Statistica Neerlandica (1992) Vol. 46, nr.2-3, pp. 165-177

A Central Limit Theorem for M-estimators by the von Mises Method

C.C. Heesterman
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

R.D. Gill*
Mathematical Institute
University of Utrecht
P.O. Box 80.010
3508 TA Utrecht
The Netherlands

Asymptotic normality of M- or maximum likelihood type estimators was established in a classic paper by HUBER (1967). REEDS (1976) argued that this could have been obtained simply as an application of the delta-method, using the tool of compactly differentiating von Mises functionals with respect to the empirical distribution function F_n. His proof however contains some errors and has been largely ignored. A corrected version of the proof is given.

Key Words & Phrases: Asymptotic normality of M-estimators, compact differentiation, Hadamard differentiation, delta-method, M-estimator, von Mises functional.

1. INTRODUCTION

Maximum likelihood type estimators, 'M-estimators', were first introduced by HUBER (1964); a statistic T_n is called an M-estimator of a parameter $\theta_0 \in \Theta \subseteq \mathbb{R}^p$ if T_n is a solution to a set of estimating equations:

$$\Phi_n(T_n) = E_{F_n}\psi(X;T_n) = n^{-1}\sum_{i=1}^{n}\psi(X_i;T_n) = 0. \tag{1}$$

Here, E_{F_n} denotes expectation over the sample space $\mathcal{X}$ with respect to the empirical probability measure F_n on $\mathcal{X}$, the empirical distribution function based on n independent and identically distributed copies $X_1,...,X_n$ of a random variable X taking values in $\mathcal{X}$ distributed according to the unknown distribution function F. In order for T_n to be a sensible estimator of θ_0, the

* Please address correspondence to the second author

function $\psi: \mathcal{X} \times \Theta \to \mathbb{R}^p$ should be chosen such that

$$\Phi(\theta_0) = E_F \psi(X; \theta_0) = 0. \tag{2}$$

As a matter of fact, Huber relaxed the definition (1) somewhat to

$$\Phi_n(T_n) = o_P(n^{-\frac{1}{2}}), \text{ as } n \to \infty. \tag{1'}$$

Let us ignore measurability problems, and suppose that both Φ_n and T_n are indeed random elements in some appropriate measurable space.

Now REEDS (1976) proved two central limit theorems for M-estimators using the von Mises method (also called the functional or generalized delta-method). The first theorem has stronger conditions and a simpler proof; the second has quite weak conditions and a more elaborate proof, to which this paper is devoted. Both theorems are close to other results in the literature but the method of proof, in a sense simply an application of the delta-method, is of interest, especially in view of recent work on this method (GILL, 1989, 1991; SHEEHY and WELLNER, 1990, 1991; and others). In particular the method gives not only a central limit theorem but with no extra work also gives results on the consistency of the bootstrap, the law of the iterated logarithm, and so on.

However Reeds' work is hard to obtain and also contains some errors which have caused many researchers to ignore it. Therefore it seems useful and timely to present a correct proof of his second (stronger) theorem. (The forthcoming book by RIEDER (1992) will also contain a complete treatment of these theorems).

First a brief introduction to Reeds' approach is in order: he observed that T_n may be treated as a von Mises functional T_ψ of F_n:

$$T_n = T_\psi(F_n) = T \circ \mu_\psi(F_n),$$

where

$$\mu_\psi(F_n) = E_{F_n} \psi(X; .) = \Phi_n,$$

and T is a functional that assigns to any $\mathbb{R}^p$-valued function on Θ a zero of this function (if a zero exists; cf. CLARKE (1986) on how to avoid ambiguity if there is more than one zero).

Now, the idea of the von Mises approach is to transfer a central limit theorem for F_n into a central limit theorem for T_n simply by approximating T_n by the first two terms of a Taylor expansion of $T_\psi(F_n)$ at F. This procedure is called a von Mises or generalised delta-method calculation, and requires a definition of differentiation. While some functionals are actually differentiable in the strong sense of Fréchet differentiation (see CLARKE, 1983), it turns out that for functionals that are only *Hadamard differentiable* (also called *compactly differentiable),* the central limit theorem for F_n may still be transferred to T_n. Since more functionals are compactly differentiable the condition of Fréchet differentiability is unnecessarily strong.

There is however one point of discussion in Reeds' approach: treating T_n as a composite functional $T_\psi = T \circ \mu_\psi$ of F_n causes unnecessary technical complications, whereas one might just as well restrict attention to the $\mathbb{R}^p$-valued functional T and consider

$$T_n = T(\Phi_n)$$

instead, since the information in F_n is only used through Φ_n. Equivalently one represents the empirical distribution of the data by the 'function indexed empirical process' Φ_n, rather than by the ordinary empirical distribution function F_n. From this point of view the change is purely cosmetic. However insisting on F_n caused Reeds to choose a much more elaborate metric on the space of distribution functions than necessary, leading to an avoidable error in his proof. (Our choice will only be a pseudo-metric but this is of no importance whatsoever).

In the next section, a heuristic approach and the basic steps of a von Mises calculation are given as well as some preliminary results. Hadamard or compact differentiation is defined and justified as a choice of differentiation to be used in a von Mises calculation in section 3. Section 4 then contains a corrected version of the proof for REEDS' (1976) second central limit theorem for M-estimators. Finally, in the last section the assumptions of Reeds' second theorem and some alternative approaches are briefly discussed.

2. PRELIMINARIES AND HEURISTICS

Let $(\mathcal{X},\mathcal{Q},P)$ be a probability space. Let $X_1,...,X_n$ be n independent and identically distributed copies of a random variable $X\in\mathcal{X}$, with distribution function F (corresponding to P). F_n is the empirical distribution function that assigns mass n^{-1} to each of the observation points. Consider estimation of the unknown parameter $\theta_0\in\Theta\subseteq\mathbb{R}^p$ that satisfies (2), i.e., $\Phi(\theta_0)=E_F\psi(X;\theta_0)=0$ for some function $\psi:\mathcal{X}\times\Theta\rightarrow\mathbb{R}^p$. Assume Θ may be chosen to be a compact subset in $\mathbb{R}^p$. Let $B^p(\Theta)$ denote the space of bounded $\mathbb{R}^p$-valued functions on Θ, let $B^1(\mathcal{X})$ denote the space of real-valued functions on $\mathcal{X}$, and let $C(\Theta)$ denote the space of continuous $\mathbb{R}^p$-valued functions on Θ. So $C(\Theta)\subseteq B^p(\Theta)$ since Θ is compact.

Now as Reeds observed, any estimator T_n that solves the estimating equations (1) may be represented as a functional $T_\psi:B^{1(}\mathcal{X})\rightarrow\mathbb{R}^p$

$$T_n = T_\psi(F_n) = T\circ\mu_\psi(F_n). \tag{3}$$

In the previous section it was already mentioned that for practical purposes the representation of T_n as a non-composite functional $T:B^p(\Theta)\rightarrow\mathbb{R}^p$ is more useful

$$T_n = T(\Phi_n). \tag{3'}$$

It is tacitly assumed that $\psi(X;.)$ is almost surely bounded in Θ. Unless mentioned otherwise, the space $C(\Theta)$ will be endowed with the convenient (though sometimes naive) choice of the supremum norm. Thus, $C(\Theta)$ will be complete and separable. Hence, weak convergence of a sequence of random variables Z_n in $C(\Theta)$ implies tightness of this sequence in $C(\Theta)$ (see BILLINGSLEY, 1968): for all $\epsilon>0$, there exists a compact $K_\epsilon\subseteq C(\Theta)$, such that

corresponds to *Fréchet or bounded differentiation.*
Obviously, $\mathbb{S}_s \subseteq \mathbb{S}_b$, so whenever a functional is Fréchet differentiable, it is also Gâteaux differentiable, and the two derivatives coincide. Also note that if B_1 and B_2 are normed vector spaces, then Fréchet differentiability of the functional T at x is equivalent to the existence of a continuous and linear mapping $dT(x;.):B_1 \to B_2$ such that

$$\|T(x+h) - T(x) - dT(x;h)\|_{B_2} = o(\|h\|_{B_1}) \text{ as } \|h\|_{B_1} \to 0. \qquad (7'')$$

Furthermore, let B_1 be $B^1(\mathfrak{X})$, endowed with sup-norm, $B_2 = \Theta$. Then Fréchet differentiability of T at F implies asymptotic normality. Indeed, by (7″) it follows that

$$n^{1/2}(T_n - \theta_0) = n^{1/2} dT(F;F_n - F) + o_P(1), \text{ as } n \to \infty \qquad (8)$$

since

$$\|F_n - F\|_\infty = O_P(n^{-1/2}).$$

Moreover, the process $n^{1/2}(F_n - F)$ converges weakly to the Brownian bridge process composed with F, hence asymptotic normality follows by Slutsky's theorem. Notice that the choice of topology is indeed crucial!

Unfortunately, not all important functionals do have a Fréchet derivative, although CLARKE (1983) actually claims that most popular functionals in fact are boundedly differentiable. In that paper he gives some general conditions for Fréchet differentiability to hold, one of which is continuity and boundedness of the function ψ on $\mathfrak{X} \times \theta$. Since the boundedness condition is necessary, 'those nonrobust estimators such as the maximum likelihood estimator in normal parametric models are excluded' as Clarke rightly admits; see BEDNARSKI, CLARKE and KOLKIEWICZ (1991) for further results in this direction. Also the median and other sample quantiles, however simple they are, are not Fréchet differentiable.

By $\mathbb{S}_c$ denote the class of all compact sets in B_1. Hence the inclusion $\mathbb{S}_s \subseteq \mathbb{S}_c \subseteq \mathbb{S}_b$ holds.

DEFINITION 3.2. A mapping $T:B_1 \to B_2$ *is called* Hadamard differentiable *(or* compactly differentiable) at $x \in B_1$ if (7) holds for all $K \in \mathbb{S}_c$.

By the inclusion above, compact differentiability is a weaker condition on the functional T; this will have to be paid for by the stochastic part of T_n: the requirement of *boundedness in probability* will have to be replaced by *tightness.*

THEOREM 3.3. *(delta-method: Reeds, 1976). Suppose $T:B_1 \to B_2$ is Hadamard differentiable at $x \in B_1$ with derivative $dT(x;.)$. Suppose furthermore that $\{Y_n\}_{n=1}^\infty$ is a sequence of random elements in B_1 that satisfies*

$$(i) \quad n^{1/2}(Y_n - x) \xrightarrow{\mathcal{D}} Z \text{ in } B_1 \text{ as } n \to \infty$$

$$(ii) \quad \text{the sequence } \{n^{1/2}(Y_n - x)\}_{n=1}^\infty \text{ is tight in } B_1 \qquad (9)$$

Then

$$n^{1/2}(T(Y_n)-T(x)) \xrightarrow{\mathcal{D}} dT(x;Z) \text{ in } B_2 \text{ as } n \to \infty.$$

In words, weak convergence of the sequence $\{n^{1/2}(Y_n - x)\}_{n=1}^{\infty}$ may be transferred to the sequence $\{n^{1/2}(T(Y_n)-T(x))\}_{n=1}^{\infty}$.

PROOF. Write $Z_n = n^{1/2}(Y_n - x)$; by compact differentiability of T at x and (9) (ii) approximate $n^{1/2}(T(Y_n)-T(x))$ by $dT(x;Z_n)$. The remainder term will be $o_P(1)$ as $n \to \infty$.

First the *analytic part.* Since T is compactly differentiable at x,

$$n^{1/2}(T(Y_n)-T(x)) = dT(x;Z_n) + n^{1/2}R_T(x;n^{-1/2}Z_n),$$

where, for all $K \in \mathfrak{S}_c$

$$n^{1/2}R_T(x;n^{-1/2}h) = o(1) \text{ as } n \to \infty, \text{ uniformly in } h \in K. \tag{10}$$

Then the *stochastic part.* Choose $\epsilon, \eta > 0$. By (9) (ii) there exists a compact K_ϵ such that

$$P(Z_n \in K_\epsilon) > 1-\epsilon, \; n = 1,2,... \tag{11}$$

Furthermore, since

$$P(\|n^{1/2}R_T(x;n^{-1/2}Z_n)\| > \eta) \leq P(\|n^{1/2}R_T(x;n^{-1/2}Z_n)\| > \eta, \; Z_n \in K_\epsilon) + P(Z_n \notin K_\epsilon),$$

(10) and (11) together imply $n^{1/2}R_T(x;n^{1/2}Z_n) = o_P(1)$ as $n \to \infty$. Hence, as $dT(x;\cdot)$ is linear and continuous, the theorem follows by (9) (i) via Slutsky's theorem. □

REMARK: The topology on B_1 will have to be chosen such that the analytic properties of T and the stochastic properties of Z_n (both depend on the topology) are attuned to each other with respect to the delta-method.

4. ASYMPTOTIC NORMALITY

In his first central limit theorem for M-estimators REEDS (1976) assumes continuous differentiability of the function $\psi : \mathcal{X} \times \Theta \to \mathbb{R}^p$ in Θ; then the implicit function theorem is used to show that a central limit theorem for $(\Phi_n(\theta), \theta \in \Theta)$ carries over to $T_n = T(\Phi_n)$, where $\Phi_n = E_{F_n}\psi(X;\cdot)$, if indeed the sequence $\{n^{1/2}(\Phi_n - \Phi)\}_{n=1}^{\infty}$ is weakly convergent and tight in some appropriate topological vector space. Here, as in (2), $\Phi = E_F\psi(X;\cdot)$. In fact in $C(\Theta)$, sufficient conditions for weak convergence and hence tightness to hold are given by the following lemma, which is a direct consequence of Theorem 2.4 by GINÉ (1974).

LEMMA 4.1. *Let $Z_1,...,Z_n$ be independent and identically distributed copies of a random variable Z in $C(\Theta)$ with zero expectation. If*

$$(i) \quad E_P|Z|_\infty^2 < \infty,$$

$$(ii) \quad E_P \sup_{\theta_1,\theta_2} \frac{|Z(\theta_1)-Z(\theta_2)|^2}{|\theta_1-\theta_2|^\lambda} < \infty, \text{for some } \lambda > 0 \tag{12}$$

then the sequence $n^{-1/2}\Sigma_{i=1}^{n} Z_i$, $n=1,2,\ldots$ *is weakly convergent in* $C(\Theta)$.

Lemma 4.1 does not characterize weak convergence in function spaces; if the conditions of the lemma are not fulfilled, for instance if $\Phi_n \notin C(\Theta)$, then tightness and weak convergence of the sequence $\{n^{1/2}(\Phi_n - \Phi)\}$ in some suitable space may be established by any other convenient means.

In his second central limit theorem for *M*-estimators Reeds drops the assumption of continuous differentiability of ψ in θ. In fact the set of conditions in this second theorem is actually weaker than the set of conditions in the first one. As a consequence the implicit function theorem cannot be invoked, and the proof will be rather more difficult. This second theorem will now be reformulated and, after we have made some remarks on it and proved two Lemmas, a corrected version of the proof will be given.

THEOREM 4.2. *Let* ψ *satisfy the conditions of Lemma 2.2 and in addition assume that the conclusion of Lemma 4.1 holds for* $n^{-1/2}\Sigma_{i=1}^{n} Z_i = n^{-1/2}\Sigma_{i=1}^{n}(\psi(X_i;\cdot) - \Phi(\cdot))$. *If the function* $\Phi:\Theta \to \mathbb{R}^p$, $\Phi(\theta) = E_F\psi(X;\theta)$, *has the following properties:*

(i) Φ *has a unique zero at* θ_0

(ii) Φ *is a local homeomorphism at* θ_0 (13)

(iii) Φ *is differentiable at* θ_0 *with nonsingular derivative* $A:\mathbb{R}^p \to \mathbb{R}^p$,

then there exists an estimator $T_n = T(\Phi_n)$ *such that*

$$(i)\quad P(\Phi_n(T_n) = 0) \to 1 \text{ as } n \to \infty$$

$$(ii)\quad n^{1/2}(T_n - \theta_0) \xrightarrow{\mathcal{D}} N(0,\Sigma) \text{ as } n \to \infty. \tag{14}$$

The covariance matrix Σ *is given by*

$$\Sigma = A^{-1}\Gamma(A^{-1})^{\mathsf{T}}, \text{ and also}$$

$$n^{1/2}(\Phi_n - \Phi)(\theta_0) \xrightarrow{\mathcal{D}} N(0,\Gamma) \text{ as } n \to \infty. \tag{15}$$

Reeds represents the estimator T_n by the composite functional $T_\psi = T\circ\mu_\psi$, i.e., $T_n = T_\psi(F_n)$. Now, consider the relatively easy situation that the true distribution function is the uniform distribution function on the unit interval in $\mathbb{R}$ so $\mathcal{X}=[0,1]$ and $F=U$ say. By U_n denote the empirical distribution function based on n independent and identically distributed observations from U. (If $F \neq U$, but $\mathcal{X}=\mathbb{R}$ then F_n and $U_n\circ F$ are identically distributed). It is a well known fact that U_n is not a random element in $D[0,1]$ equiped with the supremum norm; on the other hand, while U_n is indeed a random element in $D[0,1]$ equiped with the Skorokhod topology this is not a vector space. These two arguments illustrate the fact that the choice of B_1 is not at all trivial (cf. for instance GILL (1989) or FERNHOLZ (1979) in case $\mathcal{X}=\mathbb{R}$ and θ_0 is a location parameter). In the general case, that is $\theta\in\mathbb{R}^p$ and $\mathcal{X}$ is a separable metrizeable space, Reeds constructs the topological vector space B_1 to be isomorphic with a subspace of $\tilde{B}_1 = L_2(P) \times C(\Theta)$ equipped with the norm $\|(x,y)\|_{\tilde{B}} = \|x\|_{L_2} + \|y\|_\infty$ via the $1-1$ mapping $\bar{\alpha}(g) = (g, E_g \psi(X;\cdot))$. Since by a theorem of Prohorov for the first coordinate and Lemma 4.1 for the second coordinate $\{\bar{\alpha}(n^{1/2}(F_n - F))\}_{n=1}^{\infty}$ is random and tight in $\tilde{B}_1$, Reeds concludes that the sequence of arguments $\{n^{1/2}(F_n - F)\}_{n=1}^{\infty}$ itself is random and tight in B_1 with the topology induced by the norm $\|\cdot\|_{B_1} = \|\bar{\alpha}(\cdot)\|_{\tilde{B}_1}$.

Two remarks are in order now: firstly, properties of $\bar{\alpha}(n^{1/2}(F_n - F))$ in $\tilde{B}_1$ cannot as trivially as Reeds suggests be translated into the same properties of the argument in B_1, since $\bar{\alpha}$ is not *onto;* it maps B_1 into a proper subset of $\tilde{B}_1$, depending on ψ. Fortunately this mistake can be repaired though (VLOT, 1987). But, this is the second remark, if tightness of the sequence $n^{1/2}(\Phi_n - \Phi)$ is needed anyhow, why not apply the delta-method to the functional $T(\Phi_n)$ straightaway and forget all about the $n^{1/2}(F_n - F)$-part? Indeed, the functional T_ψ is Hadamard differentiable if and only if T is Hadamard differentiable, since μ_ψ is linear and continuous and compact differentiation follows the chain rule. So there is really less work in establishing the validity of the conditions in Theorem 3.3 if Y_n is taken to be Φ_n instead of F_n. Equivalently $\{F_n\}$ may be endowed with the *pseudonorm* $\|F_n\| = \|E_{F_n} \psi(X;\cdot)\|_\infty$ instead of the clumsy but proper norm introduced by Reeds.

Now, represent the estimator T_n by $T(\Phi_n)$. Since by the assumptions of Theorem 4.2 the sequence $\{n^{1/2}(\Phi_n - \Phi)\}_{n=1}^{\infty}$ is weakly convergent in $C(\Theta)$ and hence tight in $C(\Theta)$, the stochastic part of the delta-method applied to T_n is already settled. So it remains to prove existence and compact differentiability of a solution to the estimating equations $\Phi_n = 0$. For this purpose two lemmas will now be given:

LEMMA 4.3. *Let $\Phi:\Theta \to \mathbb{R}^p$ satisfy the conditions (13) of Theorem 4.2. Then there exists a neighbourhood V of Φ in $C(\Theta)$ and a functional $T: V \to \Theta$ such that $f(T(f)) = 0$ $\forall f \in V$ (T may not be unique).*

PROOF. By condition (13) (ii) there is a positive r and a neighbourhood W of θ_0 in Θ such that $\Phi|_W$, i.e., the restriction of Φ to $W \subseteq \Theta$, defines a homeomorphism between W and the ball $B_{0,r} = \{t \in \mathbb{R}^p : |t| \leqslant r\}$. For such r define $V_r \subseteq C(\Theta)$:

$$V_r = \{f \in C(\Theta) : \|\Phi - f\|_\infty < r\}. \tag{16}$$

Then the function $g \circ \Phi^{-1}$, with $g = \Phi - f$, maps the ball $B_{0,r}$ continuously into itself. Hence, by Brouwer's fixed point theorem, there exists for every $f \in V_r$ at least one $t_f \in B_r$ such that $g \circ \Phi^{-1}(t_f) = t_f$. Thus, the functional T defined through

$$T(f) = \begin{cases} \Phi^{-1}(t_f), & \text{if } f \in V_r \\ \theta_\infty \in \Theta, & \text{otherwise} \end{cases} \tag{17}$$

assigns to any $f \in V_r$ a zero of f, corresponding to the special fixed point t_f, since by definition, $f(T(f) = \Phi(\Phi^{-1}(t_f)) - g \circ \Phi^{-1}(t_f)$. □

COROLLARY 4.4. *Let T be defined as in (17). Under the conditions of Theorem 4.2 the M-estimator $T_n = T(\Phi_n)$ satisfies*

$$P(\Phi_n(T_n)=0) \to 1 \text{ as } n \to \infty.$$

PROOF. Tightness of the sequence $\{n^{1/2}(\Phi_n - \Phi)\}_{n=1}^{\infty}$ in $C(\Theta)$, implies

$$P(\Phi_n \in V_r) \to 1 \text{ as } n \infty. \tag{18}$$

Hence, with probability tending to $1, T_n = T(\Phi_n)$ is a solution to the estimating equations $\Phi_n = 0$. □

Let V_r be defined as in (16), and let T be defined as in (17). By $\mathbb{S}$ denote some class of bounded subsets in $C(\Theta)$. Choose $h \in K \in \mathbb{S}$. Let k be a finite norm bound for K. Let $t \in \mathbb{R}$ be such that $\Phi + th \in V_r$; so $|t| \leqslant rk^{-1}$ suffices. For ease of notation write

$$T_t = T(\Phi + th),$$

thus suppressing the dependence of T_t on $h \in K$. Also define the $\mathbb{R}^p$-valued function β_h through

$$T_t = T_0 - A^{-1}th(\theta_0) + t\beta_h(t).$$

The second lemma that will be used in the proof of Theorem 4.2 is the following:

LEMMA 4.5. *The functional T is $\mathbb{S}$-differentiable at Φ with derivative*

$$dT(\Phi;g) = -A^{-1}g(\theta_0)$$

iff all elements $K \in \mathbb{S}$ are equicontinuous.

PROOF. Since by assumption A is non-singular, the functional T is $\mathbb{S}$-differentiable at Φ iff for all $K \in \mathbb{S}$

$$A\beta_h(t) = o(1) \text{ as } t \to 0, \text{ uniformly in } h \in K. \tag{19}$$

Note that t_f and hence T may be chosen to satisfy necessary condition for bounded differentiability of T at Φ:

$$T_t - T_0 = O(t) \text{ as} \to 0, \text{uniformly in } \|h\|_\infty \leqslant k. \tag{20}$$

Indeed, since by definition $\Phi(T_t) = -th(T_t)$, it follows from assumption (13) (ii) (i.e., Φ is a local homeomorphism at $\theta_0 = T_0$) that

$$|T_t - T_0| = o(1) \text{ as } t \to 0, \text{ uniformly in } \|h\|_\infty \leqslant k$$

and also

$$|T_t - T_0| = \left[\frac{|\Phi(T_t) - \Phi(T_0)|}{|T_t - T_0|}\right]^{-1} . O(t) \text{ as } t \to 0, \text{ uniformly in } \|h\|_\infty \leqslant k.$$

Hence, since Φ is assumed to be differentiable at θ_0 with nonsingular derivative, (20) is valid.

Furthermore, by the same assumption,

$$\Phi(T_t) - \Phi(T_0) = A(T_t - T_0) + |T_t - T_0|.\epsilon(t),$$

where

$$\epsilon(t) = o(1) \text{ as } \to 0, \text{ uniformly in } \|h\|_\infty \leqslant k$$

Now by some simple algebra, using the above formulas the following expression can be derived

$$At\,\beta_h(t) = -t(h(T_t) - h(T_0)) + O(t).o(1), \text{ as } t \to 0, \text{ uniformly in } \|h\|_\infty \leqslant k,$$

hence, (19) holds iff K is equicontinuous (and of course bounded). □

COROLLARY 4.6. *The functional T defined in (17) is compactly differentiable at* Φ.

PROOF. See Proposition 2.1. □

PROOF OF THEOREM 4.2. See Corollary 4.4 for (14) (i). By Corollary 4.6, the delta-method may now be applied to obtain (14) (ii). Notice that (15) trivially holds since obviously $E_F|\psi(X;\theta_0)|^2 < \infty$. □

5. CONCLUDING REMARKS

A comparison of the conditions in Huber's central limit theorem for M-estimators and those of Theorem 4.2, i.e., the conditions that are sufficient for the delta-method to be applicable, is in order now. In fact, Huber's conditions are all but one implied by those in Theorem 4.2. Only separability of the function $\psi(x;\theta)$ in the sense of Doob (see HUBER (1983) for a precise definition of this concept) is somewhat difficult. If indeed Θ is compact and $\psi(x;\theta)$ is continuous in θ for F-almost all $x \in \chi$, then Huber's assumptions are actually weaker than those in REEDS' (1976) original theorem for M-estimators (the second one). However, since it is one of the main virtues of the delta-method, that any convenient set of conditions may be used in establishing the required properties of Φ_n, the stochastic part of T_n, a full comparison of Huber's approach and the delta-method cannot be carried out.

Our original motivation for this study was to investigate whether Reeds' approach could be generalized to the non-parametric case, i.e., θ is a function and Θ is a metric function space. The obvious generalisation to the non-parametric case is the following: Suppose $X_1,...,X_n$ have a common distribution function $F = F(x;\theta_0)$, where $\theta_0 \in \Theta$ is some unknown function. Furthermore, suppose that there exists a mapping $\Phi = \Phi(\cdot;F,\psi):\Theta \to B_2$, a function space, such that $\Phi(\theta_0) = 0 \in B_2$. Let B_1 then be some collection of mappings from Θ into B_2, such that $\Phi_n = \Phi(\cdot;F_n,\psi) \in B_1$. Define now an M-estimator T_n of θ_0 as a solution to the generalised estimating equations $\Phi_n = 0 \in B_2$, if a solution exists.

The main difficulties in extending the delta-method to the non-parametric case are the following: first, it is not at all clear that a solution to the generalised estimating equations actually exists; whereas in the parametric case Brouwer's fixed point theorem may be invoked, some other device should now be investigated or may be invented to prove existence of a solution under general conditions, not just in any ad hoc situation. Second, how should the analogue of tightness and weak convergence of the process $n^{1/2}(\Phi_n - \Phi)$ in the parametric case be defined in the non-parametric case where $n^{1/2}(\Phi_n - \Phi)$ is itself a function? Moreover, the choice of metric for B_1 will not be as easy as it was in the parametric case, where the structure of $C(\Theta)$ was such that even with the naive choice of uniform topology the conditions of the delta-method are fulfilled. Of course, the metric on B_1 should also be such that Φ_n is a random element in B_1. So it is clear that a lot of work remains to be done.

Finally, a few words should be said about the possible applications of Theorem 4.2 in the parametric case. Reeds claims that his first theorem covers maximum likelihood estimation in most parametric families used in statistics. In fact, Reeds' conditions, and the conditions in Cramér's classical theorem for maximum likelihood estimators are incommensurable: Cramér has a stronger derivative condition, whereas Reeds requires stronger moment properties. Anyway, since Theorem 4.2 in the present note is most general, i.e., the conditions in Theorem 4.2 are implied by the conditions in Reeds' second theorem, which are in turn implied by those in his first theorem, Theorem 4.2 also covers most maximum likelihood estimators in applied statistics. Furthermore, all M-estimators in the Princeton robustness study are covered by Theorem 4.2. Again, this is argued in REEDS (1976).

ACKNOWLEDGEMENTS

The authors want to thank Prof. Rieder for the priviledge of studying his lecture notes on Reeds' theory for M-estimators, the referees for their careful reading of an earlier version of this paper, and D. Vlot for his helpful and stimulating comments.

REFERENCES

ANDREWS, D.F., P.J. BICKEL, F.R. HAMPEL, P.J. HUBER, W.H. ROGERS and J.W. TUKEY (1972), *Robust estimates of location,* Princeton University Press, Princeton.

BEDNARSKI, T., B.R. CLARKE and W. KOLKIEWICZ (1991), Statistical expansions and locally uniform Fréchet differentiability, *Journal of the Australian Mathematical Society 50,* 88-97.

BILLINGSLEY, P. (1968), *Convergence of probability measures,* Wiley, New York.

CLARKE, B.R. (1983), Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations, *Annals of Statististics 11,* 1196-1205.

CLARKE, B.R. (1986), Nonsmooth analysis and Fréchet differentiability of *M*-functionals, *Probability Theory and Related Fields 73*, 197-210.

FERNHOLZ, L.T. (1979), *Von Mises calculus for statistical functionals,* Lecture Notes in Statistics 19, Springer Verlag, New York.

GILL, R.D. (1989), Non-and semi-parametric maximum likelihood estimators and the von Mises method, Part I, *Scandinavian Journal of Statistics 16,* 97-128.

GILL, R.D. (1991), Non-and semi-parametric maximum likelihood estimators and the von Mises method, Part II, Preprint 664, Department of Mathematics, University Utrecht.

GINÉ, E. (1974), On the central limit theorem for sample continuous processes, *Annals of Probability 2,* 629-641.

HUBER, P. (1964), Robust estimation of a location parameter, *Annals of Mathematical Statistics 35,* 73-101.

HUBER, P. (1967), The behaviour of maximum likelihood estimates under non-standard conditions, in: J.M. LE CAM and J. NEYMAN, (eds.), *Proceedings of the 5th Berkeley symposium on mathematical statistics and probability, Vol. 1,* University of California Press, Berkeley.

HUBER, P. (1983), *Robust statistics,* Wiley, New York.

REEDS, J.A. (1976), On the definition of von Mises functionals, Research report S-44, Department of Statistics, University of Harvard.

RIEDER, H. (1992), *Robust statistics I—asymptotic statistics,* to appear.

SHEEHY, A. and J.A. WELLNER (1990), Uniform Donsker classes of functions, Report 189, Department of Statistics, University of Washington, Seattle (to appear in *Annals of Probability).*

SHEEHY, A. and J.A. WELLNER (1991), The Delta Method and the Bootstrap for nonlinear functions of empirical distribution functions via Hadamard differentiation, Report, Department of Statistics, University of Washington, Seattle (in preparation).

VLOT, D. (1987), *Applications of a von Mises method, based on Hadamard differentiation tangentially to a subspace,* Master's thesis, Mathematical Institute, University of Utrecht.

Received July 1988, Revised May 1991.