Journal of Multivariate Analysis **74**, 193–221 (2000)
doi:10.1006/jmva.1999.1886, available online at http://www.idealibrary.com on IDEAL®

M-Estimators Converging to a Stable Limit

Miguel A. Arcones[1]

Department of Mathematical Sciences, State University of New York

Received August 1, 1996

We discuss the asymptotic linearization of multivariate M-estimators, when the limit distribution is stable. We consider two different types of kernels: VC and bracketing. When applied to the case of normal limits, our work improves the known results to obtain the limit distribution of M-estimators. We give weak conditions for the asymptotic normality of M-estimators over differentiable kernels. To obtain these results, we present an inequality on empirical processes satisfying a bracketing condition with respect to a norm smaller than the L_2 norm.

AMS 1991 subject classifications: primary 62F12, secondary 62E20.
Key words and phrases: M-estimators, delta method, stable distributions.

1. INTRODUCTION

We discuss the convergence of M-estimators to a stable (possibly normal) limit distribution. Huber (1964) introduced M-estimators as a way to obtain more robust estimators. Let $(S, \mathscr{S}, P)$ be a probability space and let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with values in S. Let X be a copy of X_1. Let Θ be a subset of $\mathbb{R}^d$. Let g: $S \times \Theta \to \mathbb{R}$ be a function such that $g(\cdot, \theta)$: $S \to \mathbb{R}$ is measurable for each $\theta \in \Theta$. Suppose that we want to estimate a parameter $\theta_0 \in \Theta$ characterized by $E[g(X, \theta) - g(X, \theta_0)] > 0$ for each $\theta \neq \theta_0$. An M-estimator $\hat{\theta}_n$ over the kernel $g(x, \theta)$ is a random variable $\hat{\theta}_n = \hat{\theta}_n(X_1, ..., X_n)$ satisfying

$$n^{-1} \sum_{i=1}^{n} g(X_i, \hat{\theta}_n) \simeq \inf_{\theta \in \Theta} n^{-1} \sum_{i=1}^{n} g(X_i, \theta). \tag{1.1}$$

Another type of M-estimator is $\hat{\theta}_n$ defined by

$$n^{-1} \sum_{i=1}^{n} h(X_i, \hat{\theta}_n) \simeq 0, \tag{1.2}$$

[1] E-mail: arcones@math.binghamton.edu. Web: http://math.binghamton.edu/arcones/.

0047-259X/00 $35.00

where $h(\cdot, \theta)$: $S \to \mathbb{R}^d$ is a measurable function for each $\theta \in \Theta$. Here, $\hat{\theta}_n$ is estimating a value θ_0 characterized by $E[h(X, \theta_0)] = 0$. In general, the two methods can give different estimates (see, for example, Bai *et al.*, 1990).

There are many estimators that fall in the previous setup. It is well known that maximum likelihood estimators are M-estimators. In this case $g(x, \theta) = -\log f(x, \theta)$, where $f(x, \theta)$, $\theta \in \Theta$, is a family of densities. Cramér (1946) obtained the asymptotic normality of maximum likelihood estimators assuming strong differentiability conditions. He even need to assume a third derivative. Weaker differentiability assumptions for the asymptotic normality of maximum likelihood estimators (or M-estimators) were imposed by Daniels (1961), Huber (1964, 1967), Le Cam (1970), Ibragimov and Has'minskii (1981), and Serfling (1981).

We will use the notation in empirical processes. For instance, we write

$$Pf = E[f(X)] \qquad \text{and} \qquad P_n f = n^{-1} \sum_{i=1}^{n} f(X_i),$$

where f is function on S. $\{\varepsilon_j\}$ will denote a sequence of i.i.d.r.v.'s with $\Pr\{\varepsilon_j = 1\} = \Pr\{\varepsilon_j = -1\} = 1/2$, which is independent of the sequence $\{X_j\}$. c will denote a constant which may vary from occurrence to occurrence. Given a vector v, $|v|$ will denote the Euclidean norm. Given a $d \times d$ matrix A, we define the following norm $\|A\| := \sup_{|b| \leqslant 1} |b'Ab|$.

To obtain the asymptotics of M-estimators we apply the delta method (Taylor expansions). In the case in (1.1), under regularity conditions, there are a function ϕ and a positive definite symmetric matrix V such that

$$a_n^2(P_n - P)(g(\cdot, \theta + a_n^{-1}\theta) - g(\cdot, \theta_0) - a_n^{-1}\theta'\phi(\cdot)) \xrightarrow{\Pr} 0,$$
$$E[g(X, \theta) - g(X, \theta_0)] = (\theta - \theta_0)' V(\theta - \theta_0) + o(|\theta - \theta_0|^2),$$

and

$$a_n(\hat{\theta}_n - \theta_0) + 2^{-1}V^{-1}a_n(P_n - P)\,\phi \xrightarrow{\Pr} 0. \tag{1.3}$$

If $a_n(P_n - P)\,\phi$ converges in distribution and (1.3) holds, then $a_n(\hat{\theta}_n - \theta_0)$ also converges in distribution. Given the kernel $g(x, \theta)$, $\phi(x)$ is chosen so that $(\theta - \theta_0)'\,\phi(x)$ is the linear part in the Taylor expansion of $g(x, \theta)$. a_n is chosen so that $a_n(P_n - P)\,\phi$ converges in distribution.

For the M-estimators defined as in (1.2), under some regularity conditions, we have that

$$a_n(\hat{\theta}_n - \theta_0) + (H'(\theta_0))^{-1}\, a_n(P_n - P)\, h(\cdot, \theta_0) \xrightarrow{\Pr} 0. \tag{1.4}$$

If either (1.3) or (1.4) holds, the rate of convergence of the M-estimator is determined by the influence function (which is either $V^{-1}\phi(x)$ or

$(H'(\theta_0))^{-1} h(\cdot, \theta_0))$. If the influence function has heavy tails, the rate of convergence of the M-estimator is of an order of magnitude smaller than $n^{1/2}$. If $g(x, \theta)$ has a bounded influence function (and some regularity conditions hold), the M-estimator converges with rate $n^{1/2}$ for any possible sample distribution. If the influence function is not bounded, we have distributions whose corresponding M-estimator converges with a rate slower than $n^{1/2}$. Hence, M-estimators over kernel with a bounded influence function are preferable.

In Section 2, we present general principles to obtain (1.3) and (1.4).

In Section 3, we see how the conditions obtained in Section 2 are satisfied for VC subgraph classes of functions. Asymptotic normality of M-estimators using that a certain class of functions is a VC subgraph class was first obtained by Pollard (1985). He considered the case when $\{|\theta-\theta_0|^{-1} (g(x, \theta)-g(x, \theta_0)-(\theta-\theta_0)' \phi(x)) : |\theta-\theta_0| \leqslant \delta\}$ is a VC subgraph class of functions, for some $\delta>0$. Here, we consider the case when for some $\delta_0>0$, $\{g(x, \theta)-g(x, \theta_0) : |\theta-\theta_0| \leqslant \delta_0\}$ is a VC subgraph class of functions. It is easier to check that these classes are VC subgraph classes than the classes of functions considered by Pollard.

In Section 4, we give some bracketing conditions in the class of functions $\{g(x, \theta): \theta \in \Theta\}$ so that (1.3) or (1.4) holds. The asymptotic normality of M-estimators under bracketing conditions has been considered by several authors (see, for example, Huber, 1967; Pollard, 1985; Hoffmann–Jørgensen, 1994; and van der Vaart and Wellner, 1996). We present an inequality on the tail of empirical processes by measuring the size of the brackets with respect to a norm smaller than the L_2 norm. Usually the brackets of size 2^{-k} are functions Δ_k such that $E[(\sum_{j=1}^n \varepsilon_j \Delta_k(X_j))^2] = nE[\Delta_k^2(X_j)] \leqslant 2^{-2k}$; we impose the weaker condition $nE[(2^k \Delta_k(X_1)) \wedge (2^{2k} \Delta_k^2(X_1))] \leqslant 1$. The norm determined by the previous inequality is a natural norm to consider. Klass (1980) proved that if $nE[(\xi_1/K(n)) \wedge (\xi_1^2/K^2(n))]=1$, where $\{\xi_j\}$ is a sequence of i.i.d.r.v.'s with mean zero, and $K(n)$ is a real number, then $cK(n) \leqslant E[|\sum_{j=1}^n \xi_j|] \leqslant 2K(n)$, where c is a universal positive constant. We apply our results to obtain the asymptotics of M-estimators under very weak differentiability conditions.

In Section 5, we apply the previous results to location estimators and to the k-means. The location estimators appear when $g(x, \theta)=\rho(x-\theta)$, where ρ is a function. The asymptotic normality of these estimators has been considered by many authors (see, for example, Huber, 1964, 1967; Serfling, 1981). We will see that, under certain conditions, these estimators converge to a stable limit. The k clusters means is the M-estimator over the kernel $g(x, \theta)=\min_{1 \leqslant i \leqslant k} |x-\theta^{(i)}|^2$. Hartigan (1978) proved the asymptotic normality of these M-estimators. Under certain conditions, these estimators can converge to a stable limit. A better estimator (k-medians) of k clusters is the M-estimator over $g(x, \theta)=\min_{1 \leqslant i \leqslant k} |x-\theta^{(i)}|$. The k-medians estimator is more robust than the k-means. We will see that the k-medians is asymptotically normal even for heavy tailed distributions.

2. A GENERAL APPROACH TO THE LINEARIZATION OF M-ESTIMATORS

Our approach to obtain the asymptotics of M-estimator is to do Taylor expansions. In this section, we give general conditions to get (1.3) and (1.4). The next theorem follows from Theorem 3 in Arcones (1998).

THEOREM 1. *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with values in S. Let Θ be a subset of $\mathbb{R}^d$. Let θ_0 be in the interior of Θ. Let $g: S \times \Theta \to \mathbb{R}$ be a function such that $g(\cdot, \theta): S \to \mathbb{R}$ is measurable for each $\theta \in \Theta$. Let $\phi: S \to \mathbb{R}^d$. Let $\{a_n\}$ be a sequence of positive numbers which converges to infinity such that $\sup_{n \geqslant 1} n^{-1} a_n^2 < \infty$. Let $\{\hat{\theta}_n = \hat{\theta}_n(X_1, ..., X_n)\}$ be a sequence of Θ-valued random variables. Suppose that*:

(A.1) $\hat{\theta}_n \xrightarrow{\Pr} \theta_0$ and $n^{-1} \sum_{i=1}^n g(X_i, \hat{\theta}_n) \leqslant \inf_{\theta \in \Theta} n^{-1} \sum_{i=1}^n g(X_i, \theta) + o_{\Pr}(a_n^{-2})$.

(A.2) *There is a positive definite symmetric $d \times d$ matrix V such that*

$$E[g(X, \theta) - g(X, \theta_0)] = (\theta - \theta_0)' V(\theta - \theta_0) + o(|\theta - \theta_0|^2),$$

as $\theta \to \theta_0$.

(A.3) $a_n(P_n - P)\phi = O_{\Pr}(1)$.

(A.4) *For each $0 < M < \infty$,*

$$\sup_{|\theta| \leqslant M a_n^{-1}} a_n^2 |(P_n - P) r(\cdot, \theta)| \xrightarrow{\Pr} 0,$$

where

$$r(x, \theta) = g(x, \theta_0 + \theta) - g(x, \theta_0) - \theta' \phi(x).$$

(A.5) *For each $\tau > 0$, there exists a $\delta > 0$ such that*

$$\lim_{M \to \infty} \limsup_{n \to \infty} \Pr\left\{ \sup_{|\theta - \theta_0| \leqslant \delta} \frac{a_n^2 |(P_n - P)(g(\cdot, \theta_0 + \theta) - g(\cdot, \theta_0))|}{\tau a_n^2 |\theta - \theta_0|^2 + M} \geqslant 1 \right\} = 0.$$

Then,

$$a_n(\hat{\theta}_n - \theta_0) + 2^{-1} a_n(P_n - P) V^{-1} \phi \xrightarrow{\Pr} 0. \tag{2.1}$$

Under (A.3), (A.5) is equivalent to the following: for each $\tau > 0$, there exists a $\delta > 0$ such that

$$\lim_{M \to \infty} \limsup_{n \to \infty} \Pr \left\{ \sup_{|\theta| \leqslant \delta} \frac{a_n^2 \, |(P_n - P)\, r(\cdot, \theta)|}{\tau a_n^2 \, |\theta|^2 + M} \geqslant 1 \right\} = 0.$$

We assume the existence and consistency of M-estimators, which could be dealt with by different methods. Conditions (A.2)–(A.3) are elementary conditions. In the next sections, we will use empirical processes to obtain manageable conditions which imply (A.4) and (A.5).

We present the following theorem for the M-estimators in (1.2):

THEOREM 2. *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with values in S. Let Θ be a subset of $\mathbb{R}^d$. Let h: $S \times \Theta \to \mathbb{R}^d$ be a function such that $h(\cdot, \theta)$: $S \to \mathbb{R}^d$ is measurable for each $\theta \in \Theta$. Let θ_0 be in the interior of Θ such that $E[h(X, \theta_0)] = 0$. Let $\{a_n\}$ be a sequence of positive numbers which converges to infinity such that $\sup_{n \geqslant 1} n^{-1} a_n^2 < \infty$. Let $\{\hat{\theta}_n = \hat{\theta}_n(X_1, ..., X_n)\}$ be a sequence of Θ–valued random variables. Suppose that:*

(B.1) $\hat{\theta}_n \xrightarrow{\Pr} \theta_0$ *and* $a_n n^{-1} \sum_{i=1}^n h(X_i, \hat{\theta}_n) \xrightarrow{\Pr} 0$.

(B.2) $H(\theta) := E[h(X, \theta)]$ *is differentiable at* θ_0 *with nonsingular derivative.*

(B.3) $a_n(P_n - P)\, h(\cdot, \theta_0) = O_{\Pr}(1)$.

(B.4) *For each* $M < \infty$,

$$\sup_{|\theta - \theta_0| \leqslant M a_n^{-1}} a_n \, |(P_n - P)(h(\cdot, \theta) - h(\cdot, \theta_0))| \xrightarrow{\Pr} 0.$$

(B.5) *For each* $\tau > 0$, *there exists a* $\delta > 0$ *such that*

$$\lim_{M \to \infty} \limsup_{n \to \infty} \Pr \left\{ \sup_{|\theta - \theta_0| \leqslant \delta} \frac{a_n \, |(P_n - P)(h(\cdot, \theta) - h(\cdot, \theta_0))|}{\tau a_n \, |\theta - \theta_0| + M} \geqslant 1 \right\} = 0.$$

Then,

$$a_n(\hat{\theta}_n - \theta_0) + (H'(\theta_0))^{-1} \, a_n(P_n - P)\, h(\cdot, \theta_0) \xrightarrow{\Pr} 0. \tag{2.2}$$

Proof. By (B.2), there are $c, \delta_0 > 0$ such that if $|\theta - \theta_0| \leqslant \delta_0$, then

$$c\, |\theta - \theta_0| \leqslant |H(\theta) - H(\theta_0)|.$$

Take $0 < \tau < c$. If $|\hat{\theta}_n - \theta_0| \leqslant \delta_0$, then

$$\begin{aligned} ca_n\, |\hat{\theta}_n - \theta_0| &\leqslant a_n\, |H(\hat{\theta}_n) - H(\theta_0)| \\ &\leqslant a_n\, |(P_n - P)(h(\cdot, \hat{\theta}_n) - h(\cdot, \theta_0))| \\ &\quad + a_n\, |(P_n - P)\, h(\cdot, \theta_0)| + a_n\, |P_n h(\cdot, \hat{\theta}_n)| \\ &= \tau a_n\, |\hat{\theta}_n - \theta_0| + O_{\mathrm{Pr}}(1). \end{aligned}$$

This implies that $a_n(\hat{\theta}_n - \theta_0) = O_{\mathrm{Pr}}(1)$. We also have that

$$\begin{aligned} a_n\, &|H'(\theta_0)(\hat{\theta}_n - \theta_0) + (P_n - P)\, h(\cdot, \theta_0)| \\ &\leqslant a_n\, |(P_n - P)(h(\cdot, \theta_0) - h(\cdot, \hat{\theta}_n))| + a_n\, |P_n h(\cdot, \hat{\theta}_n)| \\ &\quad + a_n\, |H(\hat{\theta}_n) - H(\theta_0) - H'(\theta_0)(\hat{\theta}_n - \theta_0)|, \end{aligned}$$

which converges to zero in probability. ∎

Conditions (B.1)–(B.3) are elementary conditions. Using empirical processes, we will obtain usable conditions that imply (B.4) and (B.5). It should be noted that (1.2) may have not solution when h is not continuous in θ (see, for example, Bai *et al.*, 1990).

3. CONVERGENCE OF M-ESTIMATORS OVER A VC SUBGRAPH CLASS OF FUNCTIONS

Given a set S and a collection of subsets $\mathscr{C}$, for $A \subset S$, let $\Delta^{\mathscr{C}}(A) = \operatorname{card}\{A \cap C : C \in \mathscr{C}\}$, let $m^{\mathscr{C}}(n) = \max\{\Delta^{\mathscr{C}}(A) : \operatorname{card}(A) = n\}$ and let $s(\mathscr{C}) = \inf\{n \colon m^{\mathscr{C}}(n) < 2^n\}$. $\mathscr{C}$ is said to be a VC class of sets if $s(\mathscr{C}) < \infty$. General properties of VC classes of sets can be found in Chapters 9 and 11 in Dudley (1984). Given a function $f\colon S \to \mathbb{R}$, the subgraph of f is the set $\{(x, t) \in S \times \mathbb{R} : 0 \leqslant t \leqslant f(x) \text{ or } f(x) \leqslant t \leqslant 0\}$. A class of functions $\mathscr{F}$ is a VC-subgraph class if the collection of subgraphs of $\mathscr{F}$ is a VC class. The interest of these classes of functions lies in their good properties with respect to covering numbers. Given a pseudometric space (T, d), the ε-covering number $N(\varepsilon, T, d)$ is defined as

$$N(\varepsilon, T, d) = \min\{n\colon \text{there exists a covering of T by } n \text{ balls of radius} \leqslant \varepsilon\}.$$

Given a positive measure μ on $(S, \mathscr{S})$ we define $N_2(\varepsilon, \mathscr{F}, \mu) = N(\varepsilon, \mathscr{F}, \|\cdot\|_{L_2(\mu)})$. If $\mathscr{F}$ is a VC-subgraph class (Pollard, 1984, Prop. II. 25), then there are finite constants A and v such that, for each probability measure μ with $\mu F^2 < \infty$,

$$N_2(\varepsilon, \mathscr{F}, \mu) \leqslant A((\mu F^2)^{1/2}/\varepsilon)^v, \tag{3.1}$$

where $F(x) = \sup_{f \in \mathscr{F}} |f(x)|$ and A and v can be chosen depending only on $s(\mathscr{F})$, i.e. uniformly over all the classes of functions with the same number $s(\mathscr{F})$. By the maximal inequality for subgaussian processes in Theorem 3.1 in Marcus and Pisier (1981) there exists a constant c depending only on A, v and p such that for any class of functions satisfying (3.1) and any $p > 0$,

$$E\left[\sup_{f \in \mathscr{F}} \left|\sum_{j=1}^{n} \varepsilon_j f(X_j)\right|^p\right] \leqslant cE\left[\left|\sum_{j=1}^{n} \varepsilon_j F(X_j)\right|^p\right]. \tag{3.2}$$

We apply Theorem 1 to get sufficient conditions for the convergence of the M-estimator over a VC subgraph class of functions. First, we consider condition (A.5) in Theorem 1.

LEMMA 3. *Given $\beta > 0$, suppose that:*

(i) *For some $\delta_0 > 0$, $\{g(x, \theta) - g(x, \theta_0) : |\theta - \theta_0| \leqslant \delta_0\}$ is a VC subgraph class of functions.*

(ii) *For each $M > 0$,*

$$\sup_{|\theta| \leqslant M} a_n^\beta \, |(P_n - P)(g(\cdot, \theta_0 + a_n^{-1}\theta) - g(\cdot, \theta_0))| = O_{\mathrm{Pr}}(1).$$

(iii) *There are constants $q, c > 0$ such that*

$$E[(M^{-1}G_\delta(X)) \wedge (M^{-2}G_\delta^2(X))] \leqslant c\delta^\beta M^{-1-q},$$

for each $\delta > 0$ small enough and each $M > 0$ large enough, where $G_\delta(x) = \sup_{|\theta| \leqslant \delta} |g(x, \theta_0 + \theta) - g(x, \theta_0)|$.

(iv) $a_n = O(n^{1/\beta})$.

Then, for each $\tau > 0$, there is a $\delta > 0$ such that

$$\lim_{M \to \infty} \limsup_{n \to \infty} \Pr\left\{\sup_{|\theta - \theta_0| \leqslant \delta} \frac{a_n^\beta \, |(P_n - P)(g(\cdot, \theta) - g(\cdot, \theta_0))|}{\tau a_n^\beta \, |\theta - \theta_0|^2 + M} \geqslant 1\right\} = 0.$$

Proof. We have that

$$\sup_{a_n^{-1} \leqslant |\theta-\theta_0| \leqslant \delta} \frac{a_n^{\beta} \, |(P_n - P)(g(\cdot, \theta) - g(\cdot, \theta_0))|}{\tau a_n^{\beta} \, |\theta-\theta_0|^{\beta} + M}$$
$$\leqslant \sup_{1 \leqslant j \leqslant [\log(a_n \delta)]+1} \; \sup_{e^{j-1} \leqslant a_n |\theta-\theta_0| \leqslant e^j} (\tau e^{(j-1)\beta} + M)^{-1} \, a_n^{\beta}$$
$$\times |(P_n - P)(g(\cdot, \theta) - g(\cdot, \theta_0))|.$$

Hence,

$$\Pr\left\{ \sup_{a_n^{-1} \leqslant |\theta-\theta_0| \leqslant \delta} \frac{a_n^{\beta} \, |(P_n - P)(g(\cdot, \theta) - g(\cdot, \theta_0))|}{\tau a_n^{\beta} \, |\theta-\theta_0|^{\beta} + M} \geqslant 1 \right\}$$
$$\leqslant \sum_{j=1}^{[\log(a_n \delta)]+1} \Pr\left\{ \sup_{a_n |\theta-\theta_0| \leqslant e^j} a_n^{\beta} \, |(P_n - P)(g(\cdot, \theta) - g(\cdot, \theta_0))| \right.$$
$$\left. \geqslant e^{-\beta}(\tau e^{\beta j} + M) \right\}.$$

By symmetrization, (3.2) and hypothesis (iii)

$$\Pr\left\{ \sup_{a_n |\theta-\theta_0| \leqslant e^j} a_n^{\beta} \, |(P_n - P)(g(\cdot, \theta) - g(\cdot, \theta_0))| \geqslant e^{-\beta}(\tau e^{\beta j} + M) \right\}$$
$$\leqslant e^{\beta}(\tau e^{\beta j} + M)^{-1} \, E\left[\sup_{a_n |\theta-\theta_0| \leqslant e^j} a_n^{\beta} \, |(P_n - P)(g(\cdot, \theta) - g(\cdot, \theta_0))| \right]$$
$$\leqslant 2(\tau e^{\beta j} + M)^{-1} \, a_n^{\beta} n^{-1} E\left[\sup_{a_n |\theta-\theta_0| \leqslant e^j} \left| \sum_{i=1}^{n} \varepsilon_i (g(X_i, \theta) - g(X_i, \theta_0)) \right| \right]$$
$$\leqslant c(\tau e^{\beta j} + M)^{-1} \, a_n^{\beta} n^{-1} E\left[\left| \sum_{i=1}^{n} \varepsilon_i G_{a_n^{-1} e^j}(X_i) \right| \right]$$
$$\leqslant c(\tau e^{\beta j} + M)^{-1} \, a_n^{\beta} E[G_{a_n^{-1} e^j}(X) \, I_{G_{a_n^{-1} e^j}(X) \geqslant n a_n^{-\beta}(\tau e^{\beta j} + M)}]$$
$$+ c(\tau e^{\beta j} + M)^{-1} \, a_n^{\beta} \, n^{-1} (nE[G^2_{a_n^{-1} e^j}(X) \, I_{G_{a_n^{-1} e^j}(X) \geqslant n a_n^{-\beta}(\tau e^{\beta j} + M)}])^{1/2}$$
$$\leqslant c(\tau e^{\beta j} + M)^{-1} \, e^{\beta j} (n a_n^{-\beta}(\tau e^{\beta j} + M))^{-q}$$
$$+ c(\tau e^{\beta j} + M)^{-1} \, (n^{-1} a_n^{\beta})^{1/2} \, e^{\beta j/2} (n a_n^{-\beta}(\tau e^{\beta j} + M))^{(1-q)/2}.$$

Therefore, the claim follows. ∎

The following theorem gives the asymptotic distribution of M-estimators under weak conditions:

THEOREM 4. *With the notation in Theorem* 1, *assume* (A.1), (A.2), (A.3) *and*

(A.6) *For some* $\delta_0 > 0$, $\{g(x,\theta) - g(x,\theta_0) : |\theta - \theta_0| \leqslant \delta_0\}$ *is a VC subgraph class of functions.*

(A.7) *For each* $M, \eta > 0$,

$$n \Pr\{n^{-1} a_n^2 R_{a_n^{-1} M}(X) \geqslant \eta\} \to 0,$$

where $R_M(x) = \sup_{|\theta| \leqslant M} |r(x,\theta)|$.

(A.8) *There are constants* $q, c > 0$ *such that*

$$E[(M^{-1} G_\delta(X)) \wedge (M^{-2} G_\delta^2(X))] \leqslant c\delta^2 M^{-1-q},$$

for each $\delta > 0$ *small enough and each* $M > 0$ *large enough.*

(A.9) *For each* $\theta \in \mathbb{R}^d$,

$$nE[(n^{-1} a_n^2 |r(X, a_n^{-1}\theta)|) \wedge (n^{-2} a_n^4 r^2(X, a_n^{-1}\theta))] \to 0.$$

Then, (2.1) *holds.*

Proof. We apply Theorem 1 and Lemma 3. We claim that by Theorem 2.4 in Arcones (1999), for each $0 < M\infty$, $\{a_n^2(P_n - P)(g(\cdot, \theta_0 + a_n^{-1} t) - g(\cdot, \theta_0)) : |t| \leqslant M\}$ converges weakly. We need to prove that:

(C.1) For each $|t| \leqslant M$,

$$n \Pr\{n^{-1} a_n^2 |g(X, \theta_0 + a_n^{-1} t) - g(X, \theta_0)| \geqslant 2\} = O(1).$$

(C.2) For each $\eta > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} n \Pr\left\{ \sup_{\substack{|s|, |t| \leqslant M \\ |s-t| \leqslant \delta}} n^{-1} a_n^2 |g(X, \theta_0 + a_n^{-1} t) - g(X, \theta_0 + a_n^{-1} s)| \geqslant \eta \right\}$$

$$= 0.$$

(C.3) $n^{-1} a_n^4 E[G_{a_n^{-1} M}^2(X)\, I_{n^{-1} a_n^2 G_{a_n^{-1} M}^2(X) \leqslant 1}] = O(1).$

(C.4)
$$\lim_{\delta \to 0} \limsup_{n \to \infty} \sup_{\substack{|s-t| \leqslant \delta \\ |s|, |t| \leqslant M}} n^{-1} a_n^4 E[|g(X, \theta_0 + a_n^{-1} t) - g(X, \theta_0 + a_n^{-1} s)|^2\, I_{n^{-1} a_n^2 G_{a_n^{-1} M}^2(X) \leqslant 1}] = 0.$$

(C.5)
$$\lim_{\delta \to 0} \limsup_{n \to \infty} \sup_{\substack{|s-t| \leqslant \delta \\ |s|, |t| \leqslant M}} a_n^2 |E[(g(X, \theta_0 + a_n^{-1} t) - g(X, \theta_0 + a_n^{-1} s))\, I_{n^{-1} a_n^2 G_{a_n^{-1} M}^2(X) \geqslant 1}]| = 0.$$

By the CLT for stable r.v.'s, for each $0 < t < \infty$, $n\Pr\{n^{-1}a_n|\phi(X)| \geqslant t\} = O(1)$. From this bound and (A.7), (C.1) follows.

We have that

$$n\Pr\Big\{\sup_{\substack{|s|,\,|t|\leqslant M\\ |s-t|\leqslant\delta}} n^{-1}a_n^2\,|g(X,\theta_0+a_n^{-1}t)-g(X,\theta_0+a_n^{-1}s)|\geqslant\eta\Big\}$$
$$\leqslant n\Pr\{n^{-1}a_n|\phi(X)|\geqslant 3^{-1}\delta^{-1}\eta\}+2n\Pr\{n^{-1}a_n^2R_{a_n^{-1}M}(X)\geqslant 3^{-1}\eta\}.$$

This implies (C.2).

(C.3) follows from hypothesis (A.8).

As to Condition (C.4), given $\eta > 0$, take $k > 0$ such that $\sup_{n\geqslant 1} n\Pr\{n^{-1}a_n|\phi(X)|\geqslant k\}\leqslant\eta$. Given $|s|, |t|\leqslant M$ with $|s-t|\leqslant\delta$, we have that

$$n^{-1}a_n^4E[|g(X,\theta_0+a_n^{-1}t)-g(X,\theta_0+a_n^{-1}s)|^2\,I_{n^{-1}a_n^2G_{a_n^{-1}M}(X)\leqslant 1}]$$
$$\leqslant 2\eta+n^{-1}a_n^4E[|g(X,\theta_0+a_n^{-1}t)-g(X,\theta_0+a_n^{-1}s)|^2$$
$$\times I_{n^{-1}a_n^2G_{a_n^{-1}M}(X)\leqslant 1,\,n^{-1}a_n|\phi(X)|<k}]$$
$$\leqslant 2\eta+3n^{-1}a_n^4E[|r(X,a_n^{-1}t)|^2\,I_{n^{-1}a_n^2|r(X,a_n^{-1}t)|<k+1}]$$
$$+3n^{-1}a_n^4E[|r(X,a_n^{-1}s)|^2\,I_{n^{-1}a_n^2|r(X,a_n^{-1}s)|<k+1}]$$
$$+3n^{-1}a_n^4E[|a_n^{-1}(t-s)'\,\phi(X)|^2\,I_{n^{-1}a_n|\phi(X)|<k}].$$

(A.9) implies that $3n^{-1}a_n^4E[|r(X,a_n^{-1}\,t)|^2\,I_{n^{-1}a_n^2|r(X,a_n^{-1}t)|<k+1}]\to 0$. (A.3) implies that $n^{-1}a_n^2E[|\phi(X)|^2\,I_{n^{-1}a_n|\phi(X)|<1}]=O(1)$. (C.4) follows from the previous estimations.

As to Condition (C.5), given $\eta > 0$, by (A.8), we can take $k > 0$ large so that $a_n^2E[G_{a_n^{-1}M}(X)\,I_{n^{-1}a_n^2G_{a_n^{-1}M}(X)\geqslant k}]\leqslant\eta$. Given $|s|, |t|\leqslant M$, we have that

$$a_n^2\,E[|g(X,\theta_0+a_n^{-1}t)-g(X,\theta_0+a_n^{-1}s)|\,I_{n^{-1}a_n^2G_{a_n^{-1}M}(X)\geqslant 1}]$$
$$\leqslant a_n^2E[|g(X,\theta_0+a_n^{-1}t)-g(X,\theta_0+a_n^{-1}s)|$$
$$\times I_{n^{-1}a_n^2G_{a_n^{-1}M}(X)\geqslant k,\,|g(X,\theta_0+a_n^{-1}t)-g(X,\theta_0+a_n^{-1}s)|\geqslant\eta}]$$
$$+a_n^2E[|g(X,\theta_0+a_n^{-1}t)-g(X,\theta_0+a_n^{-1}\,s)|$$
$$\times I_{k>n^{-1}a_n^2G_{a_n^{-1}M}(X)\geqslant 1,\,|g(X,\theta_0+a_n^{-1}t)-g(X,\theta_0+a_n^{-1}s)|\geqslant\eta}]$$
$$+a_n^2E[|g(X,\theta_0+a_n^{-1}t)-g(X,\theta_0+a_n^{-1}s)|$$
$$\times I_{n^{-1}a_n^2G_{a_n^{-1}M}(X)\geqslant 1,\,|g(X,\theta_0+a_n^{-1}t)-g(X,\theta_0+a_n^{-1}s)|\leqslant\eta}]$$

$$\leqslant 2a_n^2 E[G_{a_n^{-1}M}(X)\, I_{n^{-1}a_n^2 G_{a_n^{-1}M}(X) \geqslant k}]$$
$$+2kn \Pr\{|g(X, \theta_0 + a_n^{-1} t) - g(X, \theta_0 + a_n^{-1} s)| \geqslant \eta\}$$
$$+\eta a_n^2 \Pr\{n^{-1} a_n^2 G_{a_n^{-1}M}(X) \geqslant 1\}$$
$$\leqslant 2\eta + 2kn \Pr\{|g(X, \theta_0 + a_n^{-1} t) - g(X, \theta_0 + a_n^{-1} s)| \geqslant \eta\} + \eta c,$$

which implies (C.5) via (C.2). Therefore, $\{a_n^2(P_n - P)(g(\cdot, \theta_0 + a_n^{-1} t) - g(\cdot, \theta_0)) : |t| \leqslant M\}$ converges weakly. Since $\{a_n(P_n - P)\, \theta'\phi: |\theta| \leqslant M\}$ also converge weakly, so does $\{a_n^2(P_n - P)\, r(\cdot, a_n^{-1}\theta): |\theta| \leqslant M\}$. (A.9) implies that for each $\theta \in \mathbb{R}^d$ $a_n^2(P_n - P)\, r(\cdot, a_n^{-1}\theta) \xrightarrow{\Pr} 0$. Therefore, (A.4) holds. It is easy to see that the conditions in Lemma 3 are satisfied. ∎

We present the following theorem for the M-estimators in (1.2).

THEOREM 5. *With the notation in Theorem* 2, *assume* (B.1), (B.2), (B.3),

(B.6) *For some* $\delta_0 > 0$, $\{h(x, \theta) - h(x, \theta_0) : |\theta - \theta_0| \leqslant \delta_0\}$ *is a VC subgraph class of functions.*

(B.7) *There are constants* $0 < q < 1$ *and* $c > 0$ *such that*

$$E[(M^{-1} H_\delta(X)) \wedge (M^{-2} H_\delta^2(X))] \leqslant c\delta M^{-1-q},$$

for each $\delta > 0$ *small enough and each* $M > 0$ *large enough, where*

$$H_\delta(x) = \sup_{|\theta - \theta_0| \leqslant \delta} |h(x, \theta) - h(x, \theta_0)|.$$

Then, (2.2) *holds.*

Proof. By symmetrization, (3.2) and hypothesis (B.7)

$$E\left[\sup_{|\theta - \theta_0| \leqslant M a_n^{-1}} a_n\, |(P_n - P)(h(\cdot, \theta) - h(\cdot, \theta_0))|\right]$$
$$\leqslant 2n^{-1} a_n E\left[\sup_{|\theta - \theta_0| \leqslant M a_n^{-1}} \left|\sum_{i=1}^n \varepsilon_i (h(X_i, \theta) - h(X_i, \theta_0))\right|\right]$$
$$\leqslant c n^{-1} a_n E\left[\left|\sum_{i=1}^n \varepsilon_i H_{a_n^{-1}M}(X_i)\right|\right]$$
$$\leqslant c a_n E[H_{a_n^{-1}}(X)\, I_{H_{a_n^{-1}M}(X) \geqslant a_n}]$$
$$+ c(n^{-1} a_n^2 E[H_{a_n^{-1}M}^2(X)\, I_{H_{a_n^{-1}M}(X) \leqslant a_n}])^{1/2} \to 0,$$

which gives (B.4). Lemma 3 implies (B.5). ∎

4. CONVERGENCE OF M-ESTIMATORS UNDER BRACKETING CONDITIONS

In this section, we consider the case when the class of functions satisfies a bracketing condition. We present the following bound on empirical processes over classes of functions that satisfy a bracketing condition:

THEOREM 6. *Let $X_1, ..., X_n$ be independent r.v.'s with values in the measurable spaces $(S_1, \mathscr{S}_1), ..., (S_n, \mathscr{S}_n)$, respectively. Let T be a parameter set. Let $f_j(\cdot, t)$: $S_j \to \mathbb{R}$ be a measurable function for each $t \in T$ and each $1 \leqslant j \leqslant n$. Let $a \geqslant 4$. Let k_0 be an integer. Suppose that for each $k \geqslant k_0$, there exists a function π_k: $T \to T$ such that $\pi_k(\pi_k(t)) = \pi_k(t)$; $\pi_{k-1}(t) = \pi_{k-1}(s)$, if $\pi_k(t) = \pi_k(s)$ and*

$$\sum_{j=1}^{n} E[(2^k \Delta_{j,k}(X_j, \pi_k(t))) \wedge (2^{2k} \Delta_{j,k}^2(X_j, \pi_k(t)))] \leqslant 1$$

where

$$\Delta_{j,k}(x, \pi_k(t)) = \sup_{s:\, \pi_k(s) = \pi_k(t)} |f_j(x, s) - f_j(x, \pi_k(t))|.$$

Then,

$$\begin{aligned}\Pr\left\{\sup_{t \in T}\left|\sum_{j=1}^{n} \varepsilon_j(f(X_j, t) - f(X_j, \pi_{k_0}(t)))\right| \geqslant M\right\}& \\ \leqslant \sum_{j=1}^{n} \Pr\left\{\sup_{t \in T} \Delta_{j,k_0}(X_j, \pi_{k_0}(t)) \geqslant M(a+2)^{-1}\, 2^{-5-k_0}(\log N_{k_0+1})^{-1/2}\right.& \\ \left.\times\left(\sum_{k=k_0+2}^{\infty} 2^{-k}(\log N_k)^{1/2}\right)^{-1}\right\} + \sum_{k=k_0}^{\infty} 4N_k^{-a},& \qquad (4.1)\end{aligned}$$

for

$$M \geqslant 2^4(a+2) \sum_{k=k_0+2}^{\infty} 2^{-k}(\log N_k)^{1/2},$$

where N_k is the cardinality of $\pi_k(T)$.

Proof. Without loss of generality, we may assume that $N_{k_0} \geqslant 2$. First, we prove that if

$$4(a+2) \sum_{k=k_0+2}^{\infty} 2^{-k}(\log N_k)^{1/2} \leqslant \eta,$$

then

$$\Pr\left\{\sup_{t\in T}\left|\sum_{j=1}^{n}\varepsilon_j(f_j(X_j,t)-f_j(X_j,\pi_{k_0}(t)))\right|\geqslant 4\eta\right\}$$
$$\leqslant\sum_{j=1}^{n}\Pr\left\{\sup_{t\in T}\Delta_{j,k_0}(X_j,\pi_{k_0}(t))\geqslant 2^{-1-k_0}(\log N_{k_0+1})^{-1/2}\right\}+\sum_{k=k_0}^{\infty}6N_k^{-a}. \tag{4.2}$$

Let $\tau_j(t)=\inf\{k\geqslant k_0:\Delta_{j,k}(X_j,\pi_k(t))>2^{-1-k}(\log N_{k+1})^{-1/2}\}$. Take $k_1>k_0$ such that

$$n2^{-k_1}(\log 2)^{-1/2}\leqslant\eta/2. \tag{4.3}$$

We have that

$$\sum_{j=1}^{n}\varepsilon_j(f_j(X_j,t)-f_j(X_j,\pi_{k_0}(t)))$$
$$=\sum_{j=1}^{n}\varepsilon_j(f_j(X_j,t)-f(X_j,\pi_{k_0}(t)))\,I_{\tau_j(t)=k_0}$$
$$+\sum_{j=1}^{n}\varepsilon_j(f_j(X_j,t)-f_j(X_j,\pi_{k_1}(t)))\,I_{\tau_j(t)\geqslant k_1}$$
$$+\sum_{j=1}^{n}\sum_{k=k_0+1}^{k_1-1}\varepsilon_j(f_j(X_j,t)-f_j(X_j,\pi_k(t)))\,I_{\tau_j(t)=k}$$
$$+\sum_{j=1}^{n}\sum_{k=k_0+1}^{k_1}\varepsilon_j(f_j(X_j,\pi_k(t))-f_j(X_j,\pi_{k-1}(t)))\,I_{\tau_j(t)\geqslant k}$$
$$=:U_n^{(1)}(t)+U_n^{(2)}(t)+U_n^{(3)}(t)+U_n^{(4)}(t). \tag{4.4}$$

Hence, we have that

$$\Pr\left\{\sup_{t\in T}\left|\sum_{j=1}^{n}\varepsilon_j(f_j(X_j,t)-f_j(X_j,\pi_{k_0}(t)))\right|\geqslant 4\eta\right\}$$
$$\leqslant\sum_{j=1}^{4}\Pr\left\{\sup_{t\in T}|U_n^{(j)}(t)|\geqslant\eta\right\}. \tag{4.5}$$

It $\tau_j(t)=k_0$, then

$$2^{-1-k_0}(\log N_{k_0})^{-1/2}\leqslant\Delta_{j,k_0}(X,\pi_{k_0}(t)).$$

So,

$$\Pr\left\{\sup_{t\in T} |U_n^{(1)}(t)| \geqslant \eta\right\}$$

$$\leqslant \sum_{j=1}^{n} \Pr\left\{\sup_{t\in T} \Delta_{j,k_0}(X_j, \pi_{k_0}(t)) \geqslant 2^{-1-k_0}(\log N_{k_0+1})^{-1/2}\right\}. \tag{4.6}$$

By (4.3)

$$|U_n^{(2)}(t)| \leqslant \sum_{j=1}^{n} \Delta_{j,k_1}(X_j, \pi_{k_1}(t))\, I_{\Delta_{j,k_1}(X_j,\pi_{k_1}(t)) \leqslant 2^{-k_1}(\log N_{k_1})^{-1/2}}$$

$$\leqslant n2^{-k_1}(\log 2)^{-1/2} \leqslant \eta/2.$$

Hence,

$$\Pr\left\{\sup_{t\in T} |U_n^{(2)}(t)| \geqslant \eta\right\} = 0. \tag{4.7}$$

It $\tau_j(t) = k$, then

$$2^{-1-k}(\log N_{k+1})^{-1/2} < \Delta_{j,k}(X_j, \pi_k(t))$$

$$\leqslant \Delta_{j,k-1}(X_j, \pi_{k-1}(t)) \leqslant 2^{-k}(\log N_k)^{-1/2}.$$

Since $u^2 E[(u^{-1}\Delta_{j,k}(X_j, \pi_k(t))) \wedge (u^{-2}\Delta_{j,k}^2(X_j, \pi_k(t)))]$ is nondecreasing, we have that for $0 < u \leqslant 2^{-k}$,

$$\sum_{j=1}^{n} uE[\Delta_{j,k}(X_j, \pi_k(t))\, I_{\Delta_{j,k}(X_j,\pi_k(t)) \geqslant u}]$$

$$\leqslant \sum_{j=1}^{n} 2^{-2k} E[(2^k \Delta_{j,k}(X_j, \pi_k(t))) \wedge (2^{2k}\Delta_{j,k}^2(X_j, \pi_k(t)))] \leqslant 2^{-2k}.$$

Thus,

$$\sum_{j=1}^{n} E[\Delta_{j,k}(X_j, \pi_k(t))\, I_{\tau_j(t)=k}]$$

$$\leqslant \sum_{j=1}^{n} E[\Delta_{j,k}(X_j, \pi_k(t))\, I_{\Delta_{j,k}(X_j,\pi_k(t)) > 2^{-1-k}(\log N_{k+1})^{-1/2}}]$$

$$\leqslant 2^{1-k}(\log N_{k+1})^{1/2}.$$

Since $E[(u^{-1}\Delta_{j,k}(X_j, \pi_k(t))) \wedge (u^{-2}\Delta^2_{j,k}(X_j, \pi_k(t)))]$ is nonincreasing,

$$\sum_{j=1}^{n} E[\Delta^2_{j,k}(X_j, \pi_k(t))\, I_{\tau_j(t)=k}]$$
$$\leqslant \sum_{j=1}^{n} E[\Delta^2_{j,k}(X_j, \pi_k(t))\, I_{\Delta_{j,k}(X_j, \pi_k(t)) \leqslant 2^{-k}(\log 2)^{-1/2}}] \leqslant 2^{-2k}(\log 2)^{-1}.$$

From the previous estimations and the Bernstein inequality,

$$\Pr\left\{\sup_{t\in T} |U_n^{(3)}(t)| \geqslant \eta\right\}$$
$$\leqslant \sum_{k=k_0+1}^{k_1} \sum_{t\in T_k} \Pr\left\{\left|\sum_{j=1}^{n} (\Delta_{j,k}(X_j, \pi_k(t))\, I_{\tau_j(t)=k}\right.\right.$$
$$\left.\left. - E[\Delta_{j,k}(X_j, \pi_k(t))\, I_{\tau_j(t)=k}])\right| \geqslant (a+1)\, 2^{1-k}(\log N_{k+1})^{1/2}\right\}$$
$$\leqslant \sum_{k=k_0+1}^{k_1-1} \sum_{t\in T_k} 2\exp\left(\frac{-(a+1)^2\, 2^{2-2k}\log N_{k+1}}{\begin{array}{c}2\sum_{j=1}^{n} E[\Delta^2_{j,k}(X_j, \pi_k(t))\, I_{\tau_j(t)=k}]\\ +(2/3)\, 2^{1-k}(\log N_k)^{-1/2}\,(a+1)\, 2^{1-k}(\log N_{k+1})^{1/2}\end{array}}\right)$$
$$\leqslant \sum_{k=k_0+1}^{k_1-1} 2N_k \exp\left(\frac{-6(a+1)^2\,(\log N_k)^{1/2}\,(\log N_{k+1})^{1/2}}{3(\log 2)^{-1}+4(a+1)}\right) \leqslant \sum_{k=k_0+1}^{k_1} 2N_k^{-a},$$

where $T_k = \pi_k(T)$.

Since $\pi_{k-1}(t) = \pi_{k-1}(\pi_k(t))$, $|f_j(X_j, \pi_k(t)) - f_j(X_j, \pi_{k-1}(t))| \leqslant \Delta_{j,k-1}(X_j, \pi_{k-1}(t))$. Hence,

$$\Pr\left\{\sup_{t\in T} |U_n^{(4)}(t)| \geqslant \eta\right\}$$
$$\leqslant \sum_{k=k_0+1}^{k_1} \sum_{t\in T_k} 2\Pr\left\{\left|\sum_{j=1}^{n} \varepsilon_j (f_j(X_j, \pi_k(t)) - f_j(X_j, \pi_{k-1}(t)))\, I_{\tau_j(t)\geqslant k}\right|\right.$$
$$\left. \geqslant (a+2)\, 2^{1-k}(\log N_{k+1})^{1/2}\right\}$$
$$\leqslant \sum_{k=k_0+1}^{k_1} 2N_k$$
$$\times \exp\left(\frac{-(a+2)^2\, 2^{2-2k}\log N_{k+1}}{2\cdot 2^{2-2k} + (2/3)\, 2^{-k}(\log N_k)^{-1/2}\,(a+2)\, 2^{1-k}(\log N_{k+1})^{1/2}}\right)$$
$$\leqslant \sum_{k=k_0+1}^{k_1} 2N_k^{-a}.$$

(4.2) follows from (4.4)–(4.9) letting k_1 go to infinity.

Given $0<\lambda\leqslant 1$, the class of functions $\{\lambda f_j(x,t)\colon t\in T\}$ satisfies the conditions in this theorem. So, by (4.2)

$$\begin{aligned}\Pr\Big\{\sup_{t\in T}\Big|\sum_{j=1}^{n}\varepsilon_j(f_j(X_j,t)-f_j(X_j,\pi_{k_0}(t)))\Big|\\ \geqslant\lambda^{-1}2^4(a+2)\sum_{k=k_0+2}^{\infty}2^{-k}(\log N_k)^{1/2}\Big\}\\ \leqslant\sum_{j=1}^{n}\Pr\Big\{\sup_{t\in T}\Delta_{j,k_0}(X_j,\pi_{k_0}(t))\geqslant\lambda^{-1}2^{-1-k_0}(\log N_{k_0+1})^{-1/2}\Big\}\\ +\sum_{k=k_0}^{\infty}4N_k^{-a},\end{aligned}$$

for each $0<\lambda\leqslant 1$. (4.1) follows taking $M=\lambda^{-1}2^4(a+2)\sum_{k=k_0+2}^{\infty}\times 2^{-k}(\log N_k)^{1/2}$. ∎

To check conditions (A.4), (A.5), (B.4), and (B.5), we will use the following theorem:

THEOREM 7. *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with values in a measurable space $(S,\mathscr{S})$. Let X be a copy of X_1. Let $M,\delta_0>0$. Let $r\colon S\times(-\delta_0,\delta_0)^d\to\mathbb{R}$ be a measurable function such that $r(\cdot,\theta)$ is a measurable function for each $\theta\in(-\delta_0,\delta_0)^d$. Let $\{a_n\}$ and let $\{b_n\}$ be two sequences of positive numbers bounded away from zero. Let $1<q<2$ and let $v>0$. Suppose that:*

(i) *For each $M,\eta>0$,*

$$n\Pr\Big\{b_n n^{-1}\sup_{|\theta|\leqslant M}|r(X,a_n^{-1}\theta)|\geqslant\eta\Big\}\to 0.$$

(ii) *For each $\theta\in\mathbb{R}^d$,*

$$nE[(n^{-1}b_n\,|r(X,a_n^{-1}\theta)|)\wedge(n^{-2}b_n^2r^2(X,a_n^{-1}\theta))]\to 0.$$

(iii) *For each $1\leqslant a_n\,|\theta|\leqslant a_n\delta_1$ and each n large enough,*

$$nE[(n^{-1}b_n\tau_0^{-1}\,a_n^{-q}\,|\theta|^{-q}\,|r(X,\theta)|)\wedge(n^{-2}b_n^2\tau_0^{-2}a_n^{-2q}\,|\theta|^{-2q}\,r^2(X,\theta))]\leqslant 1.$$

(iv) *For each $0<\delta\leqslant 1$, each $|\theta|\leqslant M$, each $1\leqslant M\leqslant\delta_1 a_n$ and each n large enough,*

$$\begin{aligned}nE[\min(b_n n^{-1}\tau_0^{-1}M^{-q}\delta^{-1}\,\Delta_{a_n^{-1}M}(X,a_n^{-1}\theta,a_n^{-1}M\delta^v),\\ b_n^2n^{-2}\tau_0^{-2}M^{-2q}\delta^{-2}\,\Delta^2_{a_n^{-1}M}(X,a_n^{-1}\,\theta,a_n^{-1}\,M\delta^v))]\leqslant 1,\end{aligned}$$

where

$$\Delta_M(X, \theta, \delta) = \sup_{\substack{t:\ |t| \leqslant \delta \\ |\theta|,\ |\theta + t| \leqslant M}} |r(X, \theta + t) - r(X, \theta)|.$$

Then,

(a) *For each* $0 < M < \infty$,

$$\sup_{|\theta| \leqslant M} \left| b_n n^{-1} \sum_{i=1}^{n} (r(X_i, a_n^{-1}\theta) - E[r(X_i, a_n^{-1}\theta)]) \right| \xrightarrow{\Pr} 0.$$

(b) *For each* $\tau > 0$, *there exists a* $\delta > 0$ *such that*

$$\lim_{M \to \infty} \limsup_{n \to \infty} \Pr\left\{ \sup_{|\theta| \leqslant \delta} \frac{b_n |(P_n - P) r(\cdot, \theta)|}{\tau a_n^2 |\theta - \theta_0|^2 + M} \geqslant 1 \right\} = 0.$$

Proof. Define $|\theta|_\infty := \max_{1 \leqslant j \leqslant d} |\theta^{(j)}|$. To symmetrize in part (a), we need to prove that for each $\tau > 0$,

$$\sup_{|\theta| \leqslant M} \Pr\{ b_n |(P_n - P) r(\cdot, a_n^{-1}\theta)| \geqslant \tau \} \to 0. \tag{4.10}$$

Take a function π: $[-M, M]^d \to [-M, M]^d$ such that $\#(\pi([M, M]^d)) < \infty$, and $\sup_{\theta \in [-M, M]^d} |\theta - \pi(\theta)|_\infty \leqslant \delta^v$. By condition (ii),

$$\sup_{|\theta| \leqslant M} \Pr\{ b_n |(P_n - P) r(\cdot, a_n^{-1} \pi(\theta))| \geqslant 2^{-1}\tau \} \to 0.$$

Condition (iv) implies that

$$\begin{aligned} &\sup_{|\theta| \leqslant M} \Pr\{ b_n |(P_n - P)(r(\cdot, a_n^{-1} \theta) - r(\cdot, a_n^{-1} \pi(\theta)))| \geqslant 2^{-1}\tau \} \\ &\quad \leqslant \sup_{|\theta| \leqslant M} 2\tau^{-1} E\left[b_n n^{-1} \left| \sum_{j=1}^{n} (r(X_j, a_n^{-1}\theta) - r(X_j, a_n^{-1}\pi(\theta))) \right| \right] \leqslant 8\tau^{-1}\tau_0 \delta. \end{aligned}$$

So, (4.10) holds. From (4.10) and Lemma 2.5 in Giné and Zinn (1984), we may symmetrize the expression in (a). Hence, it suffices to prove that

$$\sup_{|\theta| \leqslant M} \left| b_n n^{-1} \sum_{i=1}^{n} \varepsilon_i r(X_i, a_n^{-1}\theta) \right| \xrightarrow{\Pr} 0.$$

Without loss of generality, we may assume that M and τ_0 are integers. Let k be a positive integer. Let $A_k = \{ -M + j 2^{-kv} \tau_0^{-v} : 1 \leqslant j \leqslant M \tau_0^v 2^{kv+1} \}$. It is easy to see that we may define π_k: $T \to T$ such that $\|\pi_k(\theta) - \theta\|_\infty \leqslant \tau_0^{-v} 2^{-kv}$, the coordinates of $\pi_k(\theta)$ are in A_k and $\pi_{k-1}(t_1) = \pi_{k-1}(t_2)$, if

$\pi_k(t_1)=\pi_k(t_2)$. Then, $N_k=\#\pi_k(T)=(M\tau_0^v 2^{vk+1})^d$. Take k_0 large enough. By Theorem 6,

$$\Pr\left\{\sup_{|\theta|\leqslant M}\left|b_n n^{-1}\sum_{i=1}^n \varepsilon_i r(X_i, a_n^{-1}\theta)\right|\geqslant\tau\right\}$$

$$\leqslant n\Pr\left\{\sup_{|\theta|\leqslant M} b_n n^{-1}\,|r(X, a_n^{-1}\theta)|\geqslant c\right\}+\sum_{k=k_0}^{\infty}4N_k^{-a}=:I+II,$$

where c is a finite constant. By hypothesis (i), $I\to 0$. II can be made arbitrarily small by letting $k_0\to\infty$. Therefore, (a) follows.

As to (b), by part (a), it suffices to prove that

$$\limsup_{M\to\infty}\lim_{n\to\infty}\Pr\left\{\sup_{a_n^{-1}\leqslant|\theta|_\infty\leqslant\delta}\frac{b_n\,|(P_n-P)\,r(\cdot,\theta)|}{\tau a_n^2\,|\theta|^2+M}\geqslant 1\right\}=0.$$

First we prove that we may symmetrize. Take M such that $\sup_{0<x<\infty}\tau_0 x^q/(\tau x^2+M)\leqslant 2^{-4}$. By hypothesis (iii)

$$\sup_{a_n^{-1}\leqslant|\theta|_\infty\leqslant\delta}\Pr\left\{\frac{b_n\,|(P_n-P)\,r(\cdot,\theta)|}{\tau a_n^2\,|\theta|^2+M}\geqslant 2^{-1}\right\}$$

$$\leqslant 2^3\sup_{a_n^{-1}\leqslant|\theta|_\infty\leqslant\delta}\frac{a_n^q\,|\theta|^q\,\tau_0}{\tau a_n^2\,|\theta|^2+M}\leqslant 2^{-1}.$$

From this and using the inequality $xy\leqslant\alpha^{-1}x^\alpha+\beta^{-1}y^\beta$, for $x, y>0$ and $\alpha^{-1}+\beta^{-1}=1$, it suffices to prove that

$$\lim_{M\to\infty}\limsup_{n\to\infty}\sum_{j=1}^{[\log(a_n\delta)]+1}\Pr\left\{a_n^2 n^{-1}\sup_{|\theta|_\infty\leqslant e^j a_n^{-1}}\left|\sum_{i=1}^n\varepsilon_i r(X_i,\theta)\right|\geqslant Me^{rj}\right\}=0 \tag{4.11}$$

where $q<r<2$. Let $T=\{\theta\in\mathbb{R}^d: |\theta|_\infty\leqslant 1\}$. Let k be a positive integer. Without loss of generality, we may assume that τ_0^{-1} is an integer. Let $A_k=\{-1+j2^{-kv}\tau_0^{-v}: 1\leqslant j\leqslant\tau_0^v 2^{kv+1}\}$. $\pi_k\colon T\to T$ such that $\|\pi_k(\theta)-\theta\|_\infty\leqslant\tau_0^{-v}2^{-kv}$, the coordinates of $\pi_k(\theta)$ are in A_q and $\pi_{k-1}(t_1)=\pi_{k-1}(t_2)$, if $\pi_k(t_1)=\pi_k(t_2)$. Then, $N_k=\#\pi_k(T)=\tau_0^{vd}\,2^{dvk+d}$. Let $a=4$. Take $b>(\log 2)\,dv(r-q)^{-1}$. Let $k_j=[\,j/b]$. Then, there exists an integer j_0, such that

$$2^7\sum_{k=k_{j_0}+2}^{\infty}2^{-k}(\log N_k)^{1/2}\leqslant 1.$$

By part (a),

$$\lim_{M\to\infty} \limsup_{n\to\infty} \sum_{j=1}^{j_0} \Pr\left\{ \sup_{|\theta|_\infty \leqslant e^j a_n^{-1}} n^{-1} b_n \left| \sum_{i=1}^n \varepsilon_i r(X_i, \theta) \right| \geqslant M e^{rj} \right\} = 0.$$

We apply Theorem 6 to $j > j_0$ and $f(x, \theta) = b_n n^{-1} e^{-qj} r(x, a_n^{-1} e^j \theta)$. We have that

$$\Pr\left\{ n^{-1} b_n \sup_{|\theta|_\infty \leqslant 1} \left| \sum_{i=1}^n \varepsilon_i r(X_i, e^j a_n^{-1} \theta) \right| \geqslant M e^{rj} \right\}$$
$$\leqslant \Pr\left\{ n^{-1} b_n \sup_{|\theta|_\infty \leqslant 1} \left| \sum_{i=1}^n \varepsilon_i r(X_i, e^j a_n^{-1} \pi_{q_j}(\theta)) \right| \geqslant 2^{-1} M e^{rj} \right\}$$
$$+ \Pr\left\{ n^{-1} b_n \sup_{|\theta|_\infty \leqslant 1} \left| \sum_{i=1}^n \varepsilon_i (r(X_i, e^j a_n^{-1} \theta) - r(X_i, e^j a_n^{-1} \pi_{q_j}(\theta)) \right| \geqslant 2^{-1} M e^{rj} \right\}$$
$$\leqslant N_{k_j} M^{-2} 2^2 e^{-2(r-q)j} \tau^2$$
$$+ n \Pr\left\{ b_n n^{-1} \sup_{|\theta|_\infty \leqslant 1} |r(X, e^j a_n^{-1} \theta)| \geqslant e^{rj} M 2^{-9-k_j} (\log N_{k_j+1})^{-1/2} \left(\sum_{k=k_j+2}^{\infty} 2^{-k} (\log N_k)^{1/2} \right)^{-1} \right\}$$
$$+ \sum_{k=k_j+2}^{\infty} 4 N_k^{-a}$$
$$\leqslant c M^{-2} 2^{v\, dj/b} e^{-(r-q)j} + c m^{-1} j^2 e^{-j(r-q)} + c 2^{-javd/b},$$

which implies that

$$\lim_{M\to\infty} \limsup_{n\to\infty} \sum_{j=1}^{[\log(a_n\delta)]+1} \Pr\left\{ b_n n^{-1} \sup_{|\theta|_\infty \leqslant e^j a_n^{-1}} \left| \sum_{i=1}^n \varepsilon_i r(X_i, \theta) \right| \geqslant M e^{rj} \right\}$$
$$\leqslant \sum_{j=j_0+1}^{\infty} c 2^{-javd/b}.$$

Since j_0 can be taken arbitrarily large, (4.11) follows. ∎

Theorems 1 and 7 give the following:

THEOREM 8. *Under the notation in Theorem* 1, *suppose* (A.1), (A.2), (A.3), (A.9) *and for some positive constants* v, δ_0, τ_0 *and* $1 < q < 2$, (A.10) *For each* $M, \eta > 0$,

$$n \Pr \left\{ b_n n^{-1} \sup_{|\theta| \leqslant M} |r(X, a_n^{-1}\theta)| \geqslant \eta \right\} \to 0.$$

(A.11) *For each* $1 \leqslant a_n |\theta| \leqslant \delta_0 a_n$,

$$nE[(n^{-1}a_n^2 a_n^{-q} |\theta|^{-q} r(X, \theta)|) \wedge (n^{-2}a_n^4 a_n^{-2q} |\theta|^{-2q} r^2(X, \theta))] \leqslant c.$$

(A.12)

$$nE[(a_n^2 n^{-1}\tau_0^{-1}\lambda^{-q}\delta^{-1} \Delta_{a_n^{-1}\lambda}(X, a_n^{-1}(\theta, a_n^{-1}\lambda\delta^v)) \wedge (a_n^4 n^{-2}\tau_0^{-2}\lambda^{-2q}\delta^{-2} \Delta^2_{a_n^{-1}\lambda}(X, a_n^{-1}\theta, a_n^{-1}\lambda\delta^v))] \leqslant 1,$$

for each $\delta > 0$ *small enough, each* n *large enough, each* $|\theta| \leqslant \delta$, *and each* $1 \leqslant \lambda \leqslant \delta_0 a_n$, *where*

$$\Delta_M(x, \theta, \delta) = \sup_{\substack{t:\, |t| \leqslant \delta \\ |\theta + t| \leqslant M}} |r(x, \theta + t) - r(x, \theta)|,$$

Then, (2.1) *holds.*

Pollard (1985) considered the asymptotic normality of M-estimators assuming bracketing conditions. He assumed conditions on the class $\{|\theta - \theta_0|^{-1} (g(x, \theta) - g(x, \theta_0) - (\theta - \theta_0)' \phi(x)) : |\theta - \theta_0| \leqslant \delta_0\}$. Instead, we consider the simpler class $\{g(x, \theta) - g(x, \theta_0) - (\theta - \theta_0)' \phi(x) : |\theta - \theta_0| \leqslant \delta_0\}$.

Theorem 9 gives the following theorem for differentiable kernels:

THEOREM 9. *Suppose* (A.1), (A.2) *and*

(A.13) $g(x, \cdot)\colon \Theta \to \mathbb{R}$ *is first differentiable with continuity in a neighborhood of* θ_0 *and for some* $\delta_0 > 0$,

$$E\left[\sup_{|\theta| \leqslant \delta_0} \left\| \frac{\partial g}{\partial \theta}(X, \theta_0 + \theta) \right\|^2 \right] < \infty.$$

Then, (2.1) *holds with* $a_n = n^{1/2}$.

Proof. We apply Theorem 9. Let $B(x) = \sup_{|\theta| \leqslant \delta} \| \frac{\partial g}{\partial \theta}(x, \theta_0 + \theta)\|$. It is easy to see that for $|\theta| \leqslant \delta_0$, $|r(x, \theta)| \leqslant |\theta|\, B(x)$ and for $M \leqslant \delta_0/2$, $\delta \leqslant \delta_0/2$ and $|\theta| \leqslant \delta_0/2$, $\Delta_M(x, \theta, \delta) \leqslant \delta B(x)$. From these estimations (A.3) and (A.9)–(A.12) follow. ∎

Next, we consider another way to do the bracketing.

THEOREM 10. *Let ψ be function from S into the set of $d \times d$ symmetric matrices. Assume* (A.1), (A.2), (A.3),

(A.14) $E[\|\psi(X)\|] < \infty$.

(A.15) $\lim_{\delta \to 0} E[\sup_{|\theta| \leqslant \delta} |\theta|^{-2} |r(X, \theta) - \theta' \psi(X) \theta|] = 0$.

Then, (2.1) *holds.*

Proof. We apply Theorem 1. Let $B_\delta(x) = \sup_{|\theta| \leqslant \delta} |\theta|^{-2} |r(x, \theta) - \theta' \psi(x) \theta|$. By the LLN,

$$\Pr\left\{ \sup_{|\theta| \leqslant M a_n^{-1}} a_n^2 |(P_n - P) r(\cdot, \theta)| \geqslant \eta \right\}$$
$$\leqslant \Pr\{M^2 \|(P_n - P) \psi(X)\| + M^2 (P_n + P) B_{a_n^{-1} M}(X) \geqslant \eta\} \to 0.$$

which implies (A.4).

If $E[B_\delta(X)] \leqslant 3^{-1}\tau$, then

$$\Pr\left\{ \sup_{|\theta| \leqslant \delta} \frac{a_n^2 |(P_n - P) r(\cdot, \theta)|}{\tau a_n^2 |\theta|^2 + M} \geqslant 1 \right\}$$
$$\leqslant \Pr\{\|(P_n - P) \psi(X)\| + (P_n + P) B_\delta \geqslant \tau\} \to 0,$$

which implies (A.5). ∎

The next theorem follows directly from Theorem 10.

THEOREM 11. *Assume* (A.1),

(A.15) $g(x, \cdot) \colon \Theta \to \mathbb{R}$ *is second differentiable with continuity in a neighborhood of* θ_0.

(A.16) $a_n (P_n - P) \phi = O_{\Pr}(1)$, *where* $\phi(x) = \frac{\partial g}{\partial \theta}(x, \theta_0)$.

(A.17) $E[\frac{\partial g}{\partial \theta}(X, \theta_0)] = 0$.

(A.18) $V := E[(\partial^2 g / \partial \theta^2)(X, \theta_0)]$ *is a positive definite symmetric matrix.*

(A.19) *For some* $\delta_0 > 0$,

$$E\left[\sup_{|\theta| \leqslant \delta_0} \left\| \frac{\partial^2 g}{\partial \theta^2}(X, \theta_0 + \theta) \right\| \right] < \infty,$$

Then, (2.1) *holds.*

In the particular case $a_n = n^{1/2}$ and $g(x, \theta) = -\log f(x, \theta)$, where $f(x, \theta)$, $\theta \in \Theta$ is a family of densities, previous result is very similar to Proposition 4 in Le Cam (1970). Jurečková and Sen (1996) prove the asymptotic normality of multivariate M-estimators assuming third derivatives. Previous theorem only needs second order derivatives.

For the M-estimators idefined by (1.2), Theorems 2 and 7 give the following:

THEOREM 12. *With the notation in Theorem* 2, *assume* (B.1), (B.2), (B.3),

(B.8) *For each* $M, \eta > 0$,

$$n \Pr \left\{ a_n n^{-1} \sup_{|\theta| \leqslant M} |h(X, \theta_0 + a_n^{-1}\theta) - h(X, \theta_0)| \geqslant \eta \right\} \to 0.$$

(B.9) *For each* $\theta \in \mathbb{R}^d$,

$$nE[(n^{-1}a_n \, |h(X, \theta_0 + a_n^{-1}\theta) - h(x, \theta_0)| \\ \wedge (n^{-2}a_n^2 \, |h(X, \theta_0 + a_n^{-1}\theta)| - h(X, \theta_0)|^2)] \to 0.$$

(B.10) *For each* $1 \leqslant a_n |\theta| \leqslant a_n \delta_1$ *and each n large enough,*

$$nE[\min(n^{-1}a_n \tau_0^{-1} a_n^{-q} \, |\theta|^{-q} \, |h(X, \theta_0 + a_n^{-1}\theta)| - h(X, \theta_0)|, \\ n^{-2}a_n^2 \tau_0^{-2} a_n^{-2q} \, |\theta|^{-2q} \, |h(X, \theta_0 + a_n^{-1}\theta)| - h(X, \theta_0)|^2)] \leqslant 1.$$

(B.11) *For each* $0 < \delta \leqslant 1$, *each* $|\theta| \leqslant M$, *each* $1 \leqslant M \leqslant \delta_1 a_n$ *and each n large enough,*

$$nE[\min(a_n n^{-1} \tau_0^{-1} M^{-q} \delta^{-1} \Delta_{a_n^{-1}M}(X, a_n^{-1}\,\theta, a_n^{-1}\, M\delta^v), \\ a_n^2 n^{-2} \tau_0^{-2} M^{-2q} \delta^{-2} \, \Delta_{a_n^{-1}M}^2(X, a_n^{-1}\,\theta, a_n^{-1} M\delta^v))] \leqslant 1,$$

where

$$\Delta_M(X, \theta, \delta) = \sup_{\substack{t:\, |t| \leqslant \delta \\ |\theta|,\, |\theta + t| \leqslant M}} |h(X, \theta_0 + a_n^{-1}(\theta + t)) - h(X, \theta_0 + a_n^{-1}\theta)|.$$

Then, (2.2) *holds.*

The following theorem follows similarly to Theorem 10.

THEOREM 13. *Let* ψ *be function from S into the set of symmetric* $d \times d$ *matrices. Assume* (B.1), (B.3),

(B.8) $E[\psi(X)]$ *is nonsingular.*

(B.9) $\lim_{\delta \to 0} E[\sup_{|\theta - \theta_0| \leqslant \delta} |\theta - \theta_0|^{-1} \, |h(X, \theta) - h(X, \theta_0) - \psi(X)(\theta - \theta_0)|] = 0.$

Then, (2.2) *holds.*

Proof. We apply Theorem 2. (B.2) follows from the dominated convergence theorem. Let $B_\delta(x) = \sup_{|\theta| \leqslant \delta} |\theta|^{-1} |h(x, \theta_0 + \theta) - h(x, \theta_0) - \psi(x)\,\theta|$. By the LLN,

$$\Pr\left\{\sup_{|\theta| \leqslant Ma_n^{-1}} a_n\, |(P_n - P)(h(\cdot, \theta_0 + \theta) - h(\cdot, \theta_0))| \geqslant \eta\right\}$$

$$\leqslant n \Pr\{M\, \|(P_n - P)\, \psi(X)\| + M(P_n + P)\, B_{a_n^{-1}M}(X) \geqslant \eta\} \to 0.$$

which implies (B.4). (B.5) follows similarly. ∎

It is easy to see that conditions (B.8) and (B.9) hold, if $h(x, \theta)$ is differentiable with respect to θ in a neighborhood of θ_0, the matrix derivative is continuous at θ_0 and for some $\delta_0 > 0$, $E[\sup_{|\theta| \leqslant \delta_0} \| \frac{\partial h}{\partial \theta}(X, \theta_0 + \theta)\|] < \infty$.

5. SOME APPLICATIONS

An application of previous theorem is the following is the asymptotic distribution of the M-estimators for the location parameter. Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with values in $\mathbb{R}^d$. Let ρ be a function on $\mathbb{R}^d$. We consider the M-estimator over $g(x, \theta) = \rho(x - \theta)$.

THEOREM 14. *Assume the notation in Theorem* 2 *with* $d = 1$ *and* $g(x, \theta) = \rho(x - \theta)$, *where* $\rho\colon \mathbb{R} \to \mathbb{R}$ *is a continuous function, nondecreasing on* $[0, \infty)$, *nonincreasing on* $(-\infty, 0]$ *with* $\liminf_{|x| \to \infty} \rho(x) > \rho(x_0)$, *for each* $x_0 \in \mathbb{R}$. *Let Let* $\hat{\theta}_n$ *is any sequence of r.v. satisfying*

$$n^{-1} \sum_{j=1}^{n} \rho(X_j - \hat{\theta}_n) = \inf_{\theta \in \mathbb{R}^d} n^{-1} \sum_{j=1}^{n} \rho(X_j - \theta).$$

Assume (A.2), (A.3), (A.7), (A.8), (A.9) *and*

(i) *For each* $\delta > 0$,

$$\inf_{|\theta - \theta_0| \geqslant \delta} E[\rho(X - \theta) - \rho(X - \theta_0)] > 0.$$

Then, (2.1) *holds.*

Proof. We apply Theorem 4. To prove that the classes of functions $\{\rho(x - \theta) - \rho(x - \theta_0 : \theta \in \mathbb{R}\}$ is a VC subgraph class, we have to show that

$\{A_\theta \cup B_\theta : \theta \in \mathbb{R}^d\}$ is a VC class of sets, where $A_\theta := \{(x', t)' : 0 \leqslant t \leqslant \rho(x-\theta) - \rho(x-\theta_0)\}$ and $B_\theta := \{(x', t)' : 0 \geqslant t \geqslant \rho(x-\theta) - \rho(x-\theta_0)\}$. Let $\rho_1(u) = \rho(u)$ for $u \leqslant 0$ and let $\rho_2(u) = \rho(u)$ for $x \geqslant 0$. Let $\rho_1^{-1}(t) = \sup\{u \leqslant 0 : \rho(u) \geqslant t\}$ and let $\rho_2^{-1}(t) = \inf\{u \geqslant 0 : \rho(u) \geqslant t\}$. We have that $A_\theta = A'_\theta \cup A''_\theta$, where

$$A'_\theta := \{(x', t)' : 0 \leqslant t, x-\theta, y - Z'\theta - \rho_2^{-1}(t + \rho(x-\theta_0))\}$$

and

$$A''_\theta := \{(x', t)' : 0 \leqslant t, x-\theta, x-\theta-y-\rho_1^{-1}(t+\rho(x-\theta_0))\}.$$

We have that $\{C(t_1, ..., t_m) : t_1, ..., t_m \in \mathbb{R}\}$ is a VC class, where $C(t_1, ..., t_m) = \{x \in S : \sum_{j=1}^m t_j f_j(x) \geqslant 0\}$ and $f_1, ..., f_m$ are functions on S (Dudley, 1984, Theorem 9.2.1). We also have that if $\{C_t : t \in T\}$ and $\{D_t : t \in T\}$ are VC classes, then so are $\{C_t \cap D_t : t \in T\}$ and $\{C_t \cup D_t : t \in T\}$ (Dudley, 1984, Proposition 9.2.5). Hence, $\{A_\theta : \theta \in \mathbb{R}^d\}$ is a VC class. A similar argument gives that $\{B_\theta : \theta \in \mathbb{R}^d\}$ is a VC class.

Let $l_1 = \liminf_{|x| \to \infty} \rho(x)$ and let $\tau = l_1 - E[\rho(X-\theta_0)] > 0$. Take $\varepsilon > 0$ and $M > 0$ such that

$$(l_1 - \varepsilon)(1-\varepsilon) \geqslant E[\rho(X-\theta_0)] + 2^{-1}\tau,$$

$$\Pr\{|X| \leqslant M\} \geqslant 1 - 2\varepsilon,$$

$$\min(\rho(M), \rho(-M)) \geqslant l_1 - \varepsilon.$$

For $|\theta| \geqslant 2M$, we have that

$$\begin{aligned} n^{-1} \sum_{i=1}^n \rho(X_j - \theta) &\geqslant n^{-1} \sum_{i=1}^n \rho(X_j - \theta)\, I_{|X_j| \leqslant M} \\ &\geqslant n^{-1} \sum_{i=1}^n \min(\rho(M), \rho(-M))\, I_{|X_j| \leqslant M} \geqslant (l_1 - \varepsilon)(1-2\varepsilon), \end{aligned}$$

for n large enough. Hence, $|\hat{\theta}_n| \leqslant M$, for n large enough.

By the law of the large numbers for VC subgraph classes of functions (see Theorem 8.3 in Giné and Zinn, 1984),

$$\sup_{|\theta - \theta_0| \leqslant M} |(P_n - P)(\rho(|\cdot - \theta|) - \rho(|\cdot - \theta_0|))| \to 0 \qquad \text{a.s.}$$

This and hypothesis (i) give the consistency of the M-estimator. The rest of the conditions in Theorem 4 are assumed. ∎

Given a sequence of i.i.d.r.v.'s $\{X_j\}_{j=1}^\infty$, the k-means is the M-estimator over the kernel $g(x, \theta) = \min_{1 \leqslant i \leqslant k} |x - \theta^{(i)}|^2$, where $\theta = (\theta^{(1)}, ..., \theta^{(k)})'$.

The next theorem show that the k-means can converge to a stable limit distribution.

THEOREM 15. *Let* $\{X_j\}_{j=1}^{\infty}$ *be a sequence of i.i.d.r.v.'s. Let* $g(x,\theta)=\min_{1\leqslant i\leqslant k}|x-\theta^{(i)}|^2$, *where* $\theta=(\theta^{(1)},...,\theta^{(k)})'$. *Let* $1<\alpha\leqslant 2$. *Let* $\{a_n\}$ *be a sequence regularly varying of order* $1/\alpha$. *Suppose that:*

(i) *There exists a* $\theta_0\in\mathbb{R}^k$ *with* $\theta_0^{(1)}<\cdots<\theta_0^{(k)}$ *such that* $E[g(X,\theta)-g(X,\theta_0)]>0$, *for each* $\theta\neq\theta_0$ *with* $\theta^{(1)}\leqslant\cdots\leqslant\theta^{(k)}$.

(ii) $a_n(P_n-P)\phi$ *converges in distribution, where*

$$\phi(x)=2((\theta_0^{(1)}-x)\,I_{(-\infty,2^{-1}(\theta_0^{(1)}+\theta_0^{(2)}]},$$
$$(\theta_0^{(2)}-x)\,I_{(2^{-1}(\theta_0^{(1)}+\theta_0^{(2)},2^{-1}(\theta_0^{(2)}+\theta_0^{(3)}]},...)'.$$

(iii) *The distribution function* F_X *of* X *is differentiable at* $\theta_0^{(j,j+1)}:=2^{-1}(\theta_0^{(j)}+\theta_0^{(j+1)})$, *for* $1\leqslant j\leqslant k-1$.

(iv) $F_X(\theta_0^{(1,2)})-2F'_X(\theta_0^{(1,2)})(\theta_0^{(2)}-\theta_0^{(1)})>0,$

$$F_X(\theta_0^{(j,j+1)})-F_X(\theta_0^{(j-1,j)}))-2^{-1}F'_X(\theta_0^{(j,j+1)})(\theta_0^{(j+1)}-\theta_0^{(j)})$$
$$-2^{-1}F'_X(\theta_0^{(j,j+1)})(\theta_0^{(j)}-\theta_0^{(j-1)})>0,$$

for $2\leqslant j\leqslant k-1$, *and* $1-F_X(\theta_0^{(k-1,k)})-2^{-1}F'_X(\theta_0^{(k-1,k)})(\theta_0^{(k)}-\theta_0^{(k-1)})>0$.

Then, (2.1) *holds, with* V *determined by*

$$\theta'V\theta=(\theta^{(1)})^2\,F_X(\theta_0^{(1,2)})+\sum_{j=1}^{k-1}(\theta^{(j)})^2\,(F_X(\theta_0^{(j,j+1)})-F_X(\theta_0^{(j-1,j)})$$
$$+(\theta^{(k)})^2\,(1-F_X(\theta_0^{(k-1,k)}))$$
$$-2^{-2}\sum_{j=1}^{k-1}(\theta^{(j)}+\theta^{(j+1)})^2\,F'_X(\theta_0^{(j,j+1)})(\theta_0^{(j+1)}-\theta_0^{(j)}).$$

Proof. The result follows from Theorem 4. The consistency of the k clusters means follows from Cuesta and Matran (1988).

We assume without loss of generality that

$$2^{-1}(\theta_0^{(1)}+\theta_0^{(2)})<2^{-1}(\theta^{(1)}+\theta^{(2)})<2^{-1}(\theta_0^{(2)}+\theta_0^{(3)})<2^{-1}(\theta^{(2)}+\theta^{(3)})<\cdots.$$

We have that

$$\begin{aligned} r(x,\theta) &= (\theta^{(1)}-\theta_0^{(1)})^2\, I_{(-\infty,\, 2^{-1}(\theta_0^{(1)}+\theta_0^{(2)})]} \\ &\quad + \sum_{j=2}^{k-1} (\theta^{(j)}-\theta_0^{(j)})^2\, I_{(2^{-1}(\theta^{(j-1)}+\theta^{(j)}),\, 2^{-1}(\theta_0^{(j)}+\theta_0^{(j+1)})]} \\ &\quad + (\theta^{(k)}-\theta_0^{(k)})^2\, I_{(2^{-1}(\theta^{(k-1)}+\theta^{(k)}),\, \infty)} + \sum_{j=2}^{k-1} ((\theta^{(j)}-\theta_0^{(j)})^2 \\ &\quad + (2x-\theta^{(j)}-\theta^{(j-1)})(\theta^{(j)}-\theta^{(j-1)})\, I_{(2^{-1}(\theta_0^{(j-1)}+\theta_0^{(j)}),\, 2^{-1}(\theta^{(j-1)}+\theta^{(j)})]}. \end{aligned} \tag{5.1}$$

Hence,

$$\begin{aligned} &E[r(X,\theta)] - (\theta-\theta_0)'\, V(\theta-\theta_0) \\ &\quad = -\sum_{j=2}^{k-1} (\theta^{(j)}-\theta_0^{(j)})^2\, (F_X(\theta^{(j-1,\,j)}) - F_X(\theta_0^{(j-1,\,j)})) \\ &\qquad + \sum_{j=2}^{k-1} (\theta^{(j)}-\theta_0^{(j)})^2\, (F_X(\theta^{(j-1,\,j)}) - F_X(\theta_0^{(j-1,\,j)})) \\ &\qquad + \sum_{j=2}^{k} (E[(2X-\theta^{(j)}-\theta^{(j-1)})(\theta^{(j)}-\theta^{(j-1)})\, I_{[\theta_0^{(j-1,\,j)},\, \theta^{(j-1,\,j)}]}] \\ &\qquad - 2^{-2}(\theta^{(j-1)}+\theta^{(j)}-\theta_0^{(j-1)}-\theta_0^{(j)})^2\, F'_X(\theta_0^{(j-1,\,j)})(\theta_0^{(j)}-\theta_0^{(j-1)})). \end{aligned}$$

Obviously,

$$(\theta^{(j)}-\theta_0^{(j)})^2\, (F(\theta^{(j-1,\,j)}) - F(\theta_0^{(j-1,\,j)})) = o(|\theta-\theta_0|^2)$$

and

$$(\theta^{(j)}-\theta_0^{(j)})^2\, (F_X(\theta^{(j-1,\,j)}) - F(\theta_0^{(j-1,\,j)})) = o(|\theta-\theta_0|^2).$$

By a change of variable

$$\begin{aligned} &E[(2X-\theta^{(j)}-\theta^{(j-1)})(\theta^{(j)}-\theta^{(j-1)})\, I_{[\theta_0^{(j-1,\,j)},\, \theta^{(j-1,\,j)}]}] \\ &\quad = -\int_{\theta_0^{(j-1,\,j)}}^{\theta^{(j-1,\,j)}} 2(F_X(x)-F_X(\theta_0^{(j-1,\,j)})(\theta^{(j)}-\theta^{(j-1)})\, dx \\ &\quad = -2^{-2}(\theta^{(j-1)}+\theta^{(j)}-\theta_0^{(j-1)}-\theta_0^{(j)})^2\, F'_X(\theta_0^{(j-1,\,j)})(\theta_0^{(j)}-\theta_0^{(j-1)}) \\ &\qquad + o(|\theta-\theta_0|^2). \end{aligned}$$

(A.2) in Theorem 4 follows from previous estimations.

Condition (A.3) is assumed.

Next, we consider condition (A.6). The class of functions $\{g(x, \theta)-g(x, \theta_0) : \theta \in \mathbb{R}^k\}$ is contained in the class

$$\left\{ \sum_{j=1}^{2k} (a_j x + b_j)\, I_{(c_{j-1},\, c_j]}), \text{ where } c_0 = -\infty,\ c_{2k} = \infty,\ a_j, b_j, c_j \in \mathbb{R} \right\}.$$

We claim that this is VC subgraph class. Observe that

$$\left\{ (x, t)) \in \mathbb{R}^2 : \sum_{j=1}^{k} (a_j x + b_j)\, I_{(c_{j-1},\, c_j]} \geqslant t \geqslant 0 \right\}$$
$$= \bigcup_{i=1}^{2k} (\{(x, t)\colon c_j \geqslant x\} \cap \{(x, t)\colon x > c_{j-1}\}$$
$$\cap \{(x, t)\colon a_j x + b_j \geqslant t\} \cap \{(x, t)\colon t \geqslant 0\})$$

which is obtained by unions and intersections of sets of the form

$$\left\{ (x, t) \in \mathbb{R}^2 : \sum_{j=1}^{m} s_j f_j(x, t) \geqslant 0 \right\},$$

$s_1, ..., s_m \in \mathbb{R}$, for some functions $f_1, ..., f_m$ and some $m < \infty$. So, by Proposition 9.2.5 and Theorem 9.2.1 in Dudley (1984) the class of sets

$$\left\{ (x, t)\colon \sum_{j=1}^{2k} (a_j x + b_j)\, I_{(c_{j-1},\, c_j]} \geqslant t \geqslant 0 \right\}, \qquad a_j, b_j, c_j \in \mathbb{R}$$

is a VC class. Since a similar argument applies to

$$\left\{ (x, t) \in \mathbb{R}^2 : \sum_{j=1}^{2k} (a_j x + b_j)\, I_{(c_{j-1},\, c_j]} \leqslant t \leqslant 0 \right\},$$

the class $\{g(x, \theta)-g(x, \theta_0) : \theta \in \mathbb{R}^k\}$ is a VC subgraph class.

From (5.1),

$$|r(x, \theta)| \leqslant |\theta-\theta_0|^2 + c\,|\theta-\theta_0| \sum_{j=2}^{k-1} I_{(2^{-1}(\theta_0^{(j-1)}+\theta_0^{(j)}),\, 2^{-1}(\theta^{(j-1)}+\theta^{(j)})]}$$

which gives conditions (A.7)–A.9). ∎

A better estimator of k clusters is the M-estimator over $g(x, \theta) = \min_{1 \leqslant i \leqslant k} |x-\theta^{(i)}|$. The proof of the next theorem is omitted, since it is similar to that of the previous one.

THEOREM 16. *Let* $\{X_j\}_{j=1}^{\infty}$ *be a sequence of i.i.d.r.v.'s. Let* $g(x, \theta) = \min_{1 \leqslant i \leqslant k} |x-\theta^{(i)}|$, *where* $\theta = (\theta^{(1)}, ..., \theta^{(k)})'$. *Suppose that:*

(i) *There exists a* $\theta_0 \in \mathbb{R}^k$ *with* $\theta_0^{(1)} < \cdots < \theta_0^{(k)}$ *such that* $E[g(X, \theta) - g(X, \theta_0)] > 0$, *for each* $\theta \neq \theta_0$ *with* $\theta^{(1)} \leqslant \cdots \leqslant \theta^{(k)}$.

(ii) *The distribution function* F_X *of* X *is differentiable at* $\theta_0^{(j, j+1)} := 2^{-1}(\theta_0^{(j)} + \theta_0^{(j+1)})$, *for* $1 \leqslant j \leqslant k-1$, *and at* $\theta_0^{(j)}$, *for* $1 \leqslant j \leqslant k$.

(iii) $2F'_X(\theta_0^{(1)}) + F'_X(\theta_0^{(1,2)}) > 0$, *and* $2F'_X(\theta_0^{(k)}) + F'_X(\theta_0^{(k-1,k)}) > 0$ *and, for* $2 \leqslant j \leqslant k-1$, $2F'_X(\theta_0^{(j)}) + F'_X(\theta_0^{(j-1,j)}) + F'_X(\theta_0^{(j.j+1)}) > 0$, *for* $2 \leqslant j \leqslant k-1$, *and* $2F'_X(\theta_0^{(k)}) + F'_X(\theta_0^{(k-1,k)}) > 0$.

Then, (2.1) *holds with*

$$\phi(x) = (\mathrm{sign}(\theta_0^{(1)} - x)\, I_{(-\infty, \theta_0^{(1,2)}]}, \mathrm{sign}(\theta_0^{(2)} - x)\, I_{(\theta_0^{(1,2)}, \theta_0^{(2,3)}]}, \ldots)',$$

and V *is the symmetric matrix determined by*

$$\theta' V \theta = \sum_{j=1}^{k} 2(\theta^{(j)})^2 F'_X(\theta_0^{(j)}) + \sum_{j=1}^{k-1} ((\theta^{(j)})^2 + (\theta^{(j+1)})^2)\, F'_X(\theta_0^{(j,j+1)}).$$

REFERENCES

1. M. A. Arcones, L_p-estimators as a parameter of location for a sharp-pointed symmetric density, *Scandin. J.* **25** (1998), 693–715.
2. M. A. Arcones, Weak convergence for the row sums of a triangular array of empirical processes indexed by a manageable triangular array of functions, *Elec. J. Probab.* **4**, 1–17.
3. Z. D. Bai, Y. H. Wu, X. R. Chen, and B. Q. Miao, On solvability of an equation arising in the theorem of M-estimates, *Commun. Statist. Theory Meth.* **19** (1990), 363–380.
4. H. Cramér, "Mathematical Methods of Statistics," Princeton Univ. Press, Princeton, NJ, 1946.
5. J. A. Cuesta and C. Matran, The strong law of large numbers for k-means and best possible nets of Banach valued random variables, *Probab. Theor. Rel. Fields.* **78** (1988), 523–534.
6. H. E. Daniels, The asymptotic efficiency of a maximum likelihood estimator, *in* "Proceedings of the Fourth Berkeley Simposium on Mathematical Statistics and Probability," Vol. 1, pp. 151–163, 1991.
7. R. M. Dudley, A course on empirical processes, *in* "Lecture Notes in Mathematics," Vol. 1097, pp. 1–142, Springer-Verlag, New York, 1984.
8. E. Giné and J. Zinn, Some limit theorems for empirical processes, *Ann. Probab.* **12** (1984), 929–989.
9. J. A. Hartigan, Asymptotic distributions for clustering criteria, *Ann. Statist.* **6** (1978), 117–131.
10. J. Hoffmann-Jørgensen, "Probability with a View toward Statistics," Vol. II, Chapman & Hall, New York, 1994.
11. P. J. Huber, Robust estimation of a location parameter, *Ann. Math. Statist.* **35** (1964), 73–101.
12. P. J. Huber, The behaviour of maximum likelihood estimates under non-standard conditions, *in* "Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability," Vol. 1, pp. 221–233, Univ. of California Press, Berkeley, 1967.
13. I. A. Ibragimov and R. Z. Has'minskii, "Statistical Estimation: Asymptotic Theory," Springer-Verlag, New York, 1981.

14. J. Jurečková and P. K. Sen, "Robust Statistical Procedures. Asymptotics and Interrelations," Wiley, New York, 1996.
15. M. J. Klass, Precision bounds for the relative error in the approximation of $E\,|S_n|$ and extensions, *Ann. Probab.* **8** (1980), 350–367.
16. L. Le Cam, On the assumptions used to prove the asymptotic normality of maximum likelihood estimates, *Ann. Math. Statist.* **41** (1970), 802–828.
17. M. B. Marcus and G. Pisier, "Random Fourier Series with Applications to Harmonic Analysis," Ann. Mathem. Studies, Vol. 101, Princeton Univ. Press, Princeton, NJ, 1981.
18. M. Ossiander, A central limit theorem under entropy with L_2 bracketing, *Ann. Probab.* **15** (1987), 897–919.
19. D. Pollard, "Convergence of Stochastic Processes," Springer-Verlag, New York, 1984.
20. D. Pollard, New ways to prove central limit theorems, *Econ. Theor.* **1** (1985), 295–314.
21. R. J. Serfling, "Approximation Theorems of Mathematical Statistics," Wiley, New York, 1980.
22. A. W. van der Vaart and J. A. Wellner, "Weak Convergence and Empirical Processes with Applications to Statistics," Springer-Verlag, New York, 1996.