HÖLDER PROPERTIES OF THE L_1-MEDIAN. RATE OF CONVERGENCE IN THE ESTIMATION OF THE CONDITIONAL L_1-MEDIAN (MIXING CASE)

Benot Cadre
Laboratoire de Probabilits et Statistiques
CC 051, Place Eugne Bataillon
34095 MONTPELLIER Cedex 05
FRANCE

Abstract Let us denote by $\mathcal{P}_d$ the set of all probability measures on $\mathbb{R}^d$ ($d \geq 2$), and by $\mathcal{M}(\mu)$ the set of all L_1-medians of $\mu \in \mathcal{P}_d$, i.e. $\mathcal{M}(\mu) = \text{Argmin}_{\alpha \in \mathbb{R}^d} \varphi_\mu(\alpha)$, where $\varphi_\mu(\alpha) = \int(\|t-\alpha\| - \|t\|)\mu(dt)$ for $\alpha \in \mathbb{R}^d$ ($\|.\|$ is the l_q-norm with $q \geq 2$). If $m(\mu)$ denotes any point in $\mathcal{M}(\mu)$, we shall deal with the Hölder properties of the application

$$\begin{array}{rl} \{\mu \in \mathcal{P}_d \,:\, \text{card}\,\mathcal{M}(\mu) = 1\} & \to \mathbb{R}^d \\ \mu & \mapsto m(\mu), \end{array}$$

where $\mathcal{P}_d$ is endowed with the topology induced by the metric associated with the bounded and Lipschitz functions. We apply the result to get an upper bound of the almost sure rate of convergence in the uniform estimation of the conditional L_1-median from mixing observations.

AMS 1991 subject classification : 62G07; 62J02; 60B10.
Key-words and phrases : L_1-median, nonparametric estimation, rate of convergence, mixing, BL-metric.

1 Introduction

1.1 The L_1-median. Objectives of the paper

The concept of L_1-median (sometimes called spatial median) was historically introduced by Weber [16] as follows. A company wishes to choose an appropriate location for a warehouse which will service n customers whose planar

coordinates are given to be $C_1, \cdots, C_n$. The solution suggested by Weber is to locate the warehouse so as to minimize the sum of the distances from it to C_i, $i = 1, \cdots, n$. The set of solutions is called the set of L_1-medians. This concept of L_1-median was generalized and studied later by Averous-Meste [1], Bru-Heinich [5], Gini-Galvani [10], Haldane [11], Kemperman [12], Milasevic-Ducharme [13] and Valadier [15] (see also the survey by Small [14]).
In the following, we will denote by $\mathcal{P}_d$ the set of all probability measures defined on $I\!R^d$, where $d \geq 1$. If $\mu \in \mathcal{P}_d$, let us denote by φ_μ the function defined for all $\alpha \in I\!R^d$ by

$$\varphi_\mu(\alpha) = \int (\|t - \alpha\| - \|t\|)\mu(dt),$$

where $\|.\|$ is some norm on $I\!R^d$. Note that this definition does not need the assumption $\int \|t\|\mu(dt) < \infty$. The set $\mathcal{M}(\mu)$ of all L_1-medians of μ is the set of all points where φ_μ is as small as possible, i.e.

$$\mathcal{M}(\mu) = \text{Argmin}_{\alpha \in I\!R^d} \varphi_\mu(\alpha). \tag{1.1}$$

In the special case where $d = 1$ the L_1-median is well known to reduce to the standard one dimensional median.
Kemperman [12] proved that $\mathcal{M}(\mu)$ is always non-empty, and contains only one element if $\|.\|$ is a strictly convex norm (i.e. $\|x+y\| < \|x\|+\|y\|$, whenever x and y are not proportional) and if the support of μ is not included in a straight line (which impose $d \geq 2$).
Here, and subsequently, $m(\mu)$ will denote an element of $\mathcal{M}(\mu)$, if $\mu \in \mathcal{P}_d$. In this paper, we shall deal with the continuity properties of the application

$$\begin{array}{rl} \{\mu \in \mathcal{P}_d \,:\, \text{card}\, \mathcal{M}(\mu) = 1\} & \to I\!R^d \\ \mu & \mapsto m(\mu), \end{array} \tag{1.2}$$

where $\mathcal{P}_d$ is endowed with the topology induced by the BL-metric π (BL for Bounded-Lipschitz : See Section 1.2). Let $\mu \in \mathcal{P}_d$ such that $\mathcal{M}(\mu) = \{m(\mu)\}$. Kemperman ([12], Corollary 2.26) proved that, $(\mu_n)_{n\geq 1}$ being a sequence of probability measures on $I\!R^d$ such that $\pi(\mu_n, \mu) \to_{n\to\infty} 0$, then $m(\mu_n) \to m(\mu)$, as $n \to \infty$ (see also Cadre [6] for the infinite dimensional case). Hence, the application defined by (1.2) is continuous. The first purpose of the paper is to prove that, if $\|.\|$ is the l_q-norm ($q \geq 2$), a restriction of this application is in fact Hölder of order 1/2.

We then apply the previous result to nonparametric estimation of the conditional L_1-median. Let $d \geq 2$, $p \geq 1$, (X, Y) be a random vector with values in $\mathbb{R}^p \times \mathbb{R}^d$ and $\mathcal{C}$ be a compact set in $\mathbb{R}^p$. For all $x \in \mathcal{C}$, write μ_x the law of Y given $X = x$ and assume that μ_x possesses only one L_1-median $m(\mu_x)$. We now consider the sample $(X_i, Y_i)_{i \geq 1}$, and for all $n \geq 1$, $x \in \mathcal{C}$, we denote by μ_x^n an estimator of the probability measure μ_x and by $m(\mu_x^n)$ a L_1-median of μ_x^n. In a very general context on the law of (X, Y), we use the Hölder property of the L_1-median in order to give an estimation of the almost sure rate of convergence of $\sup_{x \in \mathcal{C}} \|m(\mu_x) - m(\mu_x^n)\|$. Berlinet et al. ([2] in the i.i.d. case, [3] in the mixing case) already considered this problem but without any attempt to give an estimation of the rate of convergence (see Berlinet et al. [4] for an application of this result in nonparametric prediction). From a Statistical viewpoint, one of the interest of the L_1-median is the following : suppose that X and Y are linked by the equation

$$Y = g(X) + \varepsilon,$$

where g is defined on $\mathbb{R}^p$ and ε is a $\mathbb{R}^d$-valued random vector independent of X with known L_1-median. As easily seen, $m(\mu_x^n)$ then gives an estimator of $g(x)$. The advantage of this method, compared to the classical regression, is that it provides a robust estimator.

Section 1.2 is dedicated to fixing notations and hypotheses. In Section 2, we state the main results of the paper. Then, we set out the proofs in Section 3.

1.2 General notations and hypothese

In the following, d and q are integers greater than 2, $\|.\|$ is the l_q-norm on $\mathbb{R}^d$ (i.e. $\|x\| = (\sum_{i=1}^d |x_i|^q)^{1/q}$ if $x = (x_1, \cdots, x_d) \in \mathbb{R}^d$), and for $a \in \mathbb{R}^d$, $\delta > 0$, $B(a, \delta)$ is the closed ball $B(a, \delta) = \{x \in \mathbb{R}^d : \|x - a\| \leq \delta\}$ and $S(a, \delta)$ is the sphere $S(a, \delta) = \{x \in \mathbb{R}^d : \|x - a\| = \delta\}$. Note that $\|.\|$ is a strictly convex norm, since $q > 1$. Moreover, Hf (resp. ∇f) stands for the Hessian matrix (resp. the gradient) of a twice continuously differentiable function $f : \mathbb{R}^d \to \mathbb{R}$. The support of $\mu \in \mathcal{P}_d$ is denoted $\operatorname{supp} \mu$ and if $\mathcal{M}(\mu)$ (defined in (1.1)) contains only one element $m(\mu)$ then for all $\delta > 0$, $\varphi_{\mu,\delta}$ denotes the function defined for all $\alpha \in \mathbb{R}^d$ by

$$\varphi_{\mu,\delta}(\alpha) = \int_{B^c(m(\mu),\delta)} (\|t - \alpha\| - \|t\|)\mu(dt),$$

and $K(\mu,\delta)$ is the real number defined by

$$K(\mu,\delta) = \inf_{\alpha\in B(m(\mu),\delta/2),\, l\in \mathbb{R}^d\,:\,\|l\|=1} l^T H\varphi_{\mu,\delta}(\alpha) l,$$

where l^T stands for the transpose of $l \in \mathbb{R}^d$. Note that, since $q \geq 2$, $\varphi_{\mu,\delta}$ is a twice continuously differentiable function on $B(m(\mu),\varepsilon)$, if $\varepsilon \in]0,\delta[$.

Let us recall that the BL-metric is defined for all $\mu,\nu \in \mathcal{P}_d$ by

$$\pi(\mu,\nu) = \sup(|\int f d\mu - \int f d\nu|),$$

the supremum being taken over all $f : \mathbb{R}^d \to \mathbb{R}$ such that

$$\sup_{s,t\in\mathbb{R}^d,\, s\neq t} \frac{|f(s)-f(t)|}{\|s-t\|} + \sup_{t\in\mathbb{R}^d} |f(t)| \leq 1.$$

Our basic assumption is the following. *We say that the pair* $(\mu,\gamma) \in \mathcal{P}_d \times \mathbb{R}_+^*$ *satisfies* $H(\mu,\gamma)$ *if* $\mu(B^c(m(\mu),\gamma)) > 0$ *and* $\operatorname{supp}\mu \cap B^c(m(\mu),\gamma)$ *is not included in a straight line.* Note that if $(\mu,\gamma) \in \mathcal{P}_d \times \mathbb{R}_+^*$ satisfies $H(\mu,\gamma)$, then the support of μ is not included in a straight line, so that according to Kemperman [12], μ possesses only one L_1-median.

Remark 1.1 If $(\mu,\gamma_0) \in \mathcal{P}_d \times \mathbb{R}_+^*$ satisfies $H(\mu,\gamma_0)$, then for all $\gamma \in]0,\gamma_0]$, (μ,γ) satisfies $H(\mu,\gamma)$. Hence there is no loss of generality in assuming that $\gamma_0 \in]0,1]$.

Remark 1.2 If $(\mu,\gamma) \in \mathcal{P}_d \times]0,1]$ satisfies $H(\mu,\gamma)$, then the support of μ is not included in a straight line. However, it may happens that $\operatorname{supp}\mu$ is not included in a straight line, even though for all $\gamma > 0$, (μ,γ) does not satisfies $H(\mu,\gamma)$. Indeed, consider the case $d = 2$ and $\|.\|$ the Euclidean norm. Let $C, D, E \in \mathbb{R}^2$ such that the angle of vertex D is in the interval $[2\pi/3, \pi[$, and

$$\mu = \frac{1}{3}(\delta_C + \delta_D + \delta_E)$$

where δ_x denotes the Dirac measure on $x \in \mathbb{R}^2$. As C, D, E are not included in a straight line, $\mathcal{M}(\mu)$ contains only one element which is easily proven to be D. But, if $\gamma > 0$ is such that $\mu(B^c(D,\gamma)) > 0$, $\operatorname{supp}\mu \cap B^c(D,\gamma)$ equals, as the case may be, $\emptyset$, C, E or $\{C,E\}$. Hence, (μ,γ) does not satisfies $H(\mu,\gamma)$.

Remark 1.3 Let $\mu \in \mathcal{P}_d$. If there exists $A \subset I\!R^{d-2}$ such that $A \times I\!R^2 \subset \operatorname{supp} \mu$ then for all $\gamma > 0$, (μ, γ) satisfies $H(\mu, \gamma)$. More generally, if $\operatorname{supp} \mu$ charges the Lebesgue measure on $I\!R^d$, then there exists $\gamma > 0$ such that (μ, γ) satisfies $H(\mu, \gamma)$.

2 The results

2.1 Hölder properties of the L_1-median

Theorem 2.1 *Let $(\mu, \gamma) \in \mathcal{P}_d \times]0, 1]$ be a pair which satisfies $H(\mu, \gamma)$. Then $K(\mu, \gamma) > 0$, and for all $\nu \in \mathcal{P}_d$:*

$$i)\ \|m(\mu) - m(\nu)\|^2 \leq \frac{24(1 + \|m(\mu)\| + \|m(\nu)\|)^2}{\gamma\, K(\mu, \gamma)}\, \pi(\mu, \nu);$$

and if $A = 1 + \|m(\mu)\| + \|m(\nu)\|$,

$$ii)\ \|m(\mu) - m(\nu)\|^2 \leq \frac{8(1 + \|m(\mu)\| + \|m(\nu)\|)}{\gamma\, K(\mu, \gamma)} \sup_{\|\alpha\| \leq A} |\varphi_\mu(\alpha) - \varphi_\nu(\alpha)|.$$

Remark 2.1 It is at least necessary to assume that $\mathcal{M}(\mu)$ contains only one element. If not, and $m_1 \neq m_2$ are two L_1-medians of μ, then $\|m_1 - m_2\| \neq 0$ when $\pi(\mu, \mu) = 0$!

2.2 Nonparametric estimation of the conditional L_1-median

Let $p \geq 1$ and (X, Y) be a $I\!R^p \times I\!R^d$-valued random vector defined on the probability space $(\Omega, \mathcal{F}, P)$. In the following, $\mathcal{C}$ is a compact set in $I\!R^p$ and for all $x \in \mathcal{C}$, $\mu_x = P_{Y|X=x}$. Assumption **(A1)** below is our basic assumption on the familly $(\mu_x, x \in \mathcal{C})$. Note that under Assumption **(A1)** below, we have for all $x \in \mathcal{C}$: $\mathcal{M}(\mu_x) = \{m(\mu_x)\}$.

(A1) There exists $\gamma \in]0, 1]$ such that for all $x \in \mathcal{C}$, (μ_x, γ) satisfies $H(\mu_x, \gamma)$ and $\mu_x(S(m(\mu_x), \gamma)) = 0$.

Let $(X_i, Y_i)_{i\geq 1}$ be a sequence random vectors, with the same law as (X, Y). For all $x \in \mathcal{C}$, the probability measure μ_x is estimated from a Nadaraya-Watson type estimator. For $n \geq 1$, let μ_x^n be the probability measure defined for all borel set $A \subset I\!R^d$ by

$$\mu_x^n = \frac{\sum_{i=1}^n I_A(Y_i) K(\frac{x-X_i}{h_n})}{\sum_{i=1}^n K(\frac{x-X_i}{h_n})},$$

where K denotes a probability density on $I\!R^p$ and $(h_n)_{n\geq 1}$ (the bandwidth) is a sequence of positive real numbers such that $h_n \to 0$ as $n \to \infty$. Recall that, according to Kemperman [12], $\mathcal{M}(\mu_x^n)$ is a non empty set.

Here, and in the following $|.|$ will denote a fixed norm on $I\!R^p$ as well as the standard distance on $I\!R$. Let us introduce now the assumptions:

(A2) The sequence $(X_i, Y_i)_{i\geq 1}$ is strictly stationary and geometrically mixing.

For a definition of the term *geometrically mixing*, we refer to Doukhan [9].

(A3) X_1 admits a density g, and for some ε-neighborhood $\tilde{\mathcal{C}}_1$ of $\mathcal{C}$,

$$\inf_{x\in\tilde{\mathcal{C}}_1} g(x) > 0;$$

(A4) There exists an ε-neighborhood $\tilde{\mathcal{C}}_2$ of $\mathcal{C}$ such that, if

$$M(\theta) = \sup(\pi(\mu_x, \mu_{x'}),\, x, x' \in \tilde{\mathcal{C}}_2, |x - x'| \leq \theta) \qquad \forall\theta \geq 0,$$

then M is continuous;
(A5) K is Hölderian and $\int \exp(|t|)K(t)dt < \infty$;
(A6) $nh_n^p(\log n)^{-6} \to \infty$, as $n \to \infty$.

Remark 2.2 The first condition in **(A1)** is always satisfied if for all $x \in \mathcal{C}$, there exists $A \subset I\!R^{d-2}$ such that $A \times I\!R^2 \subset \mathrm{supp}\mu_x$. Moreover, the second condition in **(A1)** is always satisfied if for all $x \in \mathcal{C}$, the measure μ_x is absolutely continuous with respect to the Lebesgue measure.

Remark 2.3 Under assumption **(A4)** we have $M(\theta) \to 0$ if $\theta \to 0$, which implies that for all $\alpha \in I\!R^d$, the function $(\varphi_{\mu_x}(\alpha), x \in \mathcal{C})$ is continuous.

Theorem 2.2 *Assume* **(A1)-(A6)** *hold, and let* $\xi > 0$*,* $\lambda \in]0,1[$*. If for all* $x \in \mathcal{C}$*,* $n \geq 1$*,* $m(\mu_x^n) \in \mathcal{M}(\mu_x^n)$ *then w.p. 1, we have for all* n *large enough:*

$$\sup_{x\in\mathcal{C}} \|m(\mu_x^n) - m(\mu_x)\| = O(h_n^{\xi} + (nh_n^p)^{-1/4} + M(h_n^{\lambda})^{1/2}). \qquad (2.1)$$

Remark 2.4 Note that we obtained Theorem 2.2 without assuming that the law of (X_1, Y_1) is absolutely continuous with respect to the Lebesgue measure. Of course, one can not expect to get an optimal result because of the generality of the situation. Finally, it is of interest to note that the rate of convergence in Theorem 2.2 depends on the modulus of continuity M defined in **(A4)**.

Remark 2.5 When the data are i.i.d., Cadre and Gannoun [7] proved the asymptotic normality of the sequence $((nh_n^p)^{1/2}(m(\mu_x^n) - m(\mu_x)))_{n\geq 1}$. But this result is obtained under strong hypotheses.

3 Proofs

3.1 Proof of Theorem 2.1

First of all, we state two Lemmas, delaying their proofs to the end of this Section.

Lemma 3.1 *If* $(\mu, \gamma) \in \mathcal{P}_d \times]0,1]$ *satisfies* $H(\mu,\gamma)$*, then* $K(\mu,\gamma) > 0$*.*

Lemma 3.2 *Let* $\nu, \mu \in \mathcal{P}_d$*. Then,*

i) $\varphi_\mu(m(\nu)) - \varphi_\mu(m(\mu)) \leq 6(1 + \|m(\mu)\| + \|m(\nu)\|)\,\pi(\mu,\nu)$;

and if $A = 1 + \|m_\mu\| + \|m_\nu\|$,

ii) $\varphi_\mu(m(\nu)) - \varphi_\mu(m(\mu)) \leq 2 \sup_{\|\alpha\|\leq A} |\varphi_\mu(\alpha) - \varphi_\nu(\alpha)|$.

Proof Theorem 2.1 Let us denote by φ^1 the function defined for all $\alpha \in I\!R^d$ by

$$\varphi^1(\alpha) = \int_{B(m(\mu),\gamma)\setminus m(\mu)} (\|t-\alpha\| - \|t\|)\mu(dt).$$

Throughout the proof, $k = m(\nu) - m(\mu)$ (which is assumed, without loss of generality, to be different from 0) and $h = \min(1, \gamma/(2\|k\|))$.
Since φ_μ is a convex function and $h \in]0, 1]$, we have:

$$\begin{aligned}\varphi_\mu(m(\nu)) - \varphi_\mu(m(\mu)) &= \varphi_\mu(m(\mu) + k) - \varphi_\mu(m(\mu)) \\ &\geq \frac{\varphi_\mu(m(\mu) + hk) - \varphi_\mu(m(\mu))}{h}\end{aligned}$$

and the rightmost term equals

$$\frac{\varphi_{\mu,\gamma}(m(\mu) + hk) - \varphi_{\mu,\gamma}(m(\mu))}{h} + \frac{\varphi^1(m(\mu) + hk) - \varphi^1(m(\mu))}{h}$$
$$+\|k\|\mu(\{m(\mu)\}).$$

On one hand, $\varphi_{\mu,\gamma}$ being a twice continuously differentiable function on $B(m(\mu), h\|k\|)$, we get from Taylor's formula, for some $\theta \in]0, 1[$:

$$\begin{aligned}&\frac{\varphi_{\mu,\gamma}(m(\mu) + hk) - \varphi_{\mu,\gamma}(m(\mu))}{h} \\ &= k^T \nabla \varphi_{\mu,\gamma}(m(\mu)) + \frac{h}{2} k^T H \varphi_{\mu,\gamma}(m(\mu) + \theta hk) k \\ &\geq k^T \nabla \varphi_{\mu,\gamma}(m(\mu)) + \frac{h}{2} \|k\|^2 K(\mu, \gamma).\end{aligned}$$

On the other hand, since φ^1 is a convex function, we have for all $\beta \in]0, h]$:

$$\frac{\varphi^1(m(\mu) + hk) - \varphi^1(m(\mu))}{h} \geq \frac{\varphi^1(m(\mu) + \beta k) - \varphi^1(m(\mu))}{\beta}.$$

According to the Lebesgue Theorem, as $\beta \searrow 0$:

$$\frac{\varphi^1(m(\mu) + \beta k) - \varphi^1(m(\mu))}{\beta} \rightarrow k^T \nabla \varphi^1(m(\mu)).$$

Consequently,

$$\varphi_\mu(m(\nu)) - \varphi_\mu(m(\mu))$$
$$\geq k^T \nabla \varphi^1(m(\mu)) + k^T \nabla \varphi_{\mu,\gamma}(m(\mu)) + \|k\|\mu(\{m(\mu)\}) + \frac{h}{2}\|k\|^2 K(\mu, \gamma).$$

But, according to Kemperman ([12], Theorem 4.11), since $m(\mu)$ is a L_1-median for μ:

$$k^T\nabla\varphi^1(m(\mu)) + k^T\nabla\varphi_{\mu,\gamma}(m(\mu)) + \|k\|\mu(\{m(\mu)\}) \geq 0,$$

hence

$$\varphi_\mu(m(\nu)) - \varphi_\mu(m(\mu)) \geq \frac{h}{2}\|k\|^2 K(\mu,\gamma).$$

Now, by Lemma 3.1 we have $K(\mu,\gamma) > 0$, and moreover

$$\frac{1}{h} \leq \max(1, \frac{2(\|m(\mu)\| + \|m(\nu)\|)}{\gamma}) \leq \frac{2}{\gamma}(1 + \|m(\mu)\| + \|m(\nu)\|),$$

because $\gamma \in]0,1]$. Hence we have obtained the inequality

$$\|k\|^2 \leq \frac{4(1 + \|m(\mu)\| + \|m(\nu)\|)}{\gamma\, K(\mu,\gamma)}(\varphi_\mu(m(\nu)) - \varphi_\mu(m(\mu))),$$

which, according to Lemma 3.2, proves the assertions i) and ii) of Theorem 2.1.□

Proof of Lemma 3.1 Let us prove that $\varphi_{\mu,\gamma}$ is strictly convex. Obviously, $\varphi_{\mu,\gamma}$ is a convex function. If $\varphi_{\mu,\gamma}$ were not strictly convex then there exists $\alpha_0 \in I\!R^d$ and $l_0 \in I\!R^d\backslash 0$ such that $\varphi_{\mu,\gamma}(\alpha_0+2l_0) - 2\varphi_{\mu,\gamma}(\alpha_0+l_0) + \varphi_{\mu,\gamma}(\alpha_0) = 0$. But then we have, for μ-a.e. $t \in B^c(m(\mu),\gamma)$, $\|\alpha_0 - t + 2l_0\| - 2\|\alpha_0 - t + l_0\| + \|\alpha_0 - t\| = 0$. Hence from Kemperman ([12], Lemma 2.14) (since $\|.\|$ is a strictly convex norm), for μ-a.e. $t \in B^c(m(\mu),\gamma)$, $\alpha_0 - t$ is a scalar multiple of l_0. Equivalently, $\operatorname{supp}\mu \cap B^c(m(\mu),\gamma)$ is included in a straight line, which contradicts the fact that (μ,γ) satisfies $H(\mu,\gamma)$. Hence, $\varphi_{\mu,\gamma}$ is strictly convex, which implies that for all $\alpha \in B(m(\mu),\gamma/2)$, $l \in I\!R^d \setminus 0$: $l^T H\varphi_{\mu,\gamma}(\alpha) l > 0$, since $\varphi_{\mu,\gamma}$ is a twice continuously differentiable function on $B(m(\mu),\gamma/2)$. By continuity, we deduce $K(\mu,\gamma) > 0$. □

Proof of Lemma 3.2 Since $m(\mu)$ (resp. $m(\nu)$) is a minimum for φ_μ (resp. φ_ν):

$$\begin{aligned}
&\varphi_\mu(m(\nu)) - \varphi_\mu(m(\mu)) \\
&\leq |\varphi_\mu(m(\nu)) - \varphi_\nu(m(\nu))| + |\varphi_\nu(m(\nu)) - \varphi_\mu(m(\mu))| \\
&\leq \sup_{\alpha\in B}|\varphi_\mu(\alpha) - \varphi_\nu(\alpha)| + |\inf_{\alpha\in B}\varphi_\mu(\alpha) - \inf_{\alpha\in B}\varphi_\nu(\alpha)| \\
&\leq 2\sup_{\alpha\in B}|\varphi_\mu(\alpha) - \varphi_\nu(\alpha)|, \qquad (3.1)
\end{aligned}$$

where $B = B(m(\mu), 1/2) \cup B(m(\nu), 1/2)$. This gives Assertion ii). Now, by the very definition of the BL-norm :

$$
\begin{aligned}
&\sup_{\alpha\in B} |\varphi_\mu(\alpha) - \varphi_\nu(\alpha)| \\
&= 3 \sup_{\alpha\in B} \max(1, \|\alpha\|) \left| \int \frac{\|t-\alpha\| - \|t\|}{3\max(1,\|\alpha\|)} (\mu(dt) - \nu(dt)) \right| \\
&\leq 3 \sup_{\alpha\in B} \max(1, \|\alpha\|) \pi(\mu, \nu) \\
&\leq 3(1 + \|m(\mu)\| + \|m(\nu)\|) \pi(\mu, \nu),
\end{aligned}
$$

which gives Assertion i), according to (3.1). □

3.2 Proof of Theorem 2.2

First of all, we state one Lemma, delaying its proof to the end of the Section.

Lemma 3.3 *Assume* **(A1)**-**(A6)** *hold. Then w.p. 1, there exists* $S < \infty$ *such that for all* $n \geq 1$*:*

$$
\sup_{x\in\mathcal{C}} \|m(\mu_x) - m(\mu_x^n)\|^2 \leq S \sup_{\|\alpha\|\leq S} \sup_{x\in\mathcal{C}} |\varphi_{\mu_x}(\alpha) - \varphi_{\mu_x^n}(\alpha)|.
$$

Proof of Theorem 2.2 In this proof, we fix $\alpha \in \mathbb{R}^d \setminus 0$, $\xi > 0$ and $\lambda \in]0, 1[$. The constants $c_1, \cdots, c_6$ defined in this proof are independent of α. For all $n \geq 1$ and $x \in \mathbb{R}^p$, let

$$
\begin{aligned}
K_n(x) &= h_n^{-p} K(x h_n^{-1}), \\
R_n(x, \alpha) &= n^{-1} \sum_{i=1}^n (\|Y_i - \alpha\| - \|Y_i\|) K_n(x - X_i), \\
g_n(x) &= n^{-1} \sum_{i=1}^n K_n(x - X_i).
\end{aligned}
$$

Then, for all $n \geq 1$:

$$
\sup_{x\in\mathcal{C}} |\varphi_{\mu_x^n}(\alpha) - \varphi_{\mu_x}(\alpha)| \leq \frac{1}{\inf_{x\in\mathcal{C}} g_n(x)} (A_n(\alpha) + \|\alpha\| B_n + C_n(\alpha)), \qquad (3.2)
$$

where

$$\begin{aligned} A_n(\alpha) &= \sup_{x \in \mathcal{C}} |R_n(x,\alpha) - E[R_n(x,\alpha)]|; \\ B_n &= \sup_{x \in \mathcal{C}} |g_n(x) - E[g_n(x)]|; \\ C_n(\alpha) &= \sup_{x \in \mathcal{C}} |E[R_n(x,\alpha)] - \varphi_{\mu_x}(\alpha) E[g_n(x)]|. \end{aligned}$$

According to Collomb ([8], Lemma 4), there exists $\delta > 0$ such that

$$\sum_{n \geq 1} P(\inf_{x \in \mathcal{C}} g_n(x) \leq \delta) < \infty.$$

Then, the Borel-Cantelli Lemma yields that w.p. 1, for all n large enough:

$$\inf_{x \in \mathcal{C}} g_n(x) \geq \delta. \qquad (3.3)$$

Moreover, an easy modification of the proof of Lemma 3 of Collomb [8] leads to the existence of $c_1, c_2 > 0$ such that for all n large enough, and for all $\varepsilon > 0$:

$$P(A_n(\alpha) > \varepsilon) \leq c_1 n^{c_1} \exp\Big(-c_2 \frac{\varepsilon}{\|\alpha\|} \frac{n h_n^p}{(\log n)^2}\Big) + P(\|\alpha\| c_1 h_n^{2\xi} > \varepsilon).$$

Let $\varepsilon = \|\alpha\|(h_n^\xi + (n h_n^p)^{-1/2})$. Then, for n large enough, since $h_n \to 0$ as $n \to \infty$:

$$\begin{aligned} P(A_n(\alpha) > \|\alpha\|(h_n^\xi + (n h_n^p)^{-1/2})) &\leq c_1 n^{c_1} \exp\Big(-c_2 \frac{(n h_n^p)^{1/2}}{(\log n)^2}\Big) \\ &\leq c_1 \exp\Big(c_1 \log n - c_2 \frac{(n h_n^p)^{1/2}}{(\log n)^2}\Big). \end{aligned}$$

According to **(A6)**, we then have

$$\sum_{n \geq 1} P(A_n(\alpha) > \|\alpha\|(h_n^\xi + (n h_n^p)^{-1/2})) < \infty,$$

so that by the Borel-Cantelli Lemma, w.p. 1, for all n large enough:

$$A_n(\alpha) \leq \|\alpha\|(h_n^\xi + (n h_n^p)^{-1/2}). \qquad (3.4)$$

Similarly, one can also prove that w.p. 1, for all n large enough:

$$B_n \leq h_n^\xi + (nh_n^p)^{-1/2}. \qquad (3.5)$$

For all $v \geq 0$, let:

$$\Gamma(v) = \inf(x \,:\, M(x) > v),$$

M being defined by **(A4)**. Recall that, since M is continuous, we have $M(\Gamma(u)) = u$ and $\Gamma(M(u)) \geq u$ for all $u \geq 0$. Now, an easy modification of the proof of Lemma 5 of Collomb [8] leads to the existence of $c_3 > 0$ such that for all $n \geq 1$ and $\beta > 0$:

$$C_n(\alpha) \leq c_3 \max(1, \|\alpha\|)(\beta + \psi_\beta(h_n)),$$

where for all $u > 0$:

$$\psi_\beta(u) = \int I_{\{|t| > u^{-1}\Gamma(\beta/(3\max(1,\|\alpha\|)))\}} K(t)dt,$$

Let $c_4 = \int \exp(|t|)K(t)dt$, which is finite by **(A5)**. We then have for all $n \geq 1$:

$$\psi_\beta(h_n) \leq c_4 \exp(-h_n^{-1}\Gamma(\beta/(3\max(1, \|\alpha\|)))).$$

Let us choose $\beta = 3\max(1, \|\alpha\|)M(h_n^\lambda)$. Since $h_n \to 0$ as $n \to \infty$, we have for n large enough : $\psi_\beta(h_n) \leq h_n^\xi$. Consequently, there exists $c_5 > 0$ such that for all n large enough:

$$C_n(\alpha) \leq c_5 \max(1, \|\alpha\|)(h_n^\xi + M(h_n^\lambda)). \qquad (3.6)$$

Finally, we get from (3.2)-(3.6) that there exists $c_6 > 0$ such that w.p. 1, for all n large enough:

$$\sup_{x\in\mathcal{C}} |\varphi_{\mu_x^n}(\alpha) - \varphi_{\mu_x}(\alpha)| \leq c_6 \max(1, \|\alpha\|)(h_n^\xi + (nh_n^p)^{-1/2} + M(h_n^\lambda)).$$

Hence, by Lemma 3.3, w.p. 1:

$$\sup_{x\in\mathcal{C}} \|m(\mu_x) - m(\mu_x^n)\| = O(h_n^{\xi/2} + (nh_n^p)^{-1/4} + M(h_n^\lambda)^{1/2}),$$

for all n large enough. □

In the rest of this Section, Δ stands for the symetric difference, η stands for the function $\eta(x) = \|x\|$, $x \in I\!\!R^d$, and for simplicity, $\|A\|$ denotes the l_q-norm of any $d \times d$-matrix A with real coefficients. Lemma 3.4 below will be usefull for the proof of Lemma 3.3.

Lemma 3.4 *Assume* **(A1)**, **(A4)** *hold and let* $x \in \mathcal{C}$. *If* $(x_n)_{n\geq 1}$ *is any sequence in* $\mathcal{C}$ *which converges to* x, *then*

$$\mu_{x_n}(B(m(\mu_x), \gamma)\Delta B(m(\mu_{x_n}), \gamma)) \to 0, \text{ as } n \to \infty.$$

Proof Let $(\beta_n)_{n\geq 1}$ be a decreasing sequence of positive real numbers such that $\beta_1 < \gamma$. According to **(A4)**, the sequence $(\mu_{x_n})_{n\geq 1}$ weakly converges to μ_x. Then, by Kemperman ([12], Theorem 2.29), for all $p \geq 1$, there exists $N \geq 1$ such that if $n \geq N$:

$$\|m(\mu_x) - m(\mu_{x_n})\| \leq \beta_p.$$

Consequently, if $p \geq 1$, one has for all $n \geq N$:

$$B(m(\mu_x), \gamma - \beta_p) \subset B(m(\mu_{x_n}), \gamma) \subset B(m(\mu_x), \gamma + \beta_p),$$

so that

$$\mu_{x_n}(B(m(\mu_x), \gamma - \beta_p)) \leq \mu_{x_n}(B(m(\mu_{x_n}), \gamma)) \leq \mu_{x_n}(B(m(\mu_x), \gamma + \beta_p)).$$

Letting $n \to \infty$, we get from **(A4)** that for all $p \geq 1$:

$$\mu_x(B(m(\mu_x), \gamma - \beta_p)) \leq \liminf \mu_{x_n}(B(m(\mu_{x_n}), \gamma)),$$

$$\limsup \mu_{x_n}(B(m(\mu_{x_n}), \gamma)) \leq \mu_x(B(m(\mu_x), \gamma + \beta_p)).$$

By monotone convergence and **(A1)**, one deduces that

$$\mu_{x_n}(B(m(\mu_{x_n}), \gamma)) \to \mu_x(B(m(\mu_x), \gamma)), \text{ as } n \to \infty.$$

Finally, observe that by **(A4)**,

$$\mu_{x_n}(B(m(\mu_x), \gamma)) \to \mu_x(B(m(\mu_x), \gamma)), \text{ as } n \to \infty,$$

so that

$$\mu_{x_n}(B(m(\mu_x), \gamma)\Delta B(m(\mu_{x_n}), \gamma)) \to 0, \text{ as } n \to \infty,$$

hence the Lemma. □

Let us state now without proof an easy Lemma, which will be usefull for the proof of Lemma 3.3.

Lemma 3.5 *Let $a > 0$. Then, $H\eta$ is uniformly bounded on $B^c(0, a)$ and if $(\epsilon_n)_{n\geq 1}$ is a sequence in $\mathbb{R}^d$ such that $\|\epsilon_n\| \to 0$ as $n \to \infty$, then*

$$\sup_{\|u\|\geq a} \|H\eta(u) - H\eta(u + \epsilon_n)\| \to 0, \text{ as } n \to \infty.$$

Lemma 3.6 *Assume* **(A1)**, **(A4)** *hold. Then* $\inf_{x\in\mathcal{C}} K(\mu_x, \gamma) > 0$.

Proof Let

$$\Xi = \{(x, \alpha, l) \in \mathcal{C} \times \mathbb{R}^d \times \mathbb{R}^d \,:\, \alpha \in B(m(\mu_x), \gamma/2),\ \|l\| = 1\},$$

and note that

$$\inf_{x\in\mathcal{C}} K(\mu_x, \gamma) = \inf_{(x,\alpha,l)\in\Xi} l^T H\varphi_{\mu_x,\gamma}(\alpha) l.$$

Since for all $x \in \mathcal{C}$, $\varphi_{\mu_x,\gamma}$ is a twice continuously differentiable function on $B(m(\mu_x), \gamma/2)$, one deduces from Lemma 3.1 that if $(x, \alpha, l) \in \Xi$, then $l^T H\varphi_{\mu_x,\gamma}(\alpha) l > 0$. Hence, one only needs to prove that the function

$$(x, \alpha, l) \mapsto l^T H\varphi_{\mu_x,\gamma}(\alpha) l$$

defined on Ξ is continuous, since Ξ is a compact set.
In this proof, c denotes a constant which may vary from line to line. Let $(x, \alpha, l) \in \Xi$ and $((x_n, \alpha_n, l_n))_{n\geq 1}$ be a sequence in Ξ which converges to (x, α, l). For all $n \geq 1$,

$$|l^T H\varphi_{\mu_x,\gamma}(\alpha) l - l_n^T H\varphi_{\mu_{x_n},\gamma}(\alpha_n) l_n| \leq c(A_n + B_n + C_n), \qquad (3.7)$$

where

$$\begin{aligned} A_n &= \|l - l_n\| \|H\varphi_{\mu_x,\gamma}(\alpha)\|; \\ B_n &= \|H\varphi_{\mu_x,\gamma}(\alpha) - H\varphi_{\mu_{x_n},\gamma}(\alpha)\|; \\ C_n &= \|H\varphi_{\mu_{x_n},\gamma}(\alpha) - H\varphi_{\mu_{x_n},\gamma}(\alpha_n)\|. \end{aligned}$$

According to Lemma 3.5, $A_n \to 0$ as $n \to \infty$. Moreover, for all n large enough:

$$\begin{aligned} B_n &\leq \| \int H\eta(t-\alpha)(I_{B^c(m(\mu_{x_n}),\gamma)}(t) - I_{B^c(m(\mu_x),\gamma)}(t))\mu_{x_n}(dt)\| \\ &\quad + \| \int_{B^c(m(\mu_x),\gamma)} H\eta(t-\alpha)(\mu_{x_n} - \mu_x)(dt)\| \\ &\leq c\mu_{x_n}(B(m(\mu_x),\gamma)\Delta B(m(\mu_{x_n}),\gamma)) \\ &\quad + \| \int_{B^c(m(\mu_x),\gamma)} H\eta(t-\alpha)(\mu_{x_n} - \mu_x)(dt)\|, \end{aligned}$$

according to Lemma 3.5. The first term on the right hand side converges to 0 as $n \to \infty$, according to Lemma 3.4. Moreover, according to **(A4)**, the second term converges also to 0 since, by Lemma 3.5,

$$\sup_{t \in B^c(m(\mu_x),\gamma)} \|H\eta(t-\alpha)\| < \infty.$$

Consequently, $B_n \to 0$ as $n \to \infty$. Finally, observe that for all $n \geq 1$,

$$\begin{aligned} C_n &\leq \sup_{t \in B^c(m(\mu_{x_n}),\gamma)} \|H\eta(t-\alpha) - H\eta(t-\alpha_n)\| \\ &\leq \sup_{t \in B^c(\alpha_n,\gamma/2)} \|H\eta(t-\alpha) - H\eta(t-\alpha_n)\| \\ &= \sup_{\|u\| \geq \gamma/2} \|H\eta(u+\alpha_n-\alpha) - H\eta(u)\|, \end{aligned}$$

which, according to Lemma 3.5, converges to 0 as $n \to \infty$.
As a conclusion, the rightmost term in (3.7) converges to 0 as $n \to \infty$, proving that the function $(x, \alpha, l) \mapsto l^T H\varphi_{\mu_x,\gamma}(\alpha) l$ defined on Ξ is continuous. $\square$

Proof of Lemma 3.3 One wants to prove first that w.p. 1,

$$\sup_{x \in \mathcal{C}} \sup_{n \geq 1} \|m(\mu_x^n)\| + \sup_{x \in \mathcal{C}} \|m(\mu_x)\| < \infty. \qquad (3.8)$$

We only prove that the first term in the left hand side is finite. According to Collomb ([8], Theorem 2), and following the proof of Lemma 1 of Berlinet et al. [2], it is easy to obtain, under assumptions **(A2)**-**(A6)**, that w.p. 1,

$$\sup_{x \in \mathcal{C}} \sup_{n \geq 1} |\frac{\varphi_{\mu_x^n}(\alpha)}{\|\alpha\|} - 1| \to 0, \text{ as } \|\alpha\| \to \infty.$$

Consequently, w.p. 1, there exists $r > 0$ such that for all $\|\alpha\| \geq r$, $n \geq 1$ and $x \in \mathcal{C}$, $\varphi_{\mu_x^n}(\alpha) \geq \|\alpha\|/2$. Assume that for some $p \geq 1$ and $y \in \mathcal{C}$, $\|m(\mu_y^p)\| \geq r$. But then we have $\varphi_{\mu_y^p}(m(\mu_y^p)) \geq \|m(\mu_y^p)\|/2$, which is impossible because $\varphi_{\mu_y^p}(m(\mu_y^p)) \leq 0$. We have proved (3.8).
According to Theorem 2.1 ii), Lemma 3.6 and (3.8), we have w.p. 1 the existence of $S < \infty$ such that for all $n \geq 1$:

$$\sup_{x \in \mathcal{C}} \|m(\mu_x) - m(\mu_x^n)\|^2 \leq S \sup_{|\alpha\| \leq S} \sup_{x \in \mathcal{C}} |\varphi_{\mu_x}(\alpha) - \varphi_{\mu_x^n}(\alpha)|,$$

hence the result. □

References

[1] J. Averous, M. Meste. Median balls : an extension of the interquantile intervals to multivariate distributions, *J. Multivariate Anal.* **63** (1997), 222-241.
[2] A. Berlinet, B. Cadre, A. Gannoun. On the conditional L_1-median and its estimation, *Nonparametric Statist.* **13** (2001), 631-645.
[3] A. Berlinet, B. Cadre, A. Gannoun. Estimation of the conditional L_1-median from dependent observations, *Statist. Prob. Letters* **55** (2001), 353-358.
[4] A. Berlinet, B. Cadre, A. Gannoun. L'estimation non paramétrique de la médiane spatiale conditionnelle en vue de prévision, Actes du Colloque de la Société Mathématique de Tunisie (1999), 46-55.
[5] B. Bru, H. Heinich. Meilleures approximations et mdianes conditionnelles, *Ann. Inst. H. Poincar Probab. Statist.* **21** (1985), 197-224.
[6] B. Cadre. Convergent estimators for the L_1-median of a Banach valued random variable, to appear in *Statistics* (2002).
[7] B. Cadre, A. Gannoun. Asymptotic normality of consistent estimate of the conditional L_1-median, *Annales de l'I.S.U.P.* **44** (2000), 13-35.
[8] G. Collomb. Propriétés de convergence presque complète du prédicteur à noyau, *Z. Wahrsch. Verw. Gebiete* **66** (1984), 441-460.
[9] P. Doukhan. Mixing - Properties and examples, Lectures Notes in Statistics, Springer-Verlag (1994).
[10] K. Gini, L. Galvani. Di talune estensioni dei concetti di media ai caratteri qualitativi, *Metron* **8** (1929), 3-209.

[11] J.B.S. Haldane. Note on the median of a multivariate distribution, *Biometrika* **35** (1948), 414-415.
[12] J.H.D. Kemperman. The median of a finite measure of Banach space, Statistical data analysis based on the L_1-norm and related methods, North-Holland, Amsterdam-New York (1987), 217-230.
[13] P. Milasevic, G.R. Ducharme. Uniqueness of the spatial median, *Ann. Statist.* **15** (1987), 1332-1333.
[14] K.G. Small. A survey of multidimensional medians, *Internat. Statist. Rev.* **58** (1990), 263-277.
[15] M. Valadier. La multi-application médianes conditionnelles, *Z. Wahrsch. Verw. Gebiete* **67** (1984), 279-282.
[16] A. Weber. Uber den standord der industrien, tubingen, 1909. English translation by K.J. Freidrich, "Alfred Weber's theory of location of industries" (Chicago University Press, (1929)).