The Annals of Statistics
1983, Vol. 11, No. 4, 1196–1205

UNIQUENESS AND FRÉCHET DIFFERENTIABILITY OF FUNCTIONAL SOLUTIONS TO MAXIMUM LIKELIHOOD TYPE EQUATIONS

BY BRENTON R. CLARKE

Royal Holloway College

Solutions of simultaneous equations of the maximum likelihood type or M-estimators can be represented as functionals. Existence and uniqueness of a root in a local region of the parameter space are proved under conditions that are easy to check. If only one root of the equations exists, the resulting statistical functional is Fréchet differentiable and robust. When several solutions exist, conditions on the loss criterion used to select the root for the statistic ensure Fréchet differentiability. An interesting example of a Fréchet differentiable functional is the solution of the maximum likelihood equations for location and scale parameters in a Cauchy distribution. The estimator is robust and asymptotically efficient.

1. Introduction. The usefulness of writing an estimator as a statistical functional of the empirical distribution in order to make systematic descriptions of it was first deliberately exploited by Von Mises (1947). Authors who have considered the application of Fréchet differentiability of a statistical functional to proofs of asymptotic normality include Kallianpur and Rao (1955) and Boos and Serfling (1980). In another context Hampel (1968, 1971) has emphasized weak continuity and, to a lesser extent, Fréchet differentiability of an estimating functional at a parametric distribution F_θ. Such properties are relevant to the robustness of a statistical functional against the extraneous observation, rounding errors, and slight misspecification of the parametric distribution. A good description of the implications of Fréchet differentiability is given by Huber (1981).

However, as with much robustness theory, few results exist in the case of parameters other than location. A key feature of results in this paper is their applicability when the parameter space Θ is an open subset of Euclidean r-space and when more than one solution of a set of general estimating equations is known. That is, not only existence of a root which is differentiable is shown, but differentiability of a well defined statistical estimator is possible. The equations are assumed to be of a form

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

where $X_1, \cdots, X_n$ are independent, identically distributed random variables taking values in a separable metrizable space R, and ψ is an $r \times 1$ vector function with domain $R \times \Theta$ which has a continuous partial derivative. Estimators that are solutions of (1.1) are generally termed M-estimators and include maximum likelihood estimators and some minimum distance estimators.

A single functional root of equations (1.1) may be written $T[\psi, F_n]$ where F_n is the empirical distribution function that attributes atomic mass n^{-1} to each of the points $X_1, \cdots, X_n$. More generally $T[\psi, G]$ can be defined as a functional root of equations

$$K_G(\tau) = \int_R \psi(x, \tau)\, dG(x) = 0, \tag{1.2}$$

Received July 1981; revised May 1982.

AMS 1970 *subject classifications*. Primary 62E20; secondary 62G35.

Key words and phrases. Statistical functionals, weak continuity, Fréchet differentiability, M-estimators, Cauchy distribution.

$T[\psi, G] = +\infty$ if no root exists, where $G \in \mathscr{G}$ which is the space of all probability distributions on R. In the event of several roots of (1.1) a distance criterion ρ_0 is employed to select the estimator from them. Usually the roots correspond to extrema of the distance but it is not assumed so here. Specifically the general functional $T[\psi, \rho_0, \cdot]$ is defined as the solution to

$$\inf_{\tau \in I(\psi, G)} \rho_0(G, \tau) = \rho_0(G, T[\psi, \rho_0, G]) \tag{1.3}$$

where

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

if a solution exists. Otherwise $T[\psi, \rho_0, G] = \infty$. For example the mle is the functional $T[\psi, \rho_0, \cdot]$ with ψ given by the efficient score and

$$\rho_0(G, \tau) = -\int_R \log f_\tau(x)\, dG(x) \tag{1.4}$$

where f_τ is the density function which corresponds to the parametric distribution $F_\tau \in \mathscr{F} = \{F_\theta \mid \theta \in \Theta\}$.

Rao (1957) showed the mle of a univariate parameter in a multinomial distribution to be Fréchet differentiable using the Kolmogorov metric after Kallianpur and Rao (1955) had demonstrated that the class of Fisher consistent estimators that are Fréchet differentiable have asymptotic variances greater than or equal to $n\,\mathscr{I}(\theta)^{-1}$ where $\mathscr{I}(\theta)$ is Fisher information. But Kallianpur (1963) reported that in general neither author could under any reasonable set of assumptions (on the density functions in the continuous, and the probability function in the infinite discrete case) prove Fréchet differentiability of the mle. It transpires from the discussion of conditions in Section 6 that a bounded ψ function is necessary which is not true generally for the mle. An example of a Fréchet differentiable maximum likelihood functional is the root of the equations for location and scale of a Cauchy distribution, described in Section 7.

In Section 3 existence and uniqueness of a solution of equations (1.1) in a region of the parameter space for small enough neighbourhoods of F_θ is shown. This local argument complemented in Section 4 by global conditions on ρ_0 ensures continuity of the functional solution of (1.3). Fréchet differentiability subsequently follows easily from a two term Taylor expansion in Section 4.

2. Preliminaries. We will mean by a neighbourhood of a distribution G a subset of $\mathscr{G}$ for which the ordering property $0 < \varepsilon_1 \le \varepsilon_2 \Rightarrow n(\varepsilon_1, G) \subset n(\varepsilon_2, G)$ holds. Neighbourhoods may typically be formed by writing $n(\varepsilon, G) = \{F \in \mathscr{G} \mid d(F, G) < \varepsilon\}$ for metric distances on $\mathscr{G}$. The Kolmogorov metric distance between distributions on the real line is $d_k(F, G) = \sup_{x \in E} |F(x) - G(x)|$. On the more general space $(R, \mathscr{B})$ where $\mathscr{B}$ are the Borel sets of the metric space R, the Prokhorov distance is defined as $d_p(F, G) = \inf\{\delta > 0 \mid F\{A\} \le G\{A^\delta\} + \delta, G\{A\} \le F\{A^\delta\} + \delta$, for all $A \in \mathscr{B}\}$, where A^δ is the closed δ-neighbourhood of A. The Lévy distance defined on the real line is obtained by restricting sets A to $(-\infty, x]$, $x \in E$.

CONDITIONS A.

A_0: $T[\psi, \rho, F_\theta] = \theta$,

A_1: 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,

A_2: $\{\psi(x, \tau) \mid \tau \in D\}$, $\{(\partial/\partial\tau)\psi(x, \tau) \mid \tau \in D\}$ are bounded above in Euclidean norm ($\|A\| = \{\text{trace}(A'A)\}^{1/2}$) by some function g that is integrable with respect to all $G \in n(\varepsilon, F_\theta)$ for some $\varepsilon > 0$,

A_3: The matrix

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

is nonsingular,

A_4: Given $\delta > 0$ there exists an $\varepsilon > 0$ such that for all $G \in n(\varepsilon, F_\theta)$

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

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

REMARK 2.1. Fisher consistency presumes A_0 for all $\theta \in \Theta$.

REMARK 2.2. A consequence of assumption A_1 is that families of vector functions $\{\psi(x, \tau) \mid \tau \in D\}$ and matrix functions $\{(\partial/\partial\tau)\psi(x, \tau) \mid \tau \in D\}$ exist and are equicontinuous on R. See Graves (1946, page 20, Theorem 23).

REMARK 2.3. A_3 is often stated as an assumption of positive definiteness. The study of infinitesimal behaviour requires only that $M(\theta)$ be nonsingular.

To consider the notion of Fréchet derivative in a useful manner for statistical functionals, restrictions must be put on the domain of the functional and therefore also on the derivative. Let the linear space spanned by differences $F - G$ of members of $\mathscr{G}$ be denoted by $\mathscr{D}$. The real vector functional T is defined on $\mathscr{G}$ and d is a metric on $\mathscr{G}$. The statistical functional is said to be Fréchet differentiable at $G \in \mathscr{G}$ with respect to the pair $(\mathscr{G}, d)$ when it can be approximated by a linear functional $T'_G(\cdot)$ defined on $\mathscr{D}$, such that

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

as $d(F, G) \to 0$, $F \in \mathscr{G}$.

3. Uniqueness of functional solutions to equations.

LEMMA 3.1. *Let conditions* A *hold for some* ψ, ρ. *Then there is a* $\delta_1 > 0$ *and an* $\varepsilon_1 > 0$ *such that for every* $\tau \in U_{\delta_1}(\theta)$ *the open ball of radius* δ_1 *and center* θ, *and every* $G \in n(\varepsilon_1, F_\theta)$ *the matrix*

$$M(\tau, G) = \int_R \frac{\partial}{\partial\tau}\psi(x, \tau)\, dG(x)$$

is nonsingular.

PROOF. By continuity of a determinant as a function of the elements of a matrix choose $\eta > 0$ such that $\|A - M(\theta)\| < \eta$ implies $|\det\{A\}| > \frac{1}{2}|\det\{M(\theta)\}|$ for an $r \times r$ matrix A. Assumptions A_1, A_2 imply $M(\tau, F_\theta)$ is continuous in $\tau \in D$. So choose $\delta_1 > 0$ such that $\tau \in U_{\delta_1}(\theta) \subset D$ implies $\|M(\tau, F_\theta) - M(\theta)\| < \eta/2$. By A_4 let $\varepsilon_1 > 0$ be so that $G \in n(\varepsilon_1, F_\theta)$ implies $\|M(\tau, G) - M(\tau, F_\theta)\| < \eta/2$. The lemma is proved by the triangle inequality of norms.

Theorem 3.1. (*Inverse function theorem*). *Suppose f is a mapping from Θ into E^r, the partial derivatives of f exist and are continuous on Θ, and the matrix of derivatives $f'(\theta^*)$ has inverse $f'(\theta^*)^{-1}$ for some $\theta^* \in \Theta$. Write $\lambda = 1/(4\,\| f'(\theta^*)^{-1}\|)$. Use the continuity of the elements of $f'(\tau)$ to fix a neighbourhood $\cup_\delta(\theta^*)$ of sufficiently small radius $\delta > 0$ to ensure that $\| f'(\tau) - f'(\theta^*)\| < 2\lambda$, whenever $\tau \in \cup_\delta(\theta^*)$. Then*

(a) *for every $\tau_1, \tau_2 \in \cup_\delta(\theta^*)$*

$$\| f(\tau_1) - f(\tau_2)\| \geq 2\lambda \,\| \tau_1 - \tau_2 \|;$$

and

(b) *the image set $f(\cup_\delta(\theta^*))$ contains the open neighbourhood with radius $\lambda\delta$ about $f(\theta^*)$.*

Conclusion (a) ensures that f is one-to-one on $\cup_\delta(\theta^*)$ and that f^{-1} is well defined on the image set $f(\cup_\delta(\theta^*))$.

Remark 3.1. The $\| A \|$ can also be interpreted as the least upper bound of all numbers $\| Ay \|$ where y ranges over all vectors in E^r with $\| y \| \leq 1$; c.f. Foutz (1977).

In what follows the selection functional $\rho(G, \tau) = \| \tau - \theta \|$ is employed. It is emphasized that $T[\psi, \rho, G]$ is then an auxiliary functional used to discover the properties of $T[\psi, \rho_0, G]$ for suitably chosen ρ_0, although its properties are immediately applicable if only one solution to the equations exists so that $T[\psi, \rho, \cdot] = T[\psi, \cdot]$.

Theorem 3.2. *Let $\rho(G, \tau) = \| \tau - \theta \|$ and suppose conditions A hold. Then given $\kappa > 0$ there exists an $\varepsilon > 0$ such that $G \in n(\varepsilon, F_\theta)$ implies $T[\psi, \rho, G]$ exists and is an element of $\cup_\kappa(\theta)$. Further for this ε there is a $\kappa^* > 0$ such that*

$$I(\psi, G) \cap \cup_{\kappa^*}(\theta) = T[\psi, \rho, G], \tag{3.1}$$

and $M(\tau, G)$ is nonsingular for $\tau \in \cup_{\kappa^*}(\theta)$. For any null sequence of positive numbers $\{\varepsilon_k\}$ let $\{G_k\}$ be an arbitrary sequence for which $G_k \in n(\varepsilon_k, F_\theta)$. Then

$$\lim_{k\to\infty} T[\psi, \rho, G_k] = T[\psi, \rho, F_\theta] = \theta. \tag{3.2}$$

Remark 3.2. Theorem 3.2 demonstrates the uniqueness of a solution of equations (1.2) in a region $\cup_{\kappa^*}(\theta)$ by (3.1) and the continuity of the functional by (3.2).

Proof of Theorem 3.2. Write $\lambda = 1/(4\,\| M(\theta)^{-1} \|)$. By continuity of $M(\tau, F_\theta)$ in τ choose $0 < \kappa^* < \min(\delta_1, \kappa)$ such that $\tau \in \cup_{\kappa^*}(\theta)$ implies $\| M(\tau, F_\theta) - M(\theta) \| < \lambda/2$. Here δ_1, ε_1 are given by Lemma 3.1. For $G \in n(\varepsilon_1, F_\theta)$ define $\lambda(G) = 1/(4\,\| M(\theta, G)^{-1} \|)$. Choose $0 < \varepsilon^* \leq \varepsilon_1$ so that

$$\begin{aligned} \| M(\tau, G) - M(\theta, G) \| &\leq \| M(\tau, G) - M(\tau, F_\theta) \| + \| M(\theta, G) - M(\theta) \| \\ &\quad + \| M(\tau, F_\theta) - M(\theta) \| \\ &\leq \lambda < 2\lambda(G) \quad \text{whenever} \quad G \in n(\varepsilon^*, F_\theta) \end{aligned}$$

for all $\tau \in \cup_{\kappa^*}(\theta)$. Note A_1, A_2 imply $K_G(\tau)$ has continuous partial derivatives $M(\tau, G)$. Properties (a) and (b) ensure $K_G(\cdot)$ is a one-to-one function from $\cup_{\kappa^*}(\theta)$ onto $K_G(\cup_{\kappa^*}(\theta))$ and that the image set contains the open ball of radius $\lambda\kappa^*/2$ about $K_G(\theta)$. Now choose $0 < \varepsilon' \leq \varepsilon^*$ such that

$$| K_G(\theta) - 0 | < \lambda\kappa^*/2.$$

Then it is clear that $0 \in K_G(\cup_{\kappa^*}(\theta))$ for all $G \in n(\varepsilon^*, F_\theta)$ and that the image set contains the open ball of radius $\lambda\kappa^*/2$ about $K_G(\theta)$. Consider the inverse function.

$$K_G^{-1}\colon K_G(\cup_{\kappa^*}(\theta)) \to \cup_{\kappa^*}(\theta) \quad \text{for} \quad G \in n(\varepsilon', F_\theta).$$

It is well defined whenever $K_G(\tau)$ is one-to-one. Since $0 \in K_G(\cup_{\kappa^*}(\theta))$ for $G \in n(\varepsilon', F_\theta)$, we conclude that with $\varepsilon' = \varepsilon$ there exists a unique root of equations (1.2) in $\cup_{\kappa^*}(\theta)$ whenever $G \in n(\varepsilon, F_\theta)$. That is, (3.1) holds.

If we let $\{\kappa_i^*\}_{i=1}^\infty$ be a null sequence for which $0 < \kappa_i^* \le \kappa^*$, $i \ge 1$, there exists a corresponding sequence of $\{\varepsilon_i'\}$. Since $\{\varepsilon_n\}$ is null there is some $j(i)$ for which $\varepsilon_{j(i)} \le \varepsilon_i'$, whence $G_{j(i)} \in n(\varepsilon_i', F_\theta)$, $i \ge 1$. Hence

$$T[\psi, \rho, G_k] = K_{G_k}^{-1}(o) \cap \cup_{\kappa^*}(\theta), \quad k > \inf_{1 \le i < \infty} j(i),$$

and

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

4. Continuity of the functional. We may now separate the argument for continuity of the functional $T[\psi, \rho_0, \cdot]$ into local and global parts as follows:

THEOREM 4.1. *Assume conditions A hold for the functional $\rho(G, \tau) = \| \tau - \theta \|$. Suppose $\rho_0(G, \tau)$ is a selection functional such that for every neighbourhood N of θ*

$$\inf_{\tau \notin N} \rho_0(F_\theta, \tau) - \rho_0(F_\theta, \theta) > 0, \tag{4.1}$$

and for every $\eta > 0$ there exists an $\varepsilon > 0$ such that $G \in n(\varepsilon, F_\theta)$ implies $\rho_0(G, \tau)$ is continuous in $\tau \in \Theta$ and satisfies

$$\sup_{\tau \in \Theta} | \rho_0(G, \tau) - \rho_0(F_\theta, \tau) | < n. \tag{4.2}$$

Then for every $\kappa > 0$ there exists an ε_0 such that $G \in n(\varepsilon_0, F_\theta)$ implies $T[\psi, \rho_0, G]$ exists, is unique, and lies in $\cup_\kappa(\theta)$.

PROOF. From Theorem 3.2 there is $0 < \kappa^* < \kappa$ and $\varepsilon > 0$ such that $G \in n(\varepsilon, F_\theta)$ implies (3.1) holds. Denote

$$\delta(\kappa^*) = \inf\{\rho_0(F_\theta, \tau) - \rho_0(F_\theta, \theta) \mid \tau \in \Theta - \cup_{\kappa^*}(\theta)\}. \tag{4.3}$$

Choose $0 < \kappa' < \kappa^*$ so that $\tau \in \cup_{\kappa'}(\theta)$ implies

$$| \rho_0(F_\theta, \tau) - \rho_0(F_\theta, \theta) | < \delta(\kappa^*)/2.$$

For $\kappa' > 0$ choose $0 < \varepsilon_0 \le \varepsilon$ so that $G \in n(\varepsilon_0, F_\theta)$ implies $T[\psi, \rho, G] \in \cup_{\kappa'}(\theta)$ and

$$\sup_{\tau \in \Theta} | \rho(G, \tau) - \rho(F_\theta, \tau) | < \delta(\kappa^*)/4.$$

Note that (3.1) remains true for $G \in n(\varepsilon_0, F_\theta)$. Then

$$\begin{aligned}
\rho_0(G, T[\psi, \rho, G]) &< \rho_0(F_\theta, T[\psi, \rho, G]) + \delta(\kappa^*)/4 \\
&< \rho_0(F_\theta, \theta) + 3\delta(\kappa^*)/4 \\
&< \rho_0(F_\theta, \tau) - \delta(\kappa^*)/4 \quad \text{uniformly in} \quad \tau \in \Theta - \cup_{\kappa^*}(\theta) \\
&< \rho_0(G, \tau) \qquad\qquad\quad \text{uniformly in} \quad \tau \in \Theta - \cup_{\kappa^*}(\theta).
\end{aligned}$$

Hence

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

$$T[\psi, \rho_0, G] = T[\psi, \rho, G] \in \cup_{\kappa'}(\theta) \subset \cup_\kappa(\theta).$$

REMARK 4.1. The assumption of (4.1) is standard to minimum distance theory.

REMARK 4.2. The assumption (4.2) is a sufficient condition for this continuity result. It may not be necessary. Consider a Fréchet space R and set $f(G)$ to be $\int_R \| x \| \, dG(x)$ if this integral is finite. Otherwise put $f(G) = 0$. Clearly, using the selection functional $\rho_0(G, \tau)$ of Theorem 4.1 is equivalent to using $\rho_1(G, \tau) = \rho_0(G, \tau) + f(G)$, but the latter need not satisfy (4.2).

REMARK 4.3. Weak continuity of the functional follows whenever the neighbourhoods are defined either by d_L or d_p.

5. Fréchet differentiability.

THEOREM 5.1. *Let $\rho(G, \tau) = \| \tau - \theta \|$ and assume conditions A hold with respect to this functional and neighbourhoods generated by the metric d on $\mathscr{G}$. Suppose for all $G \in \mathscr{G}$*

$$\int_R \psi(x, \theta)\, d(G - F_\theta)(x) = O(d(G, F_\theta)). \tag{5.1}$$

Then $T[\psi, \rho, \cdot]$ is Fréchet differentiable at F_θ with respect to $(\mathscr{G}, d)$, and has derivative

$$T'_{F_\theta}(G - F_\theta) = -M(\theta)^{-1} \int_R \psi(x, \theta)\, d(G - F_\theta)(x).$$

PROOF. Abbreviate $T[\psi, \rho, \cdot] = T[\cdot]$ and let κ^*, ε be given by Theorem 3.2. Let $\{\varepsilon_k\}$ be so that $\varepsilon_k \downarrow 0^+$ as $k \to \infty$ and let $\{G_k\}$ be any sequence such that $G_k \in n(\varepsilon_k, F_\theta)$. Note that $n(\varepsilon_k, F_\theta)$ is the set of distributions within distance ε_k of F_θ. It is sufficient to show

$$\| T[G_k] - T[F_\theta] - T'_{F_\theta}(G_k - F_\theta) \| = o(\varepsilon_k).$$

By Theorem 3.2, $T[G_k]$ exists and is unique in $\cup_{\kappa^*}(\theta)$ for $k > k_0$ where $\varepsilon_{k_0} \le \varepsilon$. By A_4 see that

$$\| M(\tau, G_k) - M(\tau, F_\theta) \| \to_{k\to\infty} 0 \quad \text{uniformly in} \quad \tau \in D. \tag{5.2}$$

Consider the two term Taylor expansion,

$$0 = K_{G_k}(T[G_k]) = K_{G_k}(\theta) + M(\tilde{\tau}_k, G_k)(T[G_k] - \theta), \tag{5.3}$$

where $\| \tilde{\tau}_k - \theta \| \le \| T[G_k] - \theta \|$ which tends to zero as $k \to \infty$ by Theorem 3.2, and $\tilde{\tau}_\kappa$ is evaluated at different points for each component function expansion (i.e. takes different values in each row of M). See from (5.3) that

$$\| T[G_k] - \theta \| = O(K_{G_k}(\theta)) = O(\varepsilon_k).$$

Also,

$$T[G_k] - \theta = -M(\theta)^{-1}K_{G_k}(\theta) + M(\theta)^{-1}\{M(\tilde{\tau}_k, G_k) - M(\theta)\}(T[G_k] - \theta). \tag{5.4}$$

By continuity of $M(\tau, F_\theta)$ in τ and (5.2),

$$\| M(\tilde{\tau}_k, G_k) - M(\theta) \| = o(1).$$

So,

$$\| T[G_k] - \theta - T'_{F_\theta}(G_k - F_\theta) \| = o(1)O(d(G_k, F_\theta)) = o(\varepsilon_k).$$

COROLLARY 5.1. *If conditions of Theorem 5.1 hold and ρ_0 satisfies conditions of Theorem 4.1, then the functional $T[\psi, \rho_0, \cdot]$ is Fréchet differentiable at F_θ with respect to $(\mathscr{G}, d)$.*

Apart from the implications of Fréchet differentiability of the multivariate functional T with respect to the metrics d_L, d_p to robustness of the functional T is the consequence of asymptotic normality of the statistic $\sqrt{n}(T[F_n] - T[F_\theta])$ that follows from differentiability with respect to d_k. This is shown in the same manner as Lemma 1.1 of Boos and Serfling (1980) whereupon the asymptotic variance, assumed finite, is given by

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

where the integration is carried out componentwise.

6. Meeting the conditions. Conditions A can be met for large classes of ψ-functions when neighbourhoods are generated by distance metrics d_L, d_k, and d_p.

LEMMA 6.1. *Let $A \in \mathscr{B}$ be a continuity set of F_θ and $\kappa > 0$ be given. Then there exists an $\varepsilon > 0$ such that for any $G \in \mathscr{G}$,*

$$d_p(G, F_\theta) < \varepsilon \quad \textit{implies} \quad |G\{A\} - F_\theta\{A\}| < \kappa.$$

PROOF. Since A is a continuity set there is some $\varepsilon_1 > 0$ for which $F_\theta\{A^{\varepsilon_1}\} < F_\theta\{A\} + \kappa/2$. Choose $\varepsilon_1 < \kappa/2$. Then if $d_p(G, F_\theta) < \varepsilon_1$,

$$G\{A\} < F_\theta\{A^{\varepsilon_1}\} + \varepsilon_1 < F_\theta\{A\} + \kappa.$$

Similarly, since $R - A$ shares the same boundary of A, there is some ε_2 for which

$$F_\theta\{A^{-\varepsilon_2}\} = F_\theta\{R - (R - A)^{\varepsilon_2}\} > F_\theta\{A\} - \kappa/2.$$

Choose $\varepsilon_2 < \kappa/2$ and suppose $d_p(G, F_\theta) < \varepsilon_2$. Then

$$G\{A\} > F_\theta\{A^{-\varepsilon_2}\} - \varepsilon_2 > F_\theta\{A\} - \kappa.$$

The lemma follows by setting $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$.

THEOREM 6.1. *Let $\mathscr{A}$ be a class of continuous functions on separable metrizable space R possessing the following two properties: (1) $\mathscr{A}$ is uniformly bounded, that is, there exists a constant H such that $|f(x)| \le H < \infty$ for all $f \in \mathscr{A}$ and $x \in R$; and (2) $\mathscr{A}$ is equicontinuous. Let $F_\theta \in \mathscr{G}$ be given. Then,*

for every $\delta > 0$ there is an $\varepsilon > 0$ such that

$$d_p(F_\theta, G) \le \varepsilon \quad \textit{implies} \quad \sup_{f\in\mathscr{A}} \left| \int_R f \, dG - \int_R f \, dF_\theta \right| < \delta. \tag{6.1}$$

PROOF. Since R is separable and complete there exists a compact set C such that $F_\theta\{R - C\} < \delta/(16.H)$, which further can be chosen to be a continuity set of F_θ. For arbitrary $\eta > 0$, by Lemma 3.1 of Rao (1962) there exists a finite number of sets $\{A_j\}_{j=1}^n$, where $n = n(\eta)$, such that (a) $\cup_{j=1}^n A_j = C$; (b) $A_j \cap A_{j'} = \varnothing$ for $j \ne j'$; (c) for each j, A_j is a continuity set for F_θ; (d) for any $x, y \in A_j$ and $f \in \mathscr{A}$, $|f(x) - f(y)| < \eta$, for each $j = 1, \cdots, n$. Let $\eta = \delta/4$ and choose $\{y_j\}_{j=1}^n$ arbitrarily in $\{A_j\}_{j=1}^n$ respectively and let F_θ^* be the possibly improper measure attributing weight $F_\theta\{A_j\}$ to the point y_j, for each $j = 1, \cdots, n$. Then for each $f \in \mathscr{A}$

$$\left| \int_C f \, dF_\theta - \int_C f \, dF_\theta^* \right| \le \sum_{j=1}^n \int_{A_j} |f(x) - f(y_j)| \, dF_\theta < \delta/4.$$

Hence,

$$\sup_{f\in\mathscr{A}} \left| \int_C f \, dF_\theta - \int_C f \, dF_\theta^* \right| \le \delta/4.$$

Similarly, given $G \in \mathscr{G}$ let G^* be that measure attributing weight $G\{A_j\}$ to y_j for each $j =$

$1, \cdots, n$. Then,

$$\sup_{f\in\mathscr{A}} \left| \int_C f\,dG - \int_C f\,dG^* \right| < \delta/4.$$

Now,

$$\left| \int_C f\,dF_\theta^* - \int_C f\,dG^* \right| \leq H \textstyle\sum_{j=1}^n |F_\theta\{A_j\} - G\{A_j\}|.$$

By Lemma 6.1 choose ε_j such that $G \in \mathscr{G}$, $d_p(G, F_\theta) < \delta_j$ implies $|F_\theta\{A_j\} - G\{A_j\}| < \delta/(4.n.H)$. Let ε_0 be so that $G \in \mathscr{G}$, $d_p(G, F_\theta) < \varepsilon_0$ implies $|F_\theta\{R - C\} - G\{R - C\}| < \delta/(16.H)$. Set $\varepsilon = \min(\varepsilon_0, \varepsilon_1, \cdots, \varepsilon_n)$. Then $G \in \mathscr{G}$, $d_p(G, F_\theta) < \varepsilon$ implies

$$\begin{aligned}\sup_{f\in\mathscr{A}} \left| \int_R f\,dF_\theta - \int_R f\,dG \right| &\leq \sup_{f\in\mathscr{A}} \left| \int_{R-C} f\,dF_\theta - \int_{R-C} f\,dG \right| \\ &\quad + \sup_{f\in\mathscr{A}} \left| \int_C f\,dF_\theta - \int_C f\,dF_\theta^* \right| + \sup_{f\in\mathscr{A}} \left| \int_C f\,dG - \int_C f\,dG^* \right| \\ &\quad + \sup_{f\in\mathscr{A}} \left| \int_C f\,dF_\theta^* - \int_C f\,dG^* \right| \\ &< H.[F_\theta\{R - C\} + G\{R - C\}] + \tfrac14\delta + \tfrac14\delta + \tfrac14\delta < \delta.\end{aligned}$$

REMARK 6.1. When R is the real line and the decomposition of $C = [-c, c]$, say, is of the form $-c = a_0 < a_1 < \cdots < a_n = c$, for continuity points $\{a_i\}$ of F_θ, Theorem 6.1 holds for sets $A_i = (a_{i-1}, a_i]$ $i \geq 1$ and $\{a_0\}$, and d_p can be replaced by either d_k, d_L. The proof is the same.

REMARK 6.2. See from Remark 2.2 and Theorem 6.1 that for neighbourhoods generated by d_p, d_k, d_L, assumptions A_1, A_2 imply A_4 whenever g of A_2 can be chosen to be uniformly bounded.

In a sense the class $\mathscr{A}$ of Theorem 6.1 is the most general. A weaker condition than (1) is to assume $\sup_{f\in\mathscr{A}} \int_R |f|\,dF_\theta = m < +\infty$ but allow $\mathscr{A}$ to be unbounded. By choosing $\{f_n\} \subset \mathscr{A}$, $\{y_n\}$ so that $|f_n(y_n)| \to +\infty$ as $n \to \infty$, consider for any $\varepsilon > 0$, $G_n = (1 - \varepsilon)F_\theta + \varepsilon\delta_{y_n}$

$$\begin{aligned}\sup_{f\in\mathscr{A}} \left| \int f\,dF_\theta - \int f\,dG_n \right| &> \varepsilon \left| \int_R f_n\,dF_\theta - f_n(y_n) \right| \\ &> \varepsilon\,(|f_n(y_n)| - m) \to +\infty \quad \text{as} \quad n \to \infty.\end{aligned}$$

This violates (6.1) since $d_p(F_\theta, G_n) \leq \varepsilon$. If (2) does not hold, there is a $\delta > 0$ and $x \in R$ and a sequence $\{y_n\}$, $y_n \to x$ as $n \to \infty$, so that $\sup_{f\in\mathscr{A}} |f(x) - f(y_n)| > \delta$. Suppose at θ, $F_\theta = \delta_x$. Then $d_p(F_\theta, G_n) \to 0$ as $n \to \infty$ but $\sup_{f\in\mathscr{A}} |\int_R f\,dF_\theta - \int_R f\,dG_n| > \delta$, contradicting (6.1).

Finally condition (5.1) that is required for Fréchet differentiability at F_θ is also very plausible. Suppose $\psi(x, \theta)$ is a function of total bounded variation in the observation space variable and for all $G \in \mathscr{G}$ integration by parts,

$$\int_R \psi(x, \theta)\,d(G - F_\theta)(x) = -\textstyle\int_R (G - F_\theta)(x)\,d\psi(x, \theta)$$

is valid. Clearly (5.1) then holds for d_k.

Relationships between the metrics can be used to show Fréchet differentiability. Since $d_L \leq d_\kappa$ and $d_L \leq d_p$, differentiability with respect to the Lévy metric implies differentiability with respect to Kolmogorov and Prokhorov metrics. No relationship exists generally between d_κ and d_p but in the event of an absolutely continuous distribution F_θ that has a density bounded by a constant c so that $\sup_{x\in E} F_\theta(x+\delta) - F_\theta(x) < c\delta$ uniformly in $\delta > 0$, then $F_\theta(x) \le G(x+\delta) + \delta$ and $G(x) \le F_\theta(x+\delta) + \delta$ uniformly in $x \in E$ implies $\sup_{x\in E} |G(x) - F_\theta(x)| < (c+1)\delta$. Hence

$$d_k(G, F_\theta) \le (c+1)d_L(G, F_\theta) \le (c+1)d_p(G, F_\theta) \tag{6.2}$$

and Fréchet differentiability with respect to d_k implies that with respect to d_L and d_p also. This is important as the latter two metrics are difficult to work with.

7. Examples and conclusion. Fréchet differentiability of the mle in multiparameter models of the multinomial distribution is possible with Corollary 5.1, extending results of Rao (1957). Many minimum distance estimators are known to have estimating equations that can be formulated as in (1.1) with smooth bounded functions ψ. Two examples include the Cramér-Von Mises distance with weight function Lebesgue measure in estimation of location and scale of a normal distribution described in Heathcote and Silvapulle (1981) and the integrated squared error distance of Heathcote (1977). Theorem 5.1 proves existence of a Fréchet differentiable root at F_θ, indicating robustness.

M-estimators of location are frequently defined as the root closest to the median. Robustness of this selection functional follows from the next lemma.

LEMMA 7.1. *Assume F to be an absolutely continuous distribution with support on E. Let neighbourhoods be defined by one of the metrics d_κ, d_L,* or *d_p and let t be fixed. Then for all $\varepsilon > 0$ there exists a neighborhood n_δ for which*

$$\sup_{G\in n_\delta} |F^{-1}(t) - G^{-1}(t)| \le \varepsilon \quad \textit{where} \quad G^{-1}(t) = \inf\{y \mid G(y) \ge t\}.$$

PROOF. Define $a = F(F^{-1}(t) - \frac{1}{2}\varepsilon) < t$, $b = F(F^{-1}(t) + \frac{1}{2}\varepsilon) > t$. Consider neighbourhoods defined by d_κ. Let $\delta = \frac{1}{4}\min(b-t, t-a)$. Then

$$\sup_{G\in n_\delta} |G(F^{-1}(t) - \tfrac{1}{2}\varepsilon) - a| < \tfrac{1}{4}(t-a) \Rightarrow G(F^{-1}(t) - \tfrac{1}{2}\varepsilon) < t;$$

$$\sup_{G\in n_\delta} |G(F^{-1}(t) + \tfrac{1}{2}\varepsilon) - b| < \tfrac{1}{4}(t-b) \Rightarrow G(F^{-1}(t) + \tfrac{1}{2}\varepsilon) > t.$$

Hence $F^{-1}(t) - \frac{1}{2}\varepsilon < G^{-1}(t) \le F^{-1}(t) + \varepsilon/2$ which proves the result for d_k. Since F is absolutely continuous, the result holds for d_L, d_p by (6.2).

With the selection functional $\rho_0(G, \tau) = |G^{-1}(\frac{1}{2}) - \tau|$ the M-estimator of location defined by $\psi(x, \tau) = (x-\tau)\{1 - (x-\tau)^2/c^2\}$ for $|x - \tau| \le c$ and zero otherwise (Beaton and Tukey, 1974) is Fréchet differentiable at the normal distribution for metrics d_k, d_L, d_p. Here ρ_0 satisfies (4.1), (4.2). Estimating scale of an absolutely continuous distribution F, the equivalent selection functional is then $\rho_0(G, \tau) = |\{G^{-1}(\frac{3}{4}) - G^{-1}(\frac{1}{4})\} \times \{F^{-1}(\frac{3}{4}) - F^{-1}(\frac{1}{4})\}^{-1} - \tau|$. For location and scale parameters $\tau = (\tau_1, \tau_2)$ estimated simultaneously, a useful selection functional to distinguish the robust estimator from multiple roots of equations (1.1) is then

$$\rho_0(G; \tau_1, \tau_2) = (G^{-1}(\tfrac{1}{2}) - \tau_1)^2 + [\{G^{-1}(\tfrac{3}{4}) - G^{-1}(\tfrac{1}{4})\}\{F^{-1}(\tfrac{3}{4}) - F^{-1}(\tfrac{1}{4})\}^{-1} - \tau_2]^2.$$

The existence of multiple roots in the estimating equation, that gives also the solution maximizing the likelihood, has given rise to much discussion in the particular context of the location parameter of a Cauchy distribution with known scale, the density of which is $\sigma[\pi\{\sigma^2 + (x-\mu)^2\}]^{-1}$. Several papers referred to by Johnson and Kotz (1970, pages 164–165) are mainly concerned about methods for attaining a solution. As the efficient score function in this instance is uniformly bounded and continuous, it follows from the above discussion that the functional defined as the root closest to the median is Fréchet differentiable. Moreover, the functional estimator is asymptotically efficient. While the functional is not known invariably to coincide with the solution maximizing the likelihood for all samples, it can be expected in virtue of the almost uniform high relative efficiency of the median relative to the mle, as noted by Barnett (1966, page 163), that this estimator will be efficient in small samples as well as robust. Moreover the search for a root is simplified to searching for the root nearest the median. Asymptotically, consistency of the median deems that the two estimators coincide with increasing probability as sample size increases.

When both location and scale parameters are unknown, the mle is a solution of (1.1) with

$$\psi_1(x; \tau) = -\frac{x - \tau_1}{\tau_2^2 + (x - \tau_1)^2}, \quad \psi_2(x, \tau) = 1 - \frac{2(x - \tau_1)^2}{\tau_2^2 + (x - \tau_1)^2}.$$

Restricting the parameter space $\Theta = \{-\infty < \tau_1 < \infty, \eta < \tau_2 < \infty\}$ for some small positive η ensures ψ and its partial derivatives are uniformly bounded. Conditions A_1, A_2, A_3 hold. By Remark 6.2, A_4 holds. Then by Theorem 5.1 there exists a Fréchet differentiable root at the Cauchy distribution. It is noted by Copas (1975) that the Cauchy joint likelihood for both location and scale parameters is unimodal with only one stationary point. Hence the mle corresponds to the Fréchet differentiable functional evaluated at F_n and is therefore both robust and asymptotically efficient in the sense of Kallianpur and Rao (1955).

Acknowledgement. The author wishes to thank an editor for a suggestion leading to a more complete discussion in Section 7.

REFERENCES

BARNETT, V. D. (1966). Evaluation of the maximum likelihood estimator where the likelihood equation has multiple roots. *Biometrika* **53** 151–165.

BEATON, A. E. and TUKEY, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* **16** 147–185.

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.

COPAS, J. B. (1975). On the unimodality of the likelihood for the Cauchy distribution. *Biometrika* **62** 701–704.

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

GRAVES, L. M. (1946). *Theory of Functions of Real Variables*. McGraw-Hill, New York.

HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. thesis, University of California, Berkeley.

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896.

HEATHCOTE, C. R. (1977). The integrated squared error estimation of parameters. *Biometrika* **64** 255–264.

HEATHCOTE, C. R. and SILVAPULLE, M. J. (1981). Minimum mean squared estimation of location and scale parameters under misspecification of the model. *Biometrika* **68** 501–514.

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

JOHNSON, N. L. and KOTZ, S. (1970). *Continuous Univariate Distributions*, Vol 1. Houghton Mifflin, Boston.

KALLIANPUR, G. (1963). Von Mises functionals and maximum likelihood estimation. *Sankhyā, Ser. A* **23** 149–158.

KALLIANPUR, G. and RAO, C. R. (1955). On Fisher's lower bound to asymptotic variance of a consistent estimate. *Sankhyā* **15** 331–342.

RAO, C. R. (1957). Maximum likelihood estimation for the multinomial distribution. *Sankhyā* **18** 139–148.

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

VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348.

FACHGRUPPE FÜR STATISTIK
EIDGENÖSSISCHE TECHNISCHE HOCHSCHULE
CH-8092 ZÜRICH
SWITZERLAND