoulli random variables $\theta_{\mathbf{k}}$, $\mathbf{k} \in \mathcal{K}$, such that $\mathrm{P}(\theta_{\mathbf{k}} = i) = 1/2$, $i = 0, 1$. For $\mathbf{k} \in \mathcal{K}$, let $D_{\mathbf{k}} := \times_{j=1}^{d}\left[x_{\mathbf{k}j} - (b_j - a_j)h,\, x_{\mathbf{k}j}\right]$. For $\mathbf{x} \in D_{\mathbf{k}}$,

$$\begin{aligned} \max_{\boldsymbol{\theta} \in \{0, 1\}^{M^d}} \mathrm{E}_{(\mu_{\boldsymbol{\theta}} * \mu_\varepsilon)^{\otimes n}}\left[\left|\mathcal{F}_n(\mathbf{x}) - F_X(\mathbf{x})\right|\right] &\geq \mathrm{E}_{\boldsymbol{\theta}} \mathrm{E}_{(\mu_{\boldsymbol{\theta}} * \mu_\varepsilon)^{\otimes n}}\left[\left|\mathcal{F}_n(\mathbf{x}) - F_X(\mathbf{x})\right|\right] \\ &\geq \frac{1}{2}\mathrm{E}_{\boldsymbol{\theta}}\left[\left|\left(F_{\boldsymbol{\theta}_{\mathbf{k}}^1}(\mathbf{x}) - F_{\boldsymbol{\theta}_{\mathbf{k}}^0}(\mathbf{x})\right)\right| \times \int_{(\mathbb{R}^d)^n} \min\left\{(f_{\boldsymbol{\theta}_{\mathbf{k}}^1} * \mu_\varepsilon)^{\otimes n}, (f_{\boldsymbol{\theta}_{\mathbf{k}}^0} * \mu_\varepsilon)^{\otimes n}\right\} \mathbf{dy}_1 \ldots \mathbf{dy}_n\right], \end{aligned}$$

where, uniformly in $\boldsymbol{\theta}_{\mathbf{k}}^i$, $i = 0, 1$, and $\mathbf{k} \in \mathcal{K}$, the integral is bounded below using the result in (11) and

$$\left|\left(F_{\boldsymbol{\theta}_{\mathbf{k}}^1}(\mathbf{x}) - F_{\boldsymbol{\theta}_{\mathbf{k}}^0}(\mathbf{x})\right)\right| = (2h)^d c \prod_{j=1}^{d}\left|H^{(-1)}\left(\frac{x_j - x_{\mathbf{k}j}}{2h}\right)\right|,$$

where $H^{(-1)}(x) := \int_{-\infty}^{x} H(u)\,\mathrm{d}u$, $x \in \mathbb{R}$, is the primitive of H. It follows that

$$\sup_{\mu_X \in \mathcal{B}(M)} \mathrm{E}_{(\mu_X * \mu_\varepsilon)^{\otimes n}}\left[\left\|\mathcal{F}_n - F_X\right\|_1\right] \gtrsim h^d \int_{\times_{j=1}^{d}[a_j, b_j]} \prod_{j=1}^{d}\left|H^{(-1)}\left(\frac{x_j - x_{\mathbf{k}j}}{2h}\right)\right| \mathbf{dx}.$$

Choosing $(b_j - a_j)$ such that $\int_0^{(b_j - a_j)/2}\left|H^{(-1)}(-z)\right|\mathrm{d}z \geq 1/h$ for every $j = 2, \ldots, d$, we obtain that

$$\begin{aligned} h^d \int_{\times_{j=1}^{d}[a_j, b_j]} \prod_{j=1}^{d}\left|H^{(-1)}\left(\frac{x_j - x_{\mathbf{k}j}}{2h}\right)\right| \mathbf{dx} &= h^d \sum_{\mathbf{k} \in \mathcal{K}} \int_{D_{\mathbf{k}}} \prod_{j=1}^{d}\left|H^{(-1)}\left(\frac{x_j - x_{\mathbf{k}j}}{2h}\right)\right| \mathbf{dx} \\ &\geq h^d \sum_{\mathbf{k} \in \mathcal{K}} \int_{D_{\mathbf{k}}} \prod_{j=1}^{d}\left|H^{(-1)}\left(\frac{x_j - x_{\mathbf{k}j}}{2h}\right)\right| \mathbf{dx} \end{aligned}$$

$$= h^d \sum_{\mathbf{k}\in\mathcal{K}} \prod_{j=1}^{d} \int_{x_{\mathbf{k}j}-(b_j-a_j)h}^{x_{\mathbf{k}j}} \left| H^{(-1)}\left(\frac{x_j - x_{\mathbf{k}j}}{2h}\right)\right| \mathrm{d}x_j$$
$$= M^d h^{2d} \prod_{j=1}^{d} 2\int_0^{(b_j-a_j)/2} \left|H^{(-1)}(-z)\right| \mathrm{d}z$$
$$\gtrsim h^d \prod_{j=1}^{d} \int_0^{(b_j-a_j)/2} \left|H^{(-1)}(-z)\right| \mathrm{d}z \gtrsim h = n^{-1/\sum_{j=1}^{d}(2\beta_j+1)},$$

and the proof is complete. □

Remark 3. Even if the lower bound rate matches the upper bound of the distribution function deconvolution estimator F_n introduced in Section 2.2, it should be noted that the class of probability measures with independent coordinates is a subset of probability measures with uniformly bounded densities. Therefore, the lower bound could potentially be different.

4. Conclusion

In this paper, we study the problem of estimating a distribution function on $\mathbb{R}^d$, for any dimension $d \geq 1$, from data contaminated by additive noise with known density. We employ the L^1-distance between distribution functions as a discrepancy measure, which can be viewed as a multivariate extension of the L^1-Wasserstein distance between one-dimensional probability measures. First, we derive an upper bound on the convergence rate for the integrated classical deconvolution kernel density estimator, without imposing any regularity conditions on the mixing distribution beyond the existence of a Lebesgue density. We then prove a matching lower bound for the same estimation problem. Some important extensions remain to be fully investigated. These include (i) deriving adaptive minimax-optimal convergence rates for anisotropic mixing distribution functions on Hölder or Sobolev scales, (ii) developing new estimation methods for the case where the error distribution is unknown and must be estimated from data.

Acknowledgements

The author is a member of the “Gruppo Nazionale per l’Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA)” of the Istituto Nazionale di Alta Matematica (INdAM). She has been partially supported by Università di Verona and MUR–Prin 2022–Grant No. 2022CLTYP4, funded by the European Union–Next Generation EU.

References

[1] C. Caillerie, F. Chazal, J. Dedecker and B. Michel, “Deconvolution for the Wasserstein metric and geometric inference,” *Electron. J. Stat.*, vol. 5, pp. 1394-1423, 2011.
[2] R. Carroll and P. Hall, “Optimal rates of convergence for deconvolving a density,” *J. Amer. Statist. Assoc.*, vol. 83 (404), pp. 1184-1186.
[3] F. Comte and C. Lacour, “Anisotropic adaptive kernel deconvolution,” *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 49 (2), pp. 569-609, 2013.
[4] J. Dedecker, A. Fischer and B. Michel, “Improved rates for Wasserstein deconvolution with ordinary smooth error in dimension one,” *Electron. J. Statist.*, vol. 9 (1), pp. 234-265, 2015.
[5] J. Dedecker and B. Michel, “Minimax rates of convergence for Wasserstein deconvolution with supersmooth errors in any dimension,” *J. Multivar. Anal.*, vol. 122, pp. 278-291, 2013.
[6] A. Meister, *Deconvolution Problems in Nonparametric Statistics*. Lecture Notes in Statistics. Springer, 2009.