The Annals of Statistics
1982, Vol. 10, No. 2, 454–461

ON INCONSISTENT *M*-ESTIMATORS

BY D. A. FREEDMAN[1] AND P. DIACONIS[2]

University of California, Berkeley and Stanford University

If M is not convex, and the underlying density is multi-modal, M-estimators can be inconsistent. Examples are given, as well as some positive results.

1. Introduction. Consider a symmetric location problem where the parametric form of the underlying density is unknown. The use of M-estimators is often suggested. For some asymptotic theory, see Huber (1964, 1981), Collins (1976), Portnoy (1977), and Zaman (1981). Efficiency calculations are often made on the basis of Monte Carlo experiments; see, for instance, Andrews et al. (1972). Recently, M-estimators have found their way into the textbooks, with little discussion of underlying stochastic structure: see, for instance, Mosteller and Tukey (1977). The choice of the criterion function M in the M-estimator often seems slightly *ad hoc*, and the object of this paper is to point out a possible difficulty: for many M's, the corresponding estimator will be inconsistent. Roughly, this happens if M is not convex. Then, there are densities which are symmetric about zero, such that the M-estimator oscillates indefinitely between two wrong answers.

To be more specific, let M be a smooth function on $(-\infty, \infty)$ and let $\psi(t) = M'(t)$. If $\lim_{t\to\pm\infty}\psi(t) = 0$, then M is said to have a "redescending ψ-function." The inference problem for which M is used can be stated as follows. Let $X_1, X_2, \cdots$ be independent, with a common (unknown) density f, assumed symmetric about 0. Let θ be a translation parameter, to be estimated. The random variables $Y_i = \theta + X_i$ are observed for $i = 1, \cdots, n$, and θ is estimated as the location of the global minimum of

$$\sum_{i=1}^{n} M\left(\frac{Y_i - t}{k\sigma_n}\right). \tag{1.1}$$

In (1.1), the tuning constant k depends on M; often, k is chosen so that the corresponding estimator is efficient when X has one of a few artificial distributions. The expression σ_n is a scale factor computed from the data: the median absolute deviation from the median (MAD) is a common choice. To simplify the analysis, it will be assumed here that the population MAD is known; then the choice $\sigma_n \equiv$ MAD seems appropriate. (Division by the sample MAD of the data might be handled by the usual von Mises calculus.) Also, we assume throughout that $\theta = 0$ without loss of generality.

Some positive results will be stated carefully in Section 2. Essentially, the M-estimator is consistent if M is convex, or if the density f is strongly unimodal, that is, monotone decreasing on $[0, \infty)$ and increasing on $(-\infty, 0]$. The examples of inconsistent behavior will involve nonconvex M. Two of these M's will be fairly standard. The first was used in the Princeton Robustness Study (Andrews et al, 1972). It corresponds to the Cauchy likelihood

$$M(x) = \log(1 + x^2). \tag{1.2}$$

The second corresponds to Tukey's biweight:

$$M(x) = \begin{cases} -(1 - x^2)^3 & \text{for } |x| \le 1, \\ 0 & \text{for } |x| > 1. \end{cases} \tag{1.3}$$

Received July 1981; revised November 1981.

[1] Research partially supported by NSF Grant MCS-80-02535.

[2] Research partially supported by NSF Grant MCS-80-24649.

AMS 1980 *subject classifications*. Primary 62F35; secondary 62E20.

Key words and phrases. M-estimators, ψ-functions, robustness, consistency.

The third is somewhat artificial:

$$M(x) = \begin{cases} -(1 - x^2)^2 & \text{for } |x| \le 1, \\ 0 & \text{for } |x| > 1. \end{cases} \tag{1.4}$$

The underlying density f will be multi-modal, but will satisfy the following regularity conditions:

(1.5) f is symmetric about 0, infinitely differentiable, strictly positive on a compact interval, and vanishes outside this interval; f has a unique maximum at 0; and the MAD of f is 1.

With M as defined in (1.2) or (1.3), there seems to be a cutoff k_M such that if the tuning constant $k > k_M$, the M-estimator is consistent. With M as defined in (1.4), however, there is no such k_M. On the other hand, there may always be a $k_{M,f}$ such that $k > k_{M,f}$ entails consistency. (For inference, this is not so useful, since f is unknown.)

Some authorities suggest that θ be estimated as the root of the equation

$$\sum_{i=1}^{n} M'\left(\frac{Y_i - t}{k\sigma_n}\right) = 0,$$

and if there are several roots, these authorities take the one closest to the median of the data. In reply, we observe that the M-estimator is of the maximum likelihood type with the pseudo-likelihood function e^{-M}, as one sees by taking logs and changing signs. Thus, $M(x) = x^2$ corresponds to the normal; $M(x) = |x|$ to the double-exponential; $M(x) = \log(1 + x^2)$ to the Cauchy. In the examples, the root closest to the median is indeed a consistent estimator. However, it corresponds to a local maximum of (1.1), i.e., a local *minimum* of the pseudo-likelihood function Πe^{-M}, which is unattractive and which may lead to estimators with poor operating characteristics.

Our argument starts from the following proposition, whose proof is deferred to Section 3. For simplicity, the tuning constant is taken as 1; in effect, k can be absorbed into M. Let

$$h(t) = E\{M(X_i - t)\} = \int_{-\infty}^{\infty} M(x - t) f(x)\, dx, \tag{1.6}$$

$$S_n(t) = \sum_{i=1}^{n} M(X_i - t). \tag{1.7}$$

Suppose

$$M \text{ has a bounded continuous derivative } M', \tag{1.8}$$

$$h(t) < \infty \quad \text{for all} \quad t, \tag{1.9}$$

$$h \text{ does not have a global minimum at } 0. \tag{1.10}$$

PROPOSITION 1.1. *Suppose* (1.6) *through* (1.10). *For some positive* ε, *almost surely, for all sufficiently large n, the function* $S_n(\cdot)$ *does not have its global minimum in* $[-\varepsilon, \varepsilon]$. *In particular, the M-estimator is inconsistent.*

Condition (1.10) looks hard to check: the following criterion is useful. Recall that X_i has a continuous, symmetric density.

LEMMA 1.1. *Suppose* (1.6) *through* (1.9). *If* M' *is Lipschitz and* $E\{M''(X_i)\} < 0$, *then h has a local maximum at* 0.

We now present some specific examples. The tuning constant k will be dealt with explicitly.

EXAMPLE 1.1. Define M by (1.2). Let $x_0 = \sqrt{\sqrt{32} - 5} \doteq .81$. Fix k with $x_0 < 1/k < 1$. Then there is a density f satisfying (1.5) such that the M-estimator based on (1.1) is inconsistent: for some $\varepsilon > 0$ almost surely, for all sufficiently large n, the global minimum of (1.1) is not to be found in $[-\varepsilon, \varepsilon]$. Preliminary computations indicate that if $k > 1/x_0 \doteq 1.2$, then this M-estimator is consistent.

EXAMPLE 1.2. Define M by (1.3). Let $x_0 = \sqrt{3 - \sqrt{8}}/\sqrt{5} \doteq .185$. Fix k with $x_0 < 1/k < \sqrt{3/5}$. Then there is a density f satisfying (1.5) such that the M-estimator based on (1.1) is inconsistent. Preliminary computations suggest that if $k > 1/x_0 \doteq 5.41$, then this M-estimator is consistent; something of a close call, for a $k = 6$ is standard choice.

EXAMPLE 1.3. Define M by (1.4). Fix any $k > 1$. Then there is a density f satisfying (1.5) such that the M-estimator based on (1.1) is inconsistent: for some $\varepsilon > 0$, almost surely, for all sufficiently large n, the global minimum of (1.1) is not to be found in $[-\varepsilon, \varepsilon]$.

Example 1.1 can be refined as follows. Define M by (1.2) and x_0 as in the example. Select any number k with $1 < k < 1/x_0$. There is a sequence of independent random variables $\{X_i\}$, having a common infinitely differentiable density f satisfying (1.5), and the following asymptotics for the M-estimator.

a) For almost all sample sequences, for all sufficiency large n, the equation $S_n'(\theta) = 0$ has exactly three roots. Write M_{-n} for the smallest root, M_{on} for the middle root, and M_{+n} for the largest root. Then $S_n(\cdot)$ has a local maximum at M_{on}, and local minima at $M_{\pm n}$; one of the latter is the global minimum, i.e., the M-estimator of $\theta = 0$.

b) Almost surely, $M_{-n} \to -\gamma$ and $M_{on} \to 0$ and $M_{+n} \to \gamma$, where $\gamma > 0$ depends on the distribution of X.

c) For almost all sample sequences, there are subsequences n_{+j} along which $S_{n_{+j}}(\cdot)$ has its unique global minimum at M_{+n}; likewise, there are subsequences n_{-j} along which $S_{n_{-j}}(\cdot)$ has its unique global minimum at M_{-n}. Thus, the M-estimator oscillates indefinitely between $-\gamma$ and γ, and fails of almost-sure consistency.

d) For each large n, with overwhelming probability, $S_n(\cdot)$ has a unique minimum. With probability almost ½ this is at M_{+n} near γ, and with probability almost ½ this is at M_{-n} near $-\gamma$. Thus, the M-estimator fails of consistency even in probability.

Likewise for Example 1.2, except that $S_n'(\cdot) = 0$ has two additional roots, at ± 2, corresponding to endpoint maxima of $S_n(\cdot)$.

For details, see Freedman and Diaconis (1981).

The following heuristic discussion may aid in understanding the examples. Because of the strong law, $S_n(t)$ is close to $nh(t)$ defined in (1.6), and the value of t that minimizes $S_n(t)$ is close to the value of t that minimizes $h(t)$. For the M's considered in the examples, symmetric multimodal densities are constructed so that $h(t)$ takes on its minimum in two distinct places and the M-estimator hops back and forth between these two points.

2. Positive results. The following results show that M-estimators are consistent if either M is convex, or the underlying symmetric density is strongly unimodal in the sense of being nonincreasing on $[0, \infty)$, and hence nondecreasing on $(-\infty, 0]$. Weak unimodality—having a unique maximum at 0—is not enough, as the examples in Section 1 show. Throughout the following discussion, it will be assumed that

(2.1) M is bounded below, continuous, symmetric about zero, nondecreasing on $[0, \infty)$, and not identically constant.

(2.2) $X_1, X_2, X_3, \cdots$, are independent with a common probability density f which is symmetric about zero.

In Theorems 2.1 and 2.2, the following moment assumption is used:

$$\int M(x) f(x)\, dx < \infty. \tag{2.3}$$

In some cases this can be weakened, as in Huber (1981, Section 6.2). The following theorems will be proved in Section 3. Theorems 2.1 and 2.3 follow from general results in Huber (1981, Section 6.2), but the direct proofs may be of interest.

THEOREM 2.1. *Assume* (2.1) *through* (2.3). *If M is strongly convex, then for each n, $S_n(t) = \sum_1^n M(X_i - t)$ has a unique minimum at T_n, and T_n converges to zero almost surely.*

THEOREM 2.2. *Assume* (2.1) *through* (2.3). *Suppose f is strongly unimodal. Let $m = \min_t S_n(t)$. Fix $\varepsilon > 0$. For almost all ω there is an $N(\omega) < \infty$ such that $n > N$ entails $S_n(t) > m$ for $|t| > \varepsilon$.*

In Theorem 2.1, the strong convexity of M guarantees the uniqueness of the maximum, and the moment condition (2.2) guarantees consistency. Without (2.3), the M-estimator can fail: for example, take $M(x) = x^2$ and f to be Cauchy. Unfortunately, neither theorem covers the median as a location estimate for the Cauchy, since $\int |x|/(1 + x^2)\, dx = \infty$. Theorem 2.3 covers this case and others.

THEOREM 2.3. *Assume* (2.1) *through* (2.3). *Suppose M is weakly convex everywhere and strongly convex at zero, in the sense that $M(0) < M(x)$ for all $x \neq 0$. Suppose that $M(x) = O(x)$ as x tends to infinity. Then, for almost all ω, there is an $N(\omega)$ such that for all $n > N$ the minimum in t of $\sum_{i=1}^n M(X_i - t)$ is taken on over an interval I_n containing zero. The length of I_n tends to zero as n tends to infinity. Thus the midpoint of I_n is a consistent estimate.*

REMARK. The growth condition $M(x) = O(x)$ seems crucial in Theorem 2.3. If $M(x)$ is strictly convex and of order $x^{1+\varepsilon}$ at infinity, there are symmetric, long tailed densities such that the the M-estimate oscillates wildly as the sample size increases. For details, see Freedman and Diaconis (1981).

3. Proofs.

LEMMA 3.1. *Suppose g is bounded and Lipschitz, with bounded* a.e. *derivative g'. Suppose X has an absolutely continuous distribution. Let $G(t) = E\{g(X - t)\}$. Then G has a bounded continuous derivative $G'(t) = E\{g'(X - t)\}$.*

The routine proof is omitted.

PROOF OF LEMMA 1.1. Suppose (1.6) through (1.9), M' is Lipschitz, and $E\{M''(X_i)\} < 0$. Recall that X_i are independent with common symmetric density f. Now $h(t) = E\{M(X_i - t)\}$ has two continuous derivatives by Lemma 3.1, and $h''(0) = E\{M(X_i)\} < 0$. By symmetry, $h'(0) = 0$. □

LEMMA 3.2. *Let $X_1, X_2, \cdots,$ be independent with common distribution function F. Let F_n be the empirical distribution function of $X_1, \cdots, X_n$. Let $B_n = \sqrt{n}(F_n - F)$. Let g be a bounded Lipschitz function, with Lipschitz constant L. There is a finite constant A, and for almost all ω an $N = N_\omega < \infty$, such that $n > N_\omega$ entails*

a) $|B_n(t)| < A(\log\log n)^{1/2}$ for all t,

b) $\left| \int g\, dB_n \right| < AL(\log\log n)^{1/2}$.

PROOF. Claim a) follows from the law of the iterated logarithm for the invariance principle; see Chung (1949). Claim b) follows because

$$\int g\, dB_n = -\int B_n(t) g'(t)\, dt. \qquad \square$$

PROOF OF PROPOSITION 1.1. Suppose (1.6) through (1.10). Recall that X_i are independent with common symmetric density f. Define B_n as in Lemma 3.2. Now

$$S_n(t) = nh(t) + \sum_{i=1}^n \{M(X_i - t) - h(t)\} = nh(t) + \sqrt{n} \int h(t) dB_n(t).$$

But h is bounded Lipschitz by Lemma 3.1, so

$$|S_n(t) - nh(t)| \leq AL(n \log \log n)^{1/2}.$$

Suppose $h(0) > h(t_0)$. Choose $\varepsilon > 0$ and α real so $h(t) \geq \alpha > h(t_0)$ for $|t| \leq \varepsilon$, so in that interval,

$$S_n(t) \geq n\alpha - AL(n \log \log n)^{1/2}$$

while

$$S_n(t_0) \leq nh(t_0) + AL(n \log \log n)^{1/2}. \qquad \square$$

LEMMA 3.3. *Let $g(x) = (1 - x^2)/(1 + x^2)^2$. Then $g(0) = 1$, $g(1) = 0$, $g(\sqrt{3}) = 1/8$, g is strictly decreasing on* $[0, \sqrt{3}]$, *strictly increasing on* $[\sqrt{3}, \infty)$.

CONSTRUCTION FOR EXAMPLE 1.1. Recall that $M(x) = \log(1 + x^2)$, so

$$M''(x) = \frac{(1 - x^2)}{(1 + x^2)^2} = g(x). \tag{3.1}$$

The right-hand side of (3.1) takes its minimum value of $-1/8$ at $x = \pm\sqrt{3}$. Also, $(1 - x^2)/(1 + x^2)^2 = 1/8$ *when* $x = \pm x_0$, where $x_0 = \sqrt{\sqrt{32} - 5} \doteq .81$. Fix x_1 and k with $x_0 < x_1 < 1/k < 1$. Let Z take the values $\pm x_1$, $\pm\sqrt{3}$ with equal probabilities $1/4$. Plainly $E\{M''(Z)\} < 0$, and so $E\{k^{-2}M''(kZ/k)\} < 0$. Let W have density f satisfying (1.5), such that the distribution W nearly coincides with that of kZ, so $E\{k^{-2}M''(W/k)\} < 0$. Then Lemma 1.1 applies, with M replaced by $M(x/k)$. □

LEMMA 3.4. *Let $g(z) = 1 - 6z + 5z^2$. Then $g(0) = 1$, $g(1/5) = 0$, $g(1) = 0$; g is monotone decreasing on* $[0, 3/5]$ *and monotone increasing on* $[3/5, 1]$; $g(3/5) = -4/5$; $g(\cdot) = 4/5$ *has the root* $\frac{1}{5}(3 - \sqrt{8})$.

CONSTRUCTION FOR EXAMPLE 1.2. Let $M(x) = -(1 - x^2)^3$ for $|x| \leq 1$, and $M(x) = 0$ for $|x| > 1$. Then

$$M''(x) = \begin{cases} +6(1 - x^2)(1 - 5x^2) & \text{for } |x| \leq 1, \\ 0 & \text{for } |x| > 1. \end{cases}$$

As before, fix $x_0 < x_1 < 1/\sqrt{5} < 1$; and $x_1 < 1/k < \sqrt{3/5}$; let Z take the four values $\pm x_1$, $\pm\sqrt{3/5}$ with equal probabilities $1/4$. The balance of the argument is the same. □

CONSTRUCTION FOR EXAMPLE 1.3. Let $M(x) = -(1 - x^2)^2$ for $|x| \leq 1$, and $M(x) = 0$ for $|x| > 1$. Then

$$M''(x) = \begin{cases} 4 - 12x^2 & \text{for } |x| < 1, \\ 0 & \text{for } |x| > 1. \end{cases}$$

Let $Z = 0$ with probability $1/2$, and $Z = \pm z$ with probability $1/4$ each, where z is just less than 1. Then $E\{M''(Z)\} < 0$. Fix k with $1/k < z$. The balance of the argument is the same. □

We turn now to the results in Section 2. Recall that the X_i are independent with common symmetric density f, and

$$h(t) = \int M(x - t) f(x)\, dx = \int M(t - x) f(x)\, dx$$

because M is symmetric and bounded below. Clearly, h is well defined and symmetric about zero. It may take the value $+\infty$ for some t, even if $h(0) < \infty$. For example, suppose f is normal, and M is this strictly convex function: $M(x) = \exp(x^2/2 - |x|)$ for $|x| \geq 2$, with M defined in $|x| < 2$ to insure convexity.

The next two lemmas will show that h is increasing on $[0, \infty)$ in the sense that $t_1 < t_2$ and $h(t_1) < \infty$ imply $h(t_1) < h(t_2)$. In particular, when (2.3) holds, h has a unique minimum at zero.

LEMMA 3.5. *Assume* (2.1). *If M is strictly convex, then h is increasing on* $[0, \infty)$. *Conversely, if h has a unique minimum at zero for all symmetric densities with compact support, then M is strictly convex.*

PROOF. Consider a symmetric random variable with two-point support at $\pm x$, where $x > 0$. Then $E\{M(X - t)\}$ equals

$$\tfrac{1}{2}\{M(x - t) + M(x + t)\} = \tfrac{1}{2}\{M(t - x) + M(t + x)\}. \tag{3.2}$$

For strictly convex M, this is strictly monotone in t. Integration with respect to $f(x)\,dx$ over $[0, \infty)$ completes the proof of monotonicity of h. For the converse, suppose first that M is not even weakly convex. Then we can find positive x and t such that

$$\tfrac{1}{2}\{M(x - t) + M(x + t)\} < M(x).$$

The two-point distribution at $\pm x$ can be smoothed to give a symmetric density f with compact support such that $h(t) < h(0)$. Finally, suppose M is weakly convex but not strictly convex. Then there are $x > \varepsilon > 0$ such that M is linear on $[x - \varepsilon, x + \varepsilon]$ and on $[-x - \varepsilon, -x + \varepsilon]$. Let f be a continuous symmetric density concentrated on

$$[-x - \varepsilon/2, -x + \varepsilon/2] \cup [x - \varepsilon/2, x + \varepsilon/2].$$

Then h is constant on $[-\varepsilon/2, \varepsilon/2]$. □

LEMMA 3.6. *Assume* (2.1). *If f is a symmetric density which is strongly unimodal, then h is increasing on* $[0, \infty)$.

PROOF. The set of strongly unimodal densities is a convex set. Khinchine's theorem shows that the extreme points of this set are the symmetric uniform densities, and each strongly unimodal f is a unique integral average of extreme points; see page 158 of Feller (1971). So, it is enough to do the case of f uniform on $[-a, a]$. Then

$$2ah(t) = \int_{t-a}^{t+a} M(x)\,dx.$$

If $t \geq a$, monotonicity is clear. Say $0 \leq t \leq a$; take δ with $t + \delta \leq a$. Now

$$\begin{aligned} 2a\{h(t + \delta) - h(t)\} &= \int_{t+a}^{t+\delta+a} M(x)\,dx - \int_{t-a}^{t+\delta-a} M(x)\,dx \\ &= \int_{a+t}^{a+t+\delta} M(x)\,dx - \int_{a-t-\delta}^{a-t} M(x)\,dx. \end{aligned}$$

The right hand side is strictly positive because M is strictly increasing. □

The next lemma shows that for large n, any minimizing value of $S_n(t)$ is close to zero; hence M-estimators are consistent.

LEMMA 3.7. *Assume* (2.1) *through* (2.3). *Suppose that either* M *is strictly convex or* f *is strongly unimodal. Then, for any* $\varepsilon > 0$ *there is an* $N(\omega) < \infty$ *such that for* $n > N$, $|t| > \varepsilon$ *implies* $S_n(t) > S_n(0)$.

PROOF. It is convenient to argue separately for large and small values of t. It will first be shown that there are a finite positive L and N_1 such that almost surely, $n > N_1$ and $|t| > L$ imply $S_n(t) > S_n(0)$. To see this, choose positive $a < L$ so large that for some $\varepsilon > 0$,

$$P(|X_1| \le a) > 1/1 + \varepsilon, \tag{3.3}$$

$$M(L - a) > (1 + \varepsilon)^2 h(0). \tag{3.4}$$

Clearly, (3.4) implies $M(t - a) > (1 + \varepsilon)^2 h(0)$ for all $t > L$. Let $\nu_a(n)$ be the number of X_i with $|X_i| \le a$. The strong law and (3.3) imply that $\nu_a(n) > n/(1 + \varepsilon) + o(n)$, almost surely. Thus, for $t \ge L$,

$$S_n(t) \ge M(L - a)\nu_a(n) > \frac{n}{(1 + \varepsilon)} h(t) + o(n).$$

But the strong law implies that $S_n(0) = nh(0) + o(n)$, almost surely. This completes the argument for $|t| > L$.

To finish the proof, consider $t \in [\delta, L]$. For fixed large k (to be chosen later),

$$S_n(t) \ge \textstyle\sum_{i=1}^n M(X_i - t)I_{|X_i|<k} = n \displaystyle\int_{-k}^{k} M(x - t)f(x)\, dx + \sqrt{n} \int_{-k}^{k} M(u - t)\, dB_n(u).$$

Lemma 3.4 implies that the first integral on the right is monotone in t, and so at least as large as

$$n \int_{-k}^{k} M(x - \delta)f(x)\, dx > (1 + \varepsilon)nh(0),$$

for k suitably large and some $\varepsilon > 0$. For this fixed k, the second integral on the right is almost surely $O(\sqrt{n \log\log n})$ uniformly for $t \in [\delta, L]$, by Lemma 3.2.

PROOF OF THEOREM 2.1. Under the hypotheses, $S_n(\cdot)$ is a sum of strictly convex functions and so strictly convex. It therefore has a unique minimizing value t_n. Lemma 3.7 implies that θ_n converges to zero almost surely. □

PROOF OF THEOREM 2.2. This is a direct consequence of Lemma 3.7. □

Let

$$g(t) = \int_{-\infty}^{\infty} M'(x - t)f(x)\, dx. \tag{3.5}$$

LEMMA 3.8. *If* M *satisfies the condition of Theorem* 2.3, *then* g *in* (3.5) *is well-defined, antisymmetric about* 0, *and nonincreasing in* $(-\infty, \infty)$. *Also, if* 0 *is in the support of* f, *then* $g(t) < 0$ *for* $t > 0$.

PROOF. Only the last claim needs to be argued. Write $d\mu = f(x)\, dx$ and fix $t > 0$. Then $g(t)$ equals

$$\int_{-\infty}^{\infty} M'(x - t)\mu\,(dx) = \int_{t/2}^{\infty} \{M'(x - t) - M'(x + t)\}\mu\,(dx) + \int_{t/2}^{t/2} M'(x - t)\mu\,(dx)$$

$$\le \int_{-t/2}^{t/2} M'(x - t)\mu\,(dx) \le M'(-\tfrac{1}{2}t)\mu[-\tfrac{1}{2}t, \tfrac{1}{2}t].$$

□

PROOF OF THEOREM 2.3. The function $S_n(\cdot)$ is continuous, weakly convex, and bounded below, being a sum of such functions. Hence, it attains its minimum on an interval, say I_n. Almost surely, for each n, S_n is differentiable except at countably many points where it has left and right hand derivatives. Then $S_n'(\cdot)$ is nonincreasing and

$$S_n(t_2) = \int_{t_1}^{t_2} S_n'(u)\,du + S_n(t_1) \quad \text{for} \quad t_1 < t_2. \tag{3.6}$$

Finally, $S'_n(t) = -\sum_{i=1}^n M'(X_i - t) = -ng(t) + o(n)$ almost everywhere by the strong law. (The countable set of singularities for $M'(X_i - \cdot)$ does not matter, because X_i has a continuous distribution.) In principle, the null set depends on t. Given $\varepsilon > 0$, it must be shown that $I_n \subset (-\varepsilon, \varepsilon)$ for all sufficiently large n. From Lemma 3.8, $g(\varepsilon) < 0$, and so $S_n'(t) > 0$ for all sufficiently large n and all $t \geq \varepsilon$. Now, (3.6) implies that the minimum of $S_n(\cdot)$ cannot occur for $t > \varepsilon$. A similar argument works for $t < -\varepsilon$. □

Acknowledgment. We want to thank a very helpful referee.

REFERENCES

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., and TUKEY, J. W. (1972). *Robust Estimates of Location.* Princeton Univ. Press, Princeton.

CHUNG, K. L. (1949). An estimate concerning the Kolmogorov limiting distribution. *Trans. Amer. Math. Soc.* **67** 36–50.

COLLINS, J. R. (1976). Robust estimation in the presence of symmetry. *Ann. Statist.* **4** 68–85.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd ed. Wiley, New York.

FREEDMAN, D. A., and DIACONIS, P. (1981). On inconsistent M-estimates. Technical Report No. 170, Department of Statistics, Stanford Univ.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

MOSTELLER, F., and TUKEY, J. W. (1977). *Data Analysis and Regression.* Addison-Wesley, Reading, Mass.

PORTNOY, S. (1977). Robust estimation in dependent situations. *Ann. Statist.* **5** 22–43.

ZAMAN, A. (1981). Consistency of generalized M estimators. Unpublished manuscript.

STATISTICS DEPARTMENT
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305