



ELSEVIER

Contents lists available at [SciVerse ScienceDirect](http://SciVerse.ScienceDirect.com)

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistique

Propriétés asymptotiques des M-estimateurs pondérés pour des données clusterisées

Asymptotic properties of weighted M-estimators for clustered data

Mohammed El Asri¹

Université d'Avignon et des pays de Vaucluse, Laboratoire de mathématique d'Avignon, 33, rue Louis-Pasteur, 84000 Avignon, France

INFO ARTICLE

Historique de l'article :

Reçu le 21 novembre 2012

Accepté après révision le 17 avril 2013

Disponible sur Internet le 29 juillet 2013

Présenté par le Comité de rédaction

RÉSUMÉ

Nous étudions la classe des M-estimateurs pondérés, qui est une variante des M-estimateurs introduits par Huber en 1964 pour l'étude de la robustesse. Nous donnons des conditions suffisantes pour établir la convergence ainsi que la normalité asymptotique pour cette famille d'estimateurs et des variables aléatoires clusterisées.

© 2013 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

ABSTRACT

We study the class of weighted M-estimators, which is a variant of M-estimators introduced by Huber in 1964 for the study of robustness. We give sufficient conditions to establish the consistency and asymptotic normality of these estimators for clustered data.

© 2013 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

1. Introduction

La classe des M-estimateurs engendre des estimateurs classiques d'un paramètre de localisation multidimensionnel, θ , tels que l'estimateur du maximum de vraisemblance, la moyenne empirique et la médiane spatiale. Huber [1] a introduit les M-estimateurs dans le cadre de l'étude des estimateurs robustes. Parmi la littérature dédiée à ces estimateurs, on trouve en particulier les ouvrages de Huber [2] et de Hampel et al. [3] sur le comportement asymptotique et la robustesse via le point de rupture et la fonction d'influence (voir Ruiz-Gazen [4] pour une synthèse sur ces notions). Plus récemment, des résultats sur la convergence et la normalité asymptotique sont établis par Van der Vaart [5] dans le cadre multidimensionnel. Nevalainen et al. [6,7] étudient le cas particulier de la médiane spatiale pondérée et non-pondérée dans le cas clusterisé. Dans cette note, nous généralisons ces résultats aux M-estimateurs pondérés. Nous étudions leur convergence presque sûre ainsi que leur normalité asymptotique pour des données clusterisées.

2. Cadre d'étude

Soit X_1, \dots, X_n une suite de n clusters indépendants, avec $X_i = (X_{i1}, \dots, X_{im_i})$ où les variables aléatoires X_{ij} , $i = 1, 2, \dots, n$ et $j = 1, 2, \dots, m_i$, $m_i \geq 1$ sont à valeurs dans \mathbb{R}^d et sont issues de la même loi paramétrique P_θ . Le paramètre θ

Adresse e-mail : mohammed.elasri@etd.univ-avignon.fr.

¹ ANR SYSCOM, ROLSES.

appartient à Θ , avec Θ un ouvert non vide, convexe et borné de \mathbb{R}^d . Ainsi, X_{ij} est associée au j^e élément du i^e cluster. Pour chaque cluster i , nous considérons une corrélation intra-cluster identique (qui peut varier d'un cluster à un autre) : $(X_{i1}, X_{i2}) \stackrel{d}{=} (X_{ik}, X_{ik'})$ pour tout $i = 1, \dots, n$ et pour tout $k, k' = 1, \dots, m_i$, avec $k \neq k'$. Nous notons $N_n := \sum_{i=1}^n m_i$ le nombre total de variables et nous supposons que $\lim_{n \rightarrow \infty} \frac{N_n}{n} = l, l \in]0, \infty[$. Enfin, nous considérons la fonction $\rho(x, a) : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable en x pour laquelle on a $\operatorname{argmin}_{a \in \Theta} E_\theta(\rho(X_{11}, a)) = \operatorname{argmin}_{a \in \Theta} \int \rho(x, a) dP_\theta$ pour tout $\theta \in \Theta$.

Le M-estimateur pondéré associé à la fonction ρ est défini par : $\hat{\theta}_n = \operatorname{argmin}_{a \in \Theta} M_n(a)$ avec $M_n(a) = \frac{1}{N_n} \times \sum_{i=1}^n \sum_{j=1}^{m_i} w_{ij} \rho(X_{ij}, a)$ où les w_{ij} sont des poids positifs choisis par le statisticien. Si $\rho(x, a)$ est différentiable pour tout $a = (a_1, \dots, a_d)^T$ dans un voisinage de θ , alors le vecteur de ses dérivées partielles $\psi = (\frac{\partial \rho}{\partial a_1}, \dots, \frac{\partial \rho}{\partial a_d})^T$ vérifie $E_\theta(\psi(X_{11}, \theta)) = 0$. L'estimateur $\hat{\theta}_n$ correspond à la valeur de a vérifiant les relations vectorielles : $\sum_{i=1}^n \sum_{j=1}^{m_i} w_{ij} \psi(X_{ij}, a) = 0$.

Nous supposons que les hypothèses suivantes sont vérifiées pour tout $\theta \in \Theta$:

- H₁) a) Pour tout $\epsilon > 0$: $\inf_{a \in \Theta: \|a-\theta\| > \epsilon} E_\theta \rho(X_{11}, a) > E_\theta \rho(X_{11}, \theta)$,
 b) Pour tout $a \in \Theta$, $E_\theta(\rho^2(X_{11}, a)) < \infty$.
- H₂) a) $\lim_{n \rightarrow \infty} \frac{1}{N_n} \sum_{i=1}^n \sum_{j=1}^{m_i} w_{ij} = 1$,
 b) $\sum_{i \geq 1} \frac{w_i^2}{i^2} < \infty$, avec $w_i = \sum_{j=1}^{m_i} w_{ij}$.
- H₃) a) $\lim_{n \rightarrow \infty} \frac{1}{N_n} \sum_{i=1}^n \sum_{j=1}^{m_i} w_{ij}^2 = c_w$, avec $c_w < \infty$,
 b) $\lim_{n \rightarrow \infty} \frac{1}{N_n} \sum_{i=1}^n \sum_{j \neq j'}^{m_i} w_{ij} w_{ij'} C_i = C_w$, avec $C_i = E_\theta \psi^T(X_{ij}, \theta) \psi(X_{ij'}, \theta)$,
 il existe un réel $\eta > 0$ tel que :
 c) $E_\theta(\|\psi(X_{11}, \theta)\|^{2+\eta}) < \infty$,
 d) $\lim_{n \rightarrow \infty} \frac{1}{N_n} \sum_{i=1}^n w_i^{2+\eta} < \infty$.

L'hypothèse H₁a) est vérifiée si la fonction ρ est strictement convexe et si le support de la loi P_θ n'est pas concentré sur une ligne : c'est le cas de la médiane, par exemple (cf. [8]). La valeur de $C_i = E_\theta \psi^T(X_{ij}, \theta) \psi(X_{ij'}, \theta)$ ne dépend que du cluster i ; nous notons $C_i = C$ lorsque les corrélations intra-cluster sont les mêmes pour tous les clusters. Nous pouvons prendre des valeurs de pondération adaptées au cas d'étude. On peut donner, par exemple, le même poids pour tous les éléments d'un cluster donné et un poids spécifique pour chaque cluster ; dans ce cas, $w_{ij} \equiv w_i$. Si $\lim_{n \rightarrow \infty} w_n = w$, alors, d'après le théorème de Cesàro généralisé, $\lim_{n \rightarrow \infty} \frac{1}{N_n} \sum_{i=1}^n m_i w_i = w$. Si $w = 1$, les conditions H₂a) et H₃a) sont vérifiées avec $c_w = 1$. Pour ce même type de pondération, si les corrélations intra-cluster sont les mêmes pour tous les clusters, et si $\lim_{n \rightarrow \infty} m_n = m' > 1$, alors $C_w = (m' - 1)C$ et H₃d) est vérifiée. Enfin, le cas non pondéré correspond à des poids $w_{ij} \equiv 1$.

3. Résultats

Théorème 3.1. Pour tout $\theta \in \Theta$ et sous les hypothèses H₁ et H₂, si pour tout x la fonction $a \mapsto \rho(x, a)$ est $k(x)$ -lipschitzienne, avec $E_\theta(k^2(X_{11})) < \infty$, alors $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{p.s.} \theta$.

Éléments de preuve. Nous obtenons la convergence presque sûre de $\hat{\theta}_n$ vers θ en deux étapes. Tout d'abord, nous montrons la relation suivante : $\forall \eta > 0, P_\theta(\bigcap_{m \geq n} |M(\hat{\theta}_m) - M(\theta)| \leq \eta) \xrightarrow[n \rightarrow \infty]{} 1 \implies \hat{\theta}_n \xrightarrow[n \rightarrow \infty]{p.s.} \theta$ où $M(a) = E_\theta(\rho(X_{ij}, a))$. Ensuite, en utilisant la définition de $\hat{\theta}_n$ et la condition de Lipschitz sur la fonction $\rho(\cdot, a)$, nous établissons la convergence uniforme : $\sup_{a \in \Theta} |M_n(a) - M(a)| \xrightarrow[n \rightarrow \infty]{p.s.} 0$. Nous en déduisons que $M(\hat{\theta}_n) - M(\theta) \xrightarrow[n \rightarrow \infty]{p.s.} 0$, et la première étape entraîne $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{p.s.} \theta$. □

Les hypothèses du théorème 3.1 ne sont pas très contraignantes. En effet, si pour tout x , la fonction $x \mapsto \psi(x, a)$ est continue et dominée par une fonction de carré intégrable indépendante de a , alors $\rho(x, a)$ est $\sup_{a \in \Theta} \|\psi(x, a)\|$ -lipschitzienne. Dans le cas où $E_\theta \|X_{ij}\|^2 < \infty$, alors pour la moyenne empirique pondérée, la fonction ψ vérifie $\sup_{a \in \Theta} \|\psi(x, a)\| \leq 2\|x\| + 2 \sup_{a \in \Theta} \|a\|$, donc $\sup_{a \in \Theta} \|\psi(X_{11}, a)\|$ est de carré intégrable. La médiane spatiale pondérée définie par $\rho(x, a) = \|x - a\|$ et l'estimateur de Huber pondéré qui correspond à la fonction $\rho(x, a) = \frac{1}{2} \|x - a\|^2$ si $\|x - a\| \leq k$ et $\rho(x, a) = k\|x - a\| - \frac{1}{2}k^2$ si $\|x - a\| > k$ sont d'autres exemples de fonction ρ vérifiant les conditions du théorème.

Théorème 3.2. Pour tout $\theta \in \Theta$, pour tout a dans un voisinage ouvert de θ et pour tout x , nous supposons que la fonction $a \mapsto \psi(x, a)$ est deux fois différentiable, que sa dérivée seconde est continue, que les dérivées partielles secondes de $\psi(x, a)$ sont dominées par une fonction de x de carré intégrable, indépendante de a , ainsi que les éléments de la matrice $\frac{\partial \psi(X_{ij}, a)}{\partial a} /_{a=\theta}$ admettent des moments d'ordre 2 pour $i = 1, \dots, n$; $j = 1, \dots, m_i$ et que $E_\theta(\frac{\partial \psi(X_{ij}, a)}{\partial a} /_{a=\theta})$ est inversible. Alors, sous les hypothèses H₁-H₃ : $\sqrt{N_n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{L} N(0, V_\theta^{-1}(C_w B + C_w)V_\theta^{-1})$, avec $B := E_\theta \psi(X_{ij}, \theta) \psi^T(X_{ij}, \theta)$ et $V_\theta = E_\theta(\frac{\partial \psi(X_{ij}, a)}{\partial a} /_{a=\theta})$.

Éléments de preuve. Nous écrivons le développement de Taylor sur un voisinage de θ ; il existe un $\tilde{\theta}$ et un opérateur linéaire D tels que : $\psi(x, a) = \psi(x, \theta) + \dot{\psi}(x, \theta)^T(a - \theta) + \frac{1}{2}D(a - \theta)\ddot{\psi}(x, \tilde{\theta})(a - \theta)$. Cette équation nous permet d'établir l'existence d'une statistique $T_n(a)$ ainsi que ses dérivées première $\dot{T}_n(a)$ et seconde $\ddot{T}_n(a)$, ainsi que d'un vecteur $\tilde{\theta}_n$ satisfaisant : $-\{\dot{T}_n(\theta) + \frac{1}{2}D(\hat{\theta}_n - \theta)\dot{T}_n(\tilde{\theta}_n)\}^{-1}\sqrt{N_n}T_n(\theta) = \sqrt{N_n}(\hat{\theta}_n - \theta)$.

Nous vérifions les conditions du théorème de Lindeberg, Serfling [9, pp. 30–31], ce qui implique la normalité asymptotique de $\sqrt{N_n}T_n(\theta)$. Enfin, la convergence presque sûre de $\{\dot{T}_n(\theta) + \frac{1}{2}D(\hat{\theta}_n - \theta)\dot{T}_n(\tilde{\theta}_n)\}^{-1}$ vers une matrice inversible entraîne la normalité asymptotique de $\sqrt{N_n}(\hat{\theta}_n - \theta)$ par le théorème de Slutsky.

Les hypothèses du théorème 3.2 peuvent paraître contraignantes ; cependant, elles sont satisfaites par des estimateurs classiques. Par exemple, l'estimateur du maximum de vraisemblance pondéré entre dans le cadre du théorème si la fonction de densité $f_\theta(x) > 0$, avec $\psi(x, \theta) = \frac{\dot{f}_\theta(x)}{f_\theta(x)}$ et $\dot{\psi}(x, \theta) = \frac{\ddot{f}_\theta(x)}{f_\theta(x)} - \frac{\dot{f}_\theta(x)^T \dot{f}_\theta(x)}{f_\theta(x)^2}$. En supposant qu'on puisse interchanger les signes de dérivée et d'intégrale, nous obtenons $E_\theta \psi(X_{ij}, \theta) = 0$ et $E_\theta \dot{\psi}(X_{ij}, \theta) = -E_\theta \frac{\dot{f}_\theta(X_{ij})^T \dot{f}_\theta(X_{ij})}{f_\theta(X_{ij})^2} = -I_\theta$ et donc $\sqrt{N_n}(\hat{\theta}_n - \theta)$ converge vers une loi normale centrée de matrice de variance-covariance $c_w I_\theta^{-1} + I_\theta^{-1} C_w I_\theta^{-1}$, où I_θ est l'information de Fisher en θ . Sous les hypothèses H_1 – H_3 , la moyenne empirique intègre parfaitement le théorème. En affaiblissant la condition de régularité sur la fonction ρ par une condition de Lipschitz (travail en cours), on pourrait intégrer des estimateurs pertinents tels que la médiane spatiale (estimateur robuste avec un point de rupture maximal). Dans le cas particulier des variables iid (où les variables intra-clusters sont indépendantes), les résultats précédents restent valables. Nous obtenons ainsi la convergence presque sûre et la normalité asymptotique des estimateurs avec la matrice de variance-covariance $c_w V_\theta^{-1} B V_\theta^{-1}$.

Références

- [1] P. Huber, Robust estimation of a location parameter, *Ann. Math. Statist.* 35 (1) (1964) 73–101.
- [2] P. Huber, *Robust Statistics*, John Wiley & Sons Inc., New York, 1981.
- [3] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stahel, *Robust Statistics*, John Wiley & Sons Inc., New York, 1986.
- [4] A. Ruiz-Gazen, Robust statistics: a functional approach, *Ann. Inst. Statist. Univ. Paris* 56 (2–3) (2012) 49–64.
- [5] A.W. Van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge, 2000.
- [6] P. Nevalainen, D. Larocque, H. Oja, A weighted spatial median for clustered data, *Statist. Methods Appl., J. Ital. Statist. Soc.* 15 (3) (2006) 355–379.
- [7] P. Nevalainen, D. Larocque, H. Oja, On the multivariate spatial median for clustered data, *Canad. J. Statist.* 35 (2) (2007) 215–231.
- [8] P. Milasevic, Uniqueness of the spatial median, *Ann. Statist.* 15 (3) (1987) 1332–1333.
- [9] R. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons Inc., New York, 1980.