Improved rates for Wasserstein deconvolution with ordinary smooth error in dimension one

Jérôme Dedecker, Aurélie Fischer, Bertrand Michel

Improved rates for Wasserstein deconvolution with ordinary smooth error in dimension one

Jérôme Dedecker[(1)], Aurélie Fischer[(2)] and Bertrand Michel[(3)]

April 2, 2014

(1) Laboratoire MAP5 UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité

(2) LPMA UMR CNRS 7599, Université Paris Diderot, Sorbonne Paris Cité

(3) LSTA, Université Pierre et Marie Curie

Abstract

This paper deals with the estimation of a probability measure on the real line from data observed with an additive noise. We are interested in rates of convergence for the Wasserstein metric of order $p \geq 1$. The distribution of the errors is assumed to be known and to belong to a class of supersmooth or ordinary smooth distributions. We obtain in the univariate situation an improved upper bound in the ordinary smooth case and less restrictive conditions for the existing bound in the supersmooth one. In the ordinary smooth case, a lower bound is also provided, and numerical experiments illustrating the rates of convergence are presented.

1 Introduction

Consider the following convolution model: we observe n real-valued random variables $Y_1, \ldots, Y_n$ such that

$$Y_i = X_i + \varepsilon_i, \tag{1}$$

where the X_i's are independent and identically distributed according to an unknown probability μ, which we want to estimate. The random variables ε_i, $i = 1, \ldots, n$, are independent and identically distributed according to a known probability measure μ_ε, not necessarily symmetric. Moreover we assume that $(X_1, \ldots, X_n)$ is independent of $(\varepsilon_1, \ldots, \varepsilon_n)$.

The purpose of the paper is to investigate rates of convergence for the estimation of the measure μ under Wasserstein metrics. For $p \in [1, \infty)$, the Wasserstein distance W_p between μ and ν is given by

$$W_p(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \left(\int_{\mathbb{R}^2} |x - y|^p \pi(dx, dy) \right)^{\frac{1}{p}},$$

where $\Pi(\mu, \nu)$ is the set of probability measures on $\mathbb{R} \times \mathbb{R}$ with marginal distributions μ and ν (see Rachev and Rüschendorf (1998) or Villani (2008)). The distances W_p are natural metrics for comparing measures, which makes the Wasserstein deconvolution problem interesting in itself. In addition, as pointed out in Caillerie et al. (2011), they are also related to the results of Chazal et al. (2011) in geometric inference, where a distance function to measures is introduced to solve geometric inference problems in a probabilistic setting : if a known measure ν is close enough with respect to W_2 to a measure μ concentrated on a given shape, then the topological properties of the shape can be recovered by using the

distance to ν. In practice, the data can be observed with noise, which motivates in this framework the study of the Wasserstein deconvolution problem.

Rates of convergence in deconvolution have mostly been considered in density estimation, for pointwise or global convergence. Minimax rates can be found for instance in Fan (1991a), Butucea and Tsybakov (2008a), Butucea and Tsybakov (2008b) and in the monograph of Meister (2009). In this paper, however, we shall not assume that μ has a density with respect to the Lebesgue measure. In this context, rates of convergence for the W_2 Wasserstein distance have first been studied for several noise distributions by Caillerie et al. (2011). Recently, Dedecker and Michel (2013) have obtained optimal rates of convergence in the minimax sense for a class of supersmooth error distributions, in any dimension, under any Wasserstein metric W_p. The result relies on the fact that lower bounds in any dimension can be deduced in this case from the lower bounds in dimension 1. Such a method cannot be used in the ordinary smooth case, where the rate of convergence depends on the dimension. As noticed by Fan (1991a), establishing optimal rates of convergence in the ordinary smooth case is more difficult than in the supersmooth one, even for pointwise estimation.

A key fact in the univariate context is that Wasserstein metrics are linked to integrated risks between cumulative distribution functions (cdf), see the upper bound (5) below. In dimension 1, when estimating the density of μ, optimal rates of convergence for integrated risks can be found in Fan (1991b, 1993). When estimating the cdf F of μ, optimal rates for the pointwise and integrated quadratic risks are given in Hall and Lahiri (2008), where it is shown in particular that the rate $\sqrt{n}$ can be reached when the error distribution is ordinary smooth with a smoothness index less than 1/2. Concerning the pointwise estimation of $F(x_0)$, optimal rates for the quadratic risk are also given in Dattner et al. (2011), when the density of μ belongs to a Sobolev class.

The case $\beta = 0$ in the upper bound (3.9) of Hall and Lahiri (2008) corresponds to the case where no assumption (except a moment assumption) is made on the measure μ (in particular μ is not assumed to be absolutely continuous with respect to the Lebesgue measure). This is precisely the case which we want to consider in the present paper. However the results by Hall and Lahiri (2008) cannot be applied to the Wasserstein deconvolution problems for two reasons: firstly, the integrated quadratic risk for estimating a cdf is not linked to Wasserstein distances, and secondly, the estimator of the cdf of μ proposed in Hall and Lahiri (2008) is the cdf of a signed measure, and is not well defined as an estimator of μ for the Wasserstein metric.

In the present contribution, we propose in the univariate situation an improved upper bound for deconvolving μ under W_p, and a lower bound when the error is ordinary smooth. We recover the optimal rate of convergence in the supersmooth case with slightly weaker regularity conditions than in Dedecker and Michel (2013). The estimator of the cdf F of μ is built in two steps: firstly, as in Hall and Lahiri (2008), we define a preliminary estimator through a classical kernel deconvolution method, and secondly we take an appropriate isotone approximation of this estimator.

The paper is organized as follows. In Section 2, some facts about the case without error are recalled and discussed. The upper bounds for Wasserstein deconvolution with supersmooth or ordinary smooth errors are given in Section 3, and Section 4 is about lower bounds. Section 5 presents the implementation of the method and some experimental results. In particular, observed rates of convergence are compared with the theoretical bounds for the Wasserstein metrics W_1 and W_2, and we study as an illustrative example the deconvolution of the uniform measure on the Cantor set.

2 On the case without error

We begin by considering the simple case when one observes directly $X_1, \ldots, X_n$ with values in $\mathbb{R}$ without error. Let us recall some results for the quantities $W_p(\mu_n, \mu)$, where μ_n is the empirical measure, given by

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}.$$

Let F be the cdf of X_1, F_n the cdf of μ_n, and let F^{-1} and F_n^{-1} be their usual cadlag inverses. Recall that, for any $p \geq 1$,

$$W_p^p(\mu_n, \mu) = \int_0^1 |F_n^{-1}(u) - F^{-1}(u)|^p du\,, \tag{2}$$

and if $p = 1$

$$W_1(\mu_n, \mu) = \int |F_n(t) - F(t)| dt\,.$$

The case $p = 1$ is well understood since the paper by del Barrio et al. (1999). The random variable $\sqrt{n} W_1(\mu_n, \mu)$ converges in distribution to $\int |B(F(t))| dt$, where B is a standard Brownian bridge, if and only if

$$\int_0^\infty \sqrt{P(|X| > t)} dt < \infty. \tag{3}$$

For $p > 1$, the situation is not so clear. For instance, if the random variables take their values in a compact interval $[a, b]$ and if the cdf F is continuously differentiable on $[a, b]$ with strictly positive derivative f, then $n^{p/2} W_p^p(\mu_n, \mu)$ converges in distribution to $\int_0^1 |B(u)|^p / |f \circ F^{-1}(u)|^p du$ (see Lemma 3.9.23 in van der Vaart and Wellner (1996)). But in general, the rate can be much slower.

The case $p = 2$ has been studied in detail by del Barrio et al. (2005). Under additional conditions on F (see condition (2.7) in del Barrio et al. (2005), which requires in particular that F is twice differentiable), the rate of convergence depends on the behavior of F^{-1} in a neighborhood of 0 and 1. For instance, if

$$F(t) = \left(1 - \frac{1}{t^{\alpha - 1}}\right) \mathbf{1}_{\{t > 1\}},$$

where $\alpha > 3$, it follows from Theorem 4.7 in del Barrio et al. (2005) that

$$n^{(\alpha-3)/(\alpha-1)} W_2^2(\mu_n, \mu) \tag{4}$$

converges in distribution. The limiting distribution is explicitly given in del Barrio et al. (2005).

An alternative approach to obtain convergence rates is to use the following inequality, due to Èbralidze (1971): for any $p \geq 1$,

$$W_p^p(\mu, \nu) \leq \kappa_p \int |x|^{p-1} |F_\mu - F_\nu|(x) dx\,, \tag{5}$$

where $\kappa_p = 2^{p-1} p$.

Starting from (5) and arguing as in del Barrio et al. (1999), it follows that

$$\mathbb{E} W_p^p(\mu_n, \mu) \leq C n^{-1/2}$$

as soon as

$$\int_0^\infty |x|^{p-1}\sqrt{P(|X|>x)}dx < \infty\,. \tag{6}$$

For instance, taking $p=2$, a tail satisfying $P(|X|>x) = \mathcal{O}\Big(\frac{1}{x^4 \log x^{2+\varepsilon}}\Big)$ gives the rate $\sqrt{n}$. Hence, we obtain the same rate as in (4) for $\alpha = 5$, with a slightly stronger tail condition (due to the fact that we control the expectation), but without additional assumptions on the cdf F.

Since we want to estimate singular measures, we shall follow this approach in the sequel.

3 Upper bounds for W_p in deconvolution

3.1 Definitions and notation

Let us start with some notation. For μ a probability measure and ν another probability measure, with density g, we denote by $\mu \star g$ the density of $\mu \star \nu$, given by

$$\mu \star g(x) = \int_{\mathbb{R}} g(x-y)\mu(dy).$$

We further denote by μ^* (respectively f^*) the Fourier transform of the probability measure μ (respectively of the integrable function f), that is

$$\mu^*(x) = \int_{\mathbb{R}} e^{iux}\mu(du) \quad \text{and} \quad f^*(x) = \int_{\mathbb{R}} e^{iux} f(u)du.$$

Thanks to Èbralidze's inequality (5) recalled in Section 2, an upper bound on the rate of convergence of an estimator $\tilde{\mu}_n$ of the distribution μ may be derived by studying the quantity

$$\int |x|^{p-1}|\tilde{F}_n - F|(x)dx\,,$$

where $\tilde{F}_n$ is the cdf of $\tilde{\mu}_n$.

Let $\lceil p \rceil$ be the least integer greater than or equal to p. We first introduce a kernel k such that its Fourier transform k^* is $\lceil p \rceil$ times differentiable with Lipschitz $\lceil p \rceil -$th derivative and is supported on $[-1,1]$. An example of such a kernel is given by

$$k(x) = C_p \left[\frac{(2\lceil p/2 \rceil + 2)\sin\frac{x}{2\lceil p/2\rceil+2}}{x}\right]^{2\lceil p/2\rceil+2}, \tag{7}$$

where C_p is such that $\int k(x)dx = 1$.

We define now a preliminary estimator $\hat{F}_n$ of F:

$$\hat{F}_n(t) = \frac{1}{nh}\int_{-\infty}^t \sum_{k=1}^n \tilde{k}_h\left(\frac{u-Y_k}{h}\right)du \tag{8}$$

where

$$\tilde{k}_h(x) = \Re\mathrm{e}\left[\frac{1}{2\pi}\int \frac{e^{iux}k^*(u)}{\mu_\varepsilon^*(-u/h)}du\right].$$

This estimator $\hat{F}_n$, based on the standard deconvolution kernel density estimator first introduced by Carroll and Hall (1988), is not a distribution function since it is not necessarily non-decreasing.

For this reason, we choose the estimator $\tilde{F}_n$ as an approximate minimizer over all distribution functions of the quantity $\int_{\mathbb{R}} |x|^{p-1}|\hat{F}_n - G|(x)dx$. Given $\rho > 0$, let $\tilde{F}_n$ be such that, for every distribution function G,

$$\int |x|^{p-1}|\hat{F}_n - \tilde{F}_n|(x)dx \leq \int |x|^{p-1}|\hat{F}_n - G|(x)dx + \rho\,, \tag{9}$$

and let $\tilde{\mu}_n$ be the probability measure with distribution function $\tilde{F}_n$. Here, ρ may be chosen equal to $n^{-1/2}$ (or any other sequence converging faster to 0 as n tends to ∞). Denoting by K_h the function $h^{-1}k(\cdot/h)$, we have that

$$\mathbb{E}W_p^p(\tilde{\mu}_n, \mu) \leq 2^{p-1}W_p^p(\mu \star K_h, \mu) + 2^{p-1}\mathbb{E}W_p^p(\tilde{\mu}_n, \mu \star K_h). \tag{10}$$

In order to control the first term of the right-hand side, let V_h be a random variable with distribution K_h and independent of X_1, in such a way that the distribution of $X_1 + V_h$ is $\mu \star K_h$. By definition of W_p, we have

$$W_p^p(\mu \star K_h, \mu) \leq \mathbb{E}|X_1 + V_h - X_1|^p = \mathbb{E}|V_h|^p = h^p \int |x|^p k(x)dx. \tag{11}$$

Besides, since $\mathbb{E}[\hat{F}_n(t)] = \int_{-\infty}^t \mu \star K_h(x)dx$ is the cdf of $\mu \star K_h$, using inequality (5),

$$\begin{aligned} W_p^p(\tilde{\mu}_n, \mu \star K_h) &\leq \kappa_p \int |x|^{p-1}|\tilde{F}_n - \mathbb{E}[\hat{F}_n]|(x)dx \\ &\leq \kappa_p \left(\int |x|^{p-1}|\tilde{F}_n - \hat{F}_n|(x)dx + \int |x|^{p-1}|\hat{F}_n - \mathbb{E}[\hat{F}_n]|(x)dx \right) \\ &\leq \rho + 2\kappa_p \int |x|^{p-1}|\hat{F}_n - \mathbb{E}[\hat{F}_n]|(x)dx, \end{aligned} \tag{12}$$

by the definition of $\tilde{F}_n$. To get explicit rates of convergence, it remains to control the term

$$\int |x|^{p-1}|\hat{F}_n - \mathbb{E}[\hat{F}_n]|(x)dx\,.$$

3.2 Main results

Let $r_\varepsilon = 1/\mu_\varepsilon^*$, and let $r_\varepsilon^{(\ell)}$ be the ℓ-th derivative of r_ε. Let m_0 denote the least integer strictly greater than $p + \frac{1}{2}$, and m_1 be the least integer strictly greater than $p - \frac{1}{2}$.

Our first result is a general proposition which gives an upper bound for $\mathbb{E}W_p^p(\tilde{\mu}_n, \mu)$ involving a tail condition on Y and the regularity of r_ε.

Proposition 3.1. *Let $\rho \leq n^{-1/2}$, and let $\tilde{\mu}_n$ be the estimator defined in (9). Assume that r_ε is m_0 times differentiable. For any $h \leq 1$, we have*

$$\mathbb{E}W_p^p(\tilde{\mu}_n, \mu) \leq \frac{1}{\sqrt{n}} + h^p 2^{p-1} \int |x|^p k(x)dx + \frac{C}{\sqrt{n}}(A_1 + A_2 + A_3 + A_4)$$

where

$$A_1 = \Big(\sup_{t\in[-2,2]} \sum_{\ell=0}^{1} |r_\varepsilon^{(\ell)}(t)| \Big) \int_0^\infty |x|^{p-1} \sqrt{P\left(|Y| \geq x\right)} dx$$

$$A_2 = \sup_{t\in[-2,2]} \sum_{\ell=0}^{m_0} |r_\varepsilon^{(\ell)}(t)|$$

$$A_3 = \left[\mathbb{E}|Y|^{2p-\frac{1}{2}} \int_{-1/h}^{1/h} \frac{|r_\varepsilon(x)|^2}{|x|^2} \mathbf{1}_{[-1,1]^c}(x) dx \right]^{1/2}$$

$$A_4 = \left[\sum_{\ell=0}^{m_1} \int_{-1/h}^{1/h} \frac{|r_\varepsilon^{(\ell)}(x)|^2}{|x|^2} \mathbf{1}_{[-1,1]^c}(x) dx \right]^{1/2}.$$

We are now in a position to give the rates of convergence for the Wasserstein deconvolution, for a class of supersmooth error distributions, and for a class of ordinary smooth error distributions.

Theorem 3.1. *Let $\rho \leq n^{-1/2}$, and let $\tilde{\mu}_n$ be the estimator defined in (9). Assume that*

$$\int_0^\infty |x|^{p-1} \sqrt{P(|Y| \geq x)} dx < \infty \text{ and } \sup_{t\in[-2,2]} |r_\varepsilon^{(m_0)}(t)| < \infty. \tag{13}$$

1. *Assume that there exist $\beta > 0$, $\tilde{\beta} \geq 0$, $\gamma > 0$ and $c > 0$, such that for every $\ell \in \{0, 1, \ldots, m_1\}$ and every $t \in \mathbb{R}$,*

$$|r_\varepsilon^{(\ell)}(t)| \leq c(1 + |t|)^{\tilde{\beta}} \exp(|t|^\beta/\gamma). \tag{14}$$

Then, taking $h = (4/(\gamma \log n))^{1/\beta}$, there exists a positive constant C such that

$$\mathbb{E}W_p^p(\tilde{\mu}_n, \mu) \leq C(\log n)^{-p/\beta}.$$

2. *Assume that there exist $\beta > 0$ and $c > 0$, such that for every $\ell \in \{0, 1, \ldots, m_1\}$ and every $t \in \mathbb{R}$,*

$$|r_\varepsilon^{(\ell)}(t)| \leq c(1 + |t|)^{\beta}. \tag{15}$$

Then, taking $h = n^{-\frac{1}{2p+(2\beta-1)_+}}$, there exists a positive constant C such that

$$\mathbb{E}W_p^p(\tilde{\mu}_n, \mu) \leq C\psi_n\,, \tag{16}$$

where

$$\psi_n = \begin{cases} n^{-\frac{p}{2p+2\beta-1}} & \text{if } \beta > \frac{1}{2} \\ \sqrt{\frac{\log n}{n}} & \text{if } \beta = \frac{1}{2} \\ \frac{1}{\sqrt{n}} & \text{if } \beta < \frac{1}{2}. \end{cases}$$

This result requires several comments.

Remark 3.1. *In the ordinary smooth case, when $\beta < 1/2$, any bandwidth $h = \mathcal{O}(n^{-1/2p})$ leads to the rate $n^{-1/2}$. The fact that there are three different situations according as $\beta > 1/2, \beta = 1/2$ or $\beta < 1/2$ has already been pointed out in Theorem 3.2 of Hall and Lahiri (2008) and in Theorem 2.1 of Dattner et al. (2011) for the estimation of the cdf F. Note that the estimator $\hat{F}_n$ of Hall and Lahiri (2008) is exactly the estimator defined*

in (8) (with possibly a slightly different kernel). Hence it is not always non-decreasing and cannot be used directly to estimate μ with respect to Wasserstein metrics.

For instance, for a Laplace error distribution, the estimator $\hat{F}_n$ of Hall and Lahiri (2008) is such that

$$\Big(\mathbb{E}\Big[\int |\hat{F}_n(t) - F(t)|^2 dt\Big]\Big)^{1/2} \leq Cn^{-1/8}\,,$$

while the rate of convergence of our estimator for W_1 is

$$\mathbb{E}W_1(\tilde{\mu}_n, \mu) = \mathbb{E}\Big[\int |\tilde{F}_n(t) - F(t)| dt\Big] \leq Cn^{-1/5}\,.$$

Let us give another application of Theorem 3.2 of Hall and Lahiri (2008). For a Laplace error distribution and μ such that $|\mu^*(x)| \leq C(1+|x|)^{-1/2}$, the estimator $\hat{F}_n$ of Hall and Lahiri (2008) is such that

$$\Big(\sup_{x\in\mathbb{R}} \mathbb{E}|\hat{F}_n(x) - F(x)|^2\Big)^{1/2} \leq Cn^{-1/8}\,.$$

Dattner et al. (2011) focused on the pointwise estimation of $F(x_0)$. In this paper, the authors always assume that μ is absolutely continuous with respect to the Lebesgue measure, with a density f belonging to a Sobolev space of order $\alpha > -1/2$. For instance, for a density belonging to $\mathbb{L}^2$ and a Laplace error distribution, their estimator $\bar{F}_n$ is such that

$$\big(\mathbb{E}|\bar{F}_n(x_0) - F(x_0)|^2\big)^{1/2} \leq Cn^{-1/8}\,.$$

In any cases these rates are minimax (see Section 4 for our estimator).

Remark 3.2. *The tail condition*

$$\int_0^\infty |x|^{p-1}\sqrt{P(|Y| \geq x)}dx < \infty$$

in Assumption (13) is the same as the tail condition (6) obtained in Section 2 to get the rate $\mathbb{E}W_p^p(\mu_n, \mu) \leq Cn^{-1/2}$ in the case without noise. Recall that, in the case without noise when $p = 1$, this condition is necessary and sufficient for the weak convergence of $\sqrt{n}W_1(\mu_n, \mu)$.

Remark 3.3. *The rate $\mathbb{E}W_p^p(\tilde{\mu}_n, \mu) \leq C(\log n)^{-p/\beta}$ in the supersmooth case has already been given in Theorem 4 of Dedecker and Michel (2013) and is valid in any dimension. However the condition on the regularity of r_ε is more restrictive in the paper by Dedecker and Michel (2013), since it is assumed there that Condition (14) is true for $\ell \in \{0, 1, \ldots, \lceil p \rceil + 1\}$. Note that this rate is minimax, as stated in Theorem 2 of Dedecker and Michel (2013).*

Remark 3.4. *Applying Proposition 1 in Dedecker and Michel (2013), if Condition (15) is true for $\ell \in \{0, 1, \ldots, \lceil p \rceil + 1\}$, one can build an explicit estimator $\bar{\mu}_n$ such that $\mathbb{E}W_p^p(\bar{\mu}_n, \mu) \leq Cn^{-p/(2p+2\beta+1)}$, which is worse than (16). However, the procedure given in Dedecker and Michel (2013) works also when the observations Y_i are $\mathbb{R}^d$-valued, whereas the estimator $\tilde{\mu}_n$ defined in (9) is well defined for $d = 1$ only. Hence, a reasonable question is: can we improve on Proposition 1 of Dedecker and Michel (2013) in any dimension?*

Proof. We first prove Item 1. From Proposition 3.1 and Assumptions (13) and (14), we obtain the upper bound

$$\mathbb{E}W_p^p(\tilde{\mu}_n, \mu) \leq C\left(h^p + \frac{1}{\sqrt{n}}\frac{1}{h^{\tilde{\beta}}}e^{1/h^\beta\gamma}\right).$$

Taking $h = (4/(\gamma\log(n)))^{1/\beta}$ gives the result.

We now prove Item 2. From Proposition 3.1 and Assumptions (13) and (15), we obtain

$$\mathbb{E}W_p^p(\tilde{\mu}_n, \mu) \leq \begin{cases} C\left(h^p + \frac{1}{\sqrt{n}}\frac{1}{h^{\beta-1/2}}\right) & \text{if } \beta > \frac{1}{2} \\ C\left(h^p + \frac{1}{\sqrt{n}}\sqrt{\log(\frac{1}{h})}\right) & \text{if } \beta = \frac{1}{2} \\ C\left(h^p + \frac{1}{\sqrt{n}}\right) & \text{if } \beta < \frac{1}{2}. \end{cases}$$

Taking $h = n^{-\frac{1}{2p+(2\beta-1)_+}}$ gives the result.

□

3.3 Proof of Proposition 3.1

Throughout, C will denote a positive constant depending on p which may change from line to line.

We start from the basic inequality (10). Inequality (11) yields the bias term

$$h^p 2^{p-1}\int |x|^p k(x)dx\,,$$

and it remains to control the term $\mathbb{E}W_p^p(\tilde{\mu}_n, \mu \star K_h)$.

By (12), we have

$$\begin{aligned}\mathbb{E}W_p^p(\tilde{\mu}_n, \mu\star K_h) &\leq C\int |x|^{p-1}\mathbb{E}|\hat{F}_n - \mathbb{E}[\hat{F}_n]|(x)dx + \rho \\ &\leq C\int |x|^{p-1}\sqrt{\mathrm{Var}(\hat{F}_n)(x)}dx + \rho. \end{aligned} \tag{17}$$

Now, let ϕ denote a symmetric function, $\lceil p\rceil+1$ times continuously differentiable, equal to 1 on the interval $[-1,1]$ and to 0 outside $[-2,2]$. Our preliminary estimator $\hat{F}_n$ may be written

$$\begin{aligned}\hat{F}_n(t) &= \frac{1}{nh}\int_{-\infty}^t \sum_{k=1}^n \tilde{k}_h\left(\frac{u-Y_k}{h}\right)du \\ &= \frac{1}{n}\sum_{k=1}^n G_{1,h}\left(\frac{t-Y_k}{h}\right) + \frac{1}{n}\sum_{k=1}^n G_{2,h}\left(\frac{t-Y_k}{h}\right) \\ &:= \hat{F}_{1,n} + \hat{F}_{2,n},\end{aligned}$$

where

$$G_{1,h}(x) = \int_{-\infty}^x \tilde{k}_{1,h}(u)du \quad \text{and} \quad G_{2,h}(x) = \int_{-\infty}^x \tilde{k}_{2,h}(u)du.$$

Here,

$$\tilde{k}_{1,h}(u) = \Re\mathrm{e}\left[\frac{1}{2\pi}\int \frac{e^{itu}k^*(t)\phi(t/h)}{\mu_\varepsilon^*(-t/h)}dt\right], \quad \tilde{k}_{2,h}(u) = \Re\mathrm{e}\left[\frac{1}{2\pi}\int \frac{e^{itu}k^*(t)(1-\phi(t/h))}{\mu_\varepsilon^*(-t/h)}dt\right].$$

From (17), we infer that

$$\mathbb{E}W_p^p(\tilde{\mu}_n, \mu \star K_h) \leq C(I+J) + \rho, \tag{18}$$

where

$$I = \int |x|^{p-1}\sqrt{\mathrm{Var}(\hat{F}_{1,n})(x)}dx \quad \text{and} \quad J = \int |x|^{p-1}\sqrt{\mathrm{Var}(\hat{F}_{2,n})(x)}dx.$$

To prove Proposition 3.1, we shall give some upper bounds for the terms I and J.

Control of I. We first split the integral into two parts:

$$I = \int_{-\infty}^{0} |x|^{p-1}\sqrt{\mathrm{Var}(\hat{F}_{1,n})(x)}dx + \int_{0}^{\infty} |x|^{p-1}\sqrt{\mathrm{Var}(\hat{F}_{1,n})(x)}dx := I^- + I^+.$$

Now,

$$\begin{aligned} I^- &= \int_{-\infty}^{0} |x|^{p-1}\sqrt{\mathrm{Var}(\hat{F}_{1,n})(x)}dx \\ &\leq \frac{C}{\sqrt{n}}\int_{-\infty}^{0} |x|^{p-1}\sqrt{\mathbb{E}\left[G_{1,h}\left(\frac{x-Y}{h}\right)\right]^2}dx \\ &\leq \frac{C}{\sqrt{n}}\int_{-\infty}^{0} |x|^{p-1}\sqrt{\mathbb{E}\left[\int \tilde{k}_{1,h}(u)\mathbf{1}_{\left\{u\leq\frac{x-Y}{h}\right\}}du\right]^2}dx. \end{aligned}$$

Then, letting $z = uh$ and applying Cauchy-Schwarz Inequality we obtain, for any $a \in]0,1[$,

$$\begin{aligned} I^- &\leq \frac{C}{\sqrt{n}}\int_{-\infty}^{0} |x|^{p-1}\sqrt{\mathbb{E}\left[\int \frac{\tilde{k}_{1,h}(z/h)}{h}\mathbf{1}_{\{Y+z\leq x\}}dz\right]^2}dx \\ &\leq \frac{C}{\sqrt{n}}\int_{-\infty}^{0} |x|^{p-1}\sqrt{\mathbb{E}\left[\int (1+|z|^{1+a})\left(\frac{\tilde{k}_{1,h}(z/h)}{h}\right)^2\mathbf{1}_{\{Y+z\leq x\}}dz\right]}dx\,. \end{aligned}$$

Noticing that $\mathbf{1}_{\{Y+z\leq x\}} \leq \mathbf{1}_{\{Y\leq\frac{x}{2}\}} + \mathbf{1}_{\{z\leq\frac{x}{2}\}}$, we obtain that $I^- \leq I_1^- + I_2^-$, where

$$\begin{aligned} I_1^- &= \frac{C}{\sqrt{n}}\int_{-\infty}^{0} |x|^{p-1}\sqrt{\mathbb{E}\left[\int (1+|z|^{1+a})\left(\frac{\tilde{k}_{1,h}(z/h)}{h}\right)^2\mathbf{1}_{\left\{Y\leq\frac{x}{2}\right\}}dz\right]}dx \\ I_2^- &= \frac{C}{\sqrt{n}}\int_{-\infty}^{0} |x|^{p-1}\sqrt{\int (1+|z|^{1+a})\left(\frac{\tilde{k}_{1,h}(z/h)}{h}\right)^2\mathbf{1}_{\left\{z\leq\frac{x}{2}\right\}}dz}\,dx\,. \end{aligned}$$

To control the term I_1^-, note that

$$I_1^- \leq \frac{C}{\sqrt{n}}\sqrt{\int (1+|z|^2)\left(\frac{\tilde{k}_{1,h}(z/h)}{h}\right)^2 dz}\int_{-\infty}^{0} |x|^{p-1}\sqrt{\mathbb{P}\left(Y\leq\frac{x}{2}\right)}dx.$$

Here we shall use the following lemma.

Lemma 3.1. *For any nonnegative integer k and any $h \leq 1$ we have*

$$\int |z|^{2k} \left(\frac{\tilde{k}_{1,h}(z/h)}{h} \right)^2 dz \leq C \Big(\sup_{t \in [-2,2]} \sum_{\ell=0}^{k} |r_\varepsilon^{(\ell)}(t)| \Big)^2 .$$

Proof. By definition of $\tilde{k}_{1,h}$,

$$\int |z|^{2k} \left(\frac{\tilde{k}_{1,h}(z/h)}{h} \right)^2 dz \leq \frac{1}{4\pi^2} \int |z|^{2k} \left| \int \frac{e^{iuz} k^*(uh)\phi(u)}{\mu_\varepsilon^*(-u)} du \right|^2 dz \, .$$

Now, by Plancherel's identity,

$$\int |z|^{2k} \left| \int \frac{e^{iuz} k^*(uh)\phi(u)}{\mu_\varepsilon^*(-u)} du \right|^2 dz = 2\pi \int \left| \left(\frac{k^*(th)\phi(t)}{\mu_\varepsilon^*(-t)} \right)^{(k)} \right|^2 dt.$$

It can be checked that, for $h \leq 1$,

$$\left| \left(\frac{k^*(th)\phi(t)}{\mu_\varepsilon^*(-t)} \right)^{(k)} \right| \leq C \sum_{\ell=0}^{k} |r_\varepsilon^{(\ell)}(t)| \mathbf{1}_{[-2,2]}(t),$$

which concludes the proof of the Lemma. □

Applying Lemma 3.1 with $k = 1$, we obtain that

$$I_1^- \leq \frac{C}{\sqrt{n}} \Big(\sup_{t \in [-2,2]} \sum_{\ell=0}^{1} |r_\varepsilon^{(\ell)}(t)| \Big) \int_{-\infty}^{0} |x|^{p-1} \sqrt{\mathbb{P}\left(Y \leq \frac{x}{2} \right)} dx. \tag{19}$$

We now control the term I_2^-. Let $b \in]0,1[$. Applying Cauchy-Schwarz Inequality

$$I_2^- \leq \frac{C}{\sqrt{n}} \sqrt{\int_{-\infty}^{0} |x|^{2p-2}(1+|x|^{1+b}) \int_{-\infty}^{\frac{x}{2}} (1+|z|^{1+a}) \left(\frac{\tilde{k}_{1,h}(z/h)}{h} \right)^2 dz \, dx}.$$

Consequently, by Fubini's Theorem

$$\begin{aligned} I_2^- &\leq \frac{C}{\sqrt{n}} \sqrt{\int_{-\infty}^{0} (1+|z|^{1+a}) \left(\frac{\tilde{k}_{1,h}(z/h)}{h} \right)^2 \int_{2z}^{0} |x|^{2p-2}(1+|x|^{1+b}) dx \, dz} \\ &\leq \frac{C}{\sqrt{n}} \sqrt{\int (1+|z|^{2p+1+a+b}) \left(\frac{\tilde{k}_{1,h}(z/h)}{h} \right)^2 dz} \end{aligned}$$

Let m_0 be the least integer strictly greater than $p + 1/2$. Taking a and b close enough to 0, it follows that

$$I_2^- \leq \frac{C}{\sqrt{n}} \sqrt{\int (1+|z|^{2m_0}) \left(\frac{\tilde{k}_{1,h}(z/h)}{h} \right)^2 dz}$$

Applying Lemma 3.1 with $k = m_0$, it follows that

$$I_2^- \leq \frac{C}{\sqrt{n}} \Big(\sup_{t \in [-2,2]} \sum_{\ell=0}^{m_0} |r_\varepsilon^{(\ell)}(t)| \Big) . \tag{20}$$

In the same way, we have

$$\begin{aligned} I^+ &= \int_0^\infty |x|^{p-1}\sqrt{\mathrm{Var}(1-\hat{F}_{1,n})(x)}dx \\ &\leq \frac{C}{\sqrt{n}}\int_0^\infty |x|^{p-1}\sqrt{\mathbb{E}\left[1-G_{1,h}\left(\frac{x-Y}{h}\right)\right]^2}dx \\ &\leq \frac{C}{\sqrt{n}}\int_0^\infty |x|^{p-1}\sqrt{\mathbb{E}\left[\int \tilde{k}_{1,h}(u)\mathbf{1}_{\left\{u\geq\frac{x-Y}{h}\right\}}du\right]^2}dx. \end{aligned}$$

Using the same arguments as for I^-, we obtain,

$$I^+ \leq \frac{C}{\sqrt{n}}\Big(\sup_{t\in[-2,2]}\sum_{\ell=0}^{1}|r_\varepsilon^{(\ell)}(t)|\Big)\int_0^\infty |x|^{p-1}\sqrt{P\left(Y\geq\frac{x}{2}\right)}dx+\frac{C}{\sqrt{n}}\Big(\sup_{t\in[-2,2]}\sum_{\ell=0}^{m_0}|r_\varepsilon^{(\ell)}(t)|\Big). \tag{21}$$

Consequently, gathering (19), (20) and (21) we obtain that

$$I \leq \frac{C}{\sqrt{n}}\Big(\sup_{t\in[-2,2]}\sum_{\ell=0}^{1}|r_\varepsilon^{(\ell)}(t)|\Big)\int_0^\infty |x|^{p-1}\sqrt{P\left(|Y|\geq x\right)}dx + \frac{C}{\sqrt{n}}\Big(\sup_{t\in[-2,2]}\sum_{\ell=0}^{m_0}|r_\varepsilon^{(\ell)}(t)|\Big). \tag{22}$$

Control of J. Let $a\in]0,1/2[$. By definition of the term J, and applying Cauchy-Schwarz Inequality,

$$\begin{aligned} J &\leq \frac{C}{\sqrt{n}}\int |x|^{p-1}\sqrt{\mathbb{E}\left[G_{2,h}\left(\frac{x-Y}{h}\right)\right]^2}dx \\ &\leq \frac{C}{\sqrt{n}}\sqrt{\int |x|^{2p-2}(1+|x|^{1+a})\mathbb{E}\left[G_{2,h}\left(\frac{x-Y}{h}\right)\right]^2 dx}. \end{aligned}$$

Since $\int \tilde{k}_{2,h}(u)du = 0$ and $G_{2,h}$ is continuous, using the same arguments as Gurland (1948), we get that

$$G_{2,h}(x) = \Re\mathrm{e}\left[-\frac{1}{2\pi i}\int \frac{e^{-itx}k^*(t)(1-\phi(t/h))}{t\mu_\varepsilon^*(t/h)}dt\right].$$

Consequently,

$$J \leq \frac{C}{\sqrt{n}}\sqrt{\mathbb{E}\left(\int(1+|x|^{2p-1+a})\left[\Re\mathrm{e}\left(-\frac{1}{2\pi i}\int\frac{e^{-it\frac{x-Y}{h}}k^*(t)(1-\phi(t/h))}{t\mu_\varepsilon^*(t/h)}dt\right)\right]^2 dx\right)}.$$

Setting $u=t/h$ and using the fact that $|x|^q \leq 2^{q-1}|x-Y|^q + 2^{q-1}|Y|^q$ for any $q\geq 1$, we obtain that

$$\begin{aligned} J \leq \frac{C}{\sqrt{n}}\Bigg[&\mathbb{E}\left(\int|x-Y|^{2p-1+a}\left[\Re\mathrm{e}\left(-\frac{1}{2\pi i}\int\frac{e^{-iu(x-Y)}k^*(uh)(1-\phi(u))}{u\mu_\varepsilon^*(u)}du\right)\right]^2 dx\right) \\ &+\mathbb{E}\left((1+|Y|^{2p-\frac{1}{2}})\int\left[\Re\mathrm{e}\left(-\frac{1}{2\pi i}\int\frac{e^{-iu(x-Y)}k^*(uh)(1-\phi(u))}{u\mu_\varepsilon^*(u)}du\right)\right]^2 dx\right)\Bigg]^{1/2}. \end{aligned}$$

Thus,

$$J \le \frac{C}{\sqrt{n}} \left[\int (1+|x|^{2p-1+a}) \left[\Re\mathrm{e} \left(-\frac{1}{2\pi i} \int \frac{e^{-iux} k^*(uh)(1-\phi(u))}{u\mu_\varepsilon^*(u)} du \right) \right]^2 dx \right.$$
$$\left. + \mathbb{E}|Y|^{2p-\frac{1}{2}} \int \left[\Re\mathrm{e} \left(-\frac{1}{2\pi i} \int \frac{e^{-iux} k^*(uh)(1-\phi(u))}{u\mu_\varepsilon^*(u)} du \right) \right]^2 dx \right]^{1/2}.$$

Let m_1 be the least integer strictly greater than $p - \frac{1}{2}$. Taking a close enough to zero, it follows that

$$J \le \frac{C}{\sqrt{n}} \left[\int (1+|x|^{2m_1}) \left| \frac{1}{2\pi} \int \frac{e^{-iux} k^*(uh)(1-\phi(u))}{u\mu_\varepsilon^*(u)} du \right|^2 dx \right.$$
$$\left. + \mathbb{E}|Y|^{2p-\frac{1}{2}} \int \left| \frac{1}{2\pi} \int \frac{e^{-iux} k^*(uh)(1-\phi(u))}{u\mu_\varepsilon^*(u)} du \right|^2 dx \right]^{1/2}.$$

By Plancherel's identity,

$$\int \left| \int \frac{e^{-iux} k^*(uh)(1-\phi(u))}{u\mu_\varepsilon^*(u)} du \right|^2 dx = 2\pi \int \left| \frac{k^*(th)(1-\phi(t))}{t\mu_\varepsilon^*(-t)} \right|^2 dt,$$

and

$$\int |x|^{2m_1} \left| \int \frac{e^{-iux} k^*(uh)(1-\phi(u))}{u\mu_\varepsilon^*(u)} du \right|^2 dx = 2\pi \int \left| \left(\frac{k^*(th)(1-\phi(t))}{t\mu_\varepsilon^*(-t)} \right)^{(m_1)} \right|^2 dt.$$

Now, for $h \le 1$,

$$\left| \left(\frac{k^*(th)(1-\phi(t))}{t\mu_\varepsilon^*(-t)} \right)^{(m_1)} \right| \le C \sum_{j=0}^{m_1} \sum_{\ell=0}^{j} \frac{|r_\varepsilon^{(\ell)}(-t)|}{|t|^{j-\ell+1}} \mathbf{1}_{[-1,1]^c}(t)$$
$$\le C \sum_{\ell=0}^{m_1} \frac{|r_\varepsilon^{(\ell)}(-t)|}{|t|} \mathbf{1}_{[-1,1]^c}(t).$$

Finally,

$$J \le \frac{C}{\sqrt{n}} \left[\mathbb{E}|Y|^{2p-\frac{1}{2}} \int_{-1/h}^{1/h} \frac{|r_\varepsilon(x)|^2}{|x|^2} \mathbf{1}_{[-1,1]^c}(x) dx + \sum_{\ell=0}^{m_1} \int_{-1/h}^{1/h} \frac{|r_\varepsilon^{(\ell)}(x)|^2}{|x|^2} \mathbf{1}_{[-1,1]^c}(x) dx \right]^{1/2}. \tag{23}$$

Starting from (10) and gathering the upper bounds (11), (18), (22) and (23), the proof of Proposition 3.1 is complete.

4 Lower bound

For some $M > 0$ and $q \ge 1$, we denote by $\mathcal{D}(M, q)$ the set of measures μ on $\mathbb{R}$ such that $\int |x|^q d\mu(x) \le M$.

Theorem 4.1. *Let $M > 0$ and $q \geq 1$. Assume that there exist $\beta > 0$ and $c > 0$, such that for every $\ell \in \{0, 1, 2\}$ and every $t \in \mathbb{R}$,*

$$|\mu_\varepsilon^{*(\ell)}(t)| \leq c(1 + |t|)^{-\beta}. \tag{24}$$

Then, there exists a constant $C > 0$ such that, for any estimator $\hat{\mu}$,

$$\liminf_{n\to\infty} n^{\frac{p}{2\beta+1}} \sup_{\mu\in\mathcal{D}(M,q)} \mathbb{E}W_p^p(\hat{\mu}, \mu) > C.$$

Remark 4.1. *For W_1, this lower bound matches the upper bound given in Theorem 3.1 for $\beta \geq 1/2$. For W_2, we conjecture that the upper bounds given by Theorem 3.1 are appropriate under the assumed moment conditions. Getting better rates of convergence for W_2 (or more generally for W_p with $p > 1$) under stronger moment conditions is an open question.*

Proof. Let $M > 0$ and $q \geq 1$. The proof is similar to the proof of Theorem 3 in Dedecker and Michel (2013) and thus we only give here a sketch of the proof. We first define a finite family in $\mathcal{D}(M, q)$ using the densities

$$f_{0,r}(t) := C_r(1 + t^2)^{-r} \tag{25}$$

with some $r > (1 + q)/2$. Next, let b_n be the sequence

$$b_n := \left[n^{\frac{1}{2\beta+1}}\right] \vee 1\,, \tag{26}$$

where $[\cdot]$ is the integer part. For any $\theta \in \{0, 1\}^{b_n}$, let

$$f_\theta(t) = f_{0,r}(t) + C\sum_{s=1}^{b_n} \theta_s H\left(b_n(t - t_{s,n})\right), \quad t \in \mathbb{R}, \tag{27}$$

where C is a positive constant and $t_{s,n} = (s-1)/b_n$. The function H is a bounded function whose integral on the line is 0. Moreover, we may choose a function H such that (see for instance Fan (1991a) or Fan (1993)):

(A1) $\int_{-\infty}^{+\infty} H(t)\,dt = 0$ and $\int_0^1 |H^{(-1)}(t)|\,dt > 0$,

(A2) $|H(t)| \leq c(1 + t^2)^{-r_0}$ where $r_0 > \max(3/2, (1 + q)/2)$,

(A3) $H^*(z) = 0$ outside $[1, 2]$,

where $H^{(-1)}(t) := \int_{-\infty}^t H(u)\,du$ is a primitive of H. Note that by replacing H by H/C in the following, we finally can take $C = 1$ in (27). Let μ_θ be the measure of density f_θ with respect to the Lebesgue measure. Then we can find some M large enough such that for all $\theta \in \{0, 1\}^{b_n}$, $\mu_\theta \in \mathcal{D}(M, q)$. Moreover, under these assumptions the first two derivatives of H^* are continuous and bounded.

For $\theta \in \{0, 1\}^{b_n}$ and $s \in \{1, \ldots, b_n\}$, let us define the probability measures $\mu_{\theta,s,0}$ and $\mu_{\theta,s,1}$ with densities

$$f_{\theta,s,0} := f_{(\theta_1,\ldots,\theta_{s-1},0,\theta_{s+1},\ldots,\theta_{b_n})} \quad \text{and} \quad f_{\theta,s,1} := f_{(\theta_1,\ldots,\theta_{s-1},1,\theta_{s+1},\ldots,\theta_{b_n})}.$$

We also consider the densities $h_{\theta,s,u} = f_{\theta,s,u} \star g$ for $u = 0$ or 1. Since W_1 is dominated by W_p, and using Jensen's inequality, it follows that

$$\sup_{\mu\in\mathcal{D}(M,q)} \mathbb{E}_{(\mu\star\mu_\varepsilon)^{\otimes n}} W_p^p(\mu,\tilde{\mu}_n) \geq \sup_{\mu\in\mathcal{D}(M,q)} \mathbb{E}_{(\mu\star\mu_\varepsilon)^{\otimes n}} W_1^p(\mu,\tilde{\mu}_n)$$
$$\geq \left(\sup_{\mu\in\mathcal{D}(M,q)} \mathbb{E}_{(\mu\star\mu_\varepsilon)^{\otimes n}} W_1(\mu,\tilde{\mu}_n)\right)^p . \quad (28)$$

Using a standard randomization argument (see for the instance the proof of Theorem 3 in Dedecker and Michel (2013) for the multivariate case), it can be shown that there exists a constant $C > 0$ such that

$$\sup_{\mu\in\mathcal{D}(M,q)} \mathbb{E}_{(\mu\star\mu_\varepsilon)^{\otimes n}} W_1(\mu,\tilde{\mu}_n) \geq \frac{C}{b_n}\int_0^1 \left|H^{(-1)}(u)\right| du \quad (29)$$

as soon as there exists a constant $c > 0$ such that, for any $\theta \in \{0,1\}^{b_n}$,

$$\chi^2(h_{\theta,s,0}\,,\, h_{\theta,s,1}) \leq \frac{c}{n}\,, \quad (30)$$

where the χ^2 distance between two densities h_1 and h_2 on $\mathbb{R}$ is defined by

$$\chi^2(h_1,h_2) = \int \frac{\{(h_1(x)-h_2(x)\}^2}{h_1(x)}dx.$$

If (30) is satisfied, we take b_n as in (26) and the theorem is thus proved according to (28), (29) and (A1).

It remains to prove (30). Using (A2), we can find a constant $C > 0$ such that for any $t \in \mathbb{R}$ and any $s \in \{1,\ldots,b_n\}$,

$$\chi^2(h_{\theta,s,0}\,,\, h_{\theta,s,1}) \leq Cb_n^{-1}\int \frac{\left\{\int H(v-y)g(y/b_n)\,dy/b_n\right\}^2}{f_{0,r}\star g(v/b_n)}dv. \quad (31)$$

The right side of (31) is typically the kind of χ^2 divergence that is upper bounded in the proofs of Theorems 4 and 5 in Fan (1991a) for computing pointwise rates of convergence: under Assumption (24), it gives that there exists a constant C such that

$$\int \frac{\left\{\int H(v-y)g(y/b_n)\,dy/b_n\right\}^2}{f_{0,r}\star g(v/b_n)}dv \leq Cb_n^{-2\beta}$$

and (30) is proved. □

5 Numerical experiments

This section is devoted to the implementation of the deconvolution estimators. We continue the experiments of Caillerie et al. (2011) about Wasserstein deconvolution in the ordinary smooth case. In particular, we study the W_1 and W_2 univariate deconvolution problems and we compare our numerical results with the upper and lower bounds given in the previous sections. We also apply our procedure to the deconvolution of the uniform measure on the Cantor set. The deconvolution method is implemented in R.

5.1 Implementation of the deconvolution estimators

For all the experiments we use the kernel

$$k(x) = \frac{3}{16\pi}\left(\frac{8\sin(x/8)}{x}\right)^4$$

which corresponds to the kernel given by (7) with $p = 2$ and a support over $[-1/2, 1/2]$. Computing the deconvolution estimators requires to evaluate many times the function

$$\tilde{k}_h : x \mapsto \Re\mathrm{e}\left[\frac{1}{2\pi}\int \frac{e^{iux}k^*(u)}{\mu_\varepsilon^*(-u/h)}du\right].$$

In this section we consider symmetric distributions for μ_ε. Since k^* and μ_ε^* are even functions, $\tilde{k}_h$ is the real part of the Fourier transform of

$$\psi_h : u \mapsto \frac{1}{2\pi}\frac{k^*(u)}{\mu_\varepsilon^*(-u/h)}.$$

The Fourier decomposition of ψ_h is given by $\psi_h(u) = \sum_{k\in\mathbb{Z}} a_{k,h}e^{2i\pi ku}$ where $a_{k,h} = \int_{-1/2}^{1/2}\psi_h(u)e^{-2i\pi ku}du$. Thus,

$$\begin{aligned}
\tilde{k}_h(x) &= \int_{-1/2}^{1/2}\psi_h(u)e^{-2i\pi ku}du \\
&= \sum_{k\in\mathbb{Z}} a_{k,h}\int_{-1/2}^{1/2} e^{i(2\pi k - x)u}du \\
&= \sum_{k\in\mathbb{Z}} a_{k,h}\,\mathrm{sinc}\left(\frac{2\pi k - x}{2}\right).
\end{aligned}$$

For large N, the coefficient $a_{k,h}$ can be approximated by the k-th coefficient of a discrete Fourier transform taken at $(\psi_h(0), \psi_h(1/N), \ldots, \psi_h(1-1/N))$, denoted $\hat{a}_{k,h,N}$ in the sequel. Of course we use the Fast Fourier Transform algorithm to compute these quantities. For some large K, we evaluate $\tilde{k}_h$ at some point x by

$$\hat{\tilde{k}}_h(x) \approx \sum_{|k|\le K} \hat{a}_{k,h,N}\,\mathrm{sinc}\left(\frac{2\pi k - x}{2}\right). \tag{32}$$

For intensive simulation, it may be relevant to compute preliminary $\hat{\tilde{k}}_h$ on a grid of high resolution rather than calling this function each time.

We first define a discrete approximation of the function

$$\hat{\mu}_{n,h} : u \mapsto \frac{1}{nh}\sum_{k=1}^{n}\tilde{k}_h\left(\frac{u - Y_k}{h}\right).$$

Let $t_1 < \cdots < t_q$ be a finite regular grid of points in $\mathbb{R}$ with resolution η. A discrete approximation $\hat{\mu}_{n,h}^d$ of $\hat{\mu}_{n,h}$ is defined on $\mathcal{P}$ by

$$\hat{\mu}_{n,h}^d = \eta\sum_{j=1}^{q}\hat{\mu}_{n,h}(t_j)\delta_{t_j},$$

where δ_x is the Dirac distribution at x. Since $\hat{\mu}_{n,h}(t_j)$ can be negative, the first method for estimating F consists in taking the positive part of $\hat{\mu}^d_{n,h}$:

$$\hat{\mu}^{\text{naive}}_{n,h} := \frac{\sum_{j=1}^q \left(\hat{\mu}^d_{n,h}(t_j)\right)^+ \delta_{t_j}}{\sum_{j=1}^q \left(\hat{\mu}^d_{n,h}(t_j)\right)^+}.$$

This first estimator is called the "naive" deconvolution estimator henceforth. Note that it was studied in Caillerie et al. (2011) and Dedecker and Michel (2013). For implementing the alternative estimator $\tilde{\mu}_{n,h}$ proposed in this paper, we first need to find some probability distribution $\tilde{F}_{n,h}$ on $\mathbb{R}$ such that

$$\int_{\mathbb{R}} |x|^{p-1} |\hat{F}_{n,h} - \tilde{F}_{n,h}|(x)dx \\ \approx \inf\left\{ \int_{\mathbb{R}} |x|^{p-1} |\hat{F}_{n,h} - G|(x)dx\,|,\ G \text{ probability distribution on } \mathbb{R} \right\}. \quad (33)$$

In practice, this corresponds to finding a distribution function close to the step function

$$\hat{F}^d_{n,h} : t \mapsto \sum_{j=1}^q \hat{\mu}^d_{n,h}(t_j) \mathbf{1}_{\{t_j \le t\}}.$$

Since $\hat{F}^d_{n,h}$ may take its values outside $[0,1]$, we can also look for a distribution function close to $t \mapsto \hat{F}^d_{n,h}(t)\mathbf{1}_{\hat{F}^d_{n,h}(t)\in[0,1]}$. In other terms, we compute the isotone regression of $t \mapsto \hat{F}^d_{n,h}(t)\mathbf{1}_{\hat{F}^d_{n,h}(t)\in[0,1]}$ with weights t_j^{p-1}:

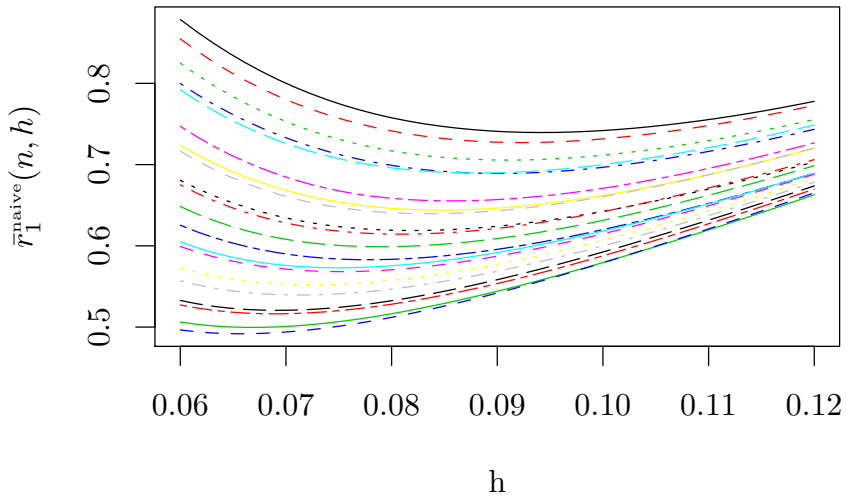
$$\hat{F}^{\text{isot},p}_{n,h} := \text{argmin} \left\{ \sum_{j=1}^q |t_j|^{p-1} \left| G(t_j) - \hat{F}^d_{n,h}(t_j)\mathbf{1}_{\hat{F}^d_{n,h}(t_j)\in[0,1]} \right|^p ,\ G \text{ non-decreasing} \right\}.$$

We compute $\hat{F}^{\text{isot},p}_{n,h}$ thanks to the function `gpava` from the R package `isotonic` (Mair et al., 2009). The measure μ is finally estimated by the absolutely continuous measure $\hat{\mu}^{\text{isot},p}_{n,h}$ whose distribution function is $\hat{F}^{\text{isot},p}_{n,h}$. We call this estimator the isotone deconvolution estimator for the metric W_p.

The construction of $\hat{\mu}^{\text{isot},p}_{n,h}$ depends on many parameters. Tuning all these paramaters is a tricky issue. For this paper we only tune K, N and η by hand. Note that one crucial point is the length N of the vector we use for computing the the $a_{k,h,N}$'s with the FFT. For ordinary smooth distributions, we observe that $\tilde{k}_h$ decreases slowly for small β for the range of bandwidths h giving minimum Wasserstein risks. Consequently, a small β requires many terms in the expansion (32), and hence a large N. For β smaller than 0.5, it was necessary to take $N \approx 10^4$.

5.2 Computation of Wasserstein risks for simulated experiments

For fixed distributions μ and μ_ε, we simulate $Y_1, \ldots, Y_n$ according to the convolution model (1). For a given bandwidth h and $p \ge 1$, we can compute $W_p^p(\hat{\mu}^{\text{naive}}_n, \mu)$ and $W_p^p(\hat{\mu}^{\text{isot},p}_{n,h}, \mu)$ using the quantile functions of the measures, thanks to the relation (2). The Wasserstein risks $\mathcal{R}^{\text{naive}}(n,h) := \mathbb{E}W_p^p(\hat{\mu}^{\text{naive}}_{n,h}, \mu)$ and $\mathcal{R}^{\text{isot}}(n,h) := \mathbb{E}W_p^p(\hat{\mu}^{\text{isot},p}_{n,h}, \mu)$ can be estimated by an elementary Monte Carlo method by repeating the simulation of the Y_i's and averaging



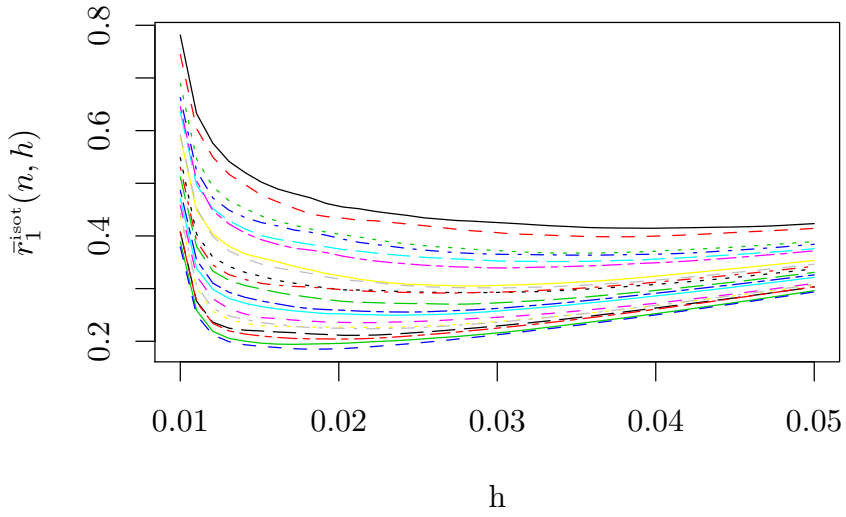


Figure 1: Estimated Wasserstein risks for the Dirac experiment. The noise distribution is the symmetrized Gamma distribution with $\beta = 2$. The twenty curves correspond to samples of length n taken between 100 and 2000.

Distribution	μ_ε^*	β
Symmetrized Gamma	$t \mapsto (1+t^2)^{-\beta/2}$	0.3, 0.5, 1.2, 2 ,3, 4
Laplace	$t \mapsto (1+t^2)^{-1}$	2
Symmetrized χ^2	$t \mapsto (1+4t^2)^{(-1/2)}$	1

Table 1: Ordinary smooth distributions used for the error.

the Wasserstein distances. Let $\bar{r}_p^{\text{isot}}(n,h)$ and $\bar{r}_p^{\text{naive}}(n,h)$ be the estimated risks obtained this way (see Figure 1 for an illustration of such curves for the Dirac experiment). For each n, an approximation of the minimal risks over the bandwidths is proposed by

$$\bar{r}_{p,*}^{\text{isot}}(n) := \min_{h\in\mathcal{H}} \bar{r}_p^{\text{isot}}(n,h)$$

and

$$\bar{r}_{p,*}^{\text{naive}}(n) := \min_{h\in\mathcal{H}} \bar{r}_p^{\text{naive}}(n,h)$$

where $\mathcal{H}$ is a grid of bandwidth values.

5.3 Estimation of the rates of convergence

In this experiment we study the rates of convergence of the estimators for the deconvolution of three distributions:

- Dirac distribution at 0,
- Uniform distribution on $[-0.5, 0.5]$,
- Mixture of the Dirac distribution at 0 and the uniform distribution on $[-0.5, 0]$.

We take for μ_ε the ordinary smooth distributions summarized in Table 1. Recall that the coefficient β of a symmetrized Gamma distribution is twice the shape parameter of the distribution. For each error distribution and for n chosen between 100 and 2000, we simulate 200 times a sample of length n from which we compute the estimated minimal risks $\bar{r}_{p,*}^{\text{isot}}(n)$ and $\bar{r}_{p,*}^{\text{naive}}(n)$. We study the Wasserstein risks W_1 and W_2. We obtain some estimation of the exponent of the rate of convergence for each deconvolution problem by computing the linear regression of $\log \bar{r}_{p,*}(n)$ by $\log n$. See Figure 2 for an illustration and Figures 7 and 8 at the end of the paper for the complete outputs of the Dirac case. A linear trend can be observed in all cases. As expected, the risks are smaller for the isotone estimators than for the naive ones.

The estimated exponents of the convergences rates are plotted in Figure 3 as functions of β. These estimated rates can be compared with the upper and lower bounds obtained in the paper. Of course the rates of convergence of the isotone estimator have no reason to match exactly the lower bounds. However it can be checked that the estimated rates we obtain are consistent with the theoretic bounds proved before. In particular we see that the parametric rate is reached for values of β close to 0, at least in the Dirac case. These results also suggest that the correct minimax rate for W_2 probably corresponds to the upper bound given in Theorem 3.1 (that is, when no further assumption is made on the unknown distribution μ).

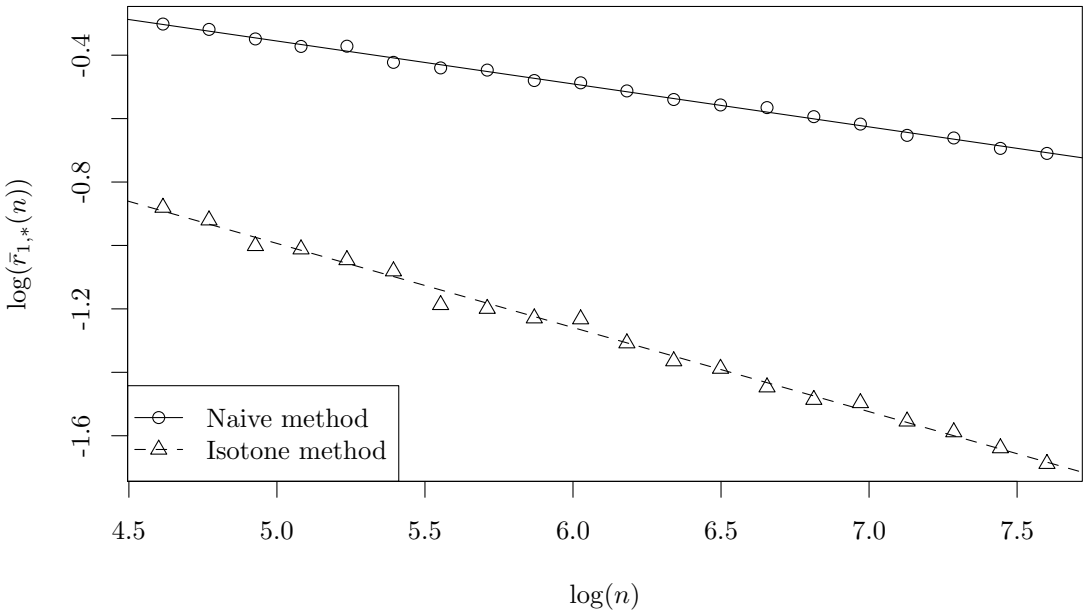


Figure 2: Estimated rates of convergence to zero of the W_1-risk for the naive method and the isotone method for μ being a Dirac distribution at 0. The noise distribution is the symmetrized Gamma distribution with $\beta = 2$.

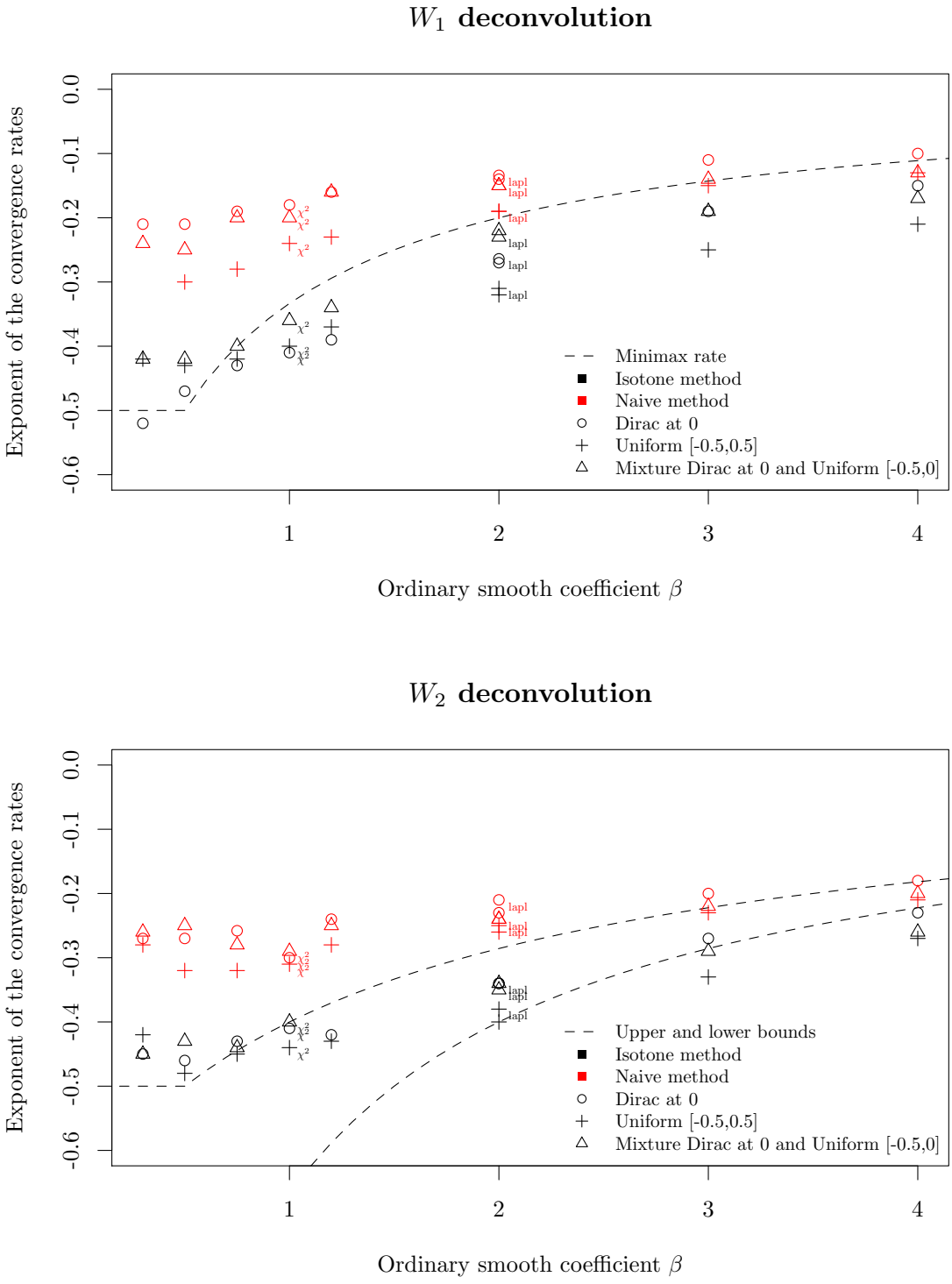


Figure 3: Estimated exponents of the convergence rates of the naive and the isotone deconvolution estimator for three distributions μ. The exponents are given as functions of the ordinary smooth coefficient β. Regarding the noise distribution, the χ^2 and the Laplace distributions are indicated directly on the graph, the others experiments have been done with the symmetrized Gamma distribution. The top graph corresponds to the W_1 deconvolution and the bottom one to the W_2 deconvolution.

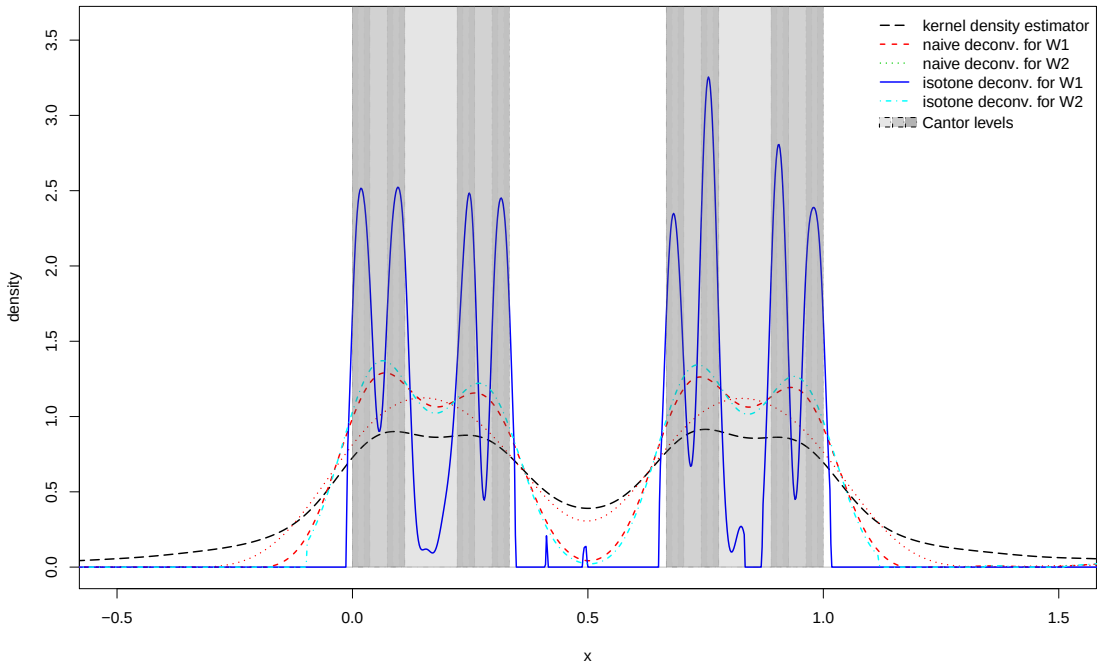


Figure 4: Deconvolution of the uniform measure on the Cantor set.

5.4 Cantor set experiment

We now illustrate the deconvolution method with a more original experiment. We take for μ the uniform distribution on the Cantor set $\mathfrak{C}$. Remember that the Cantor set can be defined by repeatedly deleting the open middle thirds of a set of line segments:

$$\mathfrak{C} = \bigcap_{m \geq 1} F_n$$

where $F_0 = [0,1]$ and F_{m+1} is obtained by cutting out the middle thirds of all the intervals of F_m: $F_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ and $F_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$, etc... The uniform measure $\mu_{\mathfrak{C}}$ on $\mathfrak{C}$ can be defined as the distribution of the random variable $X := 2\sum_{k\geq 1} 3^{-k} B_k$ where $(B_k)_{k\geq 1}$ is a sequence of independent random variables with Bernoulli distribution of parameter $1/2$. Note that the Lebesgue measure of $\mathfrak{C}$ is zero and thus the Lebesgue measure and $\mu_{\mathfrak{C}}$ are singular. The deconvolution estimators being densities for the Lebesgue measure, the Wasserstein distances are relevant metrics for comparing these with $\mu_{\mathfrak{C}}$.

Let $\mu_{\mathfrak{C},K}$ be the distribution of the random variable defined by the partial sum $\tilde{X} := 2\sum_{k=1}^{K} 3^{-k} B_k$ where the B_k's are defined as before. The distribution $\mu_{\mathfrak{C},K}$ is an approximation of $\mu_{\mathfrak{C}}$ which can be computed in practice. We simulate a sample of $n = 10^4$ observations from $\mu_{\mathfrak{C},K}$ with $K = 100$. These observations are contaminated by random variables with symmetrized Gamma distribution (the shape parameter is equal to $1/4$ (so that $\beta = 0.5$) and the scale parameter is equal to $1/2$).

In Figure 4, the isotone estimators for W_1 and W_2 and the naive estimator are plotted on the first four levels F_m of the Cantor set. The bandwidth are chosen by minimizing the Wasserstein risks as explained before. This requires to approximate the quantile functions

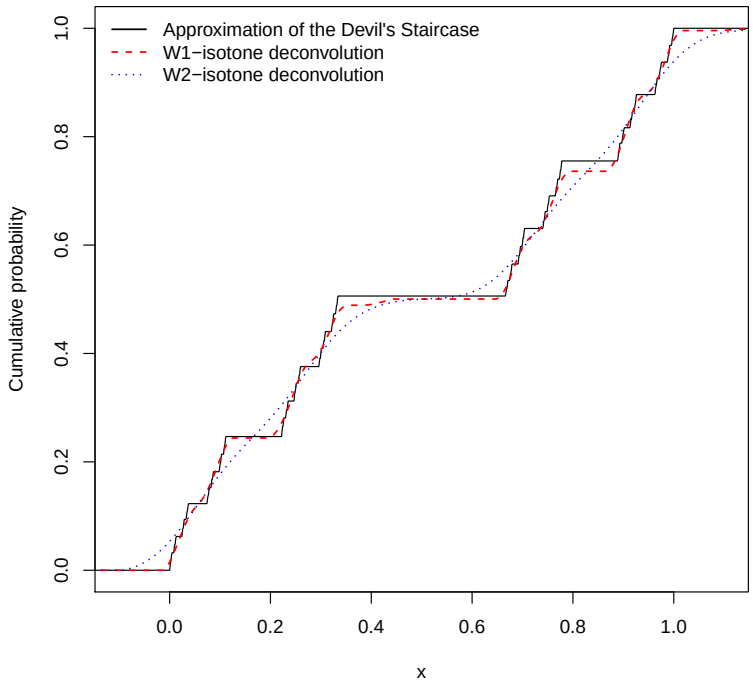


Figure 5: Approximation of the Devil's staircase and distributions functions of the W_1 and W_2 isotone deconvolution estimators.

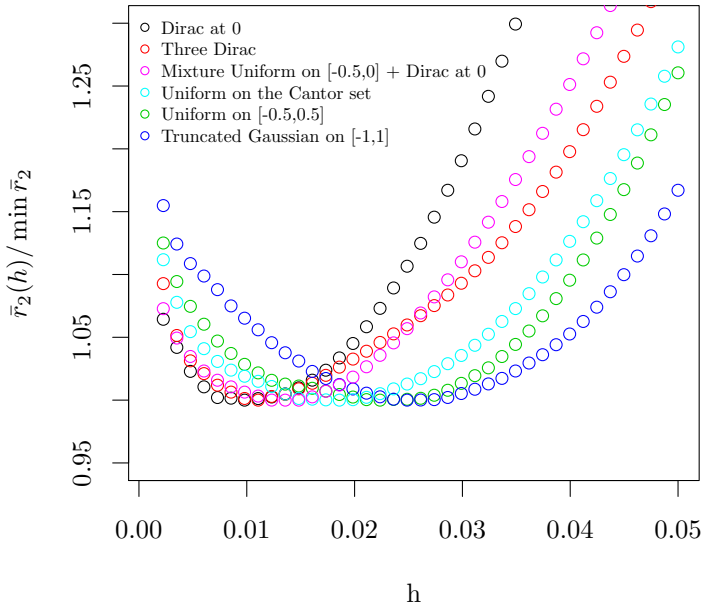


Figure 6: Comparison of the locations of the minimums of the W_2-risks for five distribution measures μ. The noise distribution μ_ε is the symmetrized Gamma distribution with $\beta = 0.75$. Each risk curve has been normalized by its minimum value for facilitating the comparison.

for the isotone deconvolution estimator and for the $\mu_{\mathfrak{C}}$. Regarding the quantile function of $\mu_{\mathfrak{C}}$, we simulate a large sample according to $\mu_{\mathfrak{C},100}$ and we compute the corresponding empirical distribution function. This last cdf is an approximation of the so called "Devil's staircase" (see Figure 5). For the naive deconvolution estimator we find $h = 0.011$ for W_1 and $h = 0.018$ for W_2. For the W_1-isotone deconvolution estimator we find $h = 0.002$ and $h = 0.01$ for the W_2-isotone estimator. Note that these values are consistent with the fact that the bandwidth increases with the parameter p of the Wasserstein metric, as shown by Theorem 3.1. On Figure 4, the W_1-isotone deconvolution estimator is able to "see" the three first levels of the Cantor set and the three others deconvolution methods recover the two first levels. A kernel density estimator (with no deconvolution) only recovers the first level.

5.5 About the bandwidth choice

In practice, we need to choose a bandwidth h for the deconvolution estimators. As was explained in Caillerie et al. (2011) (see Remark 3 in this paper), it seems that the influence of the measure μ is weak. We now propose a simple experiment to check this principle. We choose for μ_ε the symmetrized gamma distribution with a shape parameter equal to 0.375 ($\beta = 0.75$) and we simulate contaminated observations from the following various distributions:

- Truncated standard Gaussian distribution on $[-1, 1]$,

- Uniform distribution on $[-0.5, 0.5]$,
- Uniform distribution on the Cantor set,
- Mixture of the Dirac distribution at 0 and the uniform distribution on $[-0.5, 0]$,
- Mixture of Dirac distributions at -0.5, -0.2 and 0.3 with proportions 1/4, 1/4 and 1/2,
- Dirac distribution at 0.

We focus here on the study of the W_2-isotone deconvolution estimator. Figure 6 compares the locations of the minimums of the five risk curves $h \mapsto \bar{r}^{\text{isot}}_{2,h}$ by averaging over 200 samples of 1000 contaminated observations. For this experiment, the sensitivity of the minimum risk location to the distribution μ is not very large.

On another hand, from Figure 3, it seems that the rates for the mixture Dirac Uniform are quite slow (in particular, they are close to the minimax rates for W_1).

From these remarks, it seems that the bandwidth minimizing the risk computed for the mixture Dirac Uniform should be a reasonable choice for deconvolving other distributions. Of course, this is in some sense a "minimax choice", and it will not give the appropriate rate for measures which are easier to estimate (for instance measures with smooth densities).

A bootstrap method in the spirit of Delaigle and Gijbels (2004) may give a more satisfactory answer to this problem. However, note that the use of the Wasserstein metric makes difficult the asymptotical analysis of the risk. This interesting problem is out of the scope of this paper, we intend to investigate it in a future work.

Acknowledgements

The authors were supported by the ANR project TopData ANR-13-BS01-0008.

References

C. Butucea and B Tsybakov. Sharp optimality in density deconvolution with dominating bias. I. *Theory Probab. Appl.*, 52:24–39, 2008a.

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

C. Caillerie, F. Chazal, J. Dedecker, and B. Michel. Deconvolution for the Wasserstein metric and geometric inference. *Electron. J. Stat.*, 5:1394–1423, 2011.

R.J. Carroll and P. Hall. Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83:1184–1186, 1988.

F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Geometric inference for probability measures. *Found. Comput. Math.*, 11:733–751, 2011.

I. Dattner, A. Goldenshluger, and A. Juditsky. On deconvolution of distribution functions. *Ann. Statist.*, 39:2477–2501, 2011.

J. Dedecker and B. Michel. Minimax rates of convergence for Wasserstein deconvolution with supersmooth errors in any dimension. *J. Multivar. Anal.*, 122:278–291, 2013.

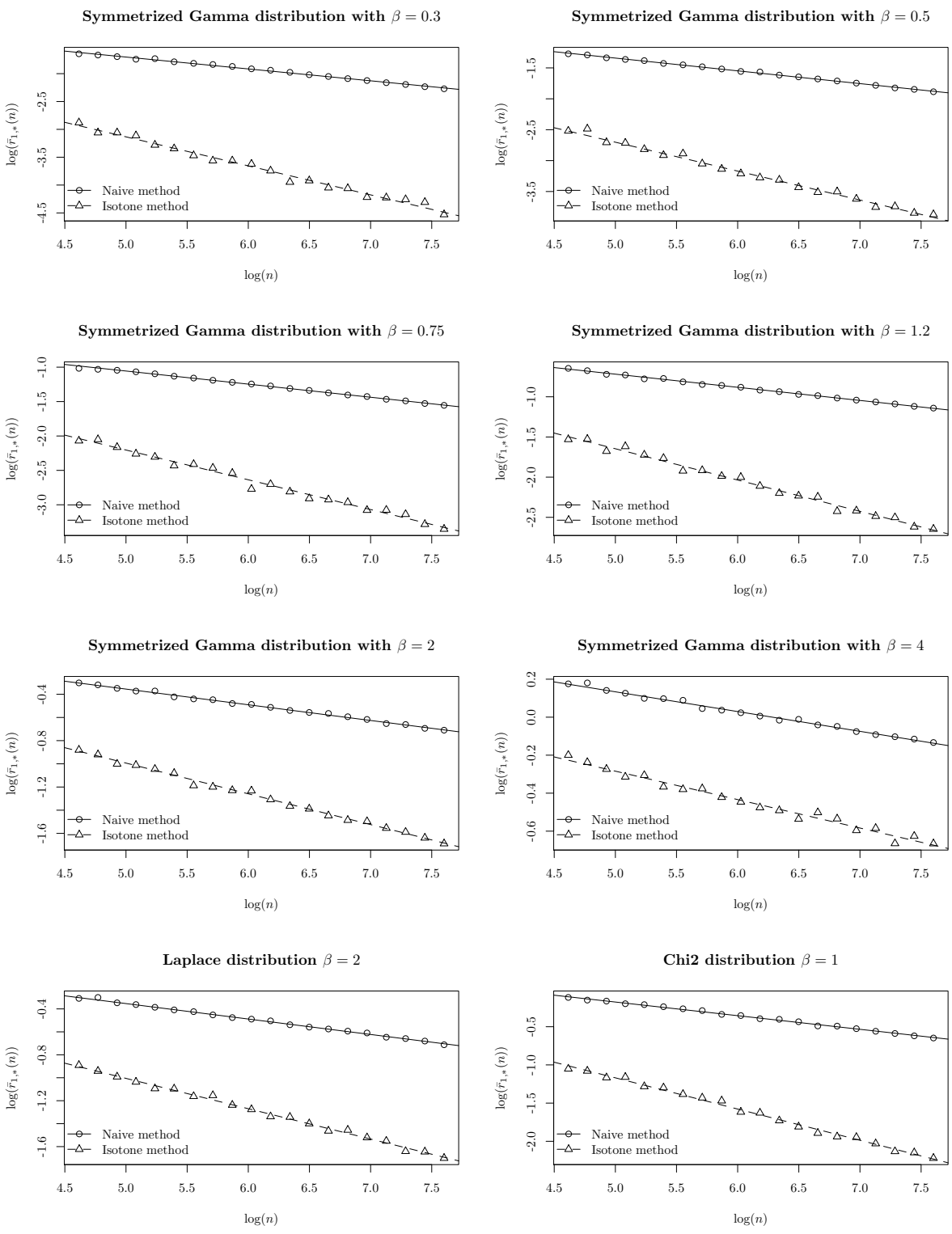


Figure 7: Deconvolution of the Dirac distribution at zero observed with one of the noise distributions listed in Table 1: log-log plots of the estimated W_1-risks for the naive method and the isotone method.

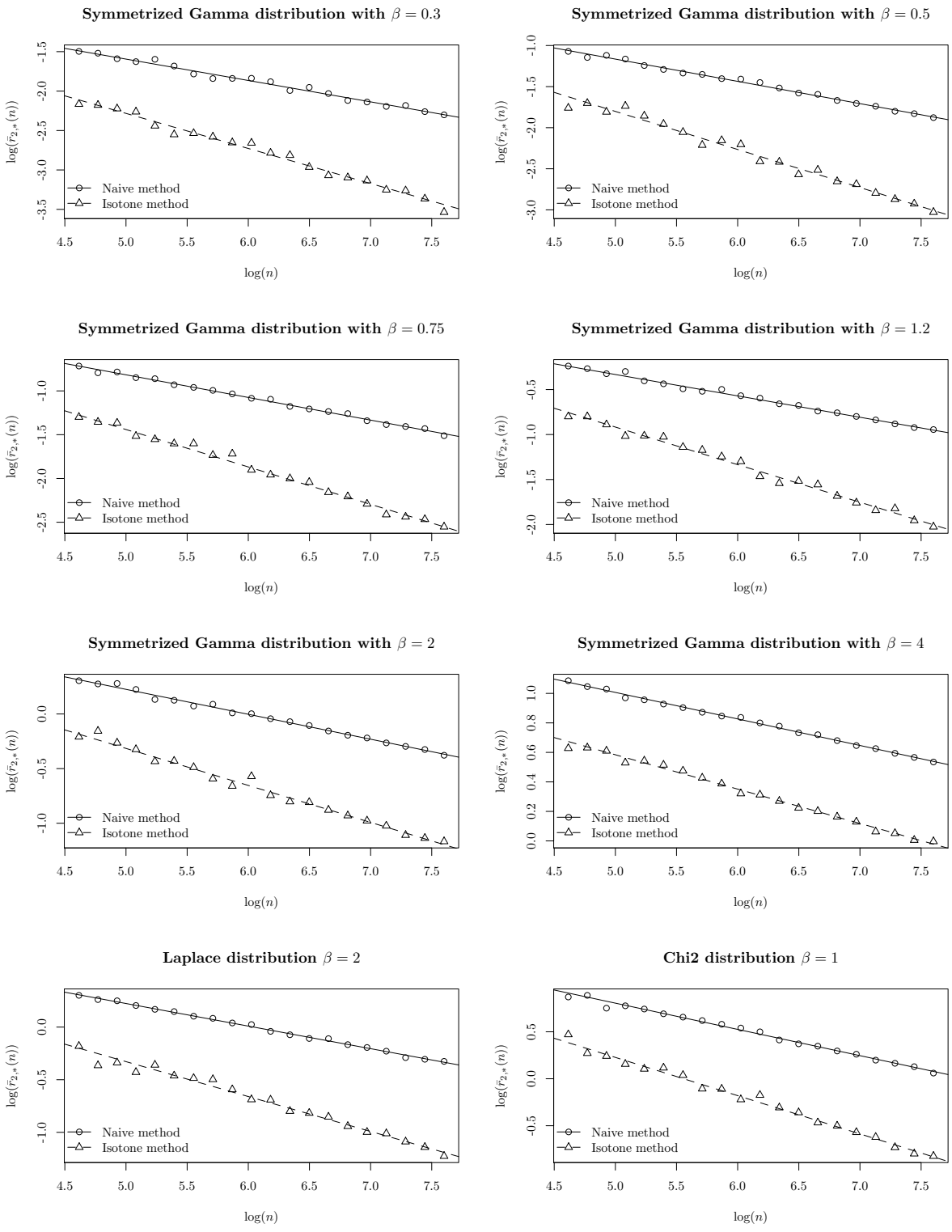


Figure 8: Deconvolution of the Dirac distribution at zero observed with one of the noise distributions listed in Table 1: log-log plots of the estimated W_2-risks for the naive method and the isotone method.

E. del Barrio, E. Giné, and C. Matrán. The central limit theorem for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.*, 27:1009–1971, 1999.

E. del Barrio, E. Giné, and F. Utzet. Asymptotics for $\mathbb{L}_2$ functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, 11:131–189, 2005.

A. Delaigle and I. Gijbels. Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. I. Stat. Math.*, 56(1):19–47, 2004.

Š.S. Èbralidze. Inequalities for the probabilities of large deviations in terms of pseudomoments. *Teor. Verojatnost. i Primenen.*, 16:760–765, 1971.

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

J. Fan. Global behavior of deconvolution kernel estimates. *Statist. Sinica*, 2:541–551, 1991b.

J. Fan. Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Stat.*, 21:600–610, 1993.

J. Gurland. Inversion formulae for the distribution of ratios. *Ann. Math. Statist.*, 19: 228–237, 1948.

P. Hall and S.N. Lahiri. Estimation of distributions, moments and quantiles in deconvolution problems. *Ann. Statist*, 36:2110–2134, 2008.

P. Mair, K. Hornik, and J. de Leeuw. Isotone optimization in R: pool-adjacent-violators algorithm (PAVA) and active set methods. *J. Stat. Softw.*, 32(5):1–24, 2009.

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

S.T. Rachev and L. Rüschendorf. *Mass transportation problems*, volume II of *Probability and its Applications*. Springer-Verlag, 1998.

A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes*. Springer series in Statistics. Springer, 1996.

C. Villani. *Optimal Transport: Old and New*. Grundlehren Der Mathematischen Wissenschaften. Springer-Verlag, 2008.