Z. Wahrscheinlichkeitstheorie verw. Gebiete
56, 127–132 (1981)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

A New Infinitesimal Approach to Robust Estimation

Peter J. Rousseeuw

Fachgruppe für Statistik, RZ H 23, ETH, CH-8092 Zürich, Switzerland

Summary. The infinitesimal robustness of the asymptotic variance of location M-estimators is investigated by means of the change-of-variance curve (CVC), which bears some resemblance to the influence curve (IC). It is proved that this CVC leads to a more stringent robustness property than the IC and that the Huber estimators are still optimal in this new sense.

1. Introduction and Definitions

In his paper (1964),, P. Huber introduced M-estimators and proved their consistency and asymptotic normality. By means of a minimax theory for the asymptotic variance $V(\psi, F)$ he then determined asymptotically most robust estimators, amounting to what is now called a "Huber estimator" at the normal distribution. The infinitesimal bias was investigated by F. Hampel (1968) by means of the influence curve (IC), and optimally robust estimators were found. At the normal, the latter coincide with Huber's solutions. The IC describes the infinitesimal behaviour of the asymptotic value of the M-estimator, whereas the aim of this paper is to study the infinitesimal behaviour of the other very important asymptotic concept, namely $V(\psi, F)$. For this purpose the change-of-variance curve (CVC) is studied. This notion, discovered by F. Hampel in 1972, was sometimes briefly referred to (Hampel, 1973, p. 98) but has not yet appeared in print explicitly. In Sect. 2 it will be shown that the CVC leads to a more stringent robustness concept than does the IC, and in Sect. 3 it is proven that the Huber estimators are still optimal.

Our investigation takes place in the classical framework of M-estimation of a location parameter when scale is known (Huber, 1964). Denote by $\mathbb{R}$ the real line with its Borel σ-algebra, by λ the Lebesgue measure, by Φ the standard normal distribution (identified with its cumulative) and by ϕ its density. We shall restrict our attention to M-estimators for which the corresponding mapping ψ belongs to Ψ. The class Ψ consists of all mappings $\psi: \mathbb{R} \to \mathbb{R}$ satisfying:

(i) ψ is continuous on $\mathbb{R}$, $\psi(-x) = -\psi(x)$ for all x and $\psi(x) \geqq 0$ for $x \geqq 0$;

0044-3719/81/0056/0127/$01.20

(ii) ψ' is defined and continuous on $\mathbb{R} \setminus D(\psi)$, where $D(\psi)$ is a finite set of points which is symmetric with regard to zero;

(iii) $\int \psi^2 \, d\Phi < \infty$;

(iv) $0 < \int \psi' d\Phi = -\int \psi \phi' dx = \int x \psi(x) \, d\Phi(x) < \infty$.

Conditions (i) and (iii) imply that $\int \psi d\Phi = 0$, reflecting Fisher-consistency of the M-estimator corresponding to ψ (see Hampel, 1974). From (i) and (iv) it follows that $0 < \int \psi^2 d\Phi$. Therefore, if

$$A(\psi) = \int \psi^2 d\Phi \quad \text{and} \quad B(\psi) = \int \psi' d\Phi$$

then $0 < A(\psi) < \infty$ and $0 < B(\psi) < \infty$ for all ψ in Ψ. To study the infinitesimal behaviour of the asymptotic value of the estimator corresponding to ψ in the vicinity of Φ, F. Hampel (1974) introduced the influence curve (IC) which equals

$$\Omega(\psi, x) = \psi(x)/B(\psi),$$

and the gross-error-sensitivity

$$\gamma^*(\psi) = \sup_{x \in \mathbb{R}} |\Omega(\psi, x)|.$$

(Of course, his definition is much more general, but here we restrict ourselves to the normal model for simplicity.) The asymptotic variance of the M-estimator corresponding to ψ at the symmetric distribution F equals the expression

$$V(\psi, F) = \int \psi^2 dF / (\int \psi' dF)^2$$

under suitable regularity conditions on ψ and F (Huber, 1967). In order to investigate the infinitesimal stability of $V(\psi, F)$ in the vicinity of Φ, the first idea would be merely to replace F by the type of contaminated normal distribution one uses in the definition of the IC, namely $(1-\varepsilon)\Phi + \varepsilon\delta_x$, where $0 \leqq \varepsilon < 1$ and δ_x is the Dirac probability measure at x. However, this distribution is not symmetric when $x \neq 0$. Following Collins (1977, Formula 2.1) we therefore prefer to evaluate $V(\psi, \Phi_{\varepsilon,x})$ where $\Phi_{\varepsilon,x}$ is the symmetric distribution $(1-\varepsilon)\Phi + \varepsilon(\frac{1}{2}\delta_x + \frac{1}{2}\delta_{-x})$. For all $\psi \in \Psi$, $x \in \mathbb{R} \setminus D(\psi)$ and $0 \leqq \varepsilon < 1$ we have $0 < V(\psi, \Phi_{\varepsilon,x}) \leqq \infty$, and for $\varepsilon = 0$ we obtain $V(\psi, \Phi) = A(\psi)/B^2(\psi)$.

Definition 1. We define the change-of-variance curve (CVC) of the M-estimator corresponding to $\psi \in \Psi$ as

$$\Xi(\psi, x) = \frac{\partial}{\partial \varepsilon}[\log V(\psi, \Phi_{\varepsilon,x})]_{\varepsilon=0},$$

defined in all x for which the right hand side exists.

It follows that $\Xi(\psi, x)$ is well-defined and continuous on $\mathbb{R} \setminus D(\psi)$, where it equals

$$\Xi(\psi, x) = 1 + \psi^2(x)/A(\psi) - 2\psi'(x)/B(\psi).$$

Therefore Ξ is symmetric, whereas Ω is skew-symmetric. For example, the arithmetic mean corresponds to $\psi(x) = x$, so $\Omega(\psi, x) = x$ and $\Xi(\psi, x) = x^2 - 1$. (Note that the CVC may also be defined for distributions other than Φ, and for other types of estimators.)

One might wonder why we work with $\log V$ instead of taking simply V, which would lead to the same curve up to a positive factor. As a matter of fact, the use of the logarithmic derivative brings along many advantages and simplifications on the mathematical level. The use of the logarithmic transformation in similar situations is not new: compare with Huber (1964, Sect. 10) and with Wegman and Carroll (1977, p. 809).

The CVC and the IC have many things in common; for example, we clearly have $\int \Omega(\psi, x)\, d\Phi(x) = 0 = \int \Xi(\psi, x)\, d\Phi(x)$ for all ψ in Ψ. However, both curves cannot be interpreted in the same way. Large positive and large negative values of the IC have qualitatively the same (unfavourable) meaning, namely a bias caused by contamination. (Therefore, $\gamma^*(\psi)$ is defined as the supremum of the *absolute value* of the IC.) On the other hand one does not have to worry about large negative values of a CVC as much as about large positive values, since the latter point to a large positive slope of V. This is in accordance with the reasoning behind Huber's minimax theory (1964): there one is concerned only about the large values of $V(\psi, F)$ (where F belongs to a neighborhood of Φ), and not about small values (see also Collins, 1977). Therefore, we define:

Definition 2. The change-of-variance sensitivity of the M-estimator corresponding to $\psi \in \Psi$ is defined as

$$\kappa^*(\psi) = \sup\{\Xi(\psi, x);\, x \in \mathbb{R} \setminus D(\psi)\}.$$

This supremum may be compared with the minimization in Theorem 4.1 and Corollary 4.1 of (Collins, 1977).

2. V-robustness and B-robustness

When $\kappa^*(\psi) < \infty$ we say that the estimator corresponding to ψ is *V-robust*; when $\gamma^*(\psi) < \infty$ we say it is *B-robust*.

Theorem 1. *For all ψ in Ψ, V-robustness implies B-robustness.*

Proof. Suppose that $\kappa^*(\psi) < \infty$ and $\gamma^*(\psi) = \infty$. There exists $K > 0$ such that $D(\psi) \subset (-K, K)$. From $\gamma^*(\psi) = \infty$ it follows that $\sup_{[K,\infty)} \psi = \infty$. Since $\kappa^*(\psi) < \infty$, there exists a constant $M > 0$ such that $\sup_{[K,\infty)} \Xi(\psi, x) \leqq M + 1$. Now there exists a point $x_0 \geqq K$ for which $\psi(x_0) > (A(\psi)\, M)^{1/2}$. If we would have $\psi'(x_0) \leqq 0$, then

$$\Xi(\psi, x_0) = 1 + \psi^2(x_0)/A(\psi) - 2\psi'(x_0)/B(\psi) > 1 + M$$

which is a contradiction; therefore $\psi'(x_0) > 0$. Hence there exists $\varepsilon > 0$ such that $\psi(z) > \psi(x_0)$ for all $z \in (x_0, x_0 + \varepsilon)$.

We now prove that $\psi(x) > \psi(x_0) > (A(\psi) M)^{1/2}$ for all $x > x_0$. Suppose there exists $y > x_0$ such that $\psi(y) \leqq \psi(x_0)$. Then $x' = \inf\{x > x_0;\ \psi(x) \leqq \psi(x_0)\}$ must satisfy $x' < \infty$. Clearly $x' \geqq x_0 + \varepsilon > x_0$. It also follows that $\psi(x') = \psi(x_0)$ and $\psi(x) > \psi(x_0)$ for all $x \in (x_0, x')$. There exists $x'' \in (x_0, x')$ such that $\psi'(x'') \leqq 0$ (otherwise $\psi(x_0) < \psi(x')$); moreover $\psi(x'') > \psi(x_0) > (A(\psi)\, M)^{1/2}$. Hence $\Xi(\psi, x'') \geqq 1 + \psi^2(x'')/A(\psi) > 1 + M$, a contradiction.

Put $d=B(\psi)/2A(\psi)>0$; it is easily checked that $\psi'(x)/[\psi^2(x)-A(\psi)M]\geqq d$ for all $x\geqq x_0$. For all $y\geqq y_0$, we define

$$Q(y)=-(A(\psi)M)^{-1/2}\coth^{-1}(\psi(y)/(A(\psi)M)^{1/2}).$$

Since $y\geqq x_0\geqq K$ and $\psi(y)>(A(\psi)M)^{1/2}$, the mapping Q is well-defined, continuous and differentiable with derivative $\psi'(y)/[\psi^2(y)-A(\psi)M]$ on $[x_0,\infty)$. Hence $Q(x)-Q(x_0)\geqq d(x-x_0)$ for all $x\geqq x_0$, so

$$\coth^{-1}(\psi(x)/(A(\psi)M)^{1/2})\leqq(A(\psi)M)^{1/2}(dx_0-Q(x_0))-d(A(\psi)M)^{1/2}x.$$

The left hand side of this inequality is strictly positive because $\psi(x)/(A(\psi)M)^{1/2}>1$, but the right hand side will tend to $-\infty$ when $x\to\infty$, which is clearly impossible. This ends the proof.

Corollary. *For monotone ψ in Ψ, V-robustness and B-robustness are equivalent.*

Proof. Observe that $\kappa^*(\psi)\leqq 1+(\gamma^*(\psi))^2B^2(\psi)/A(\psi)$.

For example, Huber estimators satisfy both properties (see next section), whereas the arithmetic mean satisfies neither. However, in general the converse of Theorem 1 is not true. Consider the following counterexample:

$$\begin{aligned}\psi(x)&=x && 0\leqq|x|\leqq 1\\ &=(2-|x|)^{1/2}\operatorname{sign}(x) && 1\leqq|x|\leqq 2\\ &=0 && 2\leqq|x|.\end{aligned}$$

Clearly $\gamma^*(\psi)<\infty$, but $\lim_{2>x\to 2}\Xi(\psi,x)=\infty$ so $\kappa^*(\psi)=\infty$.

3. Optimal V-robustness

One can ask the following question: which mappings ψ minimize $V(\psi,\Phi)$ (meaning the corresponding estimator has maximal asymptotic efficiency) under the side condition of an upper bound on $\gamma^*(\psi)$? The solution of this well-known extremal problem (which we shall refer to as "*optimal B-robustness*") has been given by Hampel (1968). In our framework it is given by Lemma 2 below. If one now replaces $\gamma^*(\psi)$ by $\kappa^*(\psi)$, one can speak of "*optimal V-robustness*"; it is shown in Theorem 2 that the Huber estimators are also optimal in this sense.

For any $b\in(0,\infty)$, the M-estimator corresponding to $\psi_b(x)=\min(b,\max(x,-b))$ is called a Huber estimator. It is clear that $\psi_b\in\Psi$, where $D(\psi_b)=\{-b,b\}$. One verifies that $A(\psi_b)=2\Phi(b)-1-2b\phi(b)+2b^2(1-\Phi(b))$ and $B(\psi_b)=2\Phi(b)-1$, so $0<A(\psi_b)<1$ and $0<B(\psi_b)<1$. For all $b\in(0,\infty)$, we put $g(b)=\gamma^*(\psi_b)=b/B(\psi_b)$ and $k(b)=\kappa^*(\psi_b)=1+b^2/A(\psi_b)$.

Lemma 1. *The mappings $g:b\to g(b)$ and $k:b\to k(b)$ are increasing continuous bijections from* $(0,\infty)$ *onto* $((\pi/2)^{1/2},\infty)$ *and from* $(0,\infty)$ *onto* $(2,\infty)$.

Proof. The mapping c: $[0,\infty)\to[0,1):t\to 2\Phi(t)-1$ is strictly concave because c' is strictly decreasing. Thus for all $0<r<s<\infty$ we have $(c(s)-c(0))/s<(c(r)$

$-c(0))/r$, so $g(r)=r/c(r)<s/c(s)=g(s)$. Hence g is strictly increasing. It is clearly continuous, and $\lim_{b\to\infty} g(b)=\infty$. By L'Hospital's rule, $\lim_{0<b\to 0} g(b)=(\pi/2)^{1/2}$. The mapping k is also strictly increasing, since $k'(b)=2b \int_{[-b,b]} x^2\,d\Phi(x)/A^2(\psi_b)>0$, and we have $\lim_{0<b\to 0} k(b)=2$ and $\lim_{b\to\infty} k(b)=\infty$.

Lemma 2. (Hampel). *For each $g>(\pi/2)^{1/2}$ there exists a unique value $b>0$ such that $\gamma^*(\psi_b)=g$, and ψ_b minimizes $V(\psi,\Phi)$ among all ψ in Ψ which satisfy $\gamma^*(\psi)\leqq g$. Any other solution of this extremal problem coincides with a positive nonzero multiple of ψ_b.*

Theorem 2. *For each $k>2$ there exists a unique value $b>0$ such that $\kappa^*(\psi_b)=k$, and ψ_b minimizes $V(\psi,\Phi)$ among all ψ in Ψ which satisfy $\kappa^*(\psi)\leqq k$. Any other solution of this extremal problem coincides with a positive nonzero multiple of ψ_b.*

Proof. We must show that ψ_b is optimal and unique in this sense.

1. Multiplication of $\psi\in\Psi$ with any factor $r>0$ changes neither $V(\psi,\Phi)$ or $\Xi(\psi,x)$. Without loss of generality we may therefore minimize $A(\psi)$, where $\psi\in\Psi$ is subject to:

$$\Xi(\psi,x)\leqq k \quad \text{for all} \quad x\in\mathbb{R}\setminus D(\psi) \tag{3.1}$$

and

$$B(\psi)=B(\psi_b). \tag{3.2}$$

Using condition (iv) of the definition of Ψ and (3.2) one verifies that $\int(x-\psi(x))^2\,d\Phi(x)=1-2B(\psi_b)+A(\psi)$, hence minimizing $A(\psi)$ is equivalent to minimizing $\int(x-\psi(x))^2\,d\Phi(x)$.

2. We must show that ψ_b is optimal, so we suppose that there exists a mapping $\psi^*\in\Psi$ which satisfies (3.1), (3.2) and

$$A(\psi^*)<A(\psi_b). \tag{3.3}$$

Then $\int(x-\psi^*(x))^2\,d\Phi(x)<\int(x-\psi_b(x))^2\,d\Phi(x)$, hence there exists $x_0\in\mathbb{R}$ such that $(x_0-\psi^*(x_0))^2<(x_0-\psi_b(x_0))^2$. Let $x_0\geqq 0$ w.l.o.g., hence $x_0>b$ and $\psi^*(x_0)>\psi_b(x_0)=b$. Since ψ^* is continuous, we may suppose that $x_0\notin D(\psi^*)$. Now it is impossible that $(\psi^*)'(x_0)\leqq 0$, because then we would have $\Xi(\psi^*,x_0)>\Xi(\psi_b,x_0)=k$ from (3.2) and (3.3), contradicting (3.1). Thus $(\psi^*)'(x_0)>0$ and there exists $\varepsilon>0$ such that $\psi^*(z)>\psi^*(x_0)$ for all $z\in(x_0,x_0+\varepsilon)$.

3. We prove that $\psi^*(x)>\psi^*(x_0)>b$ for all $x>x_0$, by means of the technique used in the proof of Theorem 1 (going from one point of $D(\psi^*)\cap(x_0,x')$ to the next where necessary.)

4. Take $K\geqq x_0$ such that ψ^* is continuously differentiable on $[K,\infty)$. We find that $(\psi^*)'(x)/[(\psi^*)^2(x)-b^2]\geqq a$ for all $x\geqq K$, where $a=B(\psi_b)/2A(\psi_b)>0$. Putting $P(y)=-\coth^{-1}(\psi^*(y)/b)/b$ for all $y\geqq K$, we obtain the final contradiction by the method used in Theorem 1.

5. Suppose another solution $\bar{\psi}$ exists. W.l.o.g. we may put $B(\bar{\psi})=B(\psi_b)$ and therefore also $A(\bar{\psi})=A(\psi_b)$, so it suffices to prove that $\bar{\psi}=\psi_b$. By means of a reasoning analogous to parts 2, 3 and 4 of the proof, we show no point $x_0>b$ can exist such that $\bar{\psi}(x_0)>\psi_b(x_0)$, hence $\bar{\psi}\leqq\psi_b$ on (b,∞) and $\bar{\psi}\geqq\psi_b$ on $(-\infty,$

$-b$). From part 1 and $A(\bar{\psi})=A(\psi_b)$ it follows that

$$\int_{|x|\geqq b}(x-\psi_b(x))^2\,d\Phi(x)=\int_{|x|<b}(x-\bar{\psi}(x))^2\,d\Phi(x)+\int_{|x|\geqq b}(x-\bar{\psi}(x))^2\,d\Phi(x).$$

Together with the above inequalities (which entail $(x-\bar{\psi}(x))^2\geqq(x-\psi_b(x))^2$ for $|x|\geqq b$) and continuity, this implies that the above inequalities can nowhere be strict, hence $\bar{\psi}=\psi_b$ on $(-\infty,-b)\cup(b,\infty)$. It then also follows that $\int_{|x|<b}(x-\bar{\psi}(x))^2\,d\Phi(x)=0$, and again using continuity we see that $\bar{\psi}(x)=x=\psi_b(x)$ on $(-b,b)$. We conclude that $\bar{\psi}=\psi_b$, which ends the proof.

Some further research, concerning the case with a finite rejection point, is presently being completed in collaboration with F. Hampel and E. Ronchetti.

Remark. Throughout this paper, it is possible to replace Φ by any distribution G having a twice continuously differentiable density g which is strictly positive and symmetric with respect to zero, and for which the mapping $\Lambda=(-\log g)'=-g'/g$ satisfies $\Lambda'(x)>0$ for all x, and $I(G)=\int\Lambda^2\,dG=\int\Lambda'\,dG<\infty$. (Here Λ corresponds to the maximum likelihood estimator, and $I(G)$ is the Fisher information.) Theorem 1 and its corollary remain unchanged. On the other hand, Huber's minimax asymptotic variance theorem (1964, p. 80) is applicable with solutions

$$\begin{aligned}\psi_t(x)&=\Lambda(x) &&\text{for } |x|\leqq t\\ &=\Lambda(t)\operatorname{sign}(x) &&\text{for } |x|>t\end{aligned}$$

where $t\in(0,\infty)$. It can be proven that the sensitivities of the ψ_t are increasing bijections from $(0,\infty)$ onto $((2g(0))^{-1},\gamma^*(\Lambda))$ and $(2,\kappa^*(\Lambda))$, and that the ψ_t are optimally B-robust as well as optimally V-robust at G.

Acknowledgments. This research was supported by the NFWO Belgium. The author wishes to thank F. Hampel for his generous guidance and encouragement, and P. Huber and a referee for suggestions for improving the paper.

References

1. Collins, J.R.: Upper bounds on asymptotic variances of M-estimators of location. Ann. Statist. **5**, 646–657 (1977)
2. Hampel, F.R.: Contributions to the theory of robust estimation. Unpublished dissertation, Berkeley: University of California 1968
3. Hampel, F.R.: Robust estimation: a condensed partial survey. Z. Wahrscheinlichkeitstheorie verw. Gebiete **27**, 87–104 (1973)
4. Hampel, F.R.: The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. **69**, 383–393 (1974)
5. Huber, P.J.: Robust estimation of a location parameter. Ann. Math. Statist. **35**, 73–101 (1964)
6. Huber, P.J.: The behaviour of maximum likelihood estimates under nonstandard conditions. Proc. Fifth Berkeley Sympos. Math. Statist. Probab. **1**, 221–233 Univ. Calif. (1967)
7. Wegman, E.J., Carroll, R.J.: A Monte Carlo study of robust estimators of location. Commun. Statist. (Theory and Methods) **A6**, 795–812 (1977)

Received June 10, 1980; in revised form September 10, 1980