A REMARK ON THE DIFFERENCE BETWEEN PROBABILITY AND ANALYSIS IN THE PROOF OF THE CENTRAL LIMIT THEOREM FOR M-ESTIMATORS

by

Brenton R Clarke

Murdoch University

Abbreviated Title: An Easy Proof of Asymptotic Normality AMS 1970 subject classifications. Primary 62E20; secondary 62G35.

Key words and phrases: Von Mises expansion, Fréchet derivative, asymptotic normality, M-estimators.

1. Introduction

In the proliferation of recent articles on expansions of the Von Mises, Hadamard or Fréchet type, including papers by Kallianpur (1963), Reeds (1976), Huber (1981), Esty, Gillette, Hamilton and Taylor (1985), Gill (1987), Heesterman (1987), the authors have often overlooked the case of the Fréchet type expansion, shown to be easily applicable to many forms of M-estimator in Clarke (1983,1986). Here the estimating equations are of the form

$$K_{F_n}(\tau) = \frac{1}{n} \sum_{i=1}^{n} \psi(X_i, \tau) = 0 \ , \tag{1.1}$$

where $X_1, \ldots, X_n$ are independent, identically distributed random variables taking values in a separable metrizeable space R, and ψ is an $r \times 1$ vector function with domain $R \times \Theta$ which has a continuous partial derivative. The common distribution

$F_\theta \in \mathcal{f} = \{F_\tau \mid \tau \in \Theta\}$, and ψ satisfies a condition of Fisher consistency $E_{F_\theta}[\psi(X,\theta)] = 0$. An estimating functional, evaluated at the empirical distribution F_n, is defined to be a root of equations (1.1), and is labeled $T[\psi, F_n]$. In the case of multiple solutions of equations (1.1), it may need further definition via a "selection functional" described in Clarke (1983,1988).

If $\mathcal{G}$ is the space of distributions on R and the linear space spanned by differences F-G of members of G is denoted by D, then T is Fréchet differentiable at $G \in \mathcal{G}$ with respect to the pair $(\mathcal{G}, d)$, where d is a metric distance on $\mathcal{G}$, when it can be approximated by a linear functional $T_G(\cdot)$ defined on D, such that

$$\|T[F] - T[G] - T'_G(F - G)\| = o(d(F,G)) \ , \tag{1.2}$$

as $d(F,G) \longrightarrow o$, $F \in \mathcal{G}$. Frequently satisfied conditions for expansion (1.2) using the Kolmogorov, Lévy and Prokhorov metric distances d are given in Clarke (1983,1986) . In particular, robustness properties of weak continuity (cf. Hampel 1971) are derived from the latter two distances, whence ψ must be a bounded continuous function. On the other hand, a proof of asymptotic normality of $\sqrt{n}(T[\psi,F_n] - T[\psi,F_\theta])$ requires fewer conditions than those required to achieve expansion (1.2) . In fact the proofs of consistency and asymptotic normality can be achieved using the same analysis as that of Clarke (1983) and using the choice of neighbourhood.

$$n(\delta,F_\theta) = \{G \in \mathcal{G} : \sup_{\tau\in D} \| \int_R \psi(x,\tau)dG(x) - \int_R \psi(x,\tau)dF_\theta(x)\| < \delta ,$$

$$\text{and} \quad \sup_{\tau\in D} \| \int_R \frac{\partial}{\partial\tau} \psi(x,\tau)dG(x) - \int_R \frac{\partial}{\partial\tau} \psi(x,\tau)dF_\theta(x)\| < \delta \} \qquad (1.3)$$

Here $D \subset \Theta$ is some nondegenerate compact set containing θ in its interior. The main implication is that the expansion

$$\|T[F_n] - T[F_\theta] - T'_{F_\theta} (F_n - F_\theta)\| = o_p\left(\frac{1}{\sqrt{n}}\right) , \qquad (1.4)$$

holds for most cases of unbounded ψ-functions, including those derived from the exponential family in maximum likelihood estimation. Though such estimating functionals will not satisfy expansion (1.2), since they are not robust, asymptotic normality is a simple consequence of (1.4).

In a sense (1.4) has been established for location M-estimators through the work of Boos and Serfling (1980) who seek to retain the framework of the expansion (1.2), though relax the error term to be in small order probability by using Kolmogorov distance $o_p(d_k(F_n,F_\theta))$, whence the expansion is appropriate only to the proof of asymptotic normality and the law of iterated logarithm. Parr (1985) also considers the expansion with a probabilistic error term for jackknifing. An advantage of considering neighbourhoods (1.3) is that multivariate observation spaces are included and conditions are easy to check. This is a consequence of the theory of Rao (1962).

2. The Main Results

A Theorem for existence of a unique consistent root of equations (1.1) follows. To describe convergence let $\{A_n\}$ be a sequence of events that

hold for all sufficiently large n , that is, $P(\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n) = 1$. Note that, $T[\psi,F_n]$ converges almost surely to θ if for arbitrary $\delta > 0$ the sequence of events $A_n = \{\omega : |T[\psi,F_n] - \theta| < \delta\}$ holds for all sufficiently large n . Conditions $A(A_0 - A_4)$ of Clarke (1983) will be referred to. By choice of the neighbourhood condition A_4 is automatically satisfied. The following theorem is analogous to the theorem of Foutz (1977), but generalizes results for maximum likelihood estimators. It is proved, for general ψ satisfying conditions A , and the assumption of uniform convergence is superseded by the conditions.

Theorem 2.1: Assume conditions A hold. Given $\kappa > 0$, $T[\psi,F_n]$ exists and is an element of $U_\kappa(\theta)$, the open ball of radius κ about θ , for sufficiently large n . Moreover, if $I(\psi,F_n) = \{$solutions of equations (1.1)$\}$, then there is a $\kappa^* > 0$ such that

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

a unique solution. Finally, $T[\psi,F_n] \xrightarrow{a.s.} T[\psi,F_\theta] = 0$.

Proof: Let $\mathcal{E} = \{\psi(\cdot,\tau) : \tau \in \Theta\}$ and $\mathcal{E}' = \left\{\frac{\partial}{\partial\tau}\psi(\cdot,\tau) : \tau \in \Theta\right\}$. By remark 2.2 of Clarke (1983) both $\mathcal{E}$ and $\mathcal{E}'$ are equicontinuous on R , as a consequence of assumption A_1 , that ψ and its partial derivatives are continuous. From condition A_2 and theorem 6.2 of Rao(1962) , it follows that for arbitrary $\epsilon > 0$, $F_n \in n(\epsilon,F_\theta)$ f.a.s.l.n. . Hence using theorem 3.2 of Clarke (1983) , and the choice of neighbourhood, since $F_n \in n(\epsilon,F_\theta)$ f.a.s.l.n., theorem 2.1 holds.

Theorem 2.2: Let conditions A hold, and let $\psi(X,\theta)$ have a finite second moment at F_θ. Then, $\sqrt{n}(T[\psi,F_n] - T[\psi,F_\theta])$ converges in distribution to a normal random variable with mean zero and variance

$$\sigma^2(T,F_\theta) = M(\theta)^{-1} \int_{-\infty}^{+\infty} \psi(x,\theta)\psi(x,\theta)' \, dF_\theta(x) \, (M(\theta)^{-1})'$$

where integration is carried out componentwise and

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

is nonsingular.

Proof: Consider the two term expansion using the mean value theorem

$$0 = K_{F_n}(T[\psi,F_n]) = K_{F_n}(\theta) + \frac{\partial}{\partial\tau} K_{F_n}(\tilde{\tau})(T[\psi,F_n] - \theta) . \tag{2.1}$$

Here $\tilde{\tau}$ is evaluated at different parameters for each component function expansion. Rewriting

$$\sqrt{n}(T[\psi,F_n] - \theta) = M(\theta)^{-1} \sqrt{n} K_{F_n}(\theta)$$
$$+ M(\theta)^{-1} \{M(\theta) - \frac{\partial}{\partial\tau} K_{F_n}(\tilde{\tau})\} \sqrt{n}(T[\psi,F_n] - \theta) . \tag{2.2}$$

Since for arbitrary $\delta > 0$, $F_n \in n(\delta,F_\theta)$ f.a.s.l.n.

$$\frac{\partial}{\partial\tau} K_{F_n}(\tilde{\tau}) = \int_R \frac{\partial}{\partial\tau} \psi(x,\tau) \, dF_n(x)\Big|_{\tau=\tilde{\tau}}$$
$$\xrightarrow{a.s.} M(\theta) .$$

The term $\sqrt{n}(T[\psi,F_n] - \theta)$ is $O_p(1)$ from the expansion (2.1). Consequently, the latter term in (2.2) is $o_p(1)O_p(1) = o_p(1)$. Hence equations (2.2) are equivalent to equations (1.4) with $T'_{F_\theta}(F_n - F_\theta) = \int M(\theta)^{-1}\psi(x,\theta)d(F_n - F_\theta)(x)$.

3. Example and Conclusion

The maximum likelihood equations for location and scale of a normal parametric family of distributions on the real line are defined by $\psi(x;\mu,\sigma) = \left(\frac{x-\mu}{\sigma}, -1 + \left\{ \frac{x-\mu}{\sigma} \right\}^2 \right)$ and equations (1.1). Without loss of generality, let $\theta = (\mu_o,\sigma_o)$ be location and scale parameters for the underlying distribution. Let $D = \{(\mu,\sigma) \mid \|(\mu,\sigma)' - (\mu_o,\sigma_o)'\| \le \sigma_o/2\}$.

Clearly ψ has continuous partial derivatives on D , whence condition A_1 is satisfied. Since uniformly on D it is true that

$$\left| \frac{x - \mu}{\sigma} \right| < (2/\sigma_o)(|x - \mu_o| + \sigma_o/2) ,$$

the vector function $\psi(x;\mu,\sigma)$ and matrix of partial derivatives of ψ are bounded in Euclidean norm by

$$g(x) = \{1 + 4\ (2/\sigma_o)(|x - \mu_o| + \sigma_o/2)\}^2 \ \max(1.,\sigma_o/2) .$$

Condition A_2 is satisfied if g is integrable with respect to each $G \in n(\epsilon,F_\theta)$, as is the case for $F_n \in n(\epsilon,F_\theta)$. That is, condition A_2 is satisfied if g is integrable with respect to each $G \in n(\epsilon,F_\theta)$, as in the case for $F_n \in n(\epsilon,F_\theta)$. That is, condition A_2 is satisfied. Thus the expansion (1.4) is true for unbounded ψ , and conditions A of Clarke (1983) are easy to check. Hence the expansion (1.4) leading to asymptotic normality is easier to achieve than is implied by the formal theory of authors seeking to use Hadamard derivatives to prove asymptotic normality, though some advantage may be gained in the discussion of quantile estimators.

The proof of asymptotic normality highlighted here is analogous to the proof of Fréchet differentiability, the difference being in the weaker form of neighbourhood and an error term that is given in probability rather than as a mathematical quantity defined in terms of a metric distance. A more direct proof of asymptotic normality for maximum likelihood estimation is given in Cramér (1946). The conditions A used here, are stated in a way which may make them easier to use. They are also applicable to M-estimators. Proofs which consider discontinuous partial derivatives are covered in Carroll (1978) and Clarke (1986) for M-estimators more generally.

REFERENCES

Anderson,T.W.(1958) An Introduction to Multivariate Statistical Analysis, Wiley, New York

Boos,D.D. and Serfling, R.J. (1980) A note on differentials and the CLT and LIL for statistical functions with application to M-estimates. Ann.Statist. 8,618-624.

Carroll, R.J. (1978) On the asymptotic distribution of multivariate M-estimates, Jour.Mult.Anal.,8,361-371

Clarke,B.R. (1983) Uniqueness and Fréchet differentiability of Functional Solutions to Maximum Likelihood type equations, Ann.Statist.,11, 1196-1205

Clarke,B.R. (1986) Nonsmooth analysis and Fréchet differentiability of M-functionals, Probab.Th.Rel.Fields,73,197-209

Clarke,B.R. (1988) The Selection Functional, submitted Poland. Builletyn Statystyczny.

Cramér,H (1946) Mathematical Methods in Statistics, Princeton University Press, Princeton, New Jersey

Esty,W.,R.Gillette, M.Hamilton and D.Taylor (1985) Asymptotic distribution theory of statistical functionals: the compact derivative approach for robust estimators, Ann.Inst.Statist.Math.,37,109-129

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

Foutz,R.V. (1977) On the unique consistent solution to the likelihood equations. J.Amer.Statist.Assoc.72,147-148

Gill,R.D. (1987) Non and semi-parametric maximum likelihood estimators and the Von Mises Method, Part 1 Research Report MS-R8709,C.W.I., Amsterdam

Hampel,F.R. (1971) A general qualitative definition of robustness. Ann. Math.Statist.42,1887-1896

Heesterman,C.C. (1987) A central limit theorem for M-estimators by the Von Mises Method. Submitted to the Annals of Statistics

Huber,P.J. (1981) Robust Statistics, Wiley, New York

Kallianpur,G. (1963) Von Mises functionals and maximum likelihood estimation, Sankhyā,Ser. A,23,149-158

Parr,W.C. (1985) Jackknifing differentiable statistical functionals, J.R.Statist.Soc.,Ser.B.,47,56-66

Rao,R.R. (1962) Relations between weak and uniform convergence of measures with applications. Ann.Math.Statist.33,659-680

Reeds,J.A. (1976) On the definition of Von Mises functions. Ph.D. Thesis Department of Statistics, Harvard University, Cambridge, MA