



The change-of-variance function of M-estimators of scale under general contamination

Marc G. Genton^a, Peter J. Rousseeuw^{b,*}

^a *Applied Statistics Group, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland*

^b *Department of Mathematics, University of Antwerp (UIA), Universiteitsplein 1, B-2610 Wilrijk, Belgium*

Received 27 January 1994; revised 3 November 1994

Abstract

In this paper we derive the change-of-variance function of M-estimators of scale under general contamination, thereby extending the formula in Hampel et al. (1986). We say that an M-estimator is B-robust if its influence function is bounded, and we call it V-robust if its change-of-variance function is bounded from above. It is shown, for a natural class of M-estimators, that the general notion of V-robustness still implies B-robustness. Several classes of M-estimators are studied closely, as well as some typical examples and their interpretation.

Keywords: Influence function; Change-of-variance function; B-robustness; V-robustness

1. Introduction

The *influence function* $IF(x, S, F)$ of a statistical functional S at a distribution F is defined as the kernel of a first-order von Mises derivative:

$$\int IF(x, S, F) dG(x) = \frac{\partial}{\partial \varepsilon} [S((1 - \varepsilon)F + \varepsilon G)]_{\varepsilon=0}, \quad (1.1)$$

where G ranges over all distributions (including point masses). Analogously, the *change-of-variance function* [3] is defined by

$$\int CVF(x, S, F) dG(x) = \frac{\partial}{\partial \varepsilon} [V(S, (1 - \varepsilon)F + \varepsilon G)]_{\varepsilon=0}, \quad (1.2)$$

* Corresponding author.

where $V(S, F)$ is the asymptotic variance of S at F . The latter formula [1, p. 128] was applied to M-estimators of location, but for M-estimators of scale only distributions G with $S(G) = S(F) = 1$ were used for simplicity. In the present paper we derive the change-of-variance function for M-estimators of scale under general contaminating distributions G .

Let us recall the definition of an M-estimator of scale. Suppose we have one-dimensional observations X_1, \dots, X_n which are independent and identically distributed according to a distribution from the parametric model $\{F_\sigma; \sigma > 0\}$, where $F_\sigma(x) = F(x/\sigma)$. An M-estimator $S_n(X_1, \dots, X_n)$ of σ is given by

$$\sum_{i=1}^n \chi(X_i/S_n) = 0$$

and corresponds to the statistical functional S defined by

$$\int \chi(x/S(F)) dF(x) = 0. \quad (1.3)$$

The influence function of S is

$$\text{IF}(u, S, F) = \frac{\chi(u/S(F))S^2(F)}{\int x \chi'(x/S(F)) dF(x)}. \quad (1.4)$$

For more information, see [1]. An important summary value of the influence function is the *gross-error sensitivity* of S at F , defined by

$$\gamma^* = \sup_u |\text{IF}(u, S, F)|. \quad (1.5)$$

It measures the worst influence that a small amount of contamination can have on the value of the estimator. Therefore, a desirable feature is that γ^* be finite, in which case S is called *B-robust* (bias-robust) at F .

Under certain regularity conditions, $\sqrt{n}(S_n - \sigma)$ is asymptotically normal with asymptotic variance

$$\begin{aligned} V(S, F) &= \int \text{IF}^2(u, S, F) dF(u) \\ &= \frac{\int \chi^2(u/S(F)) S^4(F) dF(u)}{(\int x \chi'(x/S(F)) dF(x))^2}. \end{aligned} \quad (1.6)$$

The change-of-variance function is then found by inserting (1.6) in (1.2), and the resulting expression will be given in Section 2. We then define the *change-of-variance sensitivity* κ^* as $+\infty$ if a delta function with positive factor occurs in the CVF, and otherwise as

$$\kappa^* = \sup_z \frac{\text{CVF}(z, S, F)}{V(S, F)}. \quad (1.7)$$

Note that large negative values of the CVF merely point to a decrease in V , indicating a better accuracy. If κ^* is finite then S is called *V-robust* (variance-robust) at F .

2. The change-of-variance function of M-estimators of scale

Recall that $F_\sigma(x) = F(x/\sigma)$. We need the following regularity conditions on F :

- (F1) F has a twice continuously differentiable density f (with respect to the Lebesgue measure λ) which is symmetric around zero and satisfies $f(x) > 0 \forall x \in \mathbb{R}$.
- (F2) The mapping $A = -f'/f = (-\ln f)'$ satisfies $A'(x) > 0 \forall x \in \mathbb{R}$, and $\int A' f d\lambda = -\int A f' d\lambda < \infty$.

Let us denote

$$A(\chi) = \int \chi^2(x) dF(x), \tag{2.1}$$

$$B(\chi) = \int x\chi'(x) dF(x). \tag{2.2}$$

We will assume that χ belongs to the class Ψ of all functions satisfying the following four regularity conditions:

- (R1) χ is well-defined and continuous on $\mathbb{R} \setminus D^{(0)}(\chi)$, where $D^{(0)}(\chi)$ is finite. In each point of $D^{(0)}(\chi)$ there exist finite left and right limits of χ which are different. Also $\chi(-x) = \chi(x)$ if $\{-x, x\} \subset \mathbb{R} \setminus D^{(0)}(\chi)$, and there exists $d > 0$ such that $\chi(x) \leq 0$ on $(0, d)$ and $\chi(x) \geq 0$ on (d, ∞) .
- (R2) The set $D^{(1)}(\chi)$ of points in which χ is continuous but in which χ' is not defined or not continuous, is finite.
- (R3) $\int \chi(x) dF(x) = 0$ (Fisher consistency) and $0 < A(\chi) < \infty$.
- (R4) $0 < B(\chi) = \int (xA(x) - 1)\chi(x) dF(x) < \infty$.

From (1.2) and (1.6) we obtain

$$\begin{aligned} \text{CVF}(z, S, F) &= \left(\int \chi'(x/S(F))x dF(x) \right)^{-3} \left[\left(\int \chi'(x/S(F))x dF(x) \right) \right. \\ &\quad \times \left(- \int \chi^2(u/S(F))S^4(F) dF(u) + \chi^2(z/S(F))S^4(F) \right. \\ &\quad \left. - 2\text{IF}(z, S, F) \int \chi(u/S(F))\chi'(u/S(F))(u/S^2(F))S^4(F) dF(u) \right. \\ &\quad \left. + 4\text{IF}(z, S, F) \int \chi^2(u/S(F))S^3(F) dF(u) \right) \\ &\quad \left. - 2 \left(\int \chi^2(u/S(F))S^4(F) dF(u) \right) \right. \\ &\quad \times \left(- \int \chi'(x/S(F))x dF(x) + (\chi'(z/S(F)))z \right. \\ &\quad \left. \left. - \text{IF}(z, S, F) \int \chi''(x/S(F))(x/S^2(F))x dF(x) \right) \right]. \tag{2.3} \end{aligned}$$

Making use of (2.1), (2.2), and $S(F) = 1$ at the model distribution, (2.3) becomes

$$\text{CVF}(z, S, F) = \frac{A(\chi)}{B^2(\chi)} \left[1 + \frac{\chi^2(z)}{A(\chi)} - 2 \frac{z\chi'(z)}{B(\chi)} + C(\chi) \frac{\chi(z)}{B(\chi)} \right], \quad (2.4)$$

where

$$C(\chi) = 4 - \frac{2}{A(\chi)} \int u\chi(u)\chi'(u) dF(u) + \frac{2}{B(\chi)} \int u^2 \chi''(u) dF(u). \quad (2.5)$$

Note that (2.4) differs from the expression in [1] by the addition of the last term, the integral of which is zero when $S(G) = 1$. This distinction does not exist for location, at least in the case of odd ψ , as can be seen in [1, pp. 145–146], where

$$\tilde{C}(\psi) = 2 \int \left(\frac{\psi''(u)}{B(\psi)} - \frac{\psi(u)\psi'(u)}{A(\psi)} \right) dF(u) = 0.$$

From here on we will assume that $C(\chi) \geq 0$, which is true in all practical applications. In Section 4.2 we will derive an alternative expression for $C(\chi)$ which is easier to compute than (2.5).

3. Relation between B-robustness and V-robustness

Let us define

$$\gamma^- = \sup_{u \in (0, d)} (-\text{IF}(u, S, F)), \quad (3.1)$$

$$\gamma^+ = \sup_{u \in (d, +\infty)} \text{IF}(u, S, F). \quad (3.2)$$

In the theorems below we will impose that $\gamma^+ \geq \gamma^-$ (and hence $\gamma^* = \gamma^+$). This is a very natural requirement for scale estimators. For instance, when discussing breakdown properties [2], notes that $\gamma^+ \geq \gamma^-$ in the more interesting cases. The opposite situation leads to implosion of the scale estimator, as well as to lower efficiency.

The first theorem shows that the concept of V-robustness is stronger than the concept of B-robustness.

Theorem 1. For all $\chi \in \Psi$ with $\gamma^+ \geq \gamma^-$ and $C(\chi) \geq 0$, V-robustness implies B-robustness. In fact

$$\gamma^* \leq \frac{1}{2} [\sqrt{V^2(S, F)C^2(\chi) + 4V(S, F)(\kappa^* - 1)} - V(S, F)C(\chi)].$$

Proof. Suppose that κ^* is finite and that there exists some x_0 for which

$$|\text{IF}(x_0, S, F)| > \frac{1}{2} [\sqrt{V^2(S, F)C^2(\chi) + 4V(S, F)(\kappa^* - 1)} - V(S, F)C(\chi)].$$

Without loss of generality, put $x_0 \notin D^{(1)}(\chi)$ and $x_0 > d$. It follows that

$$\chi(x_0) > \frac{1}{2} \left[\sqrt{\left(\frac{A(\chi)C(\chi)}{B(\chi)}\right)^2 + 4A(\chi)(\kappa^* - 1)} - \frac{A(\chi)C(\chi)}{B(\chi)} \right] = b.$$

If $\chi'(x_0) \leq 0$ then

$$1 + \frac{\chi^2(x_0)}{A(\chi)} - 2 \frac{x_0 \chi'(x_0)}{B(\chi)} + \frac{C(\chi)}{B(\chi)} \chi(x_0) \geq 1 + \frac{b^2}{A(\chi)} + \frac{C(\chi)}{B(\chi)} b = \kappa^*,$$

a contradiction. Therefore, $\chi'(x_0) > 0$. Since we have $\chi(x_0) > 0$, there exists $\varepsilon > 0$ such that $\chi'(t) > 0$ for all t in $[x_0, x_0 + \varepsilon]$, so $\chi(x) > \chi(x_0)$ for all x in $(x_0, x_0 + \varepsilon]$. It follows that $\chi(x) > \chi(x_0) > b$ for all $x > x_0$, $x \notin D^{(0)}(\chi)$ because only upward jumps of χ are allowed for positive x . As $D^{(0)}(\chi) \cup D^{(1)}(\chi)$ is finite, we may assume that $[x_0, +\infty) \cap (D^{(0)}(\chi) \cup D^{(1)}(\chi))$ is empty. It holds that

$$1 + \frac{\chi^2(x)}{A(\chi)} - 2 \frac{x \chi'(x)}{B(\chi)} + \frac{C(\chi)}{B(\chi)} \chi(x) \leq \kappa^*.$$

Therefore

$$\chi^2(x) - 2x \chi'(x) \frac{A(\chi)}{B(\chi)} \leq A(\chi)(\kappa^* - 1) - \frac{C(\chi)A(\chi)}{B(\chi)} \chi(x) \leq A(\chi)(\kappa^* - 1) - \frac{C(\chi)A(\chi)}{B(\chi)} b \leq b^2,$$

hence

$$\chi^2(x) - 2x \chi'(x) \frac{A(\chi)}{B(\chi)} \leq b^2$$

for all $x \geq x_0$. Hence

$$\frac{\chi'(x)}{\chi^2(x) - b^2} \geq \frac{B(\chi)}{2A(\chi)} \frac{1}{x}.$$

Putting

$$R(x) = -\frac{1}{b} \coth^{-1} \left(\frac{\chi(x)}{b} \right)$$

and

$$P(x) = \frac{B(\chi)}{2A(\chi)} \ln(x),$$

it follows that $R'(x) \geq P'(x)$ for all $x \geq x_0$. Hence $R(x) - R(x_0) \geq P(x) - P(x_0)$, and thus

$$\coth^{-1} \left(\frac{\chi(x)}{b} \right) \leq b \left[P(x_0) - R(x_0) - \frac{B(\chi)}{2A(\chi)} \ln(x) \right].$$

However, the left member is positive because $\chi(x) > b$ and the right member tends to $-\infty$ for $x \rightarrow \infty$, a contradiction. This proves the desired inequality. \square

Theorem 2. For all $\chi \in \Psi$ with $\gamma^+ \geq \gamma^-$ and $C(\chi) \geq 0$, and χ nondecreasing for $x \geq 0$, V-robustness and B-robustness are equivalent. In fact

$$\kappa^* = 1 + \frac{(\gamma^*)^2}{V(S, F)} + C(\chi)\gamma^*.$$

Proof. One of the two inequalities follows from Theorem 1. For the other, assume that S is B-robust. Because χ is monotone, the CVF can only contain negative delta functions, which do not contribute to κ^* . For all $x \geq 0$ it holds that $\chi'(x) \geq 0$, so

$$1 + \frac{\chi^2(x)}{A(\chi)} - 2 \frac{x\chi'(x)}{B(\chi)} + \frac{C(\chi)}{B(\chi)} \chi(x) \leq 1 + \frac{(\gamma^*)^2}{V(S, F)} + C(\chi)\gamma^*.$$

Hence, S is also V-robust. \square

Theorem 3. For all $\chi \in \Psi$ with $\gamma^+ \geq \gamma^-$ and $C(\chi) \geq 0$, and χ nondecreasing for $x \geq 0$, we have

$$\kappa^* \geq 2 + C(\chi)\gamma^*.$$

Proof. We have

$$V(S, F) = \int \text{IF}^2(u, S, F) dF(u) \leq (\gamma^*)^2.$$

Using Theorem 2, it follows that

$$\kappa^* = 1 + \frac{(\gamma^*)^2}{V(S, F)} + C(\chi)\gamma^* \geq 2 + C(\chi)\gamma^*. \quad \square$$

4. Examples

4.1. The L^q scale estimator

The L^q scale estimator at F is given by

$$\chi(x) = |x|^q - \int |x|^q dF(x), \quad \text{with } q > 0. \quad (4.1)$$

Theorem 4. For any distribution F and any $q > 0$, the L^q scale estimator satisfies

$$C(\chi) = 2.$$

Proof. From $\chi'(x) = q|x|^{q-1} \text{sign}(x)$ we deduce the two relations

$$x\chi'(x) = q\chi(x) + B(\chi)$$

and

$$x^2 \chi''(x) = (q - 1)x\chi'(x).$$

This yields

$$\int x\chi(x)\chi'(x)dF(x) = q \int \chi^2(x)dF(x) + B(\chi) \int \chi(x)dF(x) = qA(\chi),$$

$$\int x^2 \chi''(x)dF(x) = (q - 1)B(\chi).$$

Hence

$$C(\chi) = 4 - \frac{2}{A(\chi)} qA(\chi) + \frac{2}{B(\chi)} (q - 1)B(\chi) = 2. \quad \square$$

Theorem 5. *The L^q scale estimator is neither B-robust nor V-robust at any distribution F , that is to say*

$$\gamma^* = \infty \quad \text{and} \quad \kappa^* = \infty.$$

Proof. As χ is unbounded, the estimator is not B-robust. Moreover, as the CVF behaves like x^{2q} with a positive factor when $x \rightarrow \infty$, it is not bounded from above. \square

The maximum likelihood estimator (MLE) at $F = \Phi$ is given by $\chi(x) = x^2 - 1$, obtained by putting $q = 2$ in (4.1). This yields

$$A(\chi) = \int \chi^2(x)d\Phi(x) = 2,$$

$$B(\chi) = \int x\chi'(x)d\Phi(x) = 2,$$

$$\int \chi(x)\chi'(x)x d\Phi(x) = 4,$$

$$\int \chi''(x)x^2 d\Phi(x) = 2.$$

Hence

$$\text{IF}(u, S, \Phi) = \frac{1}{2}(u^2 - 1) \quad \text{with} \quad \gamma^* = \infty,$$

$$\text{CVF}(z, S, \Phi) = \frac{1}{4}(z^4 - 4z^2 + 1) \quad \text{with} \quad \kappa^* = \infty.$$

Both functions are plotted in Fig. 1. We see that the maximum likelihood estimator at Φ is neither B-robust nor V-robust. For $q = 1$ we obtain the mean deviation with $\chi(x) = |x| - \sqrt{2/\pi}$ which is again neither B-robust nor V-robust.

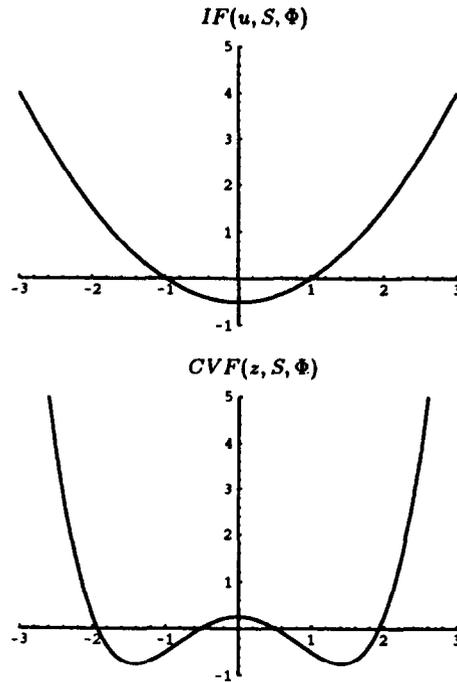


Fig. 1. The influence function and change-of-variance function of the MLE.

4.2. Computation of $C(\chi)$ at the Gaussian model

Let us recall that

$$C(\chi) = 4 - \frac{2}{A(\chi)} \int x\chi(x)\chi'(x)dF(x) + \frac{2}{B(\chi)} \int x^2\chi''(x)dF(x).$$

Theorem 6. At the Gaussian distribution $F = \Phi$ we have

$$C(\chi) = 1 - \frac{1}{A(\chi)} \int x^2\chi^2(x)d\Phi(x) + \frac{2}{B(\chi)} \int (x^4 - 3x^2)\chi(x)d\Phi(x).$$

Proof. Denoting the density of Φ by ϕ we find

$$\begin{aligned} \int x\chi(x)\chi'(x)d\Phi(x) &= \frac{1}{2} \int x(\chi^2(x))'\phi(x)dx \\ &= -\frac{1}{2} \int \chi^2(x)(\phi(x) + x\phi'(x))dx \\ &= -\frac{1}{2} \int \chi^2(x)(1 - x^2)\phi(x)dx \\ &= \frac{1}{2} \left(\int x^2\chi^2(x)d\Phi(x) - A(\chi) \right) \end{aligned} \tag{4.2}$$

and

$$\begin{aligned}
 \int x^2 \chi''(x) d\Phi(x) &= \int x^2 (\chi'(x))' \phi(x) dx \\
 &= - \int (2x \phi(x) + x^2 \phi'(x)) \chi'(x) dx \\
 &= - 2B(\chi) + \int x^3 \chi'(x) \phi(x) dx \\
 &= - 2B(\chi) - \int (x^3 \phi(x))' \chi(x) dx \\
 &= - 2B(\chi) + \int (x^4 - 3x^2) \chi(x) d\Phi(x).
 \end{aligned} \tag{4.3}$$

This yields

$$\begin{aligned}
 C(\chi) &= 4 - \frac{1}{A(\chi)} \left(\int x^2 \chi^2(x) d\Phi(x) - A(\chi) \right) + \frac{2}{B(\chi)} \left(- 2B(\chi) + \int (x^4 - 3x^2) \chi(x) d\Phi(x) \right) \\
 &= 1 - \frac{1}{A(\chi)} \int x^2 \chi^2(x) d\Phi(x) + \frac{2}{B(\chi)} \int (x^4 - 3x^2) \chi(x) d\Phi(x). \quad \square
 \end{aligned}$$

4.3. The λ th absolute deviation estimator (λ -MAD) at the Gaussian model

Consider the λ th absolute deviation estimator (λ -MAD) at $F = \Phi$ given by

$$\chi(x) = \begin{cases} (\lambda - 1)/\lambda & \text{if } -\Phi^{-1}(\frac{1}{2} + \frac{1}{2}\lambda) < x < \Phi^{-1}(\frac{1}{2} + \frac{1}{2}\lambda), \\ 1 & \text{elsewhere,} \end{cases}$$

with $0 < \lambda < 1$. Let us now look at Fig. 2, where $C(\chi)$, $\gamma^*(\chi)$ and $\kappa^*(\chi)$ are plotted as functions of λ .

First of all, we see that $C(\chi) > 0$ for all λ . Secondly, the gross-error sensitivity is minimal for $\lambda = \frac{1}{2}$, which corresponds to the usual median absolute deviation (MAD). Finally, the change-of-variance sensitivity tends to the value 2 as λ tends to zero. However, note that for $\lambda < \frac{1}{2}$ we do not have the condition $\gamma^+ \geq \gamma^-$ required by the theorems of Section 3.

Consider the special case of $\lambda = \frac{1}{2}$, which corresponds to the usual median absolute deviation at $F = \Phi$, given by $\chi(x) = \text{sign}(|x| - q)$ where $q = \Phi^{-1}(3/4)$. This yields

$$A(\chi) = \int \chi^2(x) d\Phi(x) = 1,$$

$$B(\chi) = \int x \chi'(x) d\Phi(x) = 4q \phi(q),$$

$$\int x^2 \chi^2(x) d\Phi(x) = 1,$$

$$\int (x^4 - 3x^2) \chi(x) d\Phi(x) = 4q^3 \phi(q).$$

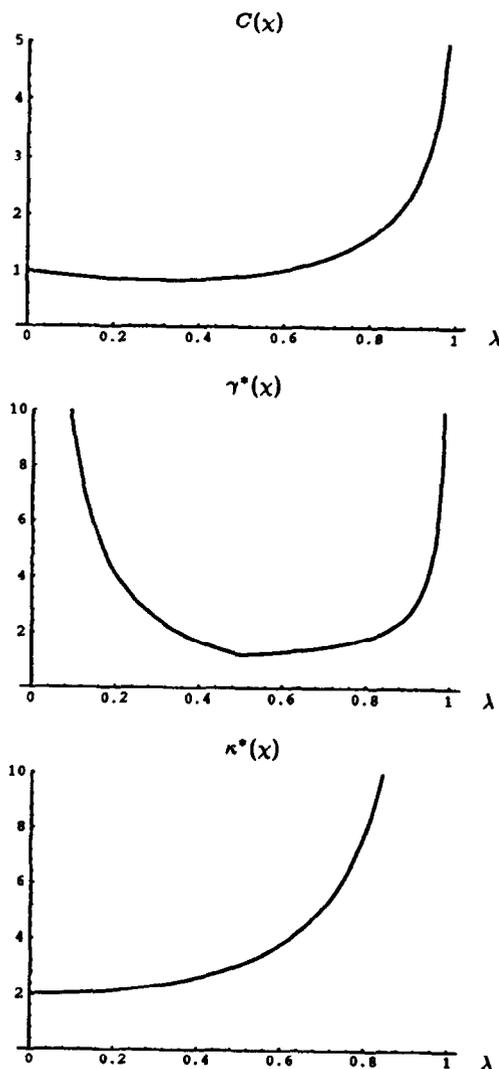


Fig. 2. The values of $C(\lambda)$, $\gamma^*(\lambda)$ and $\kappa^*(\lambda)$ as a function of λ for the λ -MAD.

Therefore

$$\text{IF}(u, S, \Phi) = \frac{\text{sign}(|u| - q)}{4q\phi(q)} \quad \text{with } \gamma^* = \frac{1}{4q\phi(q)} = 1.166,$$

$$\text{CVF}(z, S, \Phi) = \frac{1}{(4q\phi(q))^2} \left[2 - \frac{1}{q\phi(q)} (\delta_q(z) + \delta_{-q}(z)) + 2q^2 \frac{\text{sign}(|z| - q)}{4q\phi(q)} \right]$$

$$\text{with } \kappa^* = 2 + \frac{q}{2\phi(q)} = 3.061.$$

The MAD at Φ is thus both B-robust and V-robust (see Fig. 3).

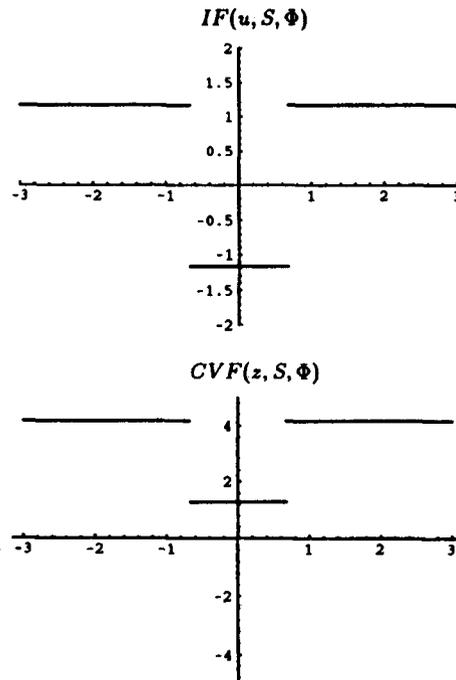


Fig. 3. The influence function and change-of-variance function of the MAD.

4.4. The Welsch estimator at the Gaussian model

Let us consider the Welsch estimator family at $F = \Phi$ given by

$$\chi(x) = \int \exp\left(-\frac{x^2}{d}\right) d\Phi(x) - \exp\left(-\frac{x^2}{d}\right) \quad \text{with } d > 0,$$

and let us look at the graphs of $C(\chi)$, $\gamma^*(\chi)$ and $\kappa^*(\chi)$ as functions of $d > 0$ in Fig. 4.

Also here we have $C(\chi) > 0$ for all $d > 0$. Secondly, the gross-error sensitivity is minimal for $d = 0.666$ which corresponds to the case $\gamma^* = \gamma^+ = \gamma^-$. Finally, the change-of-variance sensitivity is smallest for $d = 0.190$, which corresponds to a case where $\gamma^+ < \gamma^- = \gamma^*$.

5. Conclusions

In this paper we have derived the change-of-variance function of M-estimators of scale under general contamination, in which case the additional term $V(\chi) C(\chi) IF(z)$ arises. We have seen that it is still true that V-robustness implies B-robustness. The L^q scale estimators, which have a constant $C(\chi)$, are neither B-robust nor V-robust. An alternative formula for $C(\chi)$ has been obtained, and used to analyze the λ -MAD and the Welsch estimators.

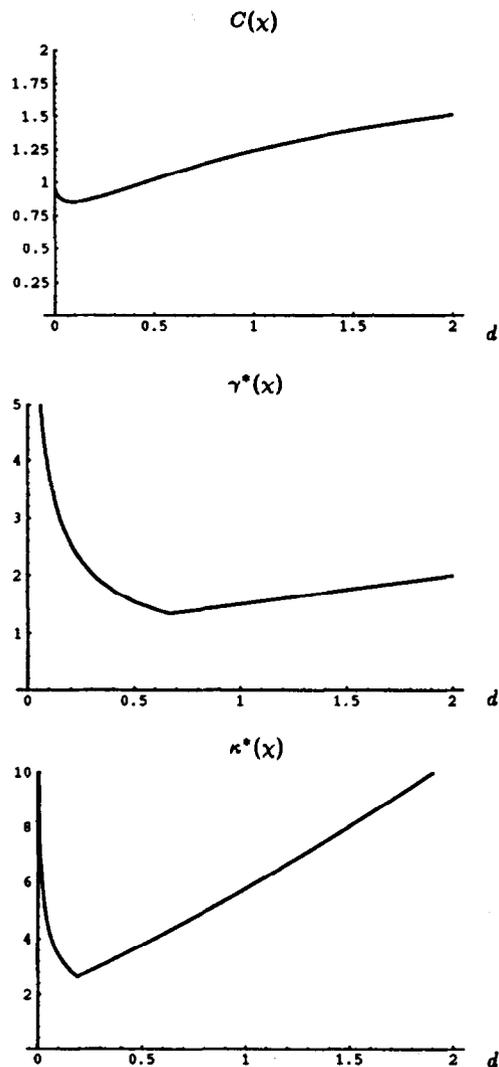


Fig. 4. The values of $C(\chi)$, $\gamma^*(\chi)$ and $\kappa^*(\chi)$ as a function of d for the Welsch estimator.

References

- [1] 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).
- [2] P.J. Huber, *Robust Statistics* (Wiley, New York, 1981).
- [3] P.J. Rousseeuw, A new infinitesimal approach to robust estimation, *Z. Wahrsch. Verw. Gebiete* **56** (1981) 127–132.