Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 344 (2007) 265–270
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistics/Probability Theory

M-processes and applications

Fateh Chebana

Université du Québec, INRS-ETE, 490, de la Couronne, G1K 9A9 Québec QC, Canada

Received 25 October 2004; accepted after revision 21 November 2006

Available online 26 January 2007

Presented by Paul Deheuvels

Abstract

We consider a class of M-estimators indexed by the criterion function ψ which belongs to a class of functions $\mathcal{F}$. Then, we obtain a process indexed by the class $\mathcal{F}$. The convergence in probability of these processes is studied uniformly on $\mathcal{F}$ when the parameter to be estimated is the same for all functions ψ. We also establish their weak convergence towards a Gaussian process. We illustrate these results on a location estimation example. ***To cite this article: F. Chebana, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

Résumé

M*-processus et applications.** Nous considérons une classe de M-estimateurs en faisant varier la fonction critère ψ dans une classe de fonctions $\mathcal{F}$. Nous obtenons ainsi un processus indexé par cette classe $\mathcal{F}$. Dans le cas où le paramètre à estimer est le même pour toutes les fonctions ψ, nous étudions la convergence en probabilité de ce processus uniformément sur la classe $\mathcal{F}$. Nous établissons également sa convergence faible vers un processus gaussien. Nous illustrons ces résultats sur un exemple d'estimation de paramètre de position. ***Pour citer cet article : F. Chebana, C. R. Acad. Sci. Paris, Ser. I 344 (2007).

Version française abrégée

Soit $X_1, \ldots, X_n$ une suite de variables aléatoires (v.a.) indépendantes et identiquement distribuées (i.i.d.) de loi P dépendant d'un paramètre inconnu $\theta \in \Theta \subseteq \mathbb{R}$ avec θ_0 comme vraie valeur. Soient ϕ et $\psi : \mathbb{R} \times \Theta \to \mathbb{R}$ deux fonctions connues. Notons par $Pf := \int f(x)\,\mathrm{d}P(x)$ et $\mathbb{P}_n f := \frac{1}{n}\sum_{i=1}^n f(X_i)$.

Rappelons qu'un M-estimateur de θ_0 est l'argument, s'il existe, du maximum de la fonction

$$\theta \mapsto \mathbb{P}_n \phi(\cdot, \theta)$$

de même un Z-estimateur $\hat{\theta}_n$ de θ est la solution, en θ, de l'équation

$$\mathbb{P}_n \psi(\cdot, \theta) = 0.$$

Lorsque la fonction ϕ définissant le M-estimateur est dérivable, le M-estimateur et le Z-estimateur coïncident. Dans la suite les deux types d'estimateurs seront confondus et appelés M-estimateurs, sauf indication contraire.

E-mail address: Fateh_chebana@ete.inrs.ca.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2006.11.029

Le M-estimateur ainsi obtenu dépend de la fonction ψ ; on le note alors $\hat{\theta}_{n,\psi}$. En faisant varier la fonction ψ dans une classe $\mathcal{F}$ on obtient une classe d'estimateurs indexés par ψ ; on la note $\{\hat{\theta}_{n,\psi};\ \psi \in \mathcal{F}\}$, où $\hat{\theta}_{n,\psi}$ est un M-estimateur de θ associé à la fonction ψ. La classe $\{\hat{\theta}_{n,\psi};\ \psi \in \mathcal{F}\}$ de M-estimateurs est appelée M-processus.

Pour une fonction ψ fixée, les propriétés asymptotiques de $\hat{\theta}_{n,\psi}$ telles la convergence en probabilité, presque sûre et la convergence faible ont été étudiées ; voir par exemple Serfling [4] et les publications recentes de van de Geer [6] et [5].

Plusieurs auteurs ont étudié la question du choix d'une meilleure fonction ψ selon des procédures et critères différents. Entre autres, Huber [2] utilise la procédure minimax sur la variance limite afin d'obtenir un estimateur robuste. On trouve un résultat récent de Fraiman et al. [1] qui optimisent une fonction reliant à la fois la variance et le biais asymptotiques des estimateurs.

Dans le cas où θ_0 est solution commune de $\theta \mapsto P\psi(\cdot,\theta) = 0$ pour toutes les fonctions ψ dans une classe $\mathcal{F}$, on considère le M-processus centré

$$\{\hat{\theta}_{n,\psi} - \theta_0;\ \psi \in \mathcal{F}\}.$$

A titre d'exemple, cela peut être obtenu si la distribution de l'échantillon est symmétrique autour de θ_0 et les fonctions ψ sont anti-symétriques.

L'objectif de cette Note est de fournir un outil permettant de construire des estimateurs de θ_0, par application des fonctionnelles sur le M-processus, en utilisant la contribution de toute la classe $\mathcal{F}$. Les propriétés asymptotiques de ces nouveaux estimateurs peuvent être obtenues par la convergence en probabilité uniformément sur la classe $\mathcal{F}$ ainsi que la convergence faible des processus $\{\hat{\theta}_{n,\psi} - \theta_0;\ \psi \in \mathcal{F}\}$.

1. Introduction

Let $X_1, \ldots, X_n$ be a sequence of independent identically distributed (iid.) random variables (r.v.) with distribution P which depends on an unknown parameter $\theta \in \Theta \subseteq \mathbb{R}$ with true value θ_0. Let ψ and $\phi : \mathbb{R} \times \Theta \to \mathbb{R}$ be two known functions. In the following we use the empirical processus notation $Pf = \int f(x)\,\mathrm{d}P(x)$ and $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$. Recall that $\hat{\theta}_n$ is a M-estimator of θ_0 if it maximizes the function

$$\theta \mapsto \mathbb{P}_n \phi(\cdot,\theta), \tag{1}$$

and is a Z-estimator if it is solution, in θ, of the equation

$$\mathbb{P}_n \psi(\cdot,\theta) = 0. \tag{2}$$

Often the maximizing value is sought by setting a derivative equal to zero. Hence, when the function ϕ which defines the M-estimator is differentiable then the M-estimator and Z-estimator coincide. In the following the two type of estimators will be called M-estimator.

The estimator $\hat{\theta}_n$ deduced from $\mathbb{P}_n\psi(\cdot,\theta) = 0$ depends on the function ψ; we denote it by $\hat{\theta}_{n,\psi}$. Varying the function ψ over a class of functions $\mathcal{F}$ we obtain a class of estimators indexed by ψ; we denote this class by $\{\hat{\theta}_{n,\psi};\ \psi \in \mathcal{F}\}$, and then we call M-process this collection of M-estimators (similarly $\{\hat{\theta}_{n,\phi};\ \phi \in \mathcal{M}\}$ for a class $\mathcal{M}$ of functions).

For fixed function ψ, the asymptotic properties of the estimator $\hat{\theta}_{n,\psi}$ such as the convergence in probability, almost sure convergence and weak convergence have been already studied; see for example Serfling [4] and the recent publications by van de Geer [5] and [6].

Many authors studied the problem of choosing the best function ψ according to different procedures and criterions. Among others, Huber [2] uses a minimax procedure in order to get a robust estimator. Recently, Fraiman et al. [1] optimized a function which links the asymptotic variance and the bias of the estimators.

In the case where θ_0 is a common solution of $P\psi(\cdot,\theta) = 0$ for all functions ψ in a class $\mathcal{F}$, we consider the centered M-process

$$\{\hat{\theta}_{n,\psi} - \theta_0;\ \psi \in \mathcal{F}\}. \tag{3}$$

This holds, for instance, when the sample has a symmetric distribution around θ_0 and when ψ is anti-symmetric (see Example 2.4 below for details).

Our goal is to provide a tool to construct estimators of θ_0, by applying real valued functionals on the M-process, using the contribution of the whole class $\mathcal{F}$. Asymptotic properties of these new estimators can be obtained by establishing the uniform convergence in probability on the class $\mathcal{F}$ as well as the weak convergence of the M-process $\{\hat{\theta}_{n,\psi} - \theta_0;\ \psi \in \mathcal{F}\}$.

2. Results

Before stating our results, we introduce some notation. Denote by $\dot{\psi}(x,\theta) := \frac{\partial}{\partial\theta}\psi(x,\theta)$ and $\ddot{\psi}(x,\theta) := \frac{\partial^2}{\partial\theta^2}\psi(x,\theta)$. Set $\mathcal{F}_{\theta_0}$ the class of functions $\psi(x,\theta)$ that belongs to $\mathcal{F}$ when $\theta = \theta_0$, i.e. $\mathcal{F}_{\theta_0} := \{\psi(\cdot,\theta_0)\colon\ \psi(\cdot,\cdot) \in \mathcal{F}\}$.

2.1. Uniform convergence in probability

The following result allows us to get the consistency of the estimators obtained via continuous functionals on the M-process. It concerns with M-estimators as defined in (1) when the function ϕ is taken in a class $\mathcal{M}$.

Theorem 2.1. *Assume that for all* $\epsilon > 0$

$$\inf_{\phi\in\mathcal{M}} \inf_{\theta\colon |\theta-\theta_0|>\epsilon} P\phi(\cdot,\theta_0) - P\phi(\cdot,\theta) > 0, \tag{4}$$

and

$$\sup_{\phi\in\mathcal{M}} \sup_{\theta\in\Theta} \left|\mathbb{P}_n\phi(\cdot,\theta) - P\phi(\cdot,\theta)\right| \xrightarrow{P} 0. \tag{5}$$

Then, we have

$$\sup_{\phi\in\mathcal{M}} |\hat{\theta}_{n,\phi} - \theta_0| \xrightarrow{P} 0. \tag{6}$$

A similar result for the M-estimators defined as a zero of (2) is given in the following theorem. Its proof can be done using the same arguments as for Theorem 2.1 with the appropriate definition of the estimator.

Theorem 2.2. *Assume that for all* $\epsilon > 0$

$$\inf_{\psi\in\mathcal{F}} \inf_{\theta\colon |\theta-\theta_0|>\epsilon} \left|P\psi(\cdot,\theta)\right| > 0, \tag{7}$$

and

$$\sup_{\psi\in\mathcal{F}} \sup_{\theta\in\Theta} \left|\mathbb{P}_n\psi(\cdot,\theta) - P\psi(\cdot,\theta)\right| \xrightarrow{P} 0. \tag{8}$$

Then, we have

$$\sup_{\psi\in\mathcal{F}} |\hat{\theta}_{n,\psi} - \theta_0| \xrightarrow{P} 0. \tag{9}$$

In Theorem 2.2, when the parameter θ is a vector, the absolute value can be replaced by a norm.

2.2. Weak convergence

In this part we formulate the weak convergence of M-processes result. Various proofs have been proposed to establish weak convergence of M-estimators (see e.g. van der Vaart [7] and van de Geer [6]). The weak convergence of the processes $\{\hat{\theta}_{n,\psi} - \theta_0;\ \psi \in \mathcal{F}\}$ uses the so-called classical hypotheses and the empirical process theory which is much involved in the study of the asymptotic properties of M-estimators (see van der Vaart [7], Section 5.6).

Hypotheses.

(A1) For any function ψ belonging to $\mathcal{F}$, assume that $P\psi^2(\cdot,\theta_0) < \infty$, where θ_0 is a common solution of $P\psi(\cdot,\theta) = 0$ independently on ψ.
(A2) For all x, the function $\theta \mapsto \psi(x,\theta)$ is twice continuously differentiable. Moreover:
  (i) $P\dot{\psi}(\cdot,\theta_0) < \infty$ and $P\dot{\psi}(\cdot,\theta_0) \neq 0$.
  (ii) There exists a P-integrable function G such that $|\ddot{\psi}(\cdot,\theta)| \leqslant G(\cdot)$ for any function $\psi(\cdot,\theta)$ in $\mathcal{F}$ and θ in a neighborhood of θ_0.
(A3) $\sup_{\psi\in\mathcal{F}} |\hat{\theta}_{n,\psi} - \theta_0| = o_P(1)$.
(B1) The class $\mathcal{F}_{\theta_0}$ is P-Donsker.
(B2) The class $\dot{\mathcal{F}}_{\theta_0}$ is P-Glivenko–Cantelli, where $\dot{\mathcal{F}}_{\theta_0} := \{\dot{\psi}(\cdot,\theta_0)\colon\ \psi(\cdot,\cdot) \in \mathcal{F}\}$.

Theorem 2.3. *Under the hypotheses* (A1)–(A3) *and* B1, B2*, we have*

$$\left\{\sqrt{n}(\hat{\theta}_{n,\psi} - \theta_0);\ \psi \in \mathcal{F}\right\} \xrightarrow{\mathcal{D}} \left\{\mathbb{G}_P(\psi);\ \psi \in \mathcal{F}\right\} \tag{10}$$

where $\mathbb{G}_P$ *is a centered Gaussian process with covariance operator*

$$\sigma(i,j) := \frac{P\psi_i(\cdot,\theta_0)\psi_j(\cdot,\theta_0)}{P\dot{\psi}_i(\cdot,\theta_0)P\dot{\psi}_j(\cdot,\theta_0)}, \tag{11}$$

for $\psi_i(\cdot,\theta_0)$ *and* $\psi_j(\cdot,\theta_0)$ *in* $\mathcal{F}_{\theta_0}$.

2.3. Comments on the hypotheses

In the case of single function ψ, the condition (7) represents an identifiability condition of the model or equivalently a condition on the separability of the solution (the estimator). In the present case this condition becomes uniform on the class $\mathcal{F}$. The hypothesis (8) turns out to assume that the class $\mathcal{F}$ is P-Glivenko–Cantelli.

The hypotheses (B1) and (B2) are also considered by Murphy and van der Vaart [3] (Theorem 3.1). Furthermore, since the classes $\mathcal{F}$ and $\dot{\mathcal{F}}$ are linked by a linear transformation, the following corollary gives the relationship between (B1) and (B2), to this end we define the following hypothesis:

(B′2) The class $\dot{\mathcal{F}}_{\theta_0}$ is P-Donsker.

Corollary 2.4. (i) *Suppose that there exists a positive constant K such that*

$$\left|\dot{\psi}_1(x,\theta_0) - \dot{\psi}_2(x,\theta_0)\right| \leqslant K\left|\psi_1(x,\theta_0) - \psi_2(x,\theta_0)\right|$$

for all x and any functions $\psi_1(\cdot,\theta_0)$ and $\psi_2(\cdot,\theta_0)$ in $\mathcal{F}_{\theta_0}$. Assume furthermore that there exists a function $\psi_0(\cdot,\theta_0)$ in $\mathcal{F}_{\theta_0}$ such that $\dot{\psi}_0(\cdot,\theta_0)$ is square integrable. Then the hypothesis (B1) *implies the hypothesis* (B2).

(ii) *If there exists a positive constant L such that*

$$\left|\psi_1(x,\theta_0) - \psi_2(x,\theta_0)\right| \leqslant L\left|\dot{\psi}_1(x,\theta_0) - \dot{\psi}_2(x,\theta_0)\right|$$

for all x and any functions $\psi_1(\cdot,\theta_0)$ and $\psi_2(\cdot,\theta_0)$ in $\mathcal{F}_{\theta_0}$, then the hypothesis (B′2) *implies both hypotheses* (B1) *and* (B2).

Remark. To obtain the weak convergence of the M-process, it is possible to assume only differentiability of $P\psi(\cdot,\theta)$ together with the Lipschitzian property of the functions $\psi(\cdot,\theta)$ rather than differentiability of $\psi(\cdot,\theta)$ (see for instance Theorem 5.21 in van der Vaart [7, p. 52]). Since the distribution P is unknown, we prefer to keep the differentiability on $\psi(\cdot,\theta)$ for a statistical point of view.

2.4. Example

The main object of this example is: first to show that all the assumptions of our results can be checked, second, to illustrate how to use the results in order to construct estimators and get their asymptotic properties.

Let $X_1, \ldots, X_n$ be a sample from a distribution P which depends on a parameter θ. Assume that this distribution is symmetric around θ_0 i.e. $P(X \geqslant x + \theta_0) = P(X \leqslant \theta_0 - x)$ for all x. Let $\tilde{\psi}$ be a regularized Huber's function given by (see Fig. 1)

$$\tilde{\psi}(u) = \operatorname{sign}(u) \begin{cases} |u|, & \text{if } |u| \leqslant 0.8, \\ p_4(|u|), & \text{if } 0.8 < |u| \leqslant 1, \\ p_4(1), & \text{if } |u| > 1, \end{cases}$$

where $p_4(u) = 38.4 - 175.0u + 300.0u^2 - 225.0u^3 + 62.5u^4$ (see Fraiman et al. [1]). It is clear that the function $\tilde{\psi}$ is twice differentiable. Now, we define the class $\mathcal{F} = \{\psi(u, \theta, t) = \tilde{\psi}(\frac{u-\theta}{t});\ t \in I\}$, where $I \subset \mathbb{R}_+^*$ is a compact subset. Thus, we obtain the M-process $\{\hat{\theta}_{n,t} - \theta_0;\ t \in I\}$, which is of the form (3).

The hypotheses of Theorems 2.2 and 2.3 are fulfilled. Consequently, the process $\{\hat{\theta}_{n,t} - \theta_0;\ t \in I\}$ converges uniformly in probability to zero and converges weakly towards a centered Gaussian process with covariance operator given by $\sigma(s,t) = st P\psi(\cdot, \theta_0, s)\psi(\cdot, \theta_0, t)/(P\psi'(\cdot, \theta_0, s)\mathbb{1}_{(\theta_0 - s, \theta_0 + s)} P\psi'(\cdot, \theta_0, t)\mathbb{1}_{(\theta_0 - t, \theta_0 + t)})$, where ψ' is the derivative of ψ w.r.t. x and $\hat{\theta}_{n,t}$ is such that $\frac{1}{n}\sum_{i=1}^{n} \psi(X_i, \hat{\theta}_{n,t}, t) = 0$.

Note that the parameter to be estimated is the same for all t, because of the distribution symmetry and the $\tilde{\psi}$ anti-symmetry.

We propose the following 'fitted' or 'weighted' mean estimator of θ_0: $\hat{\theta}_n := \frac{1}{\nu(I)} \int_I \hat{\theta}_{n,t}\, \mathrm{d}\nu(t)$, for a nonnegative and finite measure ν on the real line. Here we have taken, as a functional acting on the M-process, the linear transformation for its simplicity (we can define similar estimators by taking any regular functional e.g. supremum or infimum).

This estimator can be adjusted to be near the median (very robust) or near the mean (not robust) estimators, or an intermediate estimator between them by appropriate choices of the measure ν.

It is clear that $\hat{\theta}_n$ is a consistent estimator of θ_0, and it converges weakly to a normal distribution, i.e. $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2)$, where $\tau^2 := \frac{1}{(\nu(I))^2} \int_I \int_I \sigma(s,t)\, \mathrm{d}\nu(s)\, \mathrm{d}\nu(t)$. Since σ is a non-negative definite matrix, the limiting variance τ^2 is non-negative for any ν.

Probability measures are particular choices of the weight ν. The following illustration gives the limiting variance τ^2 when the density f is uniform on the range $[-1, 1]$. To avoid singularities on the boundary of the interval I, we take $I = [0.03, 0.97]$. The weight measure ν is taken to be a probability with density Beta which is given for positive constants a and b by $B(x, a, b) = \frac{(1-x)^{b-1} x^{a-1}}{\beta(a,b)}$, $\quad 0 \leqslant x \leqslant 1$; where $\beta(a,b)$ is the Beta function.

The choice of the Beta distribution is due to its particular shape, since it has some forms ranging from graphics like $\bigcup$ to $\bigcap$ following the values of its parameters a and b. For example to give an important weight to the mean we take small a and big b; similar for the median we take small a, very small b, big a and small b; and for less importance to both median and mean but more importance to the estimators between them it is possible with big values of both a and b.

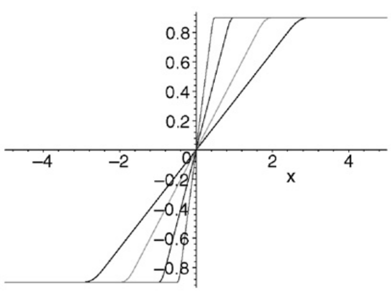

Fig. 1. Some functions $\psi(x, 0, t)$ in the class $\mathcal{F}$.

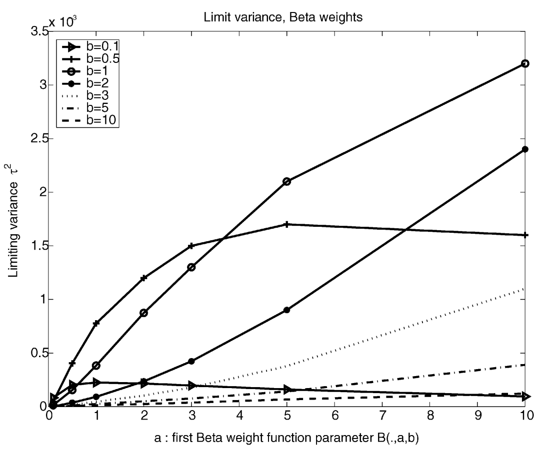

Fig. 2. Limit variance, Beta weights.

In Fig. 2, we can see that the limiting variances are small (of order 10^{-3}) and the smallest variance corresponds to small value of a and any value of b, or to the tail values of b ($b = 0.1, 5$ and 10) for all values of a.

Acknowledgements

I would like to thank Professors Michel Broniatowski, Yury A. Kutoyants, Naâmane Laïb and the anonymous referee for their helpful comments and suggestions.

References

[1] R. Fraiman, V.J. Yohai, R.H. Zamar, Optimal robust M-estimates of location, Ann. Statist. 29 (1) (2001) 194–223.

[2] P.J. Huber, Robust Statistics, in: Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1981.

[3] S.A. Murphy, A.W. van der Vaart, On profile likelihood, J. Amer. Statist. Assoc. 95 (450) (2000) 449–485. With comments and a rejoinder by the authors.

[4] R.J. Serfling, Approximation Theorems of Mathematical Statistics, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1980.

[5] S. van de Geer, Empirical processes in M-estimation, Technical report, Cowles Workshop, Yales University, http://cowles.econ.yale.edu/conferences/wkshp/lec/geer.pdf, 2003.

[6] S.A. van de Geer, Applications of Empirical Process Theory, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2000.

[7] A.W. van der Vaart, Asymptotic Statistics, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 1998.