Kybernetika

Ivan Mizera

On consistent M-estimators: Tuning constants, unimodality and breakdown

Kybernetika, Vol. 30 (1994), No. 3, 289--300

Persistent URL: http://dml.cz/dmlcz/125162

Terms of use:

© Institute of Information Theory and Automation AS CR, 1994

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* http://project.dml.cz

KYBERNETIKA – VOLUME 30 (1994), NUMBER 3, PAGES 289–300

ON CONSISTENT M-ESTIMATORS: TUNING CONSTANTS, UNIMODALITY AND BREAKDOWN[1]

IVAN MIZERA

The existence and uniqueness of values of location M-functional are explored. For the symmetric population distributions, the role of studentization is revealed – the Freedman–Diaconis conjecture is proved. The asymmetric case is studied with respect to unimodality considerations. The lower bound for the breakdown of studentized estimators is derived.

1. INTRODUCTION

The article of Freedman and Diaconis [7], see also Freedman and Diaconis [6] brought to attention some weak points of behaviour of location M-estimators. Freedman and Diaconis have presented examples of location M-functionals which are not identifiable at certain population distributions and hence not consistent for samples drawn from these distributions. Among positive results, they have shown the identifiability of location M-functionals in the convex case and in the nonconvex case for the symmetric unimodal probabilities. The objective of this note is to add some new virtues to this picture. Roughly, the problem of identifiability can be in some cases overcame by studentization with a multiple of a suitable scale estimator. This is the "big-tuning-constant-conjecture" of Freedman and Diaconis. The unimodality approach can be extended to the asymmetric case.

Recall first some facts concerning functional approach to M-estimation. The set of the values of a (location) M-functional T at a probability P (a member of the set $\mathcal{P}$ of all probabilities on $\mathbb{R}$, all defined on the underlying Borel σ-algebra) is defined as the set of those t in which a function

$$\lambda_P(t) = \int \phi\left(\frac{y-t}{kS(P)}\right) \mathrm{d}P(y) \tag{1}$$

attains its minimum. The M-estimate $T(y_1, y_2, \ldots, y_n)$ can be viewed in this framework as a value of M-functional T evaluated at the empirical probability induced by a sample $y_1, y_2, \ldots, y_n$.

[1]The research was supported by the GAS Grant A06. Some of the results constitute a part of the thesis written under supervision of Prof. Jana Jurečková, whose help and encouragement are gratefully acknowledged.

The most popular reason for including a scale functional $S(P)$ (corresponding to some suitable scale estimator, the favorite being MAD, the median of absolute deviations from the median) is a need of good invariance properties for the resulting M-estimator. The "objective function" is determined by a function ϕ together with a value of a *tuning constant* k. The function ϕ usually follows some of the known commended shapes. Among them there are convex ones (leading, among others, to the mean, the median or the Huber location estimator) and redescending ones. In the sequel we assume that every ϕ is absolutely continuous – a primitive function of a function denoted by ψ – and that ψ is bounded. We call a function ϕ *redescending* if $\psi(x) \to 0$ as $|x| \to \infty$. If ϕ is redescending, it can be bounded (the integrals of the popular Andrews sine wave, the Tukey biweight or the Hampel three-part-linear shape belong to this category, but also the less known "mean likelihood", formed by the reverted Gaussian curve) and also unbounded (with the Cauchy (log)likelihood as a representative). For the details, see Andrews et al. [1], Huber [11], Hoaglin, Mosteller and Tukey [9], Hampel et al. [8]. All functions ϕ considered are even, due to the fact that for symmetric P the centre of symmetry serves as the natural and accepted location parameter. We lack such a natural location parameter for asymmetric distributions.

The functional T is said to be *identifiable* at P if the set of its values at P contains exactly one member. Identifiability encompasses existence – thus supposing that the integral in (1) is defined for enough many values of t and the minimum of $\lambda_P(t)$ is attained in at least one t – and uniqueness. Identifiability of the M-functional at the population probability P has a great impact on the asymptotic behaviour of the sequence $T_n = T(Y_1, Y_2, \ldots, Y_n)$, when a sequence $Y_1, Y_2, \ldots$ consists of independent random variables, identically distributed according to the law P.

The most unambiguous situation is observed when ϕ is convex: then λ_P is also convex, if defined. Thus the main identifiability problem lies in the existence. The best way to save moment assumptions is to deal with a (monotone) derivative of ϕ and work with M-estimators defined as roots of equations. This approach is quite classical and its theory rather well developed, beginning from Huber [10], see also Huber [11]. Convexity itself yields only a sort of a "weak" uniqueness: the value set is an interval. Some other, relatively weak conditions are needed to make this interval a singleton. Differentiability of λ_P can be used; however, this technique, developed by Clarke [3], gives only local uniqueness in the nonconvex case.

The case of redescending ϕ is more intricate. These shapes are widely commended and used, despite the fact that the theoretical properties of the resulting M-estimators are not yet completely disclosed. It should be noted that a use of a function ϕ which falls neither under the convex nor the redescending category have not been registered.

2. THE SYMMETRIC CASE

Each example of Freedman and Diaconis consists of an even density f and an even redescending function ϕ. They are paired in such a way that the resulting M-functional T is not identifiable at the probability P represented by f. The corresponding function $\lambda_P(t)$ exhibits in these cases two equally deep minima located symmetrically around 0. This means that the values of T_n for large n almost surely oscillate between small neighbourhoods of these two values. (Freedman and Diaconis supply a sophisticated proof of this fact.) Consequently, the M-estimator is not consistent at P in any sense: it does not converge to the "true" location 0, nor to some uniquely determined asymptotic value.

The density f is in all three cases multimodal and has a unique maximum at 0. Function ϕ in the first example corresponds to the Cauchy likelihood $\phi(x)s = \log(1+x^2)$, in the second example, the biweight ($\phi(x) = -(1-x^2)^3$ for $|x| \leq 1$, $\phi(x) = 0$ otherwise) is used. Freedman and Diaconis conjecture, on the basis of preliminary computations, that in both cases the M-functionals are identifiable (for all P) if the tuning constant k (not depending on P) is set high; as $S(P)$ a functional representing MAD is used. However, their third example introduces a "somewhat artificial" (in fact, it corresponds to the first known M-estimator introduced by Smith in the nineteenth century) function $\phi(x) = -(1-x^2)^2$ for $|x| \leq 1$, $\phi(x) = 0$ otherwise. In this example, given any k there exists a probability P such that the corresponding M-functional is not identifiable at P.

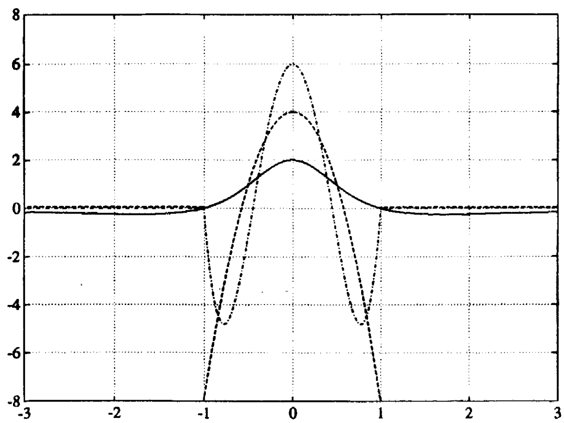

Fig. 1. The derivative of ψ for the Cauchy likelihood (—), the biweight (-·-) and the Smith ϕ (---).

To see what is happening, the derivatives ζ of ψ are plotted in Fig. 1 for the functions ψ from the all three examples. Since the functions ψ are odd ($\psi(-x) = -\psi(x)$), for a symmetric P holds that $\lambda'_P(0) = 0$. The sign of

$$\lambda'_P(0) = \int \zeta(x)\, \mathrm{d}P(x) \tag{2}$$

determines whether λ_P has a local minimum or maximum at 0. The property of MAD yields

$$P\left[\theta - S(P), \theta + S(P)\right] \geq \frac{1}{2} \tag{3}$$

for $\theta = 0$. Hence more than a half of the mass contributes to the integral (2) with a value greater than $\zeta(0) - 1/k$. This contribution for large k dominates $\min \zeta(u)$, making the integral in the first two examples positive; in the third example, masses of magnitude $1/4$ can be placed right and left such that their contribution is negative enough to make the whole integral negative.

Consequently, given any k, a local maximum at 0 can be obtained for some P, if for every $\varepsilon > 0$

$$\min_{x \in [-\varepsilon, \varepsilon]} \zeta(x) < -\min_{x \in \mathbb{R}} \zeta(x).$$

The contrary is true if the reverse inequality holds for at least one $\varepsilon > 0$: if, say, ζ is continuous at 0 and

$$\zeta(0) > \min_{x \in \mathbb{R}} \zeta(x). \tag{4}$$

This suggests a "test", which, say, by the Andrews wave or the Hampel ϕ for a certain choice of parameters is not passed. The case when ψ is not absolutely continuous can be treated in an analogous manner: involving delta functions in ζ, appropriate counterexamples can be found heuristically and then exactly checked. Proceeding this way, we obtain, for instance, that the skipped mean (see Huber [10]) fails the test as well as the skipped median.

For those ϕ which satisfy (4) and hence pass the test, like the biweight, the Cauchy likelihood or the Hampel ϕ for certain choice of parameters, the more peculiar problem is to determine whether the local minimum at 0 is the global one. Fix ϕ and define an auxiliary function

$$\Phi(x, a) = \frac{1}{2}\left[\phi(x - a) + \phi(x + a)\right] - \phi(x).$$

Note that Φ is continuous. The set of all symmetric probabilities on $\mathbb{R}$ is denoted by $\mathcal{S}$.

Theorem 1. Suppose ϕ is even and let for S the condition (3) hold for all $P \in \mathcal{S}$, with θ standing for the centre of symmetry of P. Suppose that λ_P is defined for all t. Let a constant k be given. If for all $a > 0$

$$\inf_{x \in [0, 1/k]} \Phi(x, a) > 0 \tag{5}$$

and

$$\inf_{x \in [0, 1/k]} \Phi(x, a) + \inf_{x \in (1/k, \infty)} \Phi(x, a) > 0 \tag{6}$$

then the M-functional T is identifiable at every $P \in \mathcal{S}$.

Proof. For the notational convenience, set $\theta = 0$. Since λ_P is, due to symmetry of ϕ and P, symmetric, it has a global minimum at 0 if and only if an inequality

$$\lambda_P(a) - \lambda_P(0) > 0 \tag{7}$$

holds for every $a > 0$. The assumption of existence of λ_P is trivially satisfied by bounded ϕ. If ϕ is unbounded, then

$$\phi\left(\frac{x-t}{kS(P)}\right) - \phi\left(\frac{x}{kS(P)}\right) \tag{8}$$

could be used in (1), resulting in a modified definition of T. Note at this occasion that (7) indicates the statement of Theorem is not affected by such an improvement, hence the existence assumption is not essential.

Fix a. Again by symmetry, the inequality (7) turns to

$$\int_0^\infty \frac{1}{2}\left[\phi\left(\frac{x-a}{kS(P)}\right) + \phi\left(\frac{x+a}{kS(P)}\right)\right] - \phi\left(\frac{x}{kS(P)}\right) \mathrm{d}P(x) > 0.$$

This inequality is violated by some P if the value of

$$\int_0^\infty \frac{1}{2}\left[\phi(x-a) + \phi(x+a)\right] - \phi(x)\,\mathrm{d}Q(x) = \int_0^\infty \Phi(x,a)\,\mathrm{d}Q(x)$$

is nonpositive for a measure Q defined by

$$\int f(x)\,\mathrm{d}Q(x) = \int f\left(\frac{x}{kS(P)}\right) \mathrm{d}P(x)$$

and satisfying

$$Q\left[0, \frac{1}{k}\right] = P\left[0, S(P)\right] \geq \frac{1}{4}. \tag{9}$$

By (5), (6) and (9)

$$\begin{aligned}
\int_0^\infty \Phi(x,a)\,\mathrm{d}Q(x) &\geq \int_0^{1/k} \Phi(x,a)\,\mathrm{d}Q(x) + \int_{1/k}^\infty \Phi(x,a)\,\mathrm{d}Q(x) \\
&\geq \inf_{x\in[0,1/k]} \Phi(x,a)\, Q\left[0, \frac{1}{k}\right] + \inf_{x\in(1/k,\infty)} \Phi(x,a)\, Q\left[\frac{1}{k}, \infty\right] \\
&\geq \frac{1}{4}\left(\inf_{x\in[0,1/k]} \Phi(x,a) + \inf_{x\in(1/k,\infty)} \Phi(x,a)\right) > 0,
\end{aligned}$$

and the statement follows. □

If MAD is taken for S, the conditions (5) and (6) are necessary for identifiability; if they are violated, a symmetric probability P can be found such that T is not identifiable at P. Note that if S satisfies (3), then so does also its multiple by a constant greater than 1 — this indicates why conditions (5) and (6) need not be necessary in a general case. Due to continuity, the first inf in (6) (and in (5)) can be changed to min. If ϕ is redescending, then for a fixed $\Phi(x,a)$ tends to 0 as $x \to \infty$ and the second inf can be replaced by min as well.

The conditions (5) and (6) involve only the function ϕ. Their validity can be checked numerically – resulting with $k \geq 1.23$ for the Cauchy likelihood and $k \geq 5.40$

for the biweight, the figures which are in accord to those given by Freedman and Diaconis. Another possibility is to carry an exact proof for a given ϕ of the assertion that there exists a $k > 0$ such that (5) and (6) are satisfied. We give such proofs for the Cauchy likelihood and the biweight in the Appendix. Unfortunately, a simple umbrella criterion appears not to be available – although there are some clues, it seems unavoidable in some cases to proceed just in an ad hoc manner, utilizing preliminary exploratory graphs and calculations with derivatives and inequalities.

Having established identifiability, we can truly speak about consistent M-estimators, due to fact that the M-functionals under consideration are weakly continuous – continuous as mappings from $\mathcal{S}$ with respect to topology of weak convergence on $\mathcal{S}$ (see Billingsley [2]). The following theorem is formulated just to fulfil our present needs; the greatest generality is not achieved.

Theorem 2. Suppose ϕ is even, redescending and nondecreasing on $[0, \infty)$. Suppose T is identifiable at $P \in \mathcal{P}$. If a scale functional S be weakly continuous at P, then M-functional T is weakly continuous at P.

Proof. See Mizera [15]. □

Theorem 2 can be extended to cover the case when T is multi-valued at P; in this setting, an appropriate notion of set convergence ("upper convergence") is used. For the details, see Mizera [15]. Note that for unbounded ϕ, the existence of $T(P)$ presumes the existence of the function $\lambda_P(t)$; hence some moment condition is needed. Such a condition can be eliminated, using the modified definition of T in the vein of (8); the derivative of ϕ is then supposed to be nondecreasing on $[0, c]$ for some $c > 0$, nonincreasing on $[c, +\infty)$ and tending to 0 as the argument goes to $+\infty$. This covers all practical cases; for the details, see again Mizera [15].

Weak continuity implies that T_n converges to $T(P)$ almost surely. Another consequence is qualitative robustness (see Huber [11], Hampel et al. [8]). Weak continuity of MAD can be obtained adapting the approach of Huber (1981, [11, § 3.2, § 5.2]. See also Mizera [15].

3. THE ASYMMETRIC CASE

In the asymmetric case we have no longer a natural parameter of location as the centre of symmetry. However, there is interest in identifiability also in this case: identifiability ensures that the sequence $T_1, T_2, \dots$ converges to a well defined unique quantity $T(P)$. Under *Fisher consistency*, a postulate which demands the equality of the estimated parameter to $T(P)$, consistency is achieved.

Unfortunately, Theorem 1 is not valid for asymmetric distributions – a counterexample can be constructed. Its highly asymmetric nature suggests that in the case of mild asymmetry, Theorem 1 might hold. Freedman and Diaconis imposed unimodality as a regularity condition to achieve identifiability in the redescending symmetric case. We present a polished version of their result, with a short proof adapted from Davies [4].

A probability P is called *unimodal* if it has a density f and a point c exists such that f is nondecreasing on $(-\infty, c]$ and nonincreasing on $[c, \infty)$. If only one c with this property exists, the probability P is called *strictly unimodal.* (Freedman and Diaconis use "strongly unimodal" instead of "unimodal".)

Theorem 3. Suppose ϕ is even and nondecreasing on $[0,\infty)$, with the unique point of the global minimum at 0. Suppose that λ_P is defined for all t. If $P \in \mathcal{S}$ is a strictly unimodal probability, then the M-functional T is identifiable at P.

Proof. Again, set $\theta = 0$. Let f be a density of P. By symmetry, for all $a \neq 0$

$$\int \left[\phi\left(\frac{x-a}{kS(P)}\right) - \phi\left(\frac{x}{kS(P)}\right)\right][f(x-a) - f(x)]\,\mathrm{d}x < 0,$$

the strict inequality due to the fact that zero is a common point of strict monotonicity of ϕ and f. Multiplying and using (1), the statement follows. Note that the assumption of existence of λ_P is again inessential here. □

Theorem 2 corresponds to the well-known fact that the convolution of symmetric unimodal probabilities is itself unimodal. In asymmetric case this is true only if one of the densities is log-concave – strongly unimodal (Schoenberg [16], Ibragimov [13], Karlin [14]). We call a function ϕ *extra strongly unimodal* if for all $a \geq \inf_{x\in\mathbb{R}} \phi(x)$, the logarithm of the function $\max\{a-\phi(x), 0\}$ is strictly concave (only the logarithm of positive values is considered). It turns out (for the details, see Mizera [15]) that extra strongly unimodal functions are linear transformations of functions, which are bounded from below by 0 and are either convex with unique minimum or bounded from above densities with a strictly concave logarithm. Hence, $\frac{1}{2}x^2$, $|x|$ or the Huber ϕ are extra strongly unimodal, as well as the Andrews wave, the Hampel ϕ or the biweight; but not the Cauchy likelihood.

Theorem 4. Suppose that ϕ is extra strongly unimodal. If $P \in \mathcal{P}$ is a strictly unimodal probability, then the M-functional T is identifiable at P.

Proof. See Mizera [15]. □

A numerical counterexample exists indicating that for not strongly unimodal ϕ, like the Cauchy likelihood, a unimodal probability can be found such that the resulting function λ_P attains its minimum in exactly two points. Thus bounded functions ϕ are favored against the unbounded ones. This contradicts the consequences of breakdown point investigations. In the article of Huber [10], the breakdown point $\frac{1}{2}$ is established for the redescending M-estimators with unbounded ϕ. For a bounded ϕ, the breakdown point depends on the sample configuration: choosing for each sample size the worst configuration possible, the value $\frac{2}{n}$ tending to 0 can be obtained. This is due to the a "uniform" character of the widely used definition of Donoho and Huber [5]. A definition involving the sampling distribution – like the original Hampel's definition of breakdown point — might be more appropriate in this case.

However, for the studentized M-estimators, the situation radically changes.

Theorem 5. Suppose ϕ is even, bounded and nondecreasing on $[0, \infty)$. Let the breakdown point of S be $\frac{1}{2}$ and let for every $P \in \mathcal{P}$ a location θ (depending on P) exists such that (3) holds. If $\phi(0) = 0$ and $\sup_{x \in \mathbb{R}} \phi(x) = 1$ then for the breakdown point ε^* of T an inequality

$$\varepsilon^* \geq \frac{1}{2} - \phi\left(\frac{1}{k}\right)$$

holds.

Proof. Fix a sample $y_1, y_2, \ldots, y_n$. Let s_n be a finite sample breakdown of S at the sample size n. Choose m such that

$$\frac{m}{m+n} < \min\left(\frac{1}{2} - \phi\left(\frac{1}{k}\right), s_n\right) \tag{10}$$

This means that $\frac{n}{n+m} > \frac{1}{2} + \phi\left(\frac{1}{k}\right)$. Pick $\eta > 0$ such that

$$\frac{n}{n+m}(1-\eta) > \frac{1}{2} + \phi\left(\frac{1}{k}\right).$$

Let P denote the empirical measure formed by the extended sample $y_1, y_2, \ldots, y_n$, z_1, $z_2, \ldots, z_m$. For any configuration of $z_1, z_2, \ldots, z_m$, the value of $S(P)$ remains bounded by $B < \infty$. Let $\tilde{x} = \max_i |x_i|$. Pick $A > \tilde{x} + 3B$ such that $\phi(\tilde{x} - A) \geq \eta$. The interval E of length $2S(P) \leq 2B$ covering half of the extended sample must cover some of x_i, hence $E \subseteq [0, A]$. Since ϕ is nondecreasing,

$$\lambda_P(\theta) \leq \int_E \phi\left(\frac{x-\theta}{kS(P)}\right) \mathrm{d}P(x) + P(\mathbb{R} \setminus E) \leq \phi\left(\frac{1}{k}\right) P(E) + P(\mathbb{R} \setminus E),$$

hence for $|\tilde{t}| > A$

$$\min_{t \in E} \lambda_P(t) \leq \phi\left(\frac{1}{k}\right) + \frac{1}{2} < \frac{n}{n+m} \cdot 1 + \frac{m}{n+m} \cdot 0 \leq \lambda_P(\tilde{t}) \tag{11}$$

By (10) and (11) follows that the global minimum of λ_P cannot be attained at $\tilde{t}$ satisfying $|\tilde{t}| > A$ and therefore

$$\varepsilon^* \geq \lim_{n \to \infty} \min\left(\frac{1}{2} - \phi\left(\frac{1}{k}\right), s_n\right) = \frac{1}{2} - \phi\left(\frac{1}{k}\right). \qquad \square$$

The equality is attained, if MAD is used for S: the converse inequality can be observed considering a sample with $\frac{n}{2}$ of points placed in -1, $\frac{n}{2}$ of points in 1 and the remaining m points placed in $\tilde{t}$. If $\frac{m}{m+n}$ exceeds the bound given by Theorem 5, still remaining not greater than $\frac{1}{2}$, then the estimator T breaks down to $\tilde{t}$, which can be drawn to infinity.

Theorem 5 gives $\varepsilon^* = 0.42$ for the biweight with $k = 6$ studentized by MAD. Huber [12] reports a numerically obtained value $\varepsilon^* \geq 0.49$ "in typical situations". In this context it is also interesting to quote Donoho and Huber [5], who outlined a connection between identifiability and so-called *variance breakdown point*, described roughly as the smallest portion of the sample which can cause breakdown of asymptotic variance to infinity.

4. APPENDIX

To prove that ϕ satisfies (5) and (6) for k sufficiently large, usually some additional properties of possible functions ϕ are used:

(i) ψ is a primitive function of ζ and $I = -\inf_{x\in[0,\infty)}\zeta(x) > -\infty$;

(ii) $J = \sup_{x\in[0,\infty)}\psi(x) < \infty$;

(iii) there exists $C \in (0,\infty)$ such that ψ is nondecreasing on $[0,C]$ and nonincreasing on $[C,\infty)$.

We keep assuming that ϕ is a primitive function of ψ, that ϕ is nondecreasing on $[0,\infty)$, even and $\phi(0)=0$. The value of $\zeta(0)$ is denoted by K. To satisfy (4), $K > I$ must hold.

Using this assumptions, we establish some general claims. Suppose k is given. Clearly, (6) holds for $a>0$ if

$$M(a) = \inf_{x\in[0,1/k]}\Phi(x,a) > -\inf_{x\in(1/k,\infty)}\Phi(x;a) \geq -\inf_{x\in[0,\infty)}\Phi(x,a) = m(a). \qquad (12)$$

If $m(a) > 0$, then (12) implies also (5).

Claim 1. Let (i) hold. Then for every $a > 0$, $m(a) \leq \frac{1}{2}Ia^2$.

Proof. Follows from the representation

$$\Phi(x,a) = \frac{1}{2}\int_x^{x+a}\int_{u-a}^{u}\zeta(v)\,\mathrm{d}v\,\mathrm{d}u. \qquad (13)$$

□

Claim 2. Let (ii) hold. Then for every $a > 0$, $m(a) \leq \frac{1}{2}Ja$.

Proof. Follows from the inequality

$$\Phi(x,a) = \frac{1}{2}\left[\int_x^{x+a}\psi(u)\,\mathrm{d}u - \int_{x-a}^{x}\psi(u)\,\mathrm{d}u\right] \geq \frac{1}{2}\int_x^{x+a}\psi(u)\,\mathrm{d}u.$$

Claim 3. Let (iii) hold. Then for every $a > 0$, $m(a) \leq \frac{1}{2}\phi(C+2a)$ and if $a \geq c$, then $\frac{1}{2}\phi(C+2a) \leq \frac{1}{2}\phi(2a) + \frac{1}{2}\psi(2a)c$.

Proof. Since ϕ is nondecreasing on $[0,\infty)$, $\phi(x) \leq \phi(x+a)$ for $a>0$ and

$$\Phi(x,a) \geq \frac{1}{2}\left[\phi(x-a)+\phi(x+a)\right] - \phi(x+a) = \frac{1}{2}\left[\phi(x-a)-\phi(x+a)\right].$$

The last expression attains its minimum for x satisfying $\psi(x-a) = \psi(x+a)$ (this can be seen by differentiating; but the proof without the differentiability assumption is also easy). Hence, there is $\tilde{x} \in [0,\infty)$ such that $C \in [\tilde{x}-a, \tilde{x}+a]$ and

$$\Phi(x,a) \geq \frac{1}{2}\left[\phi(\tilde{x}-a)-\phi(\tilde{x}+a)\right] \geq -\frac{1}{2}\phi(\tilde{x}+a) \geq -\frac{1}{2}\phi(C+2a).$$

The rest of Claim follows from the expression

$$\frac{1}{2}\phi(C+2a) = \frac{1}{2}\phi(a) + \frac{1}{2}\int_{2a}^{C+2a} \psi(u)\,\mathrm{d}u.$$

Note that if $\phi(x)$ is bounded from above by L, then $m(a) \leq L/2$ for all $a > 0$. □

Claim 4. Let (i) hold. Then for $a \in (0, A]$, $M(a) \geq \frac{1}{2}(K - \varepsilon_A)a^2$, where $\varepsilon_A = K - \inf_{x\in[0,A+1/k]} \zeta(x)$.

Proof. Follows from the representation (13). □

Claim 5. Let (i) hold. Then for every $a > 0$, $M(a) \geq \phi(a) - \varepsilon_k$, where $\varepsilon_k = \phi(\frac{1}{k}) + \frac{I}{2k^2}$. If $K = \sup_{x\in\mathbb{R}} \zeta(x)$, then $\varepsilon_k \geq \frac{K+I}{2k^2}$.

Proof. We have

$$\begin{aligned}
\Phi(x,a) &= \frac{1}{2}\left[\phi(-a) + \phi(a)\right] - \phi(0) \\
&+ \frac{1}{2}\int_{-a}^{-a+x} \psi(u)\,\mathrm{d}u + \frac{1}{2}\int_{a}^{a+x} \psi(u)\,\mathrm{d}u - \int_0^x \psi(u)\,\mathrm{d}u \\
&= \phi(a) - \phi(x) + \frac{1}{2}\int_a^{a+x}\int_{u-x}^{u} \zeta(v)\,\mathrm{d}v\,\mathrm{d}u \geq \phi(a) - \phi(x) + \frac{I}{2}x^2
\end{aligned}$$

and $x \in \left[0, \frac{1}{k}\right]$. For the rest of the statement, note that $\phi(x) = \int_0^x \int_0^u \zeta(v)\,\mathrm{d}v\,\mathrm{d}u$. □

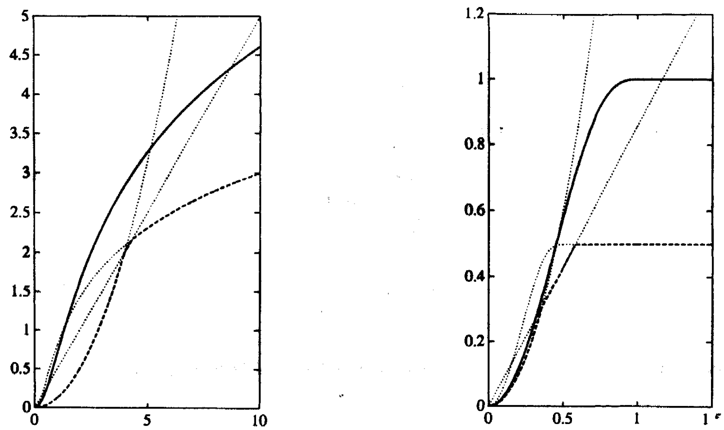

Fig. 2. Plot of $\phi(a)$, $\frac{1}{2}Ia^2$ and $\frac{1}{2}Ja$ for the Cauchy likelihood (left) and the biweight (right).

Now, consider the Cauchy likelihood. By computing the values where derivatives vanish, we obtain that $C = 1$, $J = 1$, $K = 2$, $I = \frac{1}{4}$. Since $e^0 = 1 + 0$ and $e^{16} < 3^{16} < 9^8 < 17^8$ we obtain, by convexity of e^z, that $e^z < (1+z)^8$ for $y \in [0, 16]$;

hence $\frac{1}{8}a^2 < \log(1+a^2)$. Since $K > I$, Claims 1 and 4 yield, provided k is small, (12) for a small, that is, for $a \in (0, A]$, where A is given by Claim 4. Claims 1 and 5 yield the same for $a \in [A, 4]$. Finally, for $a > 4$ (12) follows via Claims 3 and 5, since if $a \geq 4$, then elementary calculations yield $(1+a^2)^2 \geq 4(1+4a^2)$, resulting in an inequality $\log(1+a^2) \geq \frac{1}{2}\log(1+(2a)^2) + \frac{16}{2(1+16)}$. In all three cases, (5) was established as well. Note that Claim 2 was not needed here (see Fig. 2).

The biweight offers not so much space to spare (see again Fig. 2). On $[-1, 1]$ is ϕ a polynomial: hence all its derivatives are polynomials and the computations of values for well-defined fractions are exact, without numerical approximations (albeit little a bit tedious). Computation of roots of derivatives gives $C = \sqrt{1/5}$, $J = \frac{96}{125}\sqrt{5}$, $K = 6$, $I = \frac{24}{5}$. For small a is again (12) a consequence of Claims 1 and 4, since $K > I$. On the interval $\left(0, \frac{1}{2}\right]$ is ψ strictly concave, since its second derivative $24a(5a^2 - 3)$ is negative there. Let $\tilde{a} = \frac{3}{2}100$. Since $\psi(\tilde{a}) > I\tilde{a}$ and $\psi(a) > Ia$ for all $a \in [0, \tilde{a}]$, we obtain that $\phi(a) > \frac{1}{2}Ia^2$ (note that $\psi(0) = I \cdot 0 = 0$). Hence (12) by Claims 1 and 5. Since also $\psi(\tilde{a}) > \frac{J}{2}$, as well as $\psi(\frac{1}{2}) > \frac{J}{2}$ and $\phi(\tilde{a}) > \frac{1}{2}J\tilde{a}$, we have $\phi(a) > \frac{1}{2}Ja$ and hence (12) for all $a \in \left[\tilde{a}, \frac{1}{2}\right]$ by Claims 2 and 5. Finally, for $a \geq \frac{1}{2}$ is $\phi(a) > \frac{1}{2}$; hence (12) follows by Claims 3 and 5. Again all in cases (5) holds.

(Received March 3, 1994.)

REFERENCES

[1] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers and J. W. Tukey: Robust Estimates of Location: Survey and Advances. Princeton University Press, Princeton 1972.

[2] P. Billingsley: Convergence of Probability Measures. Wiley, New York 1968.

[3] B. R. Clarke: Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. Ann. Statist. *11* (1983), 1196–1205.

[4] L. Davies: The asymptotics of S-estimators in the linear regression model. Ann. Statist. *18* (1990), 1651–1675.

[5] D. L. Donoho and P. J. Huber: The notion of breakdown point. In: Festschrift in Honor of Erich Lehmann (K. Doksum and J. L. Hodges, eds.), Wadsworth, Belmont, CA 1982.

[6] D. A. Freedman and P. Diaconis: On Inconsistent M-estimates. Technical Report No. 170, Department of Statistics, Stanford University 1981.

[7] D. A. Freedman and P. Diaconis: On inconsistent M-estimators. Ann. Statist. *10* (1982), 454–461.

[8] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel: Robust Statistics – The Approach Based on Influence Functions. Wiley, New York 1986.

[9] D. C. Hoaglin, F. Mosteller and J. W. Tukey: Understanding Robust and Exploratory Data Analysis. Wiley, New York 1983.

[10] P. J. Huber: Robust estimators of a location parameter. Ann. Math. Statist. *35* (1964), 73–101.

[11] P. J. Huber: Robust Statistics. Wiley, New York 1981.

[12] P. J. Huber: Finite sample breakdown of M- and P-estimators. Ann. Statist. *12* (1984), 119–126.

[13] I. A. Ibragimov: O kompozicii odnoveršinnych raspredelenij. Teor. Veroyatnost. i Primenen. *1* (1956), 283–288. (In Russian.)

[14] S. Karlin: Total Positivity. Stanford University Press, Stanford 1968.

[15] I. Mizera: Weak continuity and identifiability of M-functionals. Thesis. Charles University, Prague 1993. (In Slovak.)
[16] I. J. Schoenberg: On Pólya frequency functions. Part I: The totally positive functions and their Laplace transforms. J. Analyse Math. *1* (1951), 331–374.

Dr. Ivan Mizera, Department of Probability and Statistics, Comenius University, Mlynská dolina, 842 15 Bratislava. Slovakia.