BOL. SOC. BRAS. MAT., VOL. 10 Nº 1 (1979), 27-42

Asymptotic behavior of iterative M-estimartors for location

Ruben Klein and Victor J. Yohai

Abstract.

In this paper, we study the Huber M-estimator for location when they are computed by a class of numerical iterative procedures. This class includes the usual method of Newton-Raphson, iterated weighted least squares and iterated winsorization. We show that under mild conditions, the numerical iterative procedures converge and the resulting estimators are consistent and asymptotically normal.

1. Introduction. Let $x_1, x_2, \ldots, x_n$ be i.i.d. random variables with distribution $F(x-\theta)$. Huber (1964) proposed the class of M-estimators for the location parameter θ, defined by the solution of the following equation:

$$\sum_{i=1}^{n} \psi(x_i - t) = 0 \tag{1.1}$$

where ψ is any function such that

$$\int \psi(x)\, dF(x) = 0. \tag{1.2}$$

In particular, if F is symmetric and ψ is odd and F-integrable, (1.2) is automatically satisfied.

Huber studied the asymptotic behavior and showed that if ψ is conveniently chosen, the resulting estimator has robustness properties. He also proved that if ψ is monotone non-decreasing, then under mild conditions, the resulting estimator is consistent a.s. and asymptotically normal with mean θ and asymptotic variance $V(\psi, F)/n$ where

$$V(\psi, F) = \int \psi^2(x) dF(x) / [\int \psi'(x) dF(x)]^2. \tag{1.3}$$

In particular, if F is unknown, but it is in some sense in a neighborhood of a normal distribution, Huber (1964) and Hampel (1968) showed that the estimators, based on the class of functions given by

---

Recebido em 27/9/78

$$\psi_k(t) = \min(|t|, k)\,\text{sgn}\, t, \tag{1.4}$$

have desirable optimal robustness properties. These functions are usually called Huber functions.

Hampel (1974) considering other robustness properties, like the influence curve, proposed to use ψ-functions that vanish outside a compact interval. Collins (1976) also suggested the use of such ψ-functions by looking at estimators that behave well under distributions F that have asymmetric tails. As examples, we hare:

(i) the Hampel functions given by

$$\psi_{a,b,c}(x) = \begin{cases} x & \text{if } |x| \le a \\ a\, sgn\, x & \text{if } a \le |x| \le b \\ a(c - |x|)\, sgn\, x/(c-b) & \text{if } b \le |x| \le c \\ 0 & \text{if } |x| \ge c, \end{cases} \tag{1.5}$$

and (ii) the sine functions given by

$$\psi_k(x) = \begin{cases} \sin(x/k) & \text{if } |x| \le k\pi \\ 0 & \text{otherwise.} \end{cases} \tag{1.6}$$

Numerical studies have shown that these estimators have good properties of efficiency and robustness, even for small sample sizes (see Andrews et al (1971)).

In these cases, the equation

$$\int \psi(x - t)\, dF(x) = 0 \tag{1.7}$$

may have solutions different from 0. Then some solutions of equation (1.1) may converge to $\theta + t_0$, where t_0 is a root of equation (1.7) different from 0. Therefore, in order to define the estimate, we have specify which solution of equation (1.1) we are considering. This can be done by indicating the numerical algorithm used to compute it.

Usually, the algorithm is an iterative procedure of the following form:

$$\hat{\theta}_{n,j+1} = \hat{\theta}_{n,j} + \left[\sum_{i=1}^{n} \psi(x_i - \hat{\theta}_{n,j}) \Big/ \sum_{i=1}^{n} r(x_i - \hat{\theta}_{n,j})\right] \tag{1.8}$$

and

$$\hat{\theta}_n = \begin{cases} \lim_{j\to\infty} \hat{\theta}_{n,j} & \text{if limit exists} \\ \hat{\theta}_{n,0} & \text{otherwise,} \end{cases} \tag{1.9}$$

where the function $r(x)$ is conveniently chosen and the iteration is started from some initial estimate $\hat{\theta}_{n,0}$, usually a simple, reasonable, robust estimate for instance, the sample median.

As examples of the above iterative procedure, we have:

(i) Newton-Raphson procedure: $r(x) = \psi'(x)$. Then

$$\hat{\theta}_{n,j+1} = \hat{\theta}_{n,j} + \left[\sum_{i=1}^{n} \psi(x_i - \hat{\theta}_{n,j}) \Big/ \sum_{i=1}^{n} \psi'(x_i - \hat{\theta}_{n,j})\right]. \tag{1.10}$$

(ii) The weighted mean iterative procedure: $r(x) = \psi(x)/x$. Then

$$\begin{aligned} \hat{\theta}_{n,j+1} &= \sum_{i=1}^{n} r(x_i - \hat{\theta}_{n,j}) x_i \Big/ \sum_{i=1}^{n} r(x_i - \hat{\theta}_{n,j}) = \\ &= \hat{\theta}_{n,j} + \left[\sum_{i=1}^{n} \psi(x_i - \hat{\theta}_{n,j}) \Big/ \sum_{i=1}^{n} r(x_i - \hat{\theta}_{n,j})\right], \end{aligned} \tag{1.11}$$

and

(iii) $r(x) = \gamma$, a constant. In the particular case of the Huber functions and $\gamma = 1$, this corresponds to iterative winsorization given by:

$$\hat{\theta}_{n,j+1} = (1/n) \sum_{i=1}^{n} (\hat{\theta}_{n,j} + \psi(x_i - \hat{\theta}_{n,j})) = \hat{\theta}_{n,j} + \sum_{i=1}^{n} \psi(x_i - \hat{\theta}_{n,j})/n. \tag{1.12}$$

The convergence of the iterative procedures (1.8)-(1.9) to a solution of equation (1.1) and the asymptotic properties of the resulting estimator were studied only (Collins (1976)) for the case of the Newton-Raphson and with strong conditions on ψ and F. F is assumed to be normal in the central part and with arbitrary tails, and ψ should be continuously differentiable in any point and vanish outside the interval where F is normal.

In this paper, we prove in Theorem 2.1, the convergence of the iterative procedure (1.8)-(1.9) and the consistency a.s. of the resulting estimator under very mild conditions on ψ, r and F, and assuming that $|\rho| = |1 - (\int \psi'(x) dF(x) / \int r(x) dF(x))| < 1$. This last assumption assures that the transformation given by (1.8) is asymptotically, a contraction, and $|\rho|$ is a measure of the speed of convergence of the algorithms. For the Newton-Raphson method, $\rho = 0$, and therefore this method has the maximum speed of convergence.

In Theorem 2.2, using a theorem of Huber (1967), we show that these estimators are asymptotically normal with mean θ and variance $V(\psi, F)/n$.

These estimators are location equivariant, i.e.,

$$\hat{\theta}(x_1 + b, \ldots, x_n + b) = \hat{\theta}(x_1, \ldots, x_n) + b,$$

but are not scale equivariant, i.e., they do not satisfy

$$\hat{\theta}(ax_1, \ldots, ax_n) = a\,\hat{\theta}(x_1, \ldots, x_n)$$

as used by Berk (1967) and Bickel (1975). So these estimators are not reasonable when scale is unknown.

One way of getting location and scale equivariant estimation of θ is by modifying equation (1.1) to

$$\sum_{i=1}^{n} \psi[(x_i - t)/s_n] = 0, \tag{1.13}$$

where s_n is a location invariant and scale equivariant estimate of scale, i.e., $s_n(ax_1 + b, \ldots, ax_n + b) = |a|\, s_n(x_1, \ldots, x_n)$. For instance, s_n may be the normalized interquartile range,

$$\hat{\sigma}_1 = (X_{(n-[n/4]+1)} - X_{([n/4])})/2\Phi^{-1}(3/4),$$

or the normalized median deviation

$$\hat{\sigma}_2 = \text{median}\,(|\, x_i - m\,|)/\Phi^{-1}(3/4),$$

where $X_{(1)} < \ldots < X_{(n)}$ are the order statistics, Φ the standard normal distribution and m the sample median.

To solve (1.13), the iterative procedure is modified to:

$$\tilde{\theta}_{n,j+1} = \tilde{\theta}_{n,j} + s_n \sum_{i=1}^{n} \psi[(x_i - \tilde{\theta}_{n,j})/s_n] \Big/ \sum_{i=1}^{n} r[(x_i - \tilde{\theta}_{n,j})/s_n] \tag{1.14}$$

and

$$\hat{\theta}_n = \begin{cases} \lim_{j\to\infty} \tilde{\theta}_{n,j} & \text{if limit exists} \\ \tilde{\theta}_{n,0} & \text{otherwise.} \end{cases} \tag{1.15}$$

In Theorems 2.3 and 2.4, we prove the consistency and asymptotic normality of these estimators.

Another possibility is Huber's proposal 2 (Huber (1964)), which estimates simultaneously location and scale, by solving the following system:

$$\sum_{i=1}^{n} \psi_1[(x_i - t)/s] = 0$$

$$\sum_{i=1}^{n} \psi_2[((x_i - t)/s)^2] = n. \tag{1.16}$$

In this case, the class of iterative algorithms for computing the solution of (1.16) is given by:

$$\tilde{\theta}_{n,j+1} = \tilde{\theta}_{n,j} + s_{n,j} \sum_{i=1}^{n} \psi_1[(x_i - \tilde{\theta}_{n,j})/s_{n,j}] \Big/ \sum_{i=1}^{n} r[(x_i - \tilde{\theta}_{n,j})/s_{n,j}]$$

$$s^2_{n,j+1} = (1/n)s^2_{n,j} \sum_{i=1}^{n} \psi_2[((x_i - \tilde{\theta}_{n,j})/s_{n,j})^2], \tag{1.17}$$

and the estimators are defined analogously to (1.9).

In this paper, the properties of these estimators are not studied but under more cumbersome regularity conditions on ψ, consistency and asymptotic normality may be proven by methods similar to those used in this paper.

2. Let $x_1, x_2, \ldots, x_n$ be i.i.d random variables with distribution $F(x-\theta)$. Let $\hat{\theta}_n$ be the estimator defined by (1.8) and (1.9).

Consider the following set of assumptions.

A.1 – ψ is continuous and has a continuous derivative ψ' except in a finite number of points. When ψ' is not defined at x, we put arbitrarily $\psi'(x) = 0$.

A.2 – F is continuous and $\int \psi(x)dF(x) = 0$.

A.3 – There exists $\delta_0 > 0$ such that

(i) $\sup_{|a| \le \delta_0} |\psi(x - a)|$ is F-integrable,

(ii) $\sup_{|a| \le \delta_0} |\psi'(x - a)|$ is F-integrable,

(iii) $\sup_{|a| \le \delta_0} |r(x - a)|$ is F-integrable.

A.4 – $r(x)$ is continuous except in a finite number of points and

$$\gamma_0 = \int r(x)dF(x) \neq 0. \tag{2.1}$$

A.5 – Put $\rho = 1 - (\int \psi'(x)dF(x)/\int r(x)dF(x))$. Then $|\rho| < 1$.

A.6 – The initial estimator $\hat{\theta}_{n,0}$ is location equivariant and consistent a.s.

A.7 – $\int \psi^2(x)dF(x) < \infty$ and ψ' is bounded.

We can state the following theorems.

Theorem 2.1. *Assume* A.1 *to* A.6. *Then*

(i) $\hat{\theta}_n$ *converges to* θ *a.s.*
(ii) *With probability one, there exists a randon number* n_0 *such that for* $n \geq n_0$ $\hat{\theta}_n = \lim_{j \to \infty} \hat{\theta}_{n,j}$ *and* $\hat{\theta}_n$ *satisfies* (1.1).

Theorem 2.2. *Assume* A.1 *to* A.7. *Then* $\sqrt{n}(\hat{\theta}_n - \theta)$ *converges in distribution to* $\mathcal{N}(0, V(\psi, F))$.

Consider now the location scale equivariant estimator $\tilde{\theta}_n$ defined by (1.14) and (1.15) and the following set of assumptions.

B.0 – s_n is location invariant and scale equivariant estimator of scale and s_n converges a.s. to $s_0 > 0$.
B.1 – The same as A.1.
B.2 – F is continuous and $\int \psi(x/s_0)dF(x) = 0$.
B.3 – There exists $\delta_0 > 0$ and $\delta_1 > 0$ such that
(i) $\sup_{|a| \leq \delta_0, |b - s_0| \leq \delta_1} |\psi((x-a)/b))|$ is F-integrable,
(ii) $\sup_{|a| \leq \delta_0, |b - s_0| \leq \delta_1} |\psi'((x-a)/b)|$ is F-integrable,
(iii) $\sup_{|a| \leq \delta_0, |b - s_0| \leq \delta_1} |r((x-a)|b)|$ is F-integrable.
B.4 – $r(x)$ is continuous except in a finite number of points and

(2.2) $$\gamma_0 = \int r(x/s_0)dF(x) \neq 0.$$

B.5 – Put $\rho = 1 - (\psi'(x/s_0)dF(x)/\int r(x/s_0)dF(x))$. Then $|\rho| < 1$.
B.6 – The initial estimator $\tilde{\theta}_{n,0}$ is location scale equivariant and is consistent a.s.
B.7 – $\int \psi^2(x/s_0)dF(x) < \infty$ and ψ is bounded.
B.8 – $\sqrt{n}(s_n - s_0)$ is bounded in probability.
B.9 – $\int \psi'(x/s_0)x\,dF = 0$.
B.10 – There exist $\delta_0 > 0$ and $\delta_1 > 0$ such that $\sup_{|a| < \delta_0, |b - s_0| < \delta_1} |\psi'((x-a)/b)|$ is F-integrable.

We have the following theorems.

Theorem 2.3. *Assume* B.0 *to* B.6. *Then*
(i) $\tilde{\theta}_n$ *converges to* θ *a.s.*
(ii) *Whit probability one, there exists a random number* n_0 *such that for* $n \geq n_0$ $\tilde{\theta}_n = \lim_{j \to \infty} \tilde{\theta}_{n,j}$ *and* $\tilde{\theta}_n$ *satisfies* (1.13).

Theorem 2.4. *Assume* B.0 *to* B.10. *Then* $\sqrt{n}(\tilde{\theta}_n - \theta)$ *converges in distribution to* $\mathcal{N}(0, V(\psi(\cdot/s_0), F))$.

Since all our estimation is location equivariant, from now on, we assume $\theta = 0$. In the case our estimation is location scale equivariant, we also assume $s_0 = 1$.

In order to prove these theorems, the following lemmas will be necessary.

Lemma 2.1. (Yohai (1974)) *Let* $U_1, U_2, \ldots$ *be a sequence of i.i.d. random variables. Let* C *be a compact space, and* $(f_k)_{k \in C}$, *a family of Borel measurable real functions such that*
(i) $|f_k| \leq f$ *where* $E(f(U_1)) < \infty$,
(ii) $\lim_{i \to k} f_i(U_1) = f_k(U_1)$ *a.s. for all* k *in* C,
(iii) $|E(f_k(U_1))| \leq A$ *for all* k *in* C.

Then

$$\limsup_{n \to \infty} \sup_{k \in C} \left| \sum_{j=1}^{n} f_k(U_j)/n \right| \leq A \text{ a.s.}$$

Lemma 2.2. *Let* $f: C \to C^*$ *be a continuous function, where* C *is a convex subset of a normed vector space and* C^* *is a normed vector space. Assume that there exists a finite family of convex sets* $(C_i)_{1 \leq i \leq m}$ *such that* $\bigcup_{i=1}^{m} C_i$ *is dense in* C *and such that* f *satisfies a Lipschitz condition in each* $C_i \cap C$, *i.e., there exists* $k_0 \geq 0$, *not depending on* C_i, *such that* $\|f(x) - f(x')\| \leq \leq k_0 \|x - x'\|$ *for all* x, x' *in* $C_i \cap C$, $1 \leq i \leq m$. *Then* f *satisfies a Lipschitz condition in* C *with same constant* k_0.

Proof. Take x, x' in $\bar{C}_i \cap C$ ($\bar{C}_i$ denotes the closure of C_i), then by the continuity of f, we have.

(2.3) $$\|f(x) - f(x')\| \leq k_0 \|x - x'\|.$$

Take now any two points x, x' in C. If x, x' belong to the same $\bar{C}_i$, (2.3) is satisfied. So let us assume that x, x' do not belong to the same $\bar{C}_i$, $1 \leq i \leq m$. Consider the line segment S joining these two points.

Define $x_0 = x$ and let C_{i_0} be any set such that $x_0 \in \bar{C}_{i_0}$. Let x_1 be the last point in S such that x_1 is in $\bar{C}_{i_0}$ and let C_{i_1} be any set different from C_{i_0} such that x_1 is also in $\bar{C}_{i_1}$. We may define in this way, a sequence $x_0, x_1, \ldots, x_k$ in S such that $x_0 = x$ belong to $\bar{C}_{i_0}$, x_j belongs to $\bar{C}_{j-1} \cap \bar{C}_j$, $2 \leq j \leq k-1$, and $x_k = x'$ belongs to $\bar{C}_{k-1}$. This sequence should be finite by the convexity of the C_i's. Then by (2.3), we have

$$\|f(x) - f(x')\| = \|f(x_0) - f(x_k)\| \leq$$
$$\leq \sum_{j=1}^{k} \|f(x_j) - f(x_{j-1})\| \leq k_0 \sum_{j=1}^{k} \|x_i - x_i'\| = k_0 \|x_i - x_i'\|,$$

the last equality holding by the alignment of $x_0, x_1, \ldots, x_k$.

Lemma 2.3. *Let (S, d) be a metric space. Assume that each γ in Γ, an arbitrary set, $\eta_\gamma : S \to S$ is a contraction uniformly in γ, i.e., there exists $0 \leq k_0 < 1$ such that*

$$d(\eta_\gamma(x), \eta_\gamma(x')) \leq k_0\, d(x, x') \;\forall\; \gamma \text{ in } \Gamma.$$

Assume also that all η_γ have the same fixed point x^*. Take any sequence $\gamma_1, \gamma_2, \ldots$ in Γ and any $x_0 \in S$. Define inductively x_k by $x_k = \eta_{\gamma_k}(x_{k-1})$. Then x_n converges to x^*.

Proof. We will prove by induction that

$$d(x_n, x^*) \leq k_0^n\, d(x_0, x^*). \tag{2.4}$$

For $n = 0$, (2.4) is automatically satisfied.

Assume (2.4) true for n, then

$$d(x_{n+1}, x^*) = d(\eta_{\gamma_{n+1}}(x_n), \eta_{\gamma_{n+1}}(x^*)) \leq k_0\, d(x_n, x^*) \leq k_0^{n+1}\, d(x_0, x^*)$$

Let us define for all t in $\mathbb{R}$, all γ in $\mathbb{R}$, all $n \geq 1$, the following transformations from $\mathbb{R}$ into $\mathbb{R}$.

$$\eta_n(t) = t + \sum_{i=1}^{n} \psi(x_i - t) \Big/ \sum_{i=1}^{n} r(x_i - t) \tag{2.5}$$

$$\eta_{n\gamma}(t) = t + \sum_{i=1}^{n} \psi(x_i - t)/n\gamma, \tag{2.6}$$

$$\eta_n^*(t) = t + s_n \sum_{i=1}^{n} \psi((x_i - t)/s_n) \Big/ \sum_{i=1}^{n} r((x_i - t)/s_n), \tag{2.7}$$

$$\eta_{n\gamma}^*(t) = t + s_n \sum_{i=1}^{n} \psi((x_i - t)/s_n)/n\gamma, \tag{2.8}$$

where s_n is a scale estimator.

Let us also define

$$\gamma_n(t) = \sum_{i=1}^{n} r(x_i - t)/n \tag{2.9}$$

$$\gamma_n^*(t) = \sum_{i=1}^{n} r((x_i - t)/s_n)/n. \tag{2.10}$$

Lemma 2.4. *Let Γ be any subset of $\mathbb{R}$ bounded away from zero.*

(a) *Assume* A.1, A.2, A.3(i) *Then*

$$\sup_{\gamma \in \Gamma} |\eta_{n\gamma}(0)| \to 0 \text{ a.s.},$$

(b) *Assume* B.0, B.1, B.2 *and* B.3(i). *Then*

$$\sup_{\gamma \in \Gamma} |\eta_{n\gamma}^*(0)| \to 0 \textit{ a.s.}$$

Proof. Call $d = \inf\{|\gamma|, \gamma \in \Gamma\}$. Then $d > 0$.

(a) $\sup_{\gamma \in \Gamma} |\eta_{n\gamma}(0)| = \sup_{\gamma \in \Gamma} \left|\sum_{i=1}^{n} \psi(x_i)/n\gamma\right| \leq d^{-1} \left|\sum_{i=1}^{n} \psi(x_i)/n\right|.$

But A.2 implies $\sum_{i=1}^{n} \psi(x_i)/n \to 0$ a.s. Then (a) follows.

(b) $\sup_{\gamma \in \Gamma} |\eta_{n\gamma}^*(0)| = \sup_{\gamma \in \Gamma} s_n \left|\sum_{i=1}^{n} \psi(x_i/s_n)/n\gamma\right| \leq d^{-1} s_n \left|\sum_{i=1}^{n} \psi(x_i/s_n)/n\right|.$

By assumption B.0, s_n is bounded by above, so it is enough to show

$$\limsup_{n \to \infty} \left|\sum_{i=1}^{n} \psi(x_i/s_n)/n\right| = 0 \textit{ a.s.}$$

Using B.0 again, it is enough to show that given $\mu > 0$, there exists $\delta > 0$ such that $|s - 1| < \delta$ implies

$$\limsup_{n \to \infty} \sup_{|s-1| \leq \delta} \left|\sum_{i=1}^{n} \psi(x_i/s)/n\right| \leq \mu \textit{ a.s.}$$

By B.1, B.2, B.3(i) and dominated convergence, given $\mu > 0$, there exists $\delta > 0$ such that $|s - 1| < \delta$ implies $|E(\psi(x_i/s))| < \mu$. Applying Lemma 1, the result follows.

Lemma 2.5. (a) *Assume* A.1, A.2, A.3 *and* A.4 *Then for every $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that*

(i) $\lim_{n \to \infty} P\left(\bigcup_{m \geq n} \left\{ \sup_{|t| \leq \delta(\varepsilon)} \left| \left(\sum_{i=1}^{m} \psi'(x_i - t)/m\right) - \int \psi'(x)dF(x) \right| \geq \varepsilon \right\}\right) = 0,$

(ii) $\lim_{n\to\infty} P\left(\bigcup_{m\geq n}\left\{\sup_{|t|\leq\delta(\varepsilon)}\left|\left(\sum_{i=1}^{m} r(x_i-t)/m\right)-\int r(x)dF(x)\right|\geq\varepsilon\right\}\right)=0.$

(b) *Assume* B.0, B.1, B.2, B.3 *and* B.4. *Then for every* $\varepsilon>0$, *there exists* $\delta(\varepsilon)>$ *such that*

(i) $\lim_{n\to\infty} P\left(\bigcup_{m\geq n}\left\{\sup_{|t|\leq\delta(\varepsilon)}\left|\left(\sum_{i=1}^{m}\psi'((x_i-t)s_m)/m\right)-\int\psi'(x)dF(x)\right|\geq\varepsilon\right\}\right)=0,$

(ii) $\lim_{n\to\infty} P\left(\bigcup_{m\geq n}\left\{\sup_{|t|\leq\delta(\varepsilon)}\left|\left(\sum_{i=1}^{m} r((x_i-t)/s_m)/m\right)-\int r(x)dF(x)\right|\geq\varepsilon\right\}\right)=0.$

Proof. (a) (i). Put $f_t(x)=\psi'(x-t)-\int\psi'(x)dF(x)$. Using A.1, A.2, A.3 and dominated convergence, given $\varepsilon>0$, there exists $\delta(\varepsilon)>0$ such that $|Ef_t(x)|\leq\varepsilon/2\ \forall\,|t|\leq\delta(\varepsilon)$. Since $f_t(x)$ satisfies other conditions of Lemma 1, we have that

$$\limsup_{n\to\infty}\ \sup_{|t|\leq\delta(\varepsilon)}\left|(1/n)\sum_{i=1}^{n}\psi'(x_i-t)-\int\psi'(x)dF(x)\right|\leq\varepsilon/2\ a.s.$$

So (i) follows.
Proof of (a)(ii) is analogous.
Proof of (b) is similar to (a). For instance, in (i), put

$$g_{t,s}(x)=\psi'((x-t)/s)-\int\psi'(x)dF(x),\ \text{instead of } f_t(x).$$

Lemma 2.6. (a) *Assume* A.1 *to* A.4. *Then for each sample* $x_1,\ldots,x_n$, *there exists a finite number of random open intervals* $D_1,\ldots,D_n$ *such that* $\bigcup_{j=1}^{u}\bar D_j=\mathbb{R}$ *and such that* $\eta_{n\gamma}(t)$ *is continuously differentiable in* D_j, $1\leq j\leq u$, *for all* $\gamma\neq 0$ *with derivative given by*

$$\eta'_{n\gamma}(t)=1-(1/n\gamma)\sum_{i=1}^{n}\psi'(x_i-t).$$

Moreover, if $k_0>|\rho|$, *there exists* $\varepsilon_1>0$ *and* $\delta_1>0$ *depending on* k_0 *such that*

$$\lim_{n\to\infty} P\left(\bigcup_{m\geq n}\{\sup\{|\eta'_{n\gamma}(t)|;|t|\leq\delta_1, t\in\bigcup_{j=1}^{u}D_j,|\gamma-\gamma_0|\leq\varepsilon_1\}>k_0\}\right)=0.$$

(b) *Assume* B.0 *to* B.4. *Then* (a) *holds with* $\eta_{n\gamma}$ *replaced by* $\eta^*_{n\gamma}$.

Proof. (a) According to A.1, there exist only a finite number of points $a_1,\ldots,a_k$ where ψ is not continuously differentiable. Put $z_{ij}=x_i-a_j$, $1\leq i\leq n$, $1\leq j\leq k$, and call $z_1\leq\ldots\leq z_{nk}$, the z_{ij}'s ordered. Put $z_0=-\infty$ and $z_u=+\infty$, where $u=nk+1$. Let $D_i=(z_{i-1},z_i)$, $1\leq i\leq u$. Then in each D_i, $\eta_{n\gamma}$ is continuously differentiable with derivate

$$\eta'_{n\gamma}(t)=1-(1/n\gamma)\sum_{i=1}^{n}\psi'(x_i-t)$$

and $\bigcup_{i=1}^{u}\bar D_i=\mathbb{R}$.

Now let $h=k_0-|\rho|$, $\varepsilon_1=\min\{|\gamma_0|/2,\ h\gamma_0^2/4\int\psi'(x)dF(x)\}$ and $\mu=h|\gamma_0|/4$. Take $\delta_1=\delta(\mu)$ as in Lemma 2.5.

Let us show that

(2.11) $$\{\sup\{|(1/n)\sum_{i=1}^{n}\psi'(x_i-t)-\int\psi'(x)dF(x)|;|t|\leq\delta_1\}\leq\mu\}\subset$$
$$\subset\{\sup\{|\eta'_{n\gamma}(t)-\rho|;|t|\leq\delta_1, t\in\bigcup_{j=1}^{u}D_j,|\gamma-\gamma_0|\leq\varepsilon_1\}\leq h\}\subset$$
$$\subset\{\sup\{|\eta'_{n\gamma}(t)|;|t|\leq\delta_1, t\in\bigcup_{j=1}^{u}D_j,|\gamma-\gamma_0|\leq\varepsilon_1\}\leq k_0\}.$$

The second inclusion is trivial, so let us show the first one. Assume the first event in (2.11) holds. Then for $t\in\bigcup_{j=1}^{u}D_j$, we have

$$|\eta'_{n\gamma}(t)-\rho|=|(1-(1/n\gamma)\sum_{i=1}^{n}\psi'(x_i-t))-(1-(1/\gamma_0)\int\psi'(x)dF(x))|\leq$$
$$\leq|\int\psi'(x)dF(x)|\,|\gamma_0-\gamma|/|\gamma_0|\,|\gamma|+(1/|\gamma|)\,|(1/n)\sum_{i=1}^{n}\psi'(x_i-t)-\int\psi'(x)dF(x)|.$$

Then using the definition of ε_1 and μ, we have

$$\sup\{|\eta_{n\gamma}(t)-\rho|;|t|\leq\delta_1, t\in\bigcup_{j=1}^{u}D_j,|\gamma-\gamma_0|\leq\varepsilon_1\}\leq$$
$$\leq 2\varepsilon_1|\int\psi'(x)dF(x)|/\gamma_0^2+2\mu/|\gamma_0|\leq(h/2)+(h/2)=h.$$

Now the result follows from Lemma 2.5 a(i).

Proof of (b). The same as (a), but using Lemma 2.5 b(i).

Lemma 2.7. *Let* $1>k_0|\rho|$. *Let*

$$A_{n,\delta,\varepsilon}=\{|\eta_{n\gamma}(t)-\eta_{n\gamma}(t')|\leq k_0|t-t'|\ \forall\,|t|\leq\delta,|t'|\leq\delta\ \text{and}\ \forall\,|\gamma-\gamma_0|\leq\varepsilon\},$$
$$B_{n,\delta,}=\{\eta_{n\gamma}\ \text{takes}\ \{|t|\leq\delta\}\ \text{into itself}\ \forall\,|\gamma-\gamma_0|\leq\varepsilon\}\ \text{and}$$
$$C_{n,\delta,\varepsilon}=\{\gamma_n\ \text{takes}\ \{|t|\leq\delta\}\ \text{into}\ \{|\gamma-\gamma_0|\leq\varepsilon\}\}.$$

Define $A^*_{n,\delta,\varepsilon}$, $B^*_{n,\delta,\varepsilon}$, $C^*_{n,\delta,\varepsilon}$ *similarly replacing* $\eta_{n\gamma}$ *by* $\eta^*_{n\gamma}$.

(a) *Assume* A.1, A.2, A.3 *and* A.4. *Then there exist* $\varepsilon_2 > 0$ *and* $\delta_2 > 0$ *such that*

(i) $$\lim_{n\to\infty} P\left(\bigcap_{m\geq n} A_{m,\delta,\varepsilon_2}\right) = 1 \;\forall\, \delta \leq \delta_2,$$

(ii) $$\lim_{n\to\infty} P\left(\bigcap_{m\geq n} B_{m,\delta,\varepsilon_2}\right) = 1 \;\forall\, \delta \leq \delta_2,$$

(iii) $$\lim_{n\to\infty} P\left(\bigcap_{m\geq n} C_{m,\delta,\varepsilon_2}\right) = 1 \;\forall\, \delta \leq \delta_2.$$

(b) *Assume* B.0, B.1, B.2, B.3 *and* B.4. *Then* (i), (ii), (iii) *hold replacing* A, B, C *by* A*, B*, C*.

Proof. (a) Let $\varepsilon_2 = \varepsilon_1$ defined in Lemma 2.6 and let $\delta_2 = \min(\delta_1, \delta(\varepsilon_2))$ where δ_1 and $\delta(\varepsilon_2)$ are defined in Lemmas 2.6 and 2.5 respectively.

(i) According to Lemma 2.6 (a), to prove (i), it is enough to show (2.12):

$$A_{n,\delta,\varepsilon_2} \supset \{\sup\{|\eta'_{n\gamma}(t)|;|t|\leq\delta_1, t\in\bigcup_{j=1}^{u} D_j, |\gamma-\gamma_0|\leq\varepsilon_1\}\leq k_0\}\;\forall\,\delta\leq\delta_2.$$

If the event in the right-hand side of (2.12) occurs, then for any $|\gamma - \gamma_0| \leq \varepsilon_1$, we have that for all $t \in D_j \cap [-\delta, \delta]$, $|\eta'_{n\gamma}(t)| \leq k_0$ and by the mean value theorem,

$$|\eta_{n\gamma}(t) - \eta_{n\gamma}(t')| \leq k_0\,|t - t'| \text{ for all } t, t' \text{ in } D_j \cap [-\delta, \delta].$$

Then by Lemma 2.2, we have that

$$|\eta_{n\gamma}(t) - \eta_{n\gamma}(t')| \leq k_0\,|t - t'| \quad \forall\, |t| \leq \delta, |t'| \leq \delta.$$

(ii) According to (a)(i) and Lemma 2.4, it is enough to show

$$B_{n,\delta,\varepsilon_2} \supset A_{n,\delta,\varepsilon_2} \cap \{\sup\{|\eta_{n\gamma}(0)|;|\gamma-\gamma_0|\leq\varepsilon_1\}\leq\delta(1-k_0)\}.$$

Assume that the event in the right-hand side occurs. Then

$$|\eta_{n\gamma}(t)| \leq |\eta_{n\gamma}(0)| + |\eta_{n\gamma}(t) - \eta_{n\gamma}(0)| \leq \delta(1-k_0) + \delta k_0 = \delta.$$

(iii) Follows immediately from Lemma 2.5 (a)(ii).

The proof of (b)(i), (ii) and (iii) are identical to the corresponding parts of (a), just replacing η by η^* and parts (a) by parts (b) of the lemmas used.

Lemma 2.8. (a) *Assume* A.1, A.2, A.3, A.4 *and* A.5. *Put* $D_{n,\delta} = \{\eta_n(t) = t$ *has a unique fixed point* t^* *in* $[-\delta, \delta]$ *and for any* t_0 *in* $[-\delta, \delta]$, $\eta_n^{(k)}(t_0) \to t^*\}$, *where* $\eta_n^{(k)}(t_0)$ *is defined inductively by* $\eta_n^{(k)}(t_0) = \eta_n(\eta_n^{(k-1)}(t_0))$. *Then, there exists* $\delta_2 > 0$ *such that* $\forall\, \delta \leq \delta_2$

$$\lim_{n\to\infty} P\left(\bigcap_{m\geq n} D_{m,\delta}\right) = 1.$$

(b) *Assume,* B.0, B.1, B.2, B.3, B.4 *and* B.5. *Define* $D^*_{n,\delta}$ *in the same way as* $D_{n,\delta}$, *just replacing* η_n *by* η_n^*. *Then, there exists* $\delta_2 > 0$ *such that* $\forall\, \delta \leq \delta_2$,

$$\lim_{n\to\infty} P\left(\bigcap_{m\geq n} D^*_{m,\delta}\right) = 1.$$

Proof. (a) Let $\varepsilon_2 > 0$ and $\delta_2 > 0$ be as in Lemma 2.7 (a).

It is enough to show that

$$D_{n,\delta} \supset A_{n,\delta,\varepsilon_2} \cap B_{n,\delta,\varepsilon_2} \cap C_{n,\delta,\varepsilon_2} \;\forall\, \delta \leq \delta_2.$$

Suppose $A_{n,\delta,\varepsilon_2} \cap B_{n,\delta,\varepsilon_2} \cap C_{n,\delta,\varepsilon_2}$ occurs. Then all the mappings $\eta_{n\gamma}$, such that $|\gamma - \gamma_0| \leq \varepsilon_2$, are contractions of $[-\delta, \delta]$ into itself, and then by Banach fixed point theorem, they have a unique fixed point. But the fixed points of $\eta_{n\gamma}$ are the same for all γ and it is equal to the fixed point of η_n since $\eta_{n\gamma}(t) = t + (1/n\gamma)\sum_{i=1}^{n} \psi(x_i - t) = t$ if and only if $\sum_{i=1}^{n} \psi(x_i - t) = 0$, and

$$\eta_n(t) = t + (1/n\gamma_n(t))\sum_{i=1}^{n} \psi(x_i - t) = t \text{ if and only if } \sum_{i=1}^{n} \psi(x_i - t) = 0.$$

Take any t_0 in $[-\delta, \delta]$. Define inductively $t_k = \eta_{n\gamma_n(t_{k-1})}(t_{k-1}) = \eta_n(t_{k-1})$.

We are going to prove by induction that $|t_k| \leq \delta$ for every k, if $A_{n,\delta,\varepsilon_2} \cap B_{n,\delta,\varepsilon_2} \cap C_{n,\delta,\varepsilon_2}$ occurs.

For $k = 0$, $|t_0| \leq \delta$ by assumption.

Suppose $|t_k| \leq \delta$. Then as $C_{n,\delta,\varepsilon_2}$ occurs, $|\gamma_n(t_k) - \gamma_0| < \varepsilon_2$ and then as $B_{n,\delta,\varepsilon_2}$ occurs,

$$|t_{k+1}| = |\eta_{n\gamma_n(t_k)}(t_k)| \leq \delta.$$

Then using Lemma 2.3 and the fact that $A_{n,\delta,\varepsilon_2} \cap B_{n,\delta,\varepsilon_2} \cap C_{n,\delta,\varepsilon_2}$ occurs, we have that $t_k \to t^*$.

(b) The proof is the same as in part (a) using η^* and part (b) of the lemmas used instead of η and part (a) of the lemmas.

Proof of Theorem 2.1. Let $\delta \le \delta_2$ as in Lemma 2.8. Then

$$\bigcap_{m\ge n}\{|\hat\theta_m|\le\delta\}\supset\bigcap_{m\ge n}\{D_{m,\delta}\cap\{|\hat\theta_{m,0}\le\delta\}\}$$

and

$$\bigcap_{m\ge n}\left\{\lim_{j\to\infty}\hat\theta_{n,j}=\hat\theta_n \text{ and } \hat\theta_m \text{ satisfies (1.1)}\right\}\supset\bigcap_{m\ge n}\{D_{m,\delta}\cap\{|\hat\theta_{m,0}|\le\delta\}.$$

So (i) and (ii) follows from Lemma 2.8 (a) and assumption A.6.

Proof of Theorem 2.3. Analogous to the proof of Theorem 2.1, but using part (b) of Lemma 2.8 and assumption B.6.

Proof of Theorem 2.2. We have to show that the assumptions of Theorem 3 and its corollary of Huber (1967) hold.

It is easy to show that our conditions A.1 to A.7 imply Huber's conditions N1 to N4, and if $\lambda(\theta)=\int\psi(x-\theta)dF(x)$, then

$$\lambda'(0)=\int\psi'(x)dF(x)\ne 0.$$

Moreover, by Theorem 2.1, $\hat\theta_n$ is consistent a.s. and

$$(1/\sqrt{n})\sum_{i=1}^{n}\psi(x_i-\hat\theta_n)\to 0 \ a.s.$$

So all conditions are satisfied and the theorem follows.

Proof of Theorem 2.4. It is enough to prove

$$\lim_{n\to\infty}(1/\sqrt{n})\sum_{i=1}^{n}\psi((x_i-\tilde\theta)/s_0)=0 \text{ in probability,} \tag{2.12}$$

since then Theorem 2.4 will follow from Huber's (1967) Theorem 3 and its corollary.

In order to prove (2.12), by Theorem 2.3 (ii), it is enough to show

$$\lim_{n\to\infty}(1/\sqrt{n})\sum_{i=1}^{n}[\psi((x_i-\tilde\theta_n)/s_0)-\psi((x_i-\hat\theta_n)/s_n)]=0 \text{ in probability.} \tag{2.13}$$

Put

$$R(x,a,b)=\begin{cases}[\psi((x-a)/s_0)-\psi((x-a)/b)]/(s_0-b) & \text{if } b\ne s_0\\ -\psi'((x-a)/s_0)(x-a)/s_0^2 & \text{if } b=s_0.\end{cases}$$

Then we have

$$(1/\sqrt{n})\sum_{i=1}^{n}[\psi((x_i-\tilde\theta_n)/s_0)-\psi((x_i-\tilde\theta_n)]=\sqrt{n}(s_n-s_0)\sum_{i=1}^{n}R(x_i,\tilde\theta_n,s_n)/n.$$

As $\sqrt{n}(s_n-s_0)$ is bounded in probability by assumption B.8, and $\tilde\theta_n\to 0$ a.s. by Theorem 2.3(i), in order to show (2.13), it is enough to show

(2.14) for every $\varepsilon>0$, there exists $\delta>0$ such that

$$\text{limp}\sup_{n\to\infty}\ \sup_{|b-s_0|<\delta,\,|a|<\delta}\left|\sum_{i=1}^{n}R(x_i,a,b)/n\right|\le\varepsilon$$

We have by B.9 that

$$E_F(R(x,0,s_0))=0. \tag{2.15}$$

$$\sup_{|a|<\delta,\,|b-s_0|<\delta}|R(x,a,b)|\le\sup_{|a|<\delta_0,\,|b-s_0|<\delta}|\psi'((x-a)/b)(x-a)/b^2|,$$

by B.10, (2.15) and the dominated convergence theorem, we can find δ such that

$$\sup_{|a|<\delta,\,|b-s_0|<\delta}|E_F(R(x,a,b))|\le\varepsilon.$$

Then Lemma 2.1 implies (2.14).

References

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). *Robust Estimates of Location: Survey and Advances*, Princeton Univ. Press.

Berk, R. (1967). *A Special Structure and Equivariant Estimation*, Ann. Math. Statist. **38**, 1436-1445.

Bickel, P. J. (1975). *One-step Huber Estimates in the Linear Model*, J. Amer. Statist. Assoc. **70**, 428-434.

Collins, J. R. (1976). *Robust Estimation of a Location Parameter in the Presence of Asymmetry*, Ann. Statist. **4**, 68-85.

Dutter, R. (1975). *Robust Regression: Different Approaches to Numerical Solutions and Algorithms*, Research Report n.° 6, Fachgruppe für Statistik, ETH, Zurich.

Hampel, F. R. (1968). *Contributions to the Theory of Robust Estimation*, Ph. D. Dissertation, Univ. of California, Berkeley.

Hampel, F. R. (1974). *The Influence Curve and Its Role in Robust Estimation*, J. Amer. Statist. Assoc. **69**, 383-393.

Huber, P. J. (1964). *Robust Estimation of a Location Parameter*, Ann. Math. Statist. **35**, 73-101.

Huber, P. J. (1977). *The Behavior of Maximum Likelihood estimates Under Nonstandard Conditions*, Proc. Fifth Berkeley Symp. Math. Statist. Prob. **1**, 221-223.

Huber, P. J. (1973). *Robust Regression: Asymptotics, Conjectures and Monte Carlo*, Ann. Statist. **1**, 799-821.

Relles, D. A. (1968). *Robust Regression by Modified Least Squares*, Ph. D. Dissertation, Yale University.

Yohai, V. J. (1974). *Robust Estimation in the Linear Model*, Ann. Statist. **2**, 562-567.
Yohai, V. J. and Maronna, R. A. (1976). *Asymptotic Behavior of M-estimators for the Linear Model*, (To appear).

Ruben Klein
Instituto de Matemática
Pura e Aplicada
Rua Luiz de Camões, 68
20.060 – Rio de Janeiro, RJ
Brasil

Victor J. Yohai
Departamento de Matematicas
Facultad de Ciencias Exactas y Naturales
Pabellón 1, Ciudad Universitaria
Buenos Aires, Argentina