A REMARK ON ROBUSTNESS AND WEAK CONTINUITY OF M-ESTIMATORS

BRENTON R. CLARKE

Abstract

Global weak continuity of M-functionals in a neighbourhood of the parametric distribution is established. This has implications for robustness of M-estimators vis a vis definitions put forward by Hampel. For instance the Tukey bisquare location estimator is robust on neighbourhoods of the parametric model, but the median is not.

2000 *Mathematics subject classification*: primary 62G35; secondary 62E20.
Keywords and phrases: M-functionals, Prohorov metric distance, weak continuity.

1. Introduction

Early in the development of robustness theory, Hampel [6] sought to define a general qualitative definition of robustness of statistical estimators. He gave two theorems (see Hampel *et al.* [7, page 99]), which linked the notion of qualitative robustness and continuity of estimators T_n. In this line of research continuity and also Fréchet differentiability of M-estimators was established in Clarke [3], going beyond the simple location models. Here estimators are written in the spirit of von Mises [12] as functionals T defined on the space of distribution functions, $\mathscr{G}$. Under Conditions A, weak continuity and Fréchet differentiability were in fact established at a parametric distribution $F_\theta \in \mathscr{F} = \{F_\tau \mid \tau \in \Theta\} \subset \mathscr{G}$. Here $\Theta \subset E^r$, Euclidean r-space. Thus using the given results of [3] to apply any of the results of Hampel we can essentially only infer from Hampel [6, Theorem 1] (see also Hampel *et al.* [7, Theorem 1]) the robustness of $T_n \equiv T[F_n]$ at the distribution F_θ, where here F_n is the empirical distribution function formed from the sample $X_1, \ldots, X_n$ of independent identically

---

distributed observations in R, a separable metrizable space. An example application of this is given in [5].

In this paper we establish a broader result, the (qualitative) robustness of $\{T_n\}$ at F, for all F in a Prohorov neighbourhood $U(F_\theta)$ of F_θ; together with the consistency of $\{T_n\}$ to $T[F]$ for all $F \in U(F_\theta)$. This is established by proving the global weak continuity of T on a $U(F_\theta)$ and then invoking Hampel [6, Theorem 2] (see also Hampel *et al.* [7, Theorem 2]). That is, for the M-functional defined to be a solution of the general estimating equations

$$(1.1) \qquad K_G(T[G]) = \int_R \psi(x, T[G])\, dG(x) = 0,$$

there exists a functional root T which is weakly continuous at each $F \in U(F_\theta)$. Here ψ is a function on $R \times \Theta$ defining the M-estimator. Examples of choices of ψ which involve bounded smooth functions with bounded partial derivatives are found increasingly in the literature, examples being given in [2, 3, 5, 8, 9] just to name some. It transpires that for such ψ this global weak continuity holds. Thus for example, the Tukey bisquare estimator of Beaton and Tukey [1] for location, defined by the ψ-function,

$$\psi_{BS}(x) = \begin{cases} x(1 - x^2/c^2)^2 & |x| \le c, \\ 0 & |x| > c, \end{cases}$$

where c is a tuning constant, can be regarded as globally robust, whereas as is noted in Hampel [6, page 1894] 'The median is robust (and continuous) at F if and only if $F^{-1}(1/2)$ contains no more than one point. It is not robust in any open set of distributions.'

It is appropriate to reexamine the fairly general Conditions A of [3] and [3, Theorem 3.2] which defines continuity of the M-estimator in terms of a general metric distance or equivalently in terms of neighbourhoods $n(\epsilon, F) = \{G \in \mathcal{G} \mid d(G, F) < \epsilon\}$. Since we are only using the Prohorov metric distance in our deliberations here these conditions are restated in a much simpler and easier to use form as Conditions W. They motivate our main theorem given in Section 3. Note that equations (1.1) need not have a unique solution. We therefore examine the roots of the estimating equations locally and then globally. For example, $T[\psi, \rho_0, G]$ is defined as the solution to

$$\inf_{\tau \in I(\psi, G)} \rho_0(G, \tau) = \rho_0(G, T[\psi, \rho_0, G]),$$

where

$$I(\psi, G) = \left\{\tau \,\middle|\, \int_R \psi(x, \tau)\, dG(x) = 0, \ \tau \in \Theta\right\},$$

if a solution exists. Otherwise, $T[\psi, \rho_0, G] = +\infty$. Here ρ_0 is some global selection functional, for instance, the root closest to the median is chosen via choosing

$$\rho_0(G, \tau) = |\tau - G^{-1}(1/2)|.$$

See [4] for example. In [3, Theorem 3.2] an auxiliary functional is used to discover the properties of $T[\psi, \rho_0, \cdot]$ simply by using $\rho(G, \tau) = \|\tau - \theta\|$, for it is shown in [3, Theorem 4.1] that $T[\psi, \rho_0, \cdot] \equiv T[\psi, \rho, \cdot]$ on small enough neighbourhoods $n(\epsilon, F_\theta)$. Of course if ψ is chosen to have only one root then $T[\psi, \rho, \cdot] \equiv T[\psi, \cdot]$ and no selection functional is needed.

2. Conditions A and Theorem 3.2 of Clarke

CONDITIONS A

A0: $T[\psi, \rho, F_\theta] = \theta$;

A1: ψ is an $r \times 1$ vector function on $R \times \Theta$ and has continuous partial derivatives on $R \times D$, where $D \subset \Theta$ is some nondegenerate compact interval containing θ in its interior;

A2: $\{\psi(x, \tau) \mid \tau \in D\}, \{(\partial/\partial\tau)\psi(x, \tau) \mid \tau \in D\}$ are bounded above in Euclidean norm ($\|A\| = \{\text{trace}(A'A)\}^{1/2}$) by some function g that is integrable with respect to all $G \in n(\epsilon, F_\theta)$ for some $\epsilon > 0$;

A3: the matrix

$$M(\theta) = \int_R \left\{ \frac{\partial}{\partial\theta} \psi(x, \theta) \right\} dF_\theta(x)$$

is nonsingular;

A4: given $\delta > 0$ there exists $\epsilon > 0$ such that for all $G \in n(\epsilon, F_\theta)$

$$\sup_{\tau \in D} \left\| \int_R \psi(x, \tau) \, dG(x) - \int_R \psi(x, \tau) \, dF_\theta(x) \right\| < \delta$$

and

$$\sup_{\tau \in D} \left\| \int_R \left\{ \frac{\partial}{\partial\tau} \psi(x, \tau) \right\} dG(x) - \int_R \left\{ \frac{\partial}{\partial\tau} \psi(x, \tau) \right\} dF_\theta(x) \right\| < \delta.$$

Using Conditions A Theorem 3.2 guarantees continuity of the functional T at F_θ. Restating the theorem here aids in our establishing global weak continuity at a more general $F \in U(F_\theta)$.

CLARKE [3, THEOREM 3.2]   *Let* $\rho(G, \tau) = \|\tau - \theta\|$ *and suppose Conditions* A *hold. Then given* $\kappa > 0$ *there exists* $\epsilon > 0$ *such that* $G \in n(\epsilon, F_\theta)$ *implies* $T[\psi, \rho, G]$

exists and is an element of $U_\kappa(\theta)$. Further for this ϵ there is a $\kappa^* > 0$ such that

$$I(\psi, G) \cap U_{\kappa^*}(\theta) = T[\psi, \rho, G]$$

and $M(\tau, G) = \int_R \{(\partial/\partial\tau)\psi(x, \tau)\}\, dG(x)$ is nonsingular for $\tau \in U_{\kappa^*}(\theta)$. For any null sequence of positive numbers $\{\epsilon_k\}$ let $\{G_k\}$ be an arbitrary sequence for which $G_k \in n(\epsilon_k, F_\theta)$. Then

$$\lim_{k\to\infty} T[\psi, \rho, G_k] = T[\psi, \rho, F_\theta] = \theta.$$

Weak continuity at F_θ follows if Conditions A hold with respect to neighbourhoods generated by the Prohorov metric, defined as $d_p(F, G) = \inf\{\delta > 0 \mid F(B) \leq G(B^\delta) + \delta, G(B) \leq F(B^\delta) + \delta,$ for all $B \in \mathscr{B}\}$, where $\mathscr{B}$ are the Borel sets, and B^δ is the closed δ-neighbourhood of B. Since we are dealing only with the Prohorov neighbourhoods the exact conditions for global weak continuity are specified as Conditions W, given in the next section. They are much simpler and guarantee robustness at all F in a Prohorov neighbourhood of F_θ.

3. Assumptions and main result

Here we give weak continuity of T at each $F \in U(F_\theta)$ for some Prohorov neighbourhood under the following assumptions which are a special case of Conditions A of [3]. Numerous examples of M-estimators satisfying them are now available.

CONDITIONS W:
   W0: $T[F_\theta] = \theta$;
 W1 $\equiv$ A1: ψ is an $r \times 1$ vector function on $R \times \Theta$ and has continuous partial derivatives on $R \times D$, where $D \subset \Theta$ is some nondegenerate compact interval containing θ in its interior;
   W2: $\{\psi(x, \tau) \mid \tau \in D\}, \{(\partial/\partial\tau)\psi(x, \tau) \mid \tau \in D\}$ are bounded above in Euclidean norm by a constant;
   W3: the matrix

$$M(\theta) = \int_R \left\{\frac{\partial}{\partial\theta}\psi(x, \theta)\right\} dF_\theta(x)$$

is nonsingular.

THEOREM 1. *Assuming Conditions W, there exists a Prohorov neighbourhood $U(F_\theta)$ of F_θ such that the functional defined via $T[\psi, \rho, \cdot]$ is weakly continuous at each $F \in U(F_\theta)$.*

COROLLARY. *Assuming Conditions* W *there exists a qualitatively robust M-functional estimator* T_n *which is robust and consistent, tending to* $T[F]$, *at all* $F \in U(F_\theta)$.

4. Essential proofs

To prove Theorem 1 we appeal to parts of the proof of [3, Theorem 3.2]. Firstly Conditions W imply Conditions A of [3] with the neighbourhoods $n(\epsilon, F_\theta)$ being the Prohorov neighbourhoods about F_θ as defined by the Prohorov metric distance d (see [3, Remark 2.2 and Remark 6.2]). From the given proof we can choose $\kappa^*/2$ so that there is an ϵ^* such that for all $F \in n(\epsilon^*, F_\theta)$

$$T[\psi, \rho, F] = \{K_F^{-1}(0) \cap U_{\kappa^*}(\theta)\} \in U_{\kappa^*/2}(\theta).$$

By [3, Lemma 3.1], which essentially follows from the continuity of the determinant as a function of the elements of a matrix, it can be assumed

$$M(\tau, F) = \int \left\{ \frac{\partial}{\partial \tau} \psi(x, \tau) \right\} dF$$

is nonsingular for $\tau \in U_{\kappa^*}(\theta)$, whence for $\theta_0 = T[\psi, \rho, F]$ the matrix $M(\theta_0, F)$ is nonsingular for any particular F in the Prohorov neighbourhood $n(\epsilon^*, F_\theta)$. Given F and θ_0 we let $\tilde{\rho}(G, \tau) = \|\tau - \theta_0\|$ be an associated selection functional.

LEMMA 1. *Let* κ^*, ϵ^* *be given by* [3, Theorem 3.2] *(see above). Let* $F \in n(\epsilon^*/2, F_\theta)$ *be given. Then Conditions* W *imply the following statements*:
CONDITIONS WA
WA0: $T[\psi, \tilde{\rho}, F] = \theta_0$,
WA1 $\equiv$ W1 $\equiv$ A1,
WA2 $\equiv$ W2 $\equiv$ A2 *(with g bounded)*,
WA3: *The matrix* $M(\theta_0, F)$ *is nonsingular*,
WA4: *Given* $\delta > 0$ *there exists* $\epsilon > 0$ *such that for all* $G \in n(\epsilon, F)$

$$\sup_{\tau \in D} \left\| \int_R \psi(x, \tau) \, dG(x) - \int_R \psi(x, \tau) \, dF(x) \right\| < \delta$$

and

$$\sup_{\tau \in D} \left\| \int_R \left\{ \frac{\partial}{\partial \tau} \psi(x, \tau) \right\} dG(x) - \int_R \left\{ \frac{\partial}{\partial \tau} \psi(x, \tau) \right\} dF(x) \right\| < \delta.$$

PROOF OF LEMMA 1. Note we can without loss of generality assume or restrict set $D \subset \Theta$ to contain θ_0 in its interior. Now the proof of [3, Theorem 6.1] follows through with F instead of F_θ, whence Remark 2.2 and Remark 6.2 of that paper imply WA4. We have noted WA3 holds already for $F \in n(\epsilon^*, F_\theta)$.   □

The following lemma and proof mimicks [3, Theorem 3.2] but with F_θ replaced by F and $\rho(G, \tau) = \|\tau - \theta\|$ replaced by $\tilde{\rho}(G, \tau) = \|\tau - \theta_0\|$.

LEMMA 2. *Let $F \in n(\epsilon^*/2, F_\theta) \subset n(\epsilon^*, F_\theta)$ be given. Then given $\xi > 0$, there exists an $\eta > 0$ such that $G \in n(\eta, F)$ implies $T[\psi, \tilde{\rho}, G]$ exists and is an element of $U_\xi(\theta_0)$. Further, for this η there exists $\tilde{\kappa} > 0$ such that*

$$I(\psi, G) \cap U_{\tilde{\kappa}}(\theta_0) = T[\psi, \tilde{\rho}, G],$$

and $M(\tau, G)$ is nonsingular for $\tau \in U_{\tilde{\kappa}}(\theta_0)$. For any null sequence of positive numbers $\{\eta_k\}$ let G_k be an arbitrary sequence for which $G_k \in n(\eta_k, F)$. Then

$$\lim_{k \to \infty} T[\psi, \tilde{\rho}, G_k] = T[\psi, \tilde{\rho}, F] = \theta_0.$$

Moreover, for $\eta < \epsilon^*/2$, $G \in n(\eta, F)$ implies

(4.1) $$T[\psi, \tilde{\rho}, G] \equiv T[\psi, \rho, G].$$

PROOF OF LEMMA 2. From Lemma 1 we have that Conditions WA hold. The proof now mimicks that of [3, Theorem 3.2] but with F replacing F_θ and Conditions WA replacing Conditions A. It remains to establish (4.1). Choosing $\xi = \kappa^*/2$ in Lemma 2 implies for $G \in n(\eta, F)$

$$K_G^{-1}(0) \in U_{\kappa^*/2}(\theta_0) \subset U_{\kappa^*}(\theta).$$

We can choose $0 < \eta \le \epsilon^*/2$ so that

$$d(G, F_\theta) \le d(G, F) + d(F, F_\theta) < \epsilon^*.$$

Hence,

$$K_G^{-1}(0) \cap U_{\kappa^*}(\theta) = T[\psi, \rho, G]$$

since

$$K_G^{-1}(0) \cap U_{\kappa^*/2}(\theta_0) = T[\psi, \tilde{\rho}, G] \in U_{\kappa^*}(\theta).$$

It follows that $T[\psi, \rho, G] = T[\psi, \tilde{\rho}, G]$. □

Now note the proof of Theorem 1 applies by choosing the neighbourhood $U(F_\theta)$ to be $n(\epsilon^*/2, F_\theta)$. By Lemma 2 $T[\psi, \rho, \cdot]$ is weakly continuous at each $F \in U(F_\theta)$. □

The corollary follows from Theorem 1 by applying Hampel [6, Theorem 2]. Note a result of Varadarajan [11] proves that the empirical distribution function F_n converges

weakly to F almost surely and a result of Prohorov [10] gives that $d_p(F_n, F) \to 0$ almost surely. Thus almost sure consistency $T[F_n] \to T[F]$ follows.

Consider now the Tukey Bisquare location functional defined to be the root of equations

$$(4.2) \qquad \int_E \psi_{BS}(x - \tau) \, dG(x) = 0.$$

Theorem 1 asserts there is a weakly continuous root $T[\psi_{BS}, \rho, \cdot]$ at each $F \in U(F_\theta)$. The fact that the selection functional $\rho_0(G, \tau) = |\tau - G^{-1}(1/2)|$ is continuous at $F_\theta(x) = \Phi(x - \theta)$, where Φ is the cumulative standard normal distribution, implies via arguments similar to [3, Theorem 4.1] that the selection functional $T[\psi_{BS}, \rho, \cdot] \equiv T[\psi_{BS}, \rho_0, \cdot]$ on small enough neighbourhoods $n(\tilde{\epsilon}, F_\theta)$. That is the root (4.2) closest to the median is weakly continuous at each F in a small enough neighbourhood of F_θ.

Acknowledgements

References

[1] A. E. Beaton and J. W. Tukey, 'The fitting of power series meaning polynomials, illustrated on band-spectroscopic data', *Technometrics* **16** (1974), 147–185.

[2] T. Bednarski and S. Zontek, 'Robust estimation of parameters in a mixed unbalanced model', *Ann. Statist.* **24** (1996), 1493–1510.

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

[4] ———, 'The selection functional', *Probab. Math. Statist.* **11** (1991), 149–156.

[5] B. R. Clarke and C. R. Heathcote, 'Robust estimation of k-component univariate normal mixtures', *Ann. Inst. Statist. Math.* **46** (1994), 83–93.

[6] F. R. Hampel, 'A general qualitative definition of robustness', *Ann. Math. Statist.* **42** (1971), 1887–1896.

[7] 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).

[8] C. R. Heathcote and M. J. Silvapulle, 'Minimum mean squared estimation of location and scale parameters under misspecification of the model', *Biometrika* **68** (1981), 501–514.

[9] R. M. Huggins, 'On the robust analysis of variance components models for pedigree data', *Austral. J. Statist.* **35** (1993), 43–57.

[10] Y. V. Prohorov, 'Convergence of random processes and limit theorems in probability', *Theory Probab. Appl.* **1** (1956), 157–214.

[11] V. S. Varadarajan, 'On the convergence of probability distributions', *Sankhyā* **19** (1958), 23–26.

[12] R. von Mises, 'On the asymptotic distribution of differentiable statistical functions', *Ann. Math. Statist.* **18** (1947), 309–348.

Mathematics and Statistics
Division of Science and Engineering
Murdoch University
Murdoch, WA 6150
Australia
e-mail: clarke@prodigal.murdoch.edu.au