Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

C.C. Heesterman

A central limit theorem for
M-estimators by the von Mises method

Department of Mathematical Statistics Report MS-R8713 November

---

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

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

Asymptotic normality of M- or maximum likelihood type estimators has long since been a result by HUBER (1967). REEDS (1976) argued that this could also have been established as an application of the δ-method, using the tool of compactly differentiating von Mises functionals w.r.t. the (empirical) distribution function F_n. If slightly adapted, this alternative approach is shown to be quite fruitful, hopefully maybe even in the non-parametric case. A corrected version of the proof by REEDS is given.

Key words & phrases: asymptotic normality of M-estimators, compact differentiation, Hadamard differentiation, δ-method, M-estimator, von Mises functional.
AMS subject classification: Primary 62F12, Secondary 62G05

1. INTRODUCTION

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

$$\Phi_n(T_n) = \mathbb{E}_{F_n}\psi(X;T_n) = 0 \tag{1.1}$$

Here, $\mathbb{E}$ denotes expectation over the sample space $\mathcal{X}$ w.r.t. some probability measure on $\mathcal{X}$, in this case w.r.t. F_n, the empirical distribution function based on n i.i.d. copies of a random variable Y taking values in $\mathcal{X}$ according to the unknown distribution function F. In order for T_n to be a sensible estimator of θ_0, the function $\psi: \mathcal{X} \times \Theta \to \mathbb{R}^p$ should be chosen such that

$$\Phi(\theta_0) = \mathbb{E}_F\psi(X;\theta_0) = 0. \tag{1.2}$$

As a matter of fact, HUBER relaxed the definition (1.1) somewhat to

$$\Phi_n(T_n) = o_{\mathbb{P}}(n^{-\frac{1}{2}}), \text{ as } n \to \infty. \tag{1.1'}$$

For the moment 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, the first of which he claims covers maximum likelihood estimation in most parametric families in applied statistics, but excludes the M-estimators in the Princeton robustness study (ANDREWS et al. 1972), since these have ψ functions which are only piecewise differentiable in θ, whereas this first theorem requires differentiability on Θ (and thus allows the implicit function theorem to be applied).

Obviously, that result does not improve HUBER's classical c.l.t. for M-estimators, which doesn't make any differentiability assumptions on ψ, but instead imposes some Lipschitz - like conditions on ψ, locally in a neighbourhood of θ_0 (c.f. HUBER 1967, 1983). However, since the proof in HUBER is based heavily on the finite dimensionality of the parameter θ (in fact this is crucial), there is little hope of extending the result to the infinite dimensional case, along the approach as taken by HUBER that is. It is a conjecture in the present note that REEDS' approach of treating M-estimators as so-called *von Mises functionals,* i.e. functionals of the empirical distribution function, does provide the right framework to extend the results for a finite dimensional parameter to the infinite dimensional

Report MS-R8713
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

case. It may be hoped that REEDS' second theorem, which doesn't assume a.s. differentiability of ψ and thus imposes comparably restrictive conditions on the functions ψ and F as in HUBER, can be generalised to also include parameters that are elements of function spaces.

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

$$\mu_\psi(F_n) = \mathbb{E}_{F_n}\psi(X;\cdot) = \Phi_n,$$

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; c.f. CLARKE (1986) on how to avoid ambiguity if there are more than one zeros). FERNHOLZ (1979) reconsidered the theory as developed by REEDS and extended it in the special case of estimating a one-dimensional location parameter. Then, if F is continuous, the functional F_ψ induces a functional τ_ψ on $D[0,1]$, the space of cadlag functions on the unit interval:

$$\tau_\psi(G) = T_\psi(G \circ F)$$
$$\tau_\psi(U_n) = T_\psi(U_n \circ F) = T_n,$$

where U_n denotes the empirical d.f. based on the uniformly distributed r.v.'s $F(X_1), \cdots, F(X_n)$. Thus at least in this special case, she managed to overcome some technical difficulties and mistakes in REEDS (1976). However, in the rest of this note it is shown that these can be solved anyway, and not only in the case of a location parameter.

Now, the central idea is to transfer a central limit theorem for F_n, or rather the stochastic part of T_n, into a c.l.t. for T_n itself 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 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* or *compactly differentiable* the c.l.t. for the stochastic part of T_n may still be transferred to T_n itself. (So since indeed more functionals are compactly differentiable the condition of Fréchet differentiability is rather too 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.

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 c.l.t. for M-estimators and also some applications. Finally then, in the last section the assumptions of REEDS' second theorem and some alternative approaches are briefly discussed.

2. PRELIMINARIES AND HEURISTICS

Let $(\mathcal{X}, \mathcal{A}, \mathbb{P})$ be a probability space. Let $X_1, \cdots, X_n$ be n i.i.d. copies of a random variable $Y \in \mathcal{X}$, with distribution function (d.f.) F corresponding to $\mathbb{P}$. F_n is the empirical d.f. that assigns mass n^{-1} to each of the observation points. Consider estimation of (or testing w.r.t.) the unknown parameter $\theta_0 \in \Theta \subset \mathbb{R}^p$ that satisfies (1.2), i.e. $\Phi(\theta_0) = \mathbb{E}_F\psi(X;\theta_0) = 0$ for some function $\psi: \mathcal{X} \times \Theta \to \mathbb{R}^p$. Assume Θ may be chosen to be a compact subset in $\mathbb{R}^p$. By $B^p(\Theta)$ ($B^1(\mathcal{X})$) denote the space of bounded $\mathbb{R}^p$-valued functions on Θ (c.q. $\mathbb{R}^1$ real-valued functions on $\mathcal{X}$), and by $C(\Theta)$ denote the space of

continuous $\mathbb{R}^p$-valued functions on Θ. So $C(\Theta) \subset B^p(\Theta)$ since Θ is compact.

Now as REEDS observed, any estimator T_n that solves the estimating equations (1.1) may be represented as a functional $T_\psi : B^1(\mathcal{X}) \to \mathbb{R}^p$

$$T_n = T_\psi(F_n) = T \circ \mu_\psi(F_n). \tag{2.1}$$

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

$$T_n = T(\Phi_n). \tag{2.1'}$$

(It is tacitly assumed that $\psi(X;\cdot)$ is a.s. 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 in $C(\Theta)$ implies tightness of this sequence in $C(\Theta)$ (see BILLINGSLEY 1968):

$$\begin{aligned} &\forall \epsilon > 0, \ \exists \text{ compact } K_\epsilon \subset C(\Theta), \text{ such that} \\ &\text{for all } n, P(Y_n \in K_\epsilon) > 1 - \epsilon; \\ &\text{for every weakly convergent series } \{Y_n\}_{n=1}^\infty \subset C(\Theta) \end{aligned} \tag{2.2}$$

A characterization of compactness in $C(\Theta)$ is given by the following well known

PROPOSITION 2.1 (ARZELA - ASCOLI). *A subset $K \subset C(\Theta)$ is compact iff K is closed, bounded and equicontinuous*

$$\begin{aligned} &\forall \epsilon > 0, \exists \delta_\epsilon > 0, \text{ such that.} \\ &\forall g \in K: \sup_{|\theta_1 - \theta_2| \leq \delta_\epsilon} |g(\theta_1) - g(\theta_2)| \leq \epsilon \end{aligned} \tag{2.3}$$

(see BILLINGSLEY (1968). Obviously, it would be very convenient if the sequence $\{\Phi_n\}_{n=1}^\infty$ is indeed a random sequence in $C(\Theta)$.

LEMMA 2.2. *If the function $\psi : \mathcal{X} \times \Theta \to \mathbb{R}^p$ satisfies the conditions* (2.5)

$$\begin{aligned} &(i)\ \psi(x;.) \text{ is continuous in } \theta \text{ for } \mathbb{P}-\text{almost all } x \in \mathcal{X} \\ &(ii)\ \psi(\cdot;\theta) \text{ is Borel measurable as a function on } \mathcal{X} \text{ for all } \theta \in \Theta. \end{aligned} \tag{2.5}$$

then $\psi(X_i;\cdot)$ defines a random element in $C(\Theta)$, for all i; hence Φ_n is a random element in $C(\Theta)$.

PROOF. Since Θ is compact, let $\{\theta_n\}_{n=1}^\infty$ be a dense subset in Θ. Let $f \in C(\Theta)$ and choose $\epsilon > 0$. Then

$$\{x \in \mathcal{X} : \|\psi(x;\cdot) - f(\cdot)\|_\infty \leq \epsilon, \psi(x,\cdot) \in C(\Theta)\} = \bigcap_{j=1}^\infty A_j \cap N_0$$

where

$$A_j = \{x \in \mathcal{X} : \ | \ \psi(x;\theta_j) - f(\theta_j) \ | \ \leq \epsilon\}$$

and

$$N_0 = \{x \in \mathcal{X} : \psi(x;\cdot) \notin C(\Theta)\}.$$

Hence, the lemma follows from (2.5) □

Of course, if Φ_n is a random element in $C(\Theta)$, and $T : C(\Theta) \to \Theta$ is measurable, then T_n can indeed be observed. However, it is not at all clear that T is actually measurable. So just assume together with those who have studied M-estimators in the past, that T_n can be observed, which in turn makes measurability of T quite plausible if Φ_n is indeed a statistic. Here is a point for further investigation

though. In the sequel it is assumed that T can be chosen to be measurable.

Now, VAN ZWET (1984) has pointed out that asymptotic calculations for a statistic T_n may be carried out by approximating T_n through a sum of U-statistics, such that the remainder term is asymptotically negligible in some appropriate sense. In the present situation, approximate $T_n-\theta_0$ by a sum of U-statistics of degree 1:

$$T_n - \theta_0 = \sum_{i=1}^{n} U_{1,n}(X_i) + R_{1,n} \tag{2.6}$$

If the T_n have finite second moments, the best approximation in terms of L_2 is obtained by projection:

$$U_n(X_i) = n^{-1}\mathbb{E}_F(T_n|X_i).$$

Then if $n^{\frac{1}{2}}R_{1,n}=o_{\mathbb{P}}(1)$ as $n\to\infty$, a central limit theorem for T_n follows immediately from (2.6).

Unfortunately, as T_n is defined implicitly through a set of estimating equations, a simple formula for T_n usually doesn't exist, and even if it does, the projection operation is very likely to mess up calculations. Moreover, $U_n(X_i)$ is for each n a different function, whereas the assumption that $X_1,\cdots,X_n$ are i.i.d. is crucial. Anyway, to get some first results for the estimator T_n, one would prefer an approach that doesn't lead to complicated calculations and if need be, improve these results later on. Such an approach was given by REEDS (1976) who adopted the representation (2.1) and carried out a so-called von Mises calculation for $T_\psi(F_n)$, thus extending the δ-method to also apply to functions with an infinite dimensional argument. (In fact, the von Mises method *is* the δ-method. See GILL 1987).

The steps of a von Mises calculation are:

(0) Suppose θ_0 may be represented as a functional of F: $\theta_0=T_1(F)$; Estimate θ_0 by $T_n = T_1(F_n)$
(1) Expand T_n in a Taylor series at F: $T_n=\theta_0+dT_1(F;F_n-F)+R(F_n)$
(2) Show that $R(F_n)=o_{\mathbb{P}}(n^{-\frac{1}{2}})$ as $n\to\infty$
(3) Show that $n^{\frac{1}{2}}dT_1(F;F_n-F)\xrightarrow{D}S$ as $n\to\infty$ and conclude that $n^{\frac{1}{2}}(T_n-\theta_0)\xrightarrow{D}S$ as $n\to\infty$

As mentioned before, it is more natural in the case of M-estimators to adopt the representation $T_n=T(\Phi_n)$, and replace F_n by Φ_n in the von Mises calculation above. Furthermore, in step 1 a concept of differentiation is needed such that in step 2 convergence of $n^{\frac{1}{2}}R(\Phi_n)$ can be shown to hold. Finally, note that it might be possible to carry out the steps (0)-(3) for infinite dimensional parameters θ_0 too.

3. DIFFERENTIATION AND THE δ-METHOD.

As mentioned in the previous section, in order to successfully carry out a von Mises calculation, differentiation of a functional should be defined so that differentiability of the functional T is just strong enough a condition for convergence of $R(F_n)$ in step 2 to hold; however, if the condition is too strong, some important functionals will not be differentiable and hence asymptotic normality can not be established along this route.

Let B_1 and B_2 be topological vector spaces. Let $\mathcal{S}$ denote a class of subsets $K\subset B_1$. A is an open subset in B_1. Consider $x\in A$, $h\in B_1$, and let t, $t_0\in\mathbb{R}$ be such that $x+th\in A$ for $|t|\leqslant t_0$. The following definition of differentiation of functions defined on topological vector spaces is classical. Details can be found for instance in REEDS (1976) or GILL (1987).

DEFINITION 3.1. A mapping $T:A\to B_2$ is called $\mathcal{S}$-differentiable at x if there exists a continuous and linear mapping $dT(x;\cdot):B_1\to B_2$ such that for all $K\in\mathcal{S}$

$$T(x+th) = T(x) + dT(x;th) + o(t) \text{ as } t\to 0, \text{ uniformly in } h\in K. \tag{3.1}$$

(3.1) is of course equivalent to

$$t^{-1}R_T(x;th) = o(1) \text{ as } t\to 0, \text{ uniformly in } h\in K \tag{3.1'}$$

where

$$R_T(x:th) = T(x+th) - T(x) - dT(x;th).$$

Special choices for $\mathcal{S}$ that have frequently been used are:

- $\mathcal{S}=\mathcal{S}_s=$ the set consisting of all singletons in B_1; this choice corresponds to *Gateaux differentiation.*
- $\mathcal{S}=\mathcal{S}_b=$ the set consisting of all bounded subsets in B_1; this choice corresponds to *Fréchet or bounded differentiation.*

Obviously, $\mathcal{S}_s\subset\mathcal{S}_b$, so whenever a functional is Fréchet differentiable, it is also Gateaux 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 $dT(x;\cdot):B_1\to B_2$ such that

$$\|T(x+h) - T(x) - dT(x;h)\|_{B_2} = o(\|h\|_{B_1}) \tag{3.1''}$$

$$\text{as } \|h\|_{B_1}\to 0.$$

Furthermore, let B_1 be $B^1(\mathcal{X})$, endowed with sup-norm, $B_2=\Theta$. Then Fréchet differentiability of T at F implies asymptotic normality. Indeed, by (3.1'') it follows that

$$n^{\frac{1}{2}}(T_n-\theta_0) = n^{\frac{1}{2}}dT(F;F_n-F) + o_{\mathbb{P}}(1), \text{ as } n\to\infty \tag{3.2}$$

since

$$\|F_n - F\|_\infty = O_{\mathbb{P}}(n^{-\frac{1}{2}}).$$

Moreover, the process $n^{\frac{1}{2}}(F_n-F)$ converges weakly to the Brownian bridge process composed with F, hence asymptotic normality follows by Slutsky's lemma (for details, see BILLINGSLEY, 1983). 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 $\mathcal{X}\times\Theta$. 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. Also, the median and other sample quantiles, however simple they are, are not Fréchet differentiable. This is shown in the following example (see also GILL (1987) or FERNHOLZ (1979)).

EXAMPLE 3.2. Let $B_1=C([0,1])$. For $G\in C([0,1])$ define the inverse G^{-1} as usual:

$$G^{-1}(q) = \inf\{1,\{x\in[0,1]:G(x)=q\}\}$$

The functional $T_q:G\to G^{-1}(q)$ is well defined and assigns to any d.f. G its q'th quantile.

PROPOSITION 3.3 *The functional $T_q:B_1\to\Theta=[0,1]$ is not Fréchet differentiable at U, the uniform d.f. on the unit interval.*

PROOF. Suppose T_q is Fréchet differentiable at U; then T_q is also Gateaux differentiable and the two

derivatives must coincide. W.l.o.g. take $q=\frac{1}{2}$ and write $T=T_q$. Let $s\in(0,\frac{1}{2})$. Define

$$G_s(x)=\begin{cases} x+s, & \text{if } 0\leqslant x<\frac{1}{2}-s \\ \frac{1}{2}, & \text{if } \frac{1}{2}-s\leqslant x<\frac{1}{2} \\ x, & \text{if } \frac{1}{2}\leqslant x\leqslant 1 \end{cases}$$

and $H_s=G_s-U$. Then, for $t<1$, $T(U+tH_s)=T(U)=\frac{1}{2}$, so both Gateaux and Fréchet derivative $dT(U;\cdot)$ should satisfy

$$dT(U;tH_s) = 0, \forall t<1,\ s\in(0,\tfrac{1}{2}).$$

Hence, by linearity of $dT(U;\cdot)$

$$dT(U;H_s) = 0.$$

But then, since $T(G_s) = \frac{1}{2}-s$

$$T(U+H_s) - T(U) - dT(U;H_s) = -s.$$

and this contradicts (3.1″), since obviously $\|H_s\|_\infty = s$. □

By $\mathcal{S}_c$ denote the class of all compact subsets in B_1. Hence the inclusion $\mathcal{S}_s\subset\mathcal{S}_c\subset\mathcal{S}_b$ holds.

DEFINITION 3.4 (REEDS 1976) A mapping $T:B_1\to B_2$ is called *Hadamard differentiable at* $x\in B_1$ if (3.1) holds for all $K\in\mathcal{S}_c$.

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

$$(i)\ n^{\frac{1}{2}}(Y_n-x)\xrightarrow{D}Y \text{ in } B_1, \text{as } n\to\infty \tag{3.3}$$

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

Then

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

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

PROOF. Write $H_n=n^{\frac{1}{2}}(Y_n-x)$; by compact differentiability of T at x and (3.3) (ii) approximate $n^{\frac{1}{2}}(T(Y_n)-T(x))$ by $dT(x;H_n)$. The remainder term will be $o_{\mathbb{P}}(1)$ as $n\to\infty$.

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

$$n^{\frac{1}{2}}(T(Y_n) - T(x)) = dT(x;H_n) + n^{\frac{1}{2}}R_T(x;n^{-\frac{1}{2}}H_n),$$

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

$$n^{\frac{1}{2}} R_T(x;n^{-\frac{1}{2}}h) = o(1) \text{ as } n\to\infty, \text{ uniformly in } h\in K \tag{3.4}$$

Then the *stochastic part.* Choose $\epsilon,\eta>0$. By (3.3 ii) there exists a compact K_ϵ such that

$$\mathbb{P}(H_n\in K_\epsilon)>1-\epsilon,\ n=1,2,\cdots \tag{3.5}$$

Furthermore, since

$$\mathbb{P}(\|n^{\frac{1}{2}} R_T(x;n^{-\frac{1}{2}}H_n)\|>\eta) \leqslant$$
$$\mathbb{P}(\|n^{\frac{1}{2}} R_T(x;n^{-\frac{1}{2}}H_n)\|>\eta | H_n\in K_\epsilon) + \mathbb{P}(H_n\notin K_\epsilon)$$

(3.4) and (3.5) together imply $n^{\frac{1}{2}} R_T(x;n^{\frac{1}{2}}H_n)= o_{\mathbb{P}}(1)$ as $n\to\infty$. Hence, as $dT(x;\cdot)$ is linear and continuous, the theorem follows by (3.3) (i) via Slutsky's lemma. □

REMARK: The topology on B_1 will have to be chosen such that the analytic properties of T and the stochastic properties of H_n (both depend on the topology) are attuned to each other w.r.t. the δ-method.

REMARK: (3.5) illustrates why Gateaux differentiation is useless in the present situation: it would require $\forall\epsilon>0$ the existence of a singleton $K_\epsilon\in\mathcal{S}_s$ such that $P(H_n\in K_\epsilon)>1-\epsilon, n=1,2,\cdots$, which is of course an absurd requirement for random variables H_n with uncountable support.

REMARK: GILL (1987) also mentioned the (generalised) δ-method, though he immediately replaced it by a theorem based on differentiation tangentially to a subspace: the problem is that some M-estimators are von Mises functionals defined on the space of cadlag functions on $\mathcal{X}$ (e.g. the quantile!) and can only be shown to be differentiable tangentially to the subspace of continuous functions on $\mathcal{X}$; this is still a weaker condition than compact differentiability. In fact there are several ways to overcome this (merely technical) problem. For instance POLLARD (1984, 1985) considers 'generalised weak convergence' instead, though of course one may also replace Φ_n by a smooth version Φ_n, say. This has the advantage that there will be a solution to the estimating equations, which of course may not be the case if Φ_n is not continuous. GILL (1987) also uses generalized weak convergence.

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 O; then the implicit function theorem proves very useful in showing that a central limit theorem for $(\Phi_n(\theta),\theta\in\Theta)$ carries ever to $T_n=T(\Phi_n)$, where $\Phi_n=\mathbb{E}_{F_n}\psi(X;\cdot)$, if indeed the sequence $\{n^{\frac{1}{2}}(\Phi_n - \Phi)\}_{n=1}^{\infty}$ is weakly convergent and tight in some appropriate topological vector space. Here, as in (1.2), $\Phi=\mathbb{E}_F\psi(X;\cdot)$. 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 a proposition by GINÉ (1974).

LEMMA 4.1. *Let $Z_1,\cdots,Z_n$ be i.i.d. copies of a random variable Z in $C(\theta)$ with zero expectation. If*

$$(i)\ \mathbb{E}_{\mathbb{P}}\|Z\|_\infty^2 <\infty\ , \tag{4.1}$$

$$(ii)\ \mathbb{E}_{\mathbb{P}} \sup_{\theta_1,\theta_2}\frac{|Z(\theta_1)-Z(\theta_2)|^2}{|\theta_1-\theta_2|^\lambda} <\infty, \text{ for some } \lambda>0$$

then the sequence $S_n\to n^{-\frac{1}{2}}\Sigma_{i=1}^n Z_i$, $n = 1,2,$ ... 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 $\{n^{\frac{1}{2}}(\Phi_n-\Phi)\}$ 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 $n^{-\frac{1}{2}}\Sigma_{i=1}^{n}Z_i = n^{-\frac{1}{2}}\Sigma_{i=1}^{n}[\psi(X_i;\cdot)-\Phi(\cdot)]$. If the function $\Phi:\Theta\to\mathbb{R}^p$, $\Phi(\theta)=\mathbb{E}_F\psi(X;\theta)$, has the following properties:*

$$\begin{aligned} &(i)\ \Phi \text{ has a unique zero at } \theta_0 \\ &(ii)\ \Phi \text{ is a local homeomorphism at } \theta_0 \\ &(iii)\ \Phi \text{ is differentiable at } \theta_0 \text{ with nonsingular derivative } A:\mathbb{R}^p\to\mathbb{R}^p, \end{aligned} \tag{4.2}$$

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

$$\begin{aligned} &(i)\ \Phi_n(T_n) = o_{\mathbf{P}}(1) \text{ as } n\to\infty \\ &(ii)\ n^{\frac{1}{2}}(T_n-\theta_0)\xrightarrow{D}N(0,\Sigma) \text{ as } n\to\infty . \end{aligned} \tag{4.3}$$

The covariance matrix Σ is given by

$$\begin{aligned} &\Sigma = A^{-1}\Gamma(A^{-1})^{\mathrm{T}}, \text{ where} \\ &n^{\frac{1}{2}}(\Phi_n-\Phi)(\theta_0)\xrightarrow{D}N(0,\Gamma) \text{ as } n\to\infty . \end{aligned} \tag{4.4}$$

REEDS consequently 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 d.f. based on n i.i.d. 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 $D[0,1]$ equipped with the supremum norm; on the other hand, while U_n is indeed a random element in $D[0,1]$ equipped 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 (1987) 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 $\tilde{B}_1 = L_2(\mathbb{P})\times C(\Theta)$ equipped with the norm $\|(x,y)\|_{\tilde{B}} = \|x\|_{L_2}+\|y\|_\infty$ via the $1-1$ mapping $\bar{\alpha}(g)=(g,\mathbb{E}_g\psi(X;.))$. Since by a theorem of Prohorov for the first coordinate and Lemma 4.1 for the second coordinate $\{\bar{\alpha}(n^{\frac{1}{2}}(F_n-F))\}_{n=1}^{\infty}$ is random and tight in $\tilde{B}_1$, REEDS concludes that the sequence of arguments $\{n^{\frac{1}{2}}(F_n-F)\}_{n=1}^{\infty}$ 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: properties of $\bar{\alpha}(n^{\frac{1}{2}}(F_n-F))$ 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 $n^{\frac{1}{2}}(\Phi_n-\Phi)$ is needed anyhow, why not apply the δ-method to the functional $T(\Phi_n)$ straightaway and forget all about the $n^{\frac{1}{2}}(F_n-F)$-part? Indeed, the functional T_ψ is Hadamard differentiable iff T is Hadamard differentiable, since μ_ψ is linear and continuous whereas compact differentiation follows the chain rule. So there is really less work in establishing the validity of the conditions in Theorem 3.5 if Y_n is taken to be Φ_n instead of

F_n. Equivalently $\{F_n\}$ may be endowed with the *pseudonorm* $\|F_n\| = \|\mathbb{E}_{F_n}\psi(X;\cdot)\|_\infty$ 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 $\{n^{\frac{1}{2}}(\Phi_n - \Phi)\}_{n=1}^{\infty}$ is weakly convergent in $C(\Theta)$ and hence tight in $C(\Theta)$, the stochastic part of the δ-method applied to T_n is already settled. So it remains to prove existence and compact differentiablity 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 (4.2) 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 (4.2) (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 $W \subset \Theta$, defines a homeomorphism between W and the ball $B_{(0,r)} = \{t \in \mathbb{R}^p : |t| \leq r\}$. For such r define $V_r \subset C(\Theta)$:

$$V_r = \{f \in C(\Theta) : \|\Phi - f\|_\infty < r\}. \tag{4.5}$$

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

$$T(f) = \begin{cases} \Phi^{-1}(t_f), & \text{if } f \in V_r \\ \theta_\infty \in \Theta, & \text{otherwise} \end{cases} \tag{4.6}$$

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 (4.6). Under the conditions of Theorem 4.2 the M-estimator $T_n = T(\Phi_n)$ satisfies*

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

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

$$\mathbb{P}(\Phi_n \in V_r) \to 1 \text{ as } n \to \infty \tag{4.7}$$

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

REMARK 4.5. Under the conditions of Theorem 4.2 existence of a solution to the estimating equations is actually quite an easy consequence of Brouwer's fixed point theorem. However, in the infinite dimensional case, i.e. θ is a function, Brouwer's fixed point theorem doesn't hold, which means that existence of a solution is possibly a very hard and deep result. Still of course some generalisation of the theory for M-estimators of a finite dimensional parameter to M-estimators of a function may very well exist. This might be an interesting project for further investigation.

Let V_r be defined as in (4.5), and let T be defined as in (4.6). 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 $|t| \leq rk^{-1}$ 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 4.6. *The functional T is $\mathbb{S}$-differentiable at Φ with derivative $T(\Phi;g)=-A^{-1}g(\theta_0)$ 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 $K\in\mathbb{S}$

$$A\beta_h(t) = o(1) \text{ as } t\to 0, \text{ uniformly in } h\in K. \tag{4.9}$$

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

$$T_t - T_0 = O(t) \text{ as } t\to 0, \text{ uniformly in } \|h\|_\infty \leqslant k. \tag{4.10}$$

Indeed, since by definition $\Phi(T_t)=-th(T_t)$, it follows from assumption (4.2) (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

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

Hence, since Φ is assumed to be differentiable at θ_0 with nonsingular derivative, (4.10) is valid. Furthermore, by the same assumption,

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

where

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

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

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

hence, (4.9) holds iff K is equicontinuous (and of course bounded). □

COROLLARY 4.7: *The functional T defined in (4.6) is compactly differentiable at Φ.*

PROOF: See proposition 2.1. □

PROOF OF THEOREM 4.2. See Corollary 4.4 for (4.3) (i). By Corollary 4.7, the δ-method may now be applied to obtain (4.3) (ii). Notice that (4.4) trivially holds since obviously $\mathbb{E}_F|\psi(X;\theta_0)|^2<\infty$. □

EXAMPLE 3.2 (continued) When estimating a quantile q of the distribution function F, the functions Φ and Φ_n are defined as

$$\Phi(\cdot)=F(\cdot)-q$$

$$\Phi_n(\cdot)=F_n(\cdot)-q$$

Assume F is differentiable at its q^{th} quantile θ_0, with strictly positive derivative $F'(\theta_0)$. Then Φ satisfies the conditions of Theorem 4.2; unfortunately, as already mentioned above, Φ_n is not continuous with probability 1, so in order to apply Theorem 4.2, a smoothed version $\tilde{\Phi}_n$ should be used instead of Φ_n itself: Define

$$\tilde{\Phi}_n(\theta) = \begin{cases} \Phi_n(\theta), & \text{if } \theta = X_{(k)} \text{ for some } 1 \leqslant k \leqslant n \\ \frac{k}{n} + \frac{\Phi_n(X_{(k+1)}) - \Phi_n(X_{(k)})}{X_{(k+1)} - X_{(k)}} \cdot (\theta - X_{(k)}), & \text{if } X_{(k)} < \theta < X_{(k+1)} \end{cases}$$

Obviously, the M-estimator that solves the estimating equation $\tilde{\Phi}_n = 0$ satisfies (1.1'). Furthermore it is easily seen that

$$n^{\frac{1}{2}}(\tilde{\Phi}_n - \Phi) \xrightarrow{D} W^0 \circ F \text{ (in } C(\Theta)) \text{ as } n \to \infty. \tag{4.11}$$

Here $W^0 \circ F$ denotes the Brownian bridge process composed with F (see BILLINGSLEY (1968) for details). Hence this sequence is also tight in $C(\Theta)$.

Now, because the analytic and stochastic aspects of the estimator T_n are so nicely separated by the δ-method, the following adapted version of Theorem 4.2 can now be formulated for M-estimators of a quantile:

THEOREM 4.2'. *Let F be a distribution function defined on a compact subset Θ in $\mathbb{R}$ and suppose that F is differentiable at its q^{th} quantile $\theta_0 \in \Theta$, with strictly positive derivative $F'(\theta_0)$. Then an M-estimator T_n that solves the smoothed estimating equation $\tilde{\Phi}_n = 0$ exists and satisfies*

$$n^{\frac{1}{2}}(T_n - \theta_0) \xrightarrow{D} N(0, \sigma^2),$$

where

$$\sigma^2 = (F'(\theta_0))^{-2} \cdot q(1-q).$$

PROOF: since $\Phi = F - q$ satisfies the conditions of Theorem 4.2, T as defined in (4.6) is Hadamard differentiable at Φ, with derivative $dT(\Phi; g) = (F'(\theta_0))^{-1} \cdot g(\theta_0)$. Hence by (4.11) the theorem follows as a direct application of the δ-method. □

REMARK. Consider now the situation that the distribution function F is defined on some arbitrary subset $\Theta \subset \mathbb{R}$, not necessarily compact, and assume that the q^{th} quantile θ_0 of F is in the interior of Θ. Obviously, if F is differentiable at θ_0 with positive derivative $F'(\theta_0)$, then the conditions (4.2) still hold for Φ. Indeed, since these are local conditions in some neighbourhood of θ_0, compactness of Θ is irrelevant in this respect. However, the sequence $\{n^{\frac{1}{2}}(\Phi_n - \Phi)\}_{n=1}^{\infty}$ will typically not be a tight random sequence in $C(\Theta)$. (See Lemma 2.2.; there it is essential that Θ be compact). So, instead of X itself, one might consider $Y = g(X)$, where g denotes some 1-1 mapping from $\mathbb{R}$ into the unit interval, and such that g is continuously differentiable at θ_0 with non-zero derivative $g'(\theta_0)$. Let T_n^g denote the M-estimator for the q^{th} quantile of $F \circ g^{-1}$, the distribution function of Y. Of course, Theorem 4.2' holds for T_n^g:

$$n^{\frac{1}{2}}(T_n^g - g(\theta_0)) \xrightarrow{D} N(0, \sigma^2(g)) \tag{4.12}$$

where

$$\sigma^2(g) = [(F \circ g^{-1})'(g(\theta_0))]^{-2} \cdot q(1-q).$$

Define then the estimator T_n for θ_0 by

$$T_n = g^{-1}(T_n^g). \tag{4.12}$$

EXTENSION TO THEOREM 4.2'. *Let F be a distribution function defined on $\Theta \subset \mathbb{R}'$ and assume that F is differentiable at its q^{th} quantile θ_0 with strictly positive derivative $F'(\theta_0)$. Then the estimator T_n of θ_0 defined in (4.13) is consistent and moreover satisfies*

$$n^{\frac{1}{2}}(T_n - \theta_0) \xrightarrow{D} N(0, \sigma^2), \tag{4.14}$$

where

$$\sigma^2 = [F'(\theta_0)]^{-2} \cdot q(1-q) \tag{4.15}$$

PROOF: Consistency of T_n is an immediate consequence of the fact that g is continuously differentiable at θ_0 with non-zero derivative $g'(\theta_0)$ and from consistency of T_n^g. Furthermore, again from consistency of T_n^g and continuous differentiability of g at θ_0 with non-zero derivative it follows that

$$n^{\frac{1}{2}}(T_n - \theta_0) - (g^{-1})'(g(\theta_0)) n^{\frac{1}{2}}(T_n^g - g(\theta_0)) \xrightarrow{P} 0 \text{ as } n \to \infty.$$

Hence, (4.14) and (4.15) follow from (4.12) □

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 δ-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 $x \in \mathcal{X}$, 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 δ-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 δ-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 $X_1, \cdots, X_n$ 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 δ-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 maybe 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 $n^{\frac{1}{2}}(\Phi_n - \Phi)$ in the parametric case be defined in the non-parametric case where $n^{\frac{1}{2}}(\Phi_n - \Phi)$ is itself a function? Moreover, the choice of metric for B_1 will not be as easy as it was in the parametric case; indeed the structure of $C(\Theta)$ is such that even with the naive choice of uniform topology the conditions of the δ-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. However, 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 m.l.e.'s 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 author wants to thank Prof. Rieder for the priviledge of studying his lecture notes on Reeds' theory for M-estimators. During the preparation of this report communication with D. Vlot has also been very helpful and stimulating.

REFERENCES

D.F. ANDREWS, et al. (1972), *Robust Estimates of Location.* Princeton University Press, Princeton.

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

B.R. CLARKE, (1983), Uniqueness and Fréchet Differentiability of Functional Solutions to maximum Likelihood Type Equations. *Ann. Statist.* **11**, 1196-1205.

B.R. CLARKE, (1986), Nonsmooth Analysis and Fréchet Differentiability of *M*-functionals. *Prob. Theory Related Fields* **73**, 197-210.

L.T. FERNHOLTZ, (1979), *Von Mises Calculus for Statistical Functionals.* Lecture Notes in Statistics **19**, Springer Verlag, New York.

R.D. GILL (1987), *Non-and semi-parametric Maximum Likelihood Estimators and the von Mises Method,* Part I. Research report MS-R8709, C.W.I., Amsterdam; Part II, Research report in preparation at the C.W.I., Amsterdam.

E. GINÉ, (1974), On the Central Limit Theorem for Sample Continuous Processes. *Ann. Probab.* **4**, 629-641.

P. HUBER, (1964), Robust Estimation of a Location Parameter. *Ann. Math. Stat.* **35**, 73-101.

P. HUBER, (1967), The Behaviour of Maximum Likelihood Estimates under Nonstandard Conditions. in: *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1.* University of California Press, Berkeley.

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

D. POLLARD, (1984), *Convergence of stochastic Processes,* Springer. New York.

D. POLLARD, (1985), New Ways to prove Central Limit Theorems. *Econometric Theory* **1**, 295-314.

J.A. REEDS, (1976), *On the Definition of von Mises Functionals.* Research report S-44, Dept. of Statistics, University of Harvard.

D. VLOT, (1987), *Applications of a von Mises Method, based on Hadamard Differentiation Tangentially to a Subspace. (Toepassingen van een von Mises Methode, gebaseerd op Tangentiele Hadamard Differentiatie.)* Unpublished B.A. Thesis, Dept. of Math. Stat., University of Utrecht.

W.R. VAN ZWET, (1984), A Berry-Esseen Bound for Symmetric Statistics. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **66**, 425-440.