Estimation of conditional L_1-median from dependent observations

Alain Berlinet, Benoît Cadre and Ali Gannoun

Department of Mathematics, University of Montpellier II, cc 051, 34 095 Montpellier Cedex 05, France

Abstract
Let Q be a conditional distribution with L_1-median μ and let $\{Q_n\}$ be a sequence of conditional distributions converging in some sense to Q. Then the L_1-medians $\{\mu_n\}$ of distributions $\{Q_n\}$ are natural estimates of μ. In the case where $\{Q_n\}$ is a sequence of kernel estimates we give conditions ensuring that $\{\mu_n\}$ is a well defined sequence of continuous functions converging to μ uniformly on compact sets.

AMS 1991 subject classification. Primary 62 G 05; secondary 60 E 05.
Keywords: Conditional L_1-median, nonparametric estimation, non-i.i.d. observations.

1. Introduction

The univariate median is well known for its robustness properties. In the multivariate case several extensions of this concept were introduced in the literature. Having in view applications to prediction problems we will consider here the conditional L_1-median which is well adapted to asymmetric or heavy tailed conditional distributions. Estimation results were recently given for independent and identically distributed (i.i.d.) observations by Berlinet, Cadre and Gannoun (1998). The aim of the present note is to consider the non-i.i.d. case.

2. Definitions, notation and main result

Let (X,Y), $\{(X_i,Y_i), i \geq 1\}$ be random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in $R^s \times R^d$ ($s \geq 1$, $d \geq 2$) and let $||.||$ be a strictly convex norm on R^d. We assume that for $x \in R^s$, the support of $Q(.\,|x)$, the unknown distribution of Y conditionally on $X = x$, is not included into a straight line. Under this assumption, the function

$$\begin{aligned} \varphi(.,x) &= E\left[||Y - .|| - ||Y||\,|X = x\right] \\ &= \int (||y - .|| - ||y||)\, dQ(y\,|x) \end{aligned}$$

[0]Correspondence to Alain Berlinet, email berlinet@helios.ensam.inra.fr

has a unique minimum over R^d. This minimum is denoted by $\mu(x)$ and called the L_1-median of the distribution $Q(.\,|x\,)$.
From the observation of $(X_1, Y_1), \ldots, (X_n, Y_n)$, we define an estimate $\mu_n(x)$ of $\mu(x)$ based on a nonparametric estimate $Q_n(.\,|x)$ of the distribution $Q(.\,|x\,)$ (Collomb (1984), Gannoun (1990), Härdle (1990), Boente and Fraiman (1995)). Let K be a probability density function on R^s and $\{h_n, n \geq 1\}$ be a sequence of positive numbers tending to zero as n tends to infinity. For $x \in R^s$, define $Q_n(.\,|x\,)$ by

$$Q_n(.\,|x\,) = \frac{\sum_{i=1}^{n} \delta_{Y_i}(.) K(\frac{x-X_i}{h_n})}{\sum_{i=1}^{n} K(\frac{x-X_i}{h_n})},$$

where δ_y denotes the Dirac measure at the point $y \in R^d$. Conditions on $(X, Y), \{(X_i, Y_i), i \geq 1\}, K, \{h_n\}$ are known to ensure the convergence of the sequence $\{Q_n\}$ to the conditional distribution Q. Now, a natural estimate of $\varphi(., x)$ is

$$\begin{aligned}\varphi_n(., x) &= \int (\|y - .\| - \|y\|)\, dQ_n(y\,|x\,) \\ &= \frac{\sum_{i=1}^{n} (\|Y_i - .\| - \|Y_i\|)\, K(\frac{x-X_i}{h_n})}{\sum_{i=1}^{n} K(\frac{x-X_i}{h_n})}.\end{aligned}$$

In the i.i.d. case Berlinet et al (1998) prove, under suitable assumptions, the existence and uniqueness (a.s. for n large enough) of a minimizer of $\varphi_n(., x)$ over R^d. This minimizer $\mu_n(x)$ is the L_1-median of $Q_n(.\,|x\,)$. It satisfies

$$\begin{aligned}\mu_n(x) &= \arg\min_{a \in R^d} \varphi_n(a, x), \\ &= \arg\min_{a \in R^d} \sum_{i=1}^{n} \|Y_i - a\|\, K(\frac{x-X_i}{h_n}),\end{aligned}$$

and the sequence $\{\mu_n(.), n \geq 1\}$ is proved to converge a.s. to $\mu(.)$, uniformly on compact sets on which $\mu(.)$ satisfies a uniform uniqueness property.

In the sequel we take a convergence property of the sequence $\{Q_n, n \geq 1\}$ to the distribution Q as hypothesis. We prove that if $\{Y_i, i \geq 1\}$ is strong mixing we can obtain the same results on the sequence $\{\mu_n(.), n \geq 1\}$ as in the i.i.d. case. For definition and properties of mixing processes the reader is referred to Györfi et al (1989), Roussas (1990) and Doukhan (1994).

Throughout the paper, $\mathcal{C}$ denotes a fixed compact subset of R^s such that $P(X \in \mathcal{C}) > 0$.

Let us introduce the following hypotheses on the kernel K, on the sequence $\{Q_n, n \geq 1\}$ and on the conditional distribution Q.
(H1) The set of L_1-medians $(\mu(x), x \in \mathcal{C})$ satisfies the following *uniform uniqueness condition*

$$\forall\, \varepsilon > 0,\ \exists\, \eta > 0, \forall\, t : \mathcal{C} \longrightarrow R^d :$$

$$\sup_{x \in \mathcal{C}} \|\mu(x) - t(x)\| \geq \varepsilon \Longrightarrow \sup_{x \in \mathcal{C}} |\varphi(\mu(x), x) - \varphi(t(x), x)| \geq \eta.$$

(H2) The kernel K is continuous and positive.
(H3) If $f : R^d \to R$ is a bounded Borel function we have, with probability1,

$$\sup_{x \in \mathcal{C}} \left| \int f(y) dQ_n(y|x) - \int f(y) dQ(y|x) \right| \to 0, \text{ if } n \to +\infty. \tag{1}$$

(H4) $\{Y_i, i \geq 1\}$ is a strong mixing process such that

$$P(Y_{k+1}, Y_{k+2}, Y_{k+3} \text{ on the same line}) \text{ is independent of } k. \tag{2}$$

Comments on hypotheses
(H1) requires that a sufficient condition for any function t to be closed to the function μ uniformly on $\mathcal{C}$ is that the function $\varphi(t(.), .)$ be closed to the function $\varphi(\mu(.), .)$ in the same sense.
The continuity of K entails the continuity of $\int f(y) dQ_n(y|.)$ for any bounded Borel function f. As a consequence of its positivity the support of the measure $Q_n(.|x)$ is exactly $\{Y_1, \ldots, Y_n\}$. These two properties are required in the proof to guaranty the existence, uniqueness and continuity of $\mu_n(.)$ for n large enough.
Condition (1) in **(H3)** simply means uniform convergence of the standard Nadaraya-Watson estimate

$$\int f(y) dQ_n(y|x) = \frac{\sum_{i=1}^n f(Y_i) K(\frac{x - X_i}{h_n})}{\sum\limits_{i=1}^n K(\frac{x - X_i}{h_n})}$$

to the regression function

$$\int f(y)dQ(y|x) = E\left[f(Y)\,|X = x\right].$$

Sufficient conditions are known in a variety of contexts (including dependent and/or non stationary variables) to get this convergence (see, among others, Bosq and Lecoutre (1987), Harel and Puri (1995)).

Condition (2) in **(H4)** is clearly satisfied if $\{Y_i, i \geq 1\}$ is strictly stationary or if all vectors $(Y_{k+1}, Y_{k+2}, Y_{k+3})$ have a density with respect to the Lebesgue measure in R^{3d}.
Let us now state the main result of this paper.

Theorem 1. *Assume that **(H1)**, **(H2)**, **(H3)**, and **(H4)** are satisfied. Then,*
i) with probability 1, one can find an integer $N \geq 1$ such that if $n \geq N$ and $x \in \mathcal{C}$, the L_1-median $\mu_n(x)$ associated with the probability measure $Q_n(.\,|x)$ exists, is unique, and the function $\mu_n(.)$ is continuous on $\mathcal{C}$.
ii) with probability 1,

$$\sup_{x\in\mathcal{C}} ||\mu_n(x) - \mu(x)|| \to 0, \textit{ if } n \to +\infty.$$

Therefore $\mu(.)$ is continuous on $\mathcal{C}$.

First of all let us prove the following Lemma on the sequence $\{Y_i, i \geq 1\}$.

Lemma 2. *Assume that **(H3)** and **(H4)** are satisfied. Then the probability of the set $\{Y_1, Y_2, ...$ not on the same line$\}$ is equal to 1.*

Proof. The proof is divided into three steps.

Step 1. One wants to prove here that there exists a subset $\Omega_0 \subset \Omega$ of $P-$measure 1, such that, if $\omega \in \Omega_0$ and $x \in \mathcal{C}$, the probability measure $Q_n(.|x)(\omega)$ converges weakly to the probability measure $Q(.|x)$.

Let $\mathcal{B}$ be the set of all bounded, real-valued, Lipschitz functions on R^d. According to Dudley (1989, Theorem 11.3.3), it is enough to prove that, P-a.s.

$$\int u(y)dQ_n(y|x) \to \int u(y)dQ(y|x) \text{ if } n \to +\infty \tag{3}$$

for all $x \in \mathcal{C}$ and all $u \in \mathcal{B}$. The difficulty lies in the fact that the P-null set on which (3) is proved to hold must not depend on u. But $\mathcal{B}$ endowed with the supremum norm $||.||_\infty$ is separable. Denote by $\mathcal{B}_1$ a countable set, dense in $\mathcal{B}$. By **(H3)** we have (3) for any $x \in \mathcal{C}$ and any $u \in \mathcal{B}_1$ since $\mathcal{B}_1$ is countable. The conclusion follows from the denseness of $\mathcal{B}_1$.

Step 2. The goal here is to prove that $P(Y_1, Y_2, Y_3 \text{ on the same line}) < 1$. Let us assume the converse. We deduce that the probability of $\Omega_1 = \{Y_1, Y_2, Y_3, \ldots \text{ on the same line}\}$ is equal to 1 by (2) and because Ω_1 is a countable intersection of sets of P-measure 1.
Let $\omega \in \Omega_1$ and $x \in \mathcal{C}$. Obviously, the support of the probability measure $Q_n(.|x)(\omega)$ is included into $\{Y_1(\omega), \ldots, Y_n(\omega)\}$. Moreover $\omega \in \Omega_1$, so that this support is included into a line $\mathcal{D}(\omega)$ joining $Y_1(\omega), Y_2(\omega), \ldots$ Let $\omega \in \Omega_0 \cap \Omega_1$ (Ω_0 was defined in Step 1). According to Step 1, $Q_n(.|x)(\omega)$ converges weakly to $Q(.|x)$. By the Portmanteau Theorem, this is equivalent to

$$\limsup_n Q_n(V|x)(\omega) \leq Q(V|x),$$

for all closed V. Letting $V = \mathcal{D}(\omega)$, we have

$$1 = \limsup_n Q_n(\mathcal{D}(\omega)|x)(\omega) \leq Q(\mathcal{D}(\omega)|x).$$

Consequently, $Q(\mathcal{D}(\omega)|x) = 1$. This contradicts the fact that the support of $Q(.|x)$ is not included into a straight line. Hence, we proved that $P(Y_1, Y_2, Y_3$ on the same line$) < 1$.

Step 3. Conclusion. Recall the notations of Step 2, denote by $\{\alpha(i), i \geq 1\}$ the strong mixing coefficients of $\{Y_i, i \geq 1\}$ and, for $i \geq 1$, denote by $\mathcal{D}_i$ the event $\{Y_i, Y_{i+1}$ and Y_{i+2} are on the same line in $R^d\}$.

Then, we have, for fixed $p \geq 4$

$$\Omega_1 \subset \bigcap_{i\geq 1} \mathcal{D}_i \subset \bigcap_{i=1}^{p} \mathcal{D}_{pi}.$$

By the mixing property of the sequence $(Y_i)_{i\geq 1}$, we have

$$\begin{aligned} P\left(\bigcap_{i=1}^{p} \mathcal{D}_{pi}\right) &\leq \alpha(p-2) + P\left(\bigcap_{i=1}^{p-1} \mathcal{D}_{pi}\right) P\left(\mathcal{D}_{p^2}\right) \\ &\leq \alpha(p-2) + \alpha(p-2) P\left(\mathcal{D}_{p^2}\right) \\ &\quad + P\left(\bigcap_{i=1}^{p-2} \mathcal{D}_{pi}\right) P\left(\mathcal{D}_{p^2}\right) \prod_{i=p-1}^{p} P\left(\mathcal{D}_{pi}\right). \end{aligned}$$

Iterating we get the following inequalities

$$\begin{aligned} P\left(\bigcap_{i=1}^{p} \mathcal{D}_{pi}\right) &\leq \alpha(p-2)\left(1+\cdots+\prod_{i=3}^{p} P\left(\mathcal{D}_{pi}\right)\right)+\prod_{i=1}^{p} P\left(\mathcal{D}_{pi}\right) \\ &\leq \frac{\alpha(p-2)}{1-P\left(\mathcal{D}_1\right)}+\left(P\left(\mathcal{D}_1\right)\right)^p, \end{aligned}$$

because $\{Y_i, i \geq 1\}$ has the stationary property (2) and $P(\mathcal{D}_1) < 1$ according to Step 2. Now, letting p tend to infinity, we get that $P(\Omega_1) = 0$.□

3. Proof of Theorem 1

According to Lemma 1, $P(Y_1, Y_2, ...$ not on the same line$) = 1$. Moreover, K is a positive function. We deduce that, P-almost surely, one can find $N \geq 1$ such that if $n \geq N$ and $x \in \mathcal{C}$, the support of the probability measure $Q_n(.|x)$ is not included in a straight line. According to Kemperman (1987, Theorem 2.17) we deduce that the conditional L_1-median $\mu_n(x)$ associated with the probability measure $Q_n(.|x)$ exists and is unique. Moreover, by continuity of K, one can prove that $\mu_n(.)$ is a continuous function on $\mathcal{C}$ following the proof of Berlinet et al (1998, Lemma 3). Then, from part *(ii)* of Theorem 1, the continuity of $\mu(.)$ over $\mathcal{C}$ is clear. Hence, it remains to prove assertion *(ii)*. The proof of Berlinet et al (1998) can be adapted to get from **(H3)** that, P-almost surely, for all $A > 0$,

$$\sup_{||\alpha|| \leq A} \sup_{x \in \mathcal{C}} |\varphi_n(\alpha, x) - \varphi(\alpha, x)| \to 0, \text{ if } n \to \infty. \tag{4}$$

Then following the proof of the first step of Theorem 1 in Berlinet et al (1998), we deduce that, P-almost surely, one can find $r > 0$ and $N_1 \geq N$ such that

$$\sup_{n \geq N_1} \sup_{x \in \mathcal{C}} ||\mu_n(x)|| \leq r, \text{ and } \sup_{x \in \mathcal{C}} ||\mu(x)|| \leq r. \tag{5}$$

Then, P-almost surely, if $n \geq N_1$

$$\begin{aligned} &\sup_{x \in \mathcal{C}} |\varphi(\mu(x), x) - \varphi(\mu_n(x), x)| \\ &\leq \sup_{x \in \mathcal{C}} |\varphi(\mu(x), x) - \varphi_n(\mu_n(x), x)| + \sup_{x \in \mathcal{C}} |\varphi_n(\mu_n(x), x) - \varphi(\mu_n(x), x)| \\ &\leq \sup_{x \in \mathcal{C}} \left| \inf_{||a|| \leq r} \varphi(a, x) - \inf_{||a|| \leq r} \varphi_n(a, x) \right| + \sup_{||a|| \leq r} \sup_{x \in \mathcal{C}} |\varphi_n(a, x) - \varphi(a, x)|, \end{aligned}$$

according to (5), and because

$$\varphi(\mu(x),x)=\inf_{||a||\leq r}\varphi(a,x),\ \text{and}\ \varphi_n(\mu_n(x),x)=\inf_{||a||\leq r}\varphi_n(a,x).$$

Hence, from 4, we get that, P-almost surely:

$$\sup_{x\in\mathcal{C}}|\varphi(\mu(x),x)-\varphi(\mu_n(x),x)|\to 0, \text{if } n\to+\infty,$$

Hence *(ii)* follows from **(H1)**.□

Acknowledgement We wish to thank a referee whose remarks helped to reorganize the paper.

References

Berlinet, A., Cadre, B., Gannoun A. (1998). On the conditional L_1-median and its estimation. *Submitted to Nonparametric Statistics.*

Boente, G., Fraiman, R. (1995). Asymptotic distribution of data driven smoothers in density and regression under dependence. *The Canadian Journal of Statistics*, **23,4**, 383-397.

Bosq, D., Lecoutre, J.P. (1987). Théorie de l'estimation fonctionnelle. *Economica.*

Bradley, R. (1983). Approximation theorems for strongly mixing random variables. *Michigan Math. J.*, **30**, 69-81.

Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. *Z. Warhs. Verw. Gebiete*, **66**, 441-460.

Doukhan, P. (1994). Mixing : properties and examples. *Lect. Notes in Stat.*, 85, *Springer-Verlag, New York.*

Dudley, R.M. (1989). Real Analysis and Probability. *Wadsworth, Belmont, CA.*

Gannoun, A. (1990). Estimation non paramétrique de la médiane conditionnelle. Médianogramme et méthode du noyau. Application à la prévision des processus. *Pub. Inst. Stat. Univer. Paris XXXV*, **1**, 11-22.

Györfi, L., Härdle, W., Sarda, P. et Vieu, P. (1989) Nonparametric curve estimation from time series. *Lect. Notes in Stat.*, 60, *Springer-Verlag, Berlin.*

Härdle, W. (1990). Applied nonparametric regression. *Cambridge University Press.*

Harel, M. and Puri, M.L. (1995). Asymptotic normality of nearest neighbor regression function estimates based on nonstationary dependent observations. *Amer. Journ. of Math. and Manag. Sc.*, **15**, 255-290.

Kemperman, J.H.D. (1987). The median of finite measure of Banach space. *Statistical Data Analysis Based on the L_1-norm and Related Methods, Y. Dodge* (ed.), *North-Holland, Amsterdam*, 217-230.

Roussas, G.G. (1990). Asymptotic normality of the kernel estimate under dependence conditions : application to the hazard rate. *J. of Statist. Plann. and Inference*, **25**, 81-104.