PINSA, **66, A**, No. 5, September 2000, pp. 467-482.
© Printed in India.

A REVIEW OF DIFFERENTIABILITY IN RELATION TO ROBUSTNESS WITH AN APPLICATION TO SEISMIC DATA ANALYSIS

BRENTON R CLARKE

Mathematics and Statistics, Division of Science and Engineering, Murdoch University, Murdoch, Western Australia, 6150 (Australia)

(Received 12 April 1999; Accepted 21 October 1999)

The introduction of the Fréchet derivative into statistics from the mathematical literature has sporned a wealth of study of the relationships between robust statistics and mathematical derivatives. The current view is that differentiability of the Fréchet type with respect to the supremum norm is a valid approach and necessarily some well known statistics such as the arithmetic mean and the median do not have the attribute of differentiability in this sense. As an illustration of a Fréchet differentiable estimating functional an L_2 minimum distance estimator is applied to some data collected from seismic analysis and this provides estimates in an area where estimation previously was *ad hoc* and estimation was not always possible. The theoretical estimation techniques known from the study of mixtures of normal distributions are applied to this data and illustrate some of the features of using robust techniques as opposed to always resorting to the maximum likelihood estimation methods.

Key Words: Fréchet Differentiability; Statistical Functional; Selection Functional; M-Estimators; Minimum Distance Estimation; Seismic Data Analysis; Maximum Likelihood Estimator

Introduction

In this paper we review the link between differentiability of statistical functionals (representing estimators) and robustness of the statistics they represent. A chronological table is given in the appendix for the statistical works on the three kinds of derivative Gâteaux, compact or Hadamard derivative, and Fréchet derivative. In the statistics literature the two areas of differentiability and robustness have grown side by side, and it is thus pertinent to see where they lie in relation to each other at this point in time. The argument for robust Fréchet differentiable functionals with respect to the supremum norm as argued for or intimated in related works by Boos[1], Boos and Serfling[2], Clarke[3], Bednarski *et al.*[4], Bednarski[5], Davies[6], Bednarski and Clarke[7], Clarke *et al.*[8] is accompanied by a graphic example from seismic data analysis taken from the author's unpublished 1980 PhD thesis at the Australian National University[9]. Here robust estimates, which are not maximum likelihood estimates, are implemented on 'rough' data, where models derived from a spherical earth, core, and inner core, may not be wholely correct in describing the data, but at least they offer a model to work with in a parametric setup.

It was the Indian statisticians Kallianpur and Rao[10] who introduced the idea of Fréchet differentiable functionals with respect to the supremum norm. Here T is a functional defined on the space of distribution functions $\mathcal{G}$ that are defined on an observation space R. The observation space can be taken to be the real line E or k-dimensional space E^k for example. The range of the functional is usually contained in $\Theta \cup \{\infty\}$ where $\Theta \subset E^r$.

Consider the distance between two distributions G, F as $\|G-F\|_\infty = \sup_{-\infty<x<\infty}|G(x)-F(x)|$ and let the linear space spanned by the differences G-F of members of $\mathcal{G}$ be denoted by $\mathcal{D}$. The statistical functional T is said to be Fréchet differentiable at $F \in \mathcal{G}$ with respect to $(\mathcal{G}, \|\cdot\|_\infty)$ when it can be approximated by a linear functional $T_F'(.)$ defined on $\mathcal{D}$, such that

$$\left|T[G]-T[F]-T_F'(G-F)\right| = o(\|G-F\|_\infty) \quad \ldots (1)$$

as $\|G-F\|_\infty \to 0, G \in \mathcal{G}$. The Fréchet derivative

goes back to Fréchet[11] who defined the derivative on Banach spaces, viz. imagine the domain of T extending to the space which is the linear span of all signed measures. Note that if X is the linear space generated by $(\mathcal{G}, \|\cdot\|_\infty)$ an equivalent statement is that for every bounded set B in X

$$\frac{\sup\{|T[F+tv]-T[F]-tT_F'(v)|:v\in B\}}{t}\to 0 \quad \ldots (2)$$

as $t\to 0$ where t is a real variable. If B is replaced by all compact sets for a topology $\mathcal{C}$ say, then T is said to be compactly or Hadamard differentiable at F[12]. Initially the arguments in Kallianpur and Rao[10] were to show that among Fréchet differentiable estimators at a parametric distribution $F_\theta \in \mathcal{F}=\{F_\tau|\tau\in\Theta\subset E^r\}$ the estimator or functional with the minimum variance was the maximum likelihood estimator (m.l.e.). Here the minimum asymptotic variance obtained was $n\mathcal{I}(\theta)\}^{-1}$, where $\mathcal{I}(\theta)$ is Fisher information. While Rao[13] was able to establish Fréchet differentiability of the m.l.e. of a multinomial distribution Kallianpur[14] reported that neither author could prove Fréchet differentiability of the m.l.e. for any absolutely continuous or infinitely discrete distribution. Kallianpur introduced a weaker Gâteaux derivative. It was made clear in Clarke[3] that the m.l.e. which is a solution of equations

$$\sum_{i=1}^{n}\frac{(\partial/\partial\tau)f(X_i;\tau)}{f(X_i,\tau)}=0, \quad \ldots (3)$$

where $X_1,\ldots,X_n$ are independent identically distributed random variables distributed as F_θ and $f(x,\tau)=F_\tau'(x)$, will only be Fréchet differentiable in the sense of (1) at F_θ with respect to the supremum norm (also called the Kolmogorov metric), if the efficient score function is bounded in the observation space variable. Thus, for instance the location and scale parametric family defined by a Cauchy distribution leads to a Fréchet differentiable m.l.e., but for many parametric families including the popular normal parametric family the efficient score function is unbounded and the m.l.e. is not Fréchet differentiable with respect to the supremum norm. The m.l.e. is only a special case of estimators defined as solutions of M-estimating equations

$$\sum_{i=1}^{n}\psi(X_i,\tau)=0, \quad \ldots (4)$$

where ψ may vary from being the efficient score function of the m.l.e. and since we can write eq. (4) as

$$\int\psi(x,\tau)dF_n(x)=0, \quad \ldots (5)$$

where $F_n(x)=\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}(x)$ where δ_{X_i} represents the distribution function attributing mass one at X_i, the estimator can be represented by the functional $T[\psi,F_n]$. For estimation purposes we require a Fisher consistency condition $T[\psi,F_\theta]=\theta$ for all $\theta\in\Theta$. It can be noted that if there is a set of solutions $I(\psi,F_n)$ of eq. (5) we then apply a selection functional[3] ρ_0 to select the consistent root via

$$\inf_{\tau\in I(\psi,F_n)}\rho_0(F_n,\tau)=\rho_0(F_n,T[F_n]) \quad \ldots (6)$$

For instance, if we were seeking the m.l.e. solution we would use $\rho_0(F_n,\tau)=-\int \ell n\, f(x,\tau)dF_n(x)$ with $\psi(x,\tau)=\{(\partial/\partial\tau)f(x,\tau)\}/f(x,\tau)$. (See also Clarke[15]).

Some well known estimators that can be easily represented as functionals are the mean functional

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=\int x\,dF_n(x)=T[F_n],$$

whereupon, more generally for $G\in\mathcal{G}$ the space of distributions on R we have

$$T[G]=\int x\,dG(x).$$

Another well known example is the median functional defined on E by $T[F_n]=F_n^{-1}(1/2)$, the more general median functional defined to be $T[G]=G^{-1}(1/2)=\inf\{y|G(y)\geq 1/2\}$ for $G\in\mathcal{G}$. The idea of writing statistics as functionals is generally attributed to Von Mises[16]. In that paper a calculus is given to describe asymptotic properties of $T[F_n]$ at $F_\theta\in\mathcal{F}$.

The mean and the median functionals are also natural examples of the nonrobust estimator and the robust estimator respectively. To envisage this consider one observation chosen randomly from the sample and let it travel to plus infinity. This could represent some contamination for example. It is clear that the contamination causes the arithmetic mean to travel to plus infinity, but the median will not vary farther than say the middle three or four observations of the original sample for example $(n \geq 5)$. The mean is the m.l.e. of location of the normal parametric family. It corresponds to an M-estimator with an unbounded psi-function $\psi(x,\tau) = x - \tau$, and is not Fréchet differentiable. While it satisfies a linear expansion of the form (1) at the distributions $G = F_n$ and $F = F_\theta$, more generally the mean is not even continuous at F with respect to the supremum norm. But neither is the median Fréchet differentiable with respect to the supremum norm. In the development of functional derivatives in the statistical literature, the compact or Hadamard derivative was introduced in the thesis by Reeds[17] and later in the monograph by Fernholz[18] and in the review paper by Gill[19] to cover the asymptotic theory of functionals such as the median and quantiles. It is interesting that the chronology of the derivatives in the statistical literature is Fréchet derivative first, then Gâteaux, then Hadamard, whereas in the mathematical literature it was Hadamard first, then Gâteaux and Fréchet, for Gâteaux and Fréchet were students of Hadamard. Excellent surveys of the whole field are given by Averbukh and Smolyanov[20]. A more recent review of the weaker Gâteaux derivative is given in Fabian[21].

The more general M-estimators $T[\psi, F_n]$ which are both continuous at F_θ and Fréchet differentiable at F_θ with respect to the supremum norm are necessarily defined by a continuous bounded ψ function[3]. Fréchet differentiability of a functional with respect to the supremum norm is an advantage in that this imples both compact differentiability and Gâteaux differentiability[19,22]. We say a functional T is Gâteaux differentiable at F if there is a linear functional T_F' such that for all $G \in \mathcal{G}$

$$\lim_{t \to 0} \frac{T[F_t] - T[F]}{t} = T_F'[G - F]$$

with $F_t = (1-t)F + tG$. The introduction of M-estimators in Huber[23] and variations of them, for instance as in Hampel's three part redescender given in Hampel[24] and the Tukey bisquare estimator of Beaton and Tukey[25], where incidentally a selection functional $\rho_0(F,\tau) = \left|F^{-1}(1/2) - \tau\right|$ is implemented selecting that root which is closest to the median, meant that there was a need to compare estimators. Moreover recognition that there were classes of estimators such as L-estimators and R-estimators as in Huber[26] meant that a uniform way of comparing estimators was sought. In a significant step to a comparison of all these 'popular' estimators Hampel[24,27] introduced the influence curve, later called the influence function in Hampel *et al.*[22]. The influence function is a special case of the Gâteaux derivative obtained by writing

$$IF(x,F,T) = T_F'[\delta_x - F]$$

Huber[22] points out that the Fréchet derivative when it exists is given by

$$T_F'[G - F] = \int_R IF(x,F,T)\, d(G - F)(x) \qquad \ldots (7)$$

Hampel[27] argues that IF(x,F,T) measures the infinitesimal influence of a contaminating observation at the point $x \in R$ on the functional T at F. For instance, the typical calculation of the influence function is via

$$IF(x,F,T) = \lim_{t \to 0} \frac{T[(1-t)F + t\delta_x] - T[F]}{t}.$$

This provides a path to compare different estimators through the quantity known as the 'gross error sensitivity'

$$\gamma^* = \sup_{x \in R} \left|IF(x,F,T)\right| \qquad \ldots (8)$$

Solutions of eq. (4) have influence functions at F_θ given by

$$IF(x,F_\theta,T) = -M(\theta)^{-1}\psi(x,\theta), \qquad \ldots (9)$$

where it is assumed that $M(\theta)$ is nonsingular and

$$M(\theta)=\int_R\{(\partial/\partial\theta)\psi(x,\theta)\}dF_\theta(x).$$

It is possible to define for a given upper bound $\gamma^*<\infty$ an optimal M-estimator achieving minimum variance among all Fréchet differentiable estimators subject to this bound. This is argued in Hampel *et al.*[22]. For example, the optimal M-estimator for location at the normal distribution takes the form $\psi(x,\tau)=\psi(x-\tau)$ with

$$\psi(x)=\min(k,|x|)sign(x) \qquad \ldots (10)$$

for some constant $0<k<\infty$ which coincides with Huber's[23] minimax proposal for M-estimation of location. Note that the m.l.e. at the normal model corresponds to a choice of $\psi(x)=x$ with $\gamma^*=\infty$ and so is excluded from consideration. In Huber[22,26] a stronger interpretation of γ^* is asserted, which assumes Fréchet differentiability with respect to $(\mathcal{G},\|\cdot\|_\infty)$. One can argue that for $G_t=(1-t)F+tH$ as $t\to 0$

$$\begin{aligned}&\sup_H|T[G_t]-T[F]|\\&=\sup_H\left|T_F'(G_t-F)\right|+o(t)\\&=\sup_H\left|\int IF(x,F,T)d(G_t-F)\right|+o(t)\\&\le\gamma^*t+o(t)\end{aligned}$$

In 1998 on page 1855 Davies[6] points out as t sufficiently small

$$\sup_H|T[G_t]-T[F]|\le 2\gamma^*t.$$

This gives essentially the asymptotic bias in 'ε'-contaminated neighbourhoods as being proportional to 'γ^*'. The fact the smaller this is chosen, the larger the 'optimal' variance is, mitigates us from letting $\gamma^*\to 0$. That is, there is a tradeoff between maximum asymptotic bias and optimal asymptotic variance. This is elegantly argued in Hampel[24], Hampel *et al.*[22]. The fact that for more general parametric families other than location γ^* depends also on θ and one also needs to be sure that one can obtain solutions to eqs. (4), (5) for finite samples has led researchers to seek M-estimators which are solutions of minimum distance estimating equations[28,29]. An important distance that is discussed in amongst these works is the L_2 distance

$$J_n(\theta)=\int_{-\infty}^{+\infty}(F_n(x)-F_\theta(x))^2dx \qquad \ldots (11)$$

For the normal location family the estimator is an M-estimator with $\psi(x)=\Phi(x)-1/2$, where $\Phi(x)$ is the standard normal cumulative distribution function. The estimator has bounded gross error sensitivity and remarkably high efficiency at the normal model (95.5%). Clarke and Heathcote[29] give the minimizing equations by defining the function ψ when F_θ is a finite mixture of normals. The defining ψ function is bounded and smooth giving a Fréchet differentiable M-estimator which has high efficiency in the case of asymptotic efficiency and in numerical simulation studies. Thus one obtains solutions to eq. (4) in finite samples since one is minimizing a distance, a finite $\gamma^*(\theta)$, high relative efficiency at the parametric model, and Fréchet differentiability with respect to $(\mathcal{G},\|\cdot\|_\infty)$.

The compact derivative choosers can claim that many seemingly robust functionals are not Fréchet differentiable with respect to the supremum norm but are compactly differentiable. Yet even the mean functional is compactly differentiable[30]. Thus compact differentiability encompasses both robust and nonrobust estimators. Naturally expansions for nonrobust estimates are made only at the parametric model, with the empirical distribution only being generated by F_θ.

The case for considering the Fréchet derivative at the supremum norm is elegantly argued in Bednarski[5]. Other norms while they can be used to obtain asymptotic results, usually have no implications for robustness. For instance Dudley[12] shows the median and quantiles are Fréchet differentiable with respect to a smaller norm. Shao[31] considers Fréchet differentiability with respect to the supremum norm and also the L_2 distance

$$\|G-F\|_2=\{\int[G(x)-F(x)]^2dx\}^{1/2}.$$

Shao and Tu[32] introduce other norms whereby they show various estimators are continuously Fréchet differentiable. It is clear now, that if one changes

the neighbourhood in which the Fréchet expansion of the solutions to eq. (5) is to take place, that one can gain an expansion of the form

$$T[F_n]-T[F_\theta]-T'_{F_\theta}[F_n-F_\theta]=o_p(1/\sqrt{n}), \quad \ldots (12)$$

where $o_p(1)$ is a small order probability term assuming F_n is generated from the parametric model distribution F_θ. See for example Clarke *et al.*[33] and Davies[6]. Indeed the work of Boos and Serfling[2] while it begins with the expansion (1) essentially considers the expansion stochastically for $G \equiv F_n$ thus arriving at eq. (12) The form for the derivative is

$$T'_{F_\theta}[F_n-F_\theta]=-M(\theta)^{-1}\int \psi(x,\theta)d(F_n-F_\theta) \quad \ldots (13)$$

where indeed ψ may be an unbounded function in the observation space variable. Thus the general arguments in Clarke[3] can be used to prove (1) when ψ is bounded and smooth, or to prove (12) when ψ is unbounded and suitably regular. Note that eq. (1) leads to eq. (12) since the well known Dvoretzky-Kiefer-Wolfowitz inequality (Theorem 2.1.3A of Serfling[34]) implies $o\|(F_n-F_\theta\|_\infty)=o_p(1/\sqrt{n})$, when F_n is generated by $F_\theta \in \mathcal{F}$. Early discussions of Fréchet differentiability see the form eq. (12) as in Reid[35], but strictly speaking this is only a stochastic expansion obtained at the model parametric family. It has relatively little to say when the underlying distribution is considered to be in a neighbourhood of F_θ. Note that it is from eq. (12) that one has that $\sqrt{n}(T[F_n]-T[F_\theta])$ is asymptotically normally distributed like

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} IF(X_i,F_\theta,T)+o_p(1)$$

the main term of which is a normed sum of independent identically distributed random variables with a normal asymptotic distribution with mean zero and variance $\int IF(x,F_\theta,T)^2 dF_\theta(x)$.

Bednarski[5] argues however that it is quite essential for robustness reasons that we use the supremum norm. We quote ".... neighbourhoods induced by this norm are large compared for instance to Hellinger or total variation distances, thus giving high smoothness to the functional..... The notion of infinitesimal neighbourhood $\{G:\|G-F\|_\infty \le \varepsilon/\sqrt{n}\}$ complies then with the ability of testing goodness of fit. More precisely if F is a member of a parametric family and a Kolmogorov type statistics is used to verify goodness of fit, then the hypothesis can be accepted with asymptotically positive probability for distributions from these infinitesimal neighbourhoods only."

Expansion (1) leads naturally to

$$T[F_n]-T[F]-\int IF(x,F,T)dF_n(x)=o_p(\|F_n-F\|_\infty)$$

whence one obtains asymptotic normality of $\sqrt{n}(T[F_n]-T[F_\theta])$ when F_n is generated by F_θ or equivalently under $F_\theta^{\otimes n}$ (product distribution). Note that if in fact F_n is the empirical distribution based on the n-sample from the distribution G_n for which $\sqrt{n}\|G_n-F_\theta\|_\infty$ is bounded then the stochastic expansion implied by the above formula holds also under $G_n^{\otimes n}$. Thus the asymptotic behaviour of $T[F_n]$ can be established for the infinitesimal neighbourhoods of the form $\{G:\|G-F\|_\infty \le \varepsilon/\sqrt{n}\}$ directly from the expansion. Bednarski *et al.*[4] show an estimator which has an expansion valid for infinitesimal neighbourhoods is asymptotically equivalent to an M-estimator and the M-functional corresponding to the M-estimator is Fréchet differentiable. They also demonstrate that the median functional is such that it does not follow a strong expansion under $1/\sqrt{n}$ shrinking neighbourhoods. That is illustrated by noting that the distribution of $\sqrt{n}(T[F_n]-T[\Phi])$ at the distribution $G_n=(1-\varepsilon/\sqrt{n})\Phi(x)+(\varepsilon/\sqrt{n})\delta_0$ is different from the first term of the expansion $\int IF(x,F,T)\ dG_n$ (cf. Bednarski *et al.*[4], Lemma 3). Thus the fact that the median is not Fréchet differentiable is accompanied by a breakdown of the usual expansion under $1/\sqrt{n}$ shrinking neighbourhoods. This is thus demonstrated by putting

contamination at the point of discontinuity of the influence function.

Lack of continuity of the influence function for the median is an indicator of the lack of Fréchet differentiability. Huber[22] points out that the influence function of a Fréchet differentiable functional is continuous. Clarke[3] gave conditions by which M-estimators with smooth ψ -functions are Fréchet differentiable, indeed the ψ functions also need to be bounded for the estimators to be differentiable with respect to the supremum norm. The robust M-estimator defined by eq. (10) which is bounded and continuous, but not smooth is shown to be Fréchet differentiable at the normal distribution in Clarke[3b]. However more recent proposals for M-estimators have been deliberately linked to smooth forms for ψ [36].

Compact differentiability is useful for considering asymptotic normality of $\sqrt{n}(T[F_n]-T[F])$ for functionals with discontinuous influence functions, since Fréchet differentiability is unattainable. A compact like derivative is discussed for the trimmed likelihood estimator[37], where necessarily the influence function of the estimating functional is discontinuous for example. Note that in a sense robustness is inferred since the influence function is bounded, but since it is not continuous the rigorous arguments of Huber[22,26] and Bednarski *et al.*[4] which require Fréchet differentiability do not hold. In fact Hampel *et al.*[22] argue for not only boundedness of the influence function but augmenting this condition with a finite breakdown point. The local bias is reflected by the influence function and the global bias by the breakdown point. Davies[6] in 1993 gives an example where this does not hold, namely the middle of the shortest half sample. At the normal distribution the influence function exists and is bounded. In spite of this the influence function does not reflect the local behaviour of the functional. Another example where without Fréchet differentiability one cannot deduce the local behaviour of the estimator is the normal scores estimator. Huber[26] points out that this estimator has an unbounded influence function and is efficient at the normal model. Unboundedness of the influence function at the normal model precludes Fréchet differentiability. However it has a positive breakdown point. In essence the influence functions at F in a neighbourhood of Φ become bounded.

Fréchet Differentiability and the Jackknife

The development of robustness sought also to obtain robust estimates of variance of estimators, an early example of a unified approach being the jackknife[38]. Regarding $T_n(X_1,...,X_n)=T[F_n]$, the i'th pseudovalue is defined as

$$T_{ni}^* = nT_n(X_1,...,X_n) - (n-1)T_{n-1}(X_1,...,X_{i-1},X_{i+1},...,X_n)$$

and a "corrected" estimator is

$$T_n^* = \frac{1}{n}\sum_{i=1}^{n} T_{ni}^*$$

often has a smaller bias than the original T_n. On page 95 of Hampel *et al.*[22] it is shown that

$$V_n = \frac{1}{n-1}\sum_{i=1}^{n}(T_{ni}^* - T_n^*)^2 \approx \mathrm{var}_{F_n}[IF(X,F_n,T)]$$

estimates

$$V(T,F) = \int IF(x,F,T)^2 dF(x) \qquad \ldots (14)$$

the asymptotic variance of the estimator. Interest in the differentiability approach to jackknifing is given in Parr *et al.*[39], both using a continuous form of Fréchet differentiability. The functional T is said to possess a strong Fréchet differential at F_θ iff there exists a $\psi : E \to E$ depending only on T and F_θ such that

$$\frac{\left|T[G]-T[H]-\int \psi(x)d[G-H](x)\right|}{\|G-H\|}=0 \qquad \ldots (15)$$

as $\|G-F_\theta\|+\|H-F_\theta\|\to 0$ for $G,H\in\Lambda$ where Λ is a convex set of distribution functions containing F_θ and all discrete distributions with finite support. Parr[39] does not assume the function ψ to be bounded. Hence T can possess a strong Fréchet differential at F_θ without being a continuous functional (consider the mean functional). The quest for consistent jackknifed

variance estimates stems from the knowledge that sometimes the jacknifed variance estimator is inconsistent (see the warning of Huber[22], p.16), for example when $T[F_n]=F_n^{-1}(1/2)$. But jackknifing non-robust estimators such as the mean leads to the variance estimator

$$S_n^2=\frac{1}{n(n-1)}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

which is again a very sensitive estimator in the presence of contamination, although it is consistent at the parametric model.

On the other hand it is interesting that robust M-estimators with bounded smooth ψ functions satisfy strong Fréchet differentials, whereby theorem 1 of Parr[39] gives the consistency of the jacknifed variance estimator. For instance Parr states that if T is differentiable in the sense of (1) in a neighbourhood of F_θ and that the derivative is continuous at F then the functional has a strong Fréchet differential. Essentially when both ψ and its partial derivative are bounded and continuous the resulting M-estimator will have a strong Fréchet differential with respect to $(G,\|\|_\infty)$[3] and the jackknifed variance estimator will be consistent.

However, jackknifing implicitly defined functionals such as $T[\psi,F_n]$ is not necessarily numerically easy when possibly there can be more than one solution to the eq. (4). A far better estimator of the variance of $T[F_n]$ is $V(T,F_n)$ given in (14) (see Huber[22], p.149), Hampel *et al.*[22], p. 95), and Staudte and Sheather[40] (p.79). Bednarski[5] stresses that if a functional is Fréchet differentiable with respect to the supremum norm, then the asymptotic variance for the resulting estimator can be approximated via the influence function for the whole infinitesimal neighbourhood. It is clear from Huber[22], Clarke[3] and Bednarski[5] that the Fréchet differentiable functionals in this case have bounded influence functions, whereby robustness of the variance estimates therefore follows.

The arguments for jackknifing at the parametric model by Shao[31], Shao and Tu[32] use more generally applicable norms such as

$$\|F-G\|_{(\infty+\ell)}=\|F-G\|_\infty+\left[\int|F-G|^\ell\,dx\right]^{1/\ell}$$

and it is clear that if one has a strong Fréchet differential with respect to the supremum norm then one automatically has a strong Fréchet differential with respect to $\|\|_{(\infty+\ell)}$. Thus if one has M-estimators defined by ψ functions which satisfy smoothness and boundedness thus leading to Fréchet differentiable functionals with respect to the supremum norm, as afforded by conditions A in Clarke[3] automatically one has consistency of the jackknife.

Besides the jackknife is the bootstrap, also useful for estimating the variance of the estimator $T[F_n]$. The bootstrap has close connections to the jackknife[41]. Bootstrapping involves taking random samples from the empirical distribution function F_n. Again with implicitly defined estimators solved iteratively such as solutions of eq. (4) this can lead to numerical problems, for instance if there are multiple solutions of the equations. Nevertheless Gill[22] relates consistency of the bootstrap to Hadamard differentiability and Shao and Tu[32], p83 consider more general differentiability for bootstrap estimates to be consistent. But again it is the case that since a statistic which is Fréchet differentiable with respect to the supremum norm is also Hadamard differentiable, it follows that bootstrap estimates of Fréchet differentiable functionals with respect to the supremum norm will be consistent.

Recent Developments

Bednarski[5] tackles the problem of robust estimation in Cox's regression model relying on the equivalence established by Bednarski *et al.*[4] of estimators which satisfy an expansion valid for infinitesimal neighbourhoods of the form $\{G:\|G-F\|\le\varepsilon/\sqrt{n}\}$ to M-estimators which are Fréchet differentiable. In another development Bednarski and Zontek[36] discuss robust Fréchet differentiable estimates which are arrived at in variance component models. Davies[6] considers Fréchet differentiability for multivariate location and dispersion estimators and regression estimators. Field and Smith[42] show asymptotics of a weighted likelihood estimator by showing Fréchet differentiability of their estimator. Davies[6] also considers robust location and scale functionals. Clarke, Gamble, and Bednarski[8]

discuss the Fréchet differentiability of the β-trimmed mean in the exponential parametric family, while in 1999 Bednarski and Clarke[43] discusses asymptotics of the adaptive trimmed likelihood location estimator by making use of the compact like derivative of Bednarski and Clarke[37].

Bickel *et al.*[4] give a discussion of Gâteaux, Fréchet and Hadamard differentiability and link the discussion to functionals differentiable in quadratic mean as in Le Cam. Also see Van der Vaart[4]. Again the mean functional is noted to be Hadamard differentiable, and hence this approach to differentiability encompasses non robust estimators.

However in a largely theoretical development Bednarski and Clarke[7] demonstrate that an estimating functional which is regular at a rich family of smooth parametric models containing the underlying F and for which the asymptotic limits satisfy a natural continuity property over shrinking Cramér von Mises neighbourhoods of F, is Fréchet differentiable at F for the Cramér von Mises norm. Since the Cramér von Mises norm is smaller than the supremum norm it follows that such functionals must at least be Fréchet differentiable with respect to the supremum norm. This is another argument for functionals to be at least Fréchet differentiable with respect to the supremum norm.

Application of a Fréchet Differentiable Functional to Earth Science

In this section, data from seismological recordings which may be considered "rough" is analysed. Influences of many extraneous factors contribute to the "roughness" of the data. These are exemplified by the nonhomogeneous nature of the earth and also measuring techniques used[45]. The measuring techniques can be summarised as least squares methods, notorious for being nonrobust. A rudimentary analysis of a larger data set containing the sample of data analysed here was presented in

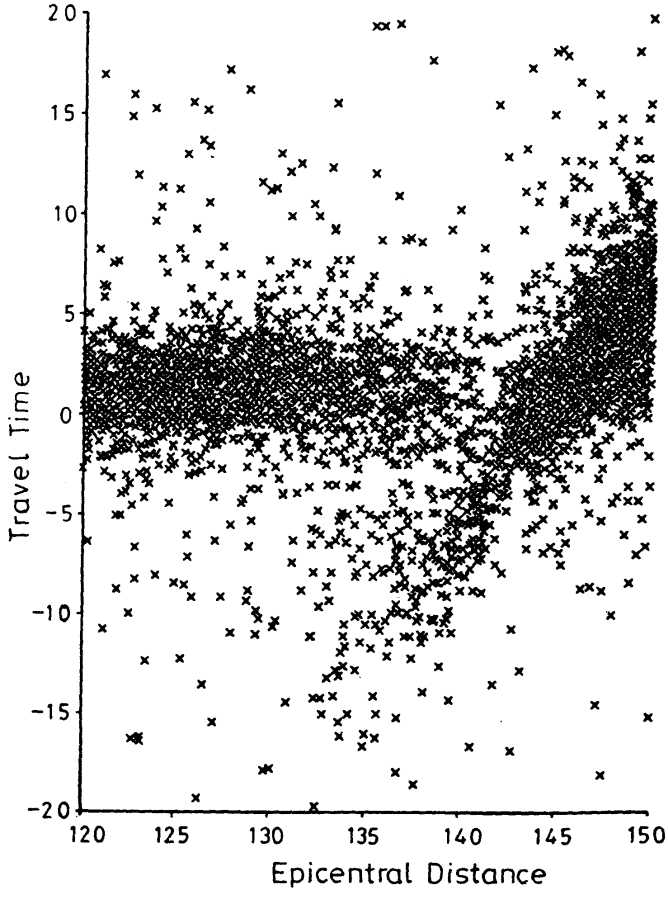

Fig. 1 Precursor arrivals for PKIKP earthquakes of March, April and May 1966

Anderssen and Cleary[46] and a summary description for the uninitiated in seismological nomenclature follows as an aid to describing the data.

Recordings of all earthquakes of reasonable magnitude are normally kept of date, for it is known that their attributes give insight into the structure of the earth. The focus of a quake is deduced, and time taken for a primary or longitudinal wave to travel an epicentral distance to an observation point is recorded. Fig. 1 illustrates a plot of 4303 residuals of observations for data collected over the months March-May of 1966, recorded by the International Seismological Center (ISC). These are the residuals based on Jeffreys-Bullen PKIKP times which use the revised depth correction of Shimshoni[47]. The figure exhibits two main streams of points, so that at any given epicentral distance, ω, there are two component populations. The horizontal and major stream corresponds to PKIKP waves. In fact this is how all the waves of the ISC data have been classified[46], even though some PKP_1 waves and precursors to PKP are also recorded. PKIKP waves are waves that have travelled to the surface through the outer layer of the earth, the core of the earth and then the inner core. They then travel back to the surface. At each interface they are refracted due to changes in velocity as the waves travel through material of different densities. The diagonal stream ascending upwards from 145° epicentral distance corresponds to PKP_2 and PKP_1 waves, which are waves which have travelled in the same way as the PKIKP waves, but on reaching the surface are reflected once more through the outer layer to the observation point. The interested reader is referred to Jeffreys[45] for a more detailed account. Fig. 16 of the book "The Earth: ..." gives plots of established paths of these distance time plots. Assuming a model of a perfectly spherical earth and core, both with uniform density, and assuming perfect observation, the data would follow continuous curves, these being indicated by the continuous lines of the sketch in Fig. 2. But actual irregularities in the earth, and in particular the interface between the mantle and core, allow distortions to appear. These add to observational error, and thus we have the scatter of Fig. 1 about the horizontal stream.

Until recently, it was believed that the so-called precursors to PKP, usually observed at distances

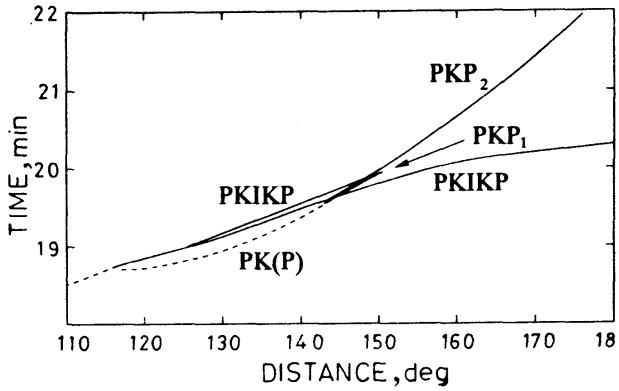

Fig. 2 Established theoretical paths of earthquake travel times and the investigated precursors to PKIKP indicated by dashed line, as in Anderssen and Cleary[46]

between about 130° and 140° were indicative of transition regions in the earth's outer core, though Anderssen and Cleary[46] claim that Cleary and Haddens'[48] hypothesis that the observations are the result of scattering PKP waves from irregularities in the vicinity of the core mantle boundary is generally accepted. Anderssen and Cleary take a large sample of 30,000 observations, identifying the points due to three sources, those points due to PKIKP and scattered PKP signals as well as a contribution from error. Equally well the process could be modelled as a mixture of two populations, where the scattered PKP and PKIKP signals are modelled as component normal populations and any further error is incorporated in departure from the model mixture of normals; rather than ignoring the error altogether, consider estimating the component population parameters robustly. The identification of the locations of the two streams of data and the proportion of each is of interest, though scale parameters for each population are also estimated. The estimated proportion of scattered PKP waves will indicate the amount of false classification for example. A sketch of the paths of scattered PKP waves is indicated by the dashed line of Fig. 2, originally proffered in Anderssen and Cleary[46]. The data is examined at fixed epicentral distances using the mixture model, the assumed model for the cumualtive distribution of points being

$$F_{\theta_\omega}(x) = \varepsilon_\omega \Phi\left(\frac{x-\mu_\omega+\Delta_\omega}{\sigma_{1\omega}}\right); \quad \ldots (16)$$
$$+(1-\varepsilon_\omega)\Phi\left(\frac{x-\mu_\omega}{\sigma_{2\omega}}\right)$$

Here $\theta = (\varepsilon, \mu, \Delta, \sigma_1 \sigma_2)$ is indexed by the epicentral distance ω. The parameter Δ represents the separation between the two component populations which have means $\mu_1 = \mu - \Delta$ $\mu_2 = \mu$ and variance σ_1^2 and σ_2^2 respectively. The parameter θ is then estimated at each epicentral distance ω over the region $133.5^o \leq \omega \leq 143^o$. By taking increments, $(\omega - \delta, \omega + \delta)$ at each value of ω, for $\delta = 0.5$, and letting the empirical distribution function $F_n (\equiv F_{n,\omega})$ be formed from all observation travel times for earthquakes with epicentral distance falling inside this region, an estimate $\hat{\theta}$ can be obtained by solving the minimizing equations of the distance $J_n(\theta)$ given in eq. (11). The equations are similarly derived to those in Clarke and Heathcote[29] but with two component populations and with the current parameterization. The equations are in the form of eqs. (4), (5) with a bounded ψ function and there exists a robust Fréchet differentiable root of the equations[3,29]. Due to multiple solutions of the resulting equations which are of the form of eqs. (4), (5) the set $I(\psi, \tau)$ is not a single element. Some solutions can be discounted since they are due to nonidentifiability, for example $\theta = (\varepsilon, \mu, \Delta, \sigma_1, \sigma_2)$ and $\bar{\theta} = (1 - \varepsilon, \mu - \Delta, -\Delta, \sigma_2, \sigma_1)$ generate the same distribution. This is overcome by restricting the parameter space to $\Theta = \{\theta : 0 < \varepsilon < 1, -\infty < \mu < \infty, \Delta > 0, \sigma_1 > 0, \sigma_2 > 0\}$. Nevertheless even with this restriction there can be multiple solutions of the equations and a selection functional has to be implemented. A suggestion is to use the Cramér von Mises distance as a selection functional as this is a well known goodness of fit statistic[29].

The approach using the robust estimating equations as indicated above can be compared to the approach where one solves eq. (3). Eq. (3) for mixtures have been studied by various authors[49]. The approach is to solve eq. (3) and use the selection functional $\rho_0(G, \tau) = -\int \log f_\tau dG(x)$. This selects the local extrema of the likelihood surface with maximum value of the likelihood. This will not be a global maximum of the likelihood due to singularities in the likelihood surface.

Fitting the Equations to the Data

The data of Fig. 1 yield 21 subsamples between epicentral distances 133° and 143°. For each subsample corresponding to an epicentral distance ω, eq. (4) were solved using the bounded function corresponding to minimizing eq. (11) using nonlinear equation solving algorithms. Also solutions of the mle estimating eq. (3) were sought using the parametric density $f_\theta(x) = F_\theta'(x)$ again using nonlinear equation solving algorithms. In the region of epicentral distance $120^o < \omega < 132.5^o$ there appears to be only one component population which is modelled by a parametric distribution $\Phi((x - \mu)/\sigma)$ and allowing for slight departures from this parametric distribution the parameters are estimated following the approach of Heathcote and Silvapulle[28] using the resulting robust minimizing equations derived from minimizing eq. (11) with respect to location and scale. Figs. 3-5 plot the subsequent estimates of location for this region of epicentral distance.

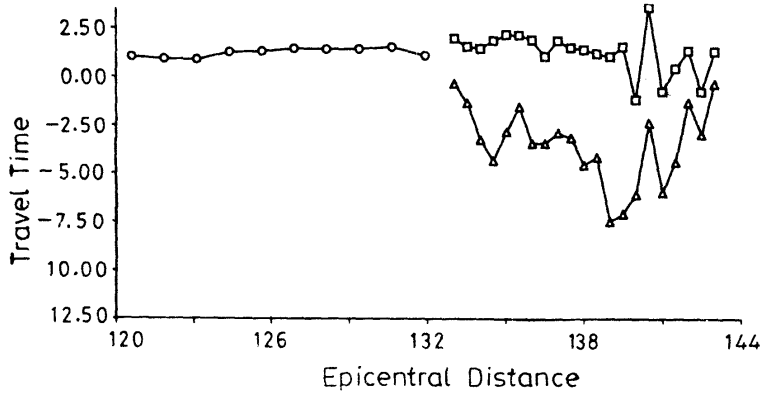

Fig. 3 M.l.e. estimates for PKIKP earthquakes of March, April and May 1966

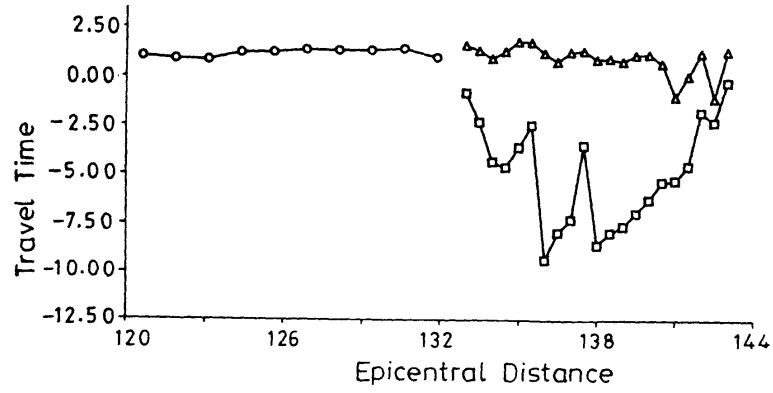

Fig. 4 Cramér von Mises selected L_2 estimates for PKIKP earthquakes of March, April and May 1966

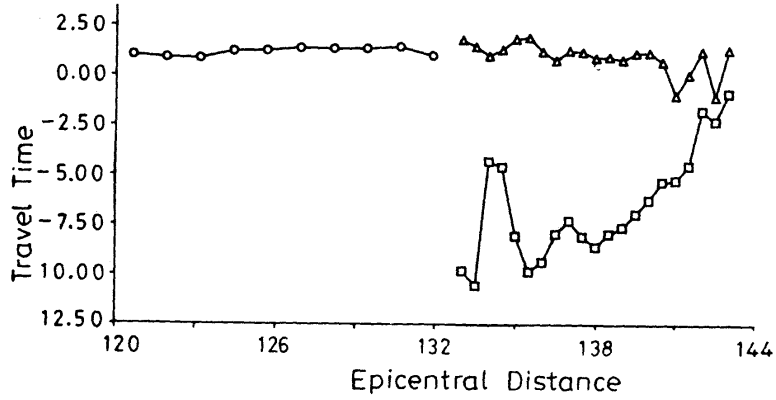

Fig. 5 Author selected L_2 estimates for PKIKP earthquakes of March, April and May 1966 based on robust estimating eq. (4)

Solving the eq. (3) by using a grid of starting points confirmed the observations of Fowlkes and Woodward *et al.*[49], who note in a slightly different context that good starting values are crucial for acceptable performance of maximum likelihood methods when the mixture distribution eq. (16) is employed. Those authors minimized minus the log of the likelihood directly. However numerous starting points led to the an algorithm for solving eq. (3) iteratively using a routine written by Powell and described in Rabinowitz[50] to report bad convergence criterion and other initial starting points were required for solutions of these equations. Moreover, more than one solution of the equations would frequently exist. This may be a function of the data, but has also been reported in Fowlkes[49].

In comparison the solutions to eqs. (4), (5) with the choice of a bounded ψ function derived from minimizing eq. (11) where F_θ is given by eq. (16) appeared to be less numerous and more often obtained by the numerical nonlinear equation solving algorithms. Table I summarises the number of solutions for the two methods, that of solving the maximum likelihood equations and that of solving eq. (4) minimizing the L_2 distance of eq. (11). It can be noted the number of local extrema is fewer for the latter robust eq. (4) than for the solutions of eqs. (3) for most subsamples. In particular note for the region $139^o \leq \omega \leq 143^o$ the robust equations solving eq. (4) for minimizing eq. (11) led to only one solution. Grids with as many as 500 initial starting points for the nonlinear equation solving algorithms were used. In the authors opinion all local extrema of $J_n(\theta)$ were found, though more solutions could be found for the sensitive likelihood equations.

Fig. 3 exhibits the resultant locations of the component populations of PKIKP and scattered PKP waves as given by solving the maximum

Table I

Numbers of solutions of the m.l.e. eq. (3) and eq. (4) with a ψ obtained from the partial derivative of $J_n(\theta)$.

Epicentral Distance ω	Solutions of (3)	Solutions of (4) (robust ψ)	Epicentral Distance ω	Solutions of (3)	Solutions of (4) (robust ψ)
133	3	3	139	4	1
133.5	2	3	139.5	5	1
134	5	1	140	4	1
134.5	3	1	140.5	1	1
135	3	3	141	3	1
135.5	2	3	141.5	2	1
136	4	4	142	2	1
136.5	4	3	142.5	3	1
137	2	3	143	2	1
137.5	4	3			
138	2	3			
138.5	7	2			

likelihood equations and solving eq. (6) using minus the log of the likelihood as a selection functional. This method has difficulty in separating the two population locations between $133^o \le \omega \le 138.5^o$. For example the estimate of standard deviation of the component population representing the scattered PKP wave is greater than 5 on average. The maximum likelihood estimates are erratic for $140^o \le \omega \le 143^o$. Essentially this is implored by the fact that the influence function of the m.l.e. is unbounded in the observation space variable.

As is explained in Clarke and Heathcote[29] a robust selection functional being the Cramér von Mises distance

$$\rho(F_n,\tau)=\frac{1}{n}\sum_{i=1}^{n}\{F_\tau(X_{(i)})-\frac{(i-1/2)}{n}\}^2+(12n^2)^{-1},$$

is employed to select the roots of eq. (4), even though the equations are the minimizing equations of the L2-distance $J_n(\theta)$ in eq. (11). That is we solve eq. (6) using the Cramér von Mises distance and the bounded ψ from minimizing $J_n(\theta)$. Note that $J_n(\tau)$ affords another selection functional but one which is not robust. The Cramér von Mises statistic can however be used as a robust selection functional, since it satisfies the criterion (4.1) and (4.2) of Clarke[3] for a restricted parameter set $D \subset \Theta$. It can be observed in Fig. 4 that this leads to some continuity in the locations of the component populations.

While there is relative continuity in the estimates, say as compared to maximum likelihood estimation, a perusal of the solutions $I(\psi, F_n)$ showed that among the roots of the equations there were solutions that separated the component distributions more fully. For instance, at 133^o there were three solutions for the parameter $\theta=(\varepsilon,\mu,\Delta,\sigma_1,\sigma_2)$. These were

.1595	1.9488	9.8792	5.7126	2.2529
.1248	1.8685	11.6981	3.9046	2.3383
.3899	1.6778	2.4494	8.2291	1.4085

The first and third solution have component populations with at least one of the component populations having a standard deviation which is inflated so that the spread of this component distribution easily covers the other component population. The second estimate however nicely separates the two component populations and this is the preferred estimate of the author. For we know eq. (4) have a Fréchet differentiable root in a local neighbourhood of the true θ and we would like to select that root which is physically meaningful.

Thus the author's selected estimates from among

Table II

Comparison of Anderssen and Cleary's[46] fitted locations and those computed from solutions of equations (4) using a ψ obtained from the minimizing equations of $J_n(\theta)$. The proportion $\hat{\varepsilon}$ of misclassified waves is also estimated.

Epicentral Distance ω	Fitted Scattered PKP	Estimated PKP Scatter	Fitted PKIKP	Estimated PKIKP	Estimated Proportion $\hat{\varepsilon}$	n
133.5	-	-11.581	-	1.513	0.170	104
134.5	-10.690	-4.592	1.614	1.369	0.379	81
135.5	-10.030	-9.860	1.651	2.011	0.208	78
136.5	-9.225	-7.962	1.684	0.842	0.268	98
137.5	-8.285	-8.090	1.712	1.274	0.304	84
138.5	-7.205	-7.952	1.734	1.020	0.359	81
139.5	-5.985	-6.948	-	1.212	0.580	74
140.5	-4.627	-5.318	-	0.779	0.568	76
141.5	-3.129	-4.497	-	0.123	0.551	78
142.5	-1.491	-2.242	-	-0.998	0.085	88

the solutions $I(\psi, F_{n\omega})$ are selected for each ω and the subsequent locations are plotted in Fig. 5.

The method of modelling a normal mixture affords solutions which are not otherwise obtainable for the analysis of such seismic data. Anderssen and Cleary[46] used arbitrary separation of population points before taking 10% trimmed means. Table II shows that the mixture estimation with approximately 4,000 points in the study compares excellently with a study that used 30,000 points, and moreover it provides estimates where the other cruder methods do not. Sample sizes for each parameter estimate are also listed. In only one instance does the estimate of the mean of the scattered PKP appear to deviate greatly. The estimate of -4.592 at epicentral distance $\omega = 134.5$ is accompanied by an estimate of $\hat{\sigma}_1 = 6.05$ suggesting that in this instance the estimating eq. (4) using the robust bounded ψ from minimizing the distance $J_n(\theta)$ could not separate the two populations.

Conclusion

The previous sections analysis highlights some of the problems associated with practical implementation of both robust and non-robust estimators. The existence of multiple roots of ones estimating equations is but just one facet. No confidence intervals were proffered, since the locations of the PKIKP and scattered PKP waves were just two parameters in a multivariate parameter estimation, and while theoretically $Var(T, F_n)$ is obtainable it is not a simple expression to calculate. Neither do we offer a simple goodness of fit test to see whether the mixture model is acceptable, for instance, it is rejected in favour of another model. The structure of the seismic data lends itself to interpretation when there are two component populations at each epicentral distance and even if there were more data, this would only lead to eventually the rejection of the mixture model, say as the true distribution may no longer to be judged as being in a $1/\sqrt{n}$ shrinking neighbourhood of a mixture distribution as argued by Bednarski[5]. As is well known model fit tends to become poorer rather than better in larger samples[51].

The appropriateness of a mixture model as being an approximate description for the seismic data is reflected in the relative continuity of the robust estimates of locations of scattered PKP and PKIKP waves respectively, as is exhibited in Fig. 5. This plot has clear superiority as opposed to the plot of mle estimates in Fig. 3. In larger samples one then becomes more interested in reducing the asymptotic bias through having a finite gross error sensitivity, and fitting mixture models as we have

done here since they offer some interpretation, using an estimator with a bounded influence function.

Acknowledgements

The author has benefited from early discussions with Professors Chip Heathcote, Bob Anderssen, and more recently discussions with Tadeusz Bednarski, all who have helped in honing the present article. Thanks are owing to David Schibeci for help with the figures and the author is indebted to Professor Anderssen for the data of 4,303 points being made available. The author will gladly pass on the data on request. Thanks are owing to the Mathematics and Statistics Department at the University of Western Australia where the author spent part of his sabbatical leave in the preparation of this paper.

Chronological Chart

Year	Approach	Author	Description
1947	VM	[16]VonMises	Definition of VM functionals
1955	F	[10]Kallianpur&Rao	Asymptotic efficiency of mle.
1957	F	[13]Rao	Multinomial mle
1962	VM	[52]Fillipova	Mle and goodness of fit statistics
1963	G&VM	[14]Kallianpur	Asymptotic normality
1974	G	[24]Hampel	Definition of influence function (curve)
1976	C	[17]Reeds	Introduction of compact derivative
1977	G,F	[26]Huber	Influence function asymptotic efficiency; robustness
	O	[53]Beran	Fréchet differentials using hellinger neighbourhoods
1979	F	[1]Boos	*L*-estimators
	G	[54]Papantoni-Kazakos & Gray	Stationary Observations
	F	[55]Pollak	Class of robust estimators
1980	F	[2]Boos & Serfling	Quasi Derivative for *M*-estimators
	F,VM	[34]Serfling	General limit theorems, asymptotic normality & *LIL*
1981	VM	[56]Bickel & Freedman	The bootstrap
1982	G	[57]Cook and Weisberg	Influence function in regression
	F	[53]Beran	The bootstrap
	F	[35]Reid	Kaplan Meir estimator
1983	F	[3]Clarke	*M*-estimators
	C	[18]Fernholz	*M, L &R*-estimators
	VM	[58]Withers	Quantiles
	G	[59]Parr	Hajek projections and the influence function (curve)
1984	G	[60]Künsch	Influence curves for autoregression
1985	F	[59]Parr	Bootstrapping
	F	[39]Parr	Jackknifing
	C	[61]Esty,Gillette, Hamilton & Taylor	Asymptotic normality
	C	[61]Taylor	*M, L & R* statistics
	F	[62]Yang	Bootstrapping L,M statistics differentiability
	O	[63]Pollard	Stochastic differential
1986	F	[3]Clarke	*M*-estimators
	G	[22]Hampel *et al.*	Bounded influence estimation
	G	[64]Martin & Yohai	Time series influence function
1987	C	[65]Lohse	Bootstrap

(Chronological Chart continued)

Key to Approach
VM : Von Mises; F : Fréchet; G : Gâteaux; C : Compact or hadamard; O : Other

Chronological Chart—*Contd*

Year	Approach	Author	Description
1988	C	[66]Sen	Jacknifing using a 2nd order compact derivative
1989	C	[19]Gill	Nonparametric mle
	F	[31]Shao	Limit theorems M & L Estimators
	F	[39]Shao & Wu	Jacknifing
1991	C	[44]Van der Vaart	Efficiency & the Hadamard derivative
	O	[44]Van der Vaart	Differentiable Functionals
	F	[4]Bednarski, Clarke & Kolkiewicz	Asymptotically equivalent M-estimators in $1/\sqrt{n}$ shrinking neighbourhoods
1992	C	[30]Heesterman & Gill	M-estimators(correction of Reeds' work)
	F	[67]Cheng	Dependent Data
	F	[6]Davies	Multivariate location and dispersion
	F	[12]Dudley	p-variation norms
1993	F	[5]Bednarski	Robust Cox Regression
	C	[27]Bednarski & Clarke	Trimmed likelihood estimators
	CFG	[44]Bickel *et al.*	Efficient estimators
	F	[33]Clarke, Milne & Yeo	Mle in ion channel models
	F	[6]Davies	Robust Regression
	F	[31]Shao	Consistency of the jacknife
1994	F	[29]Clarke & Heathcote	Robust mixture estimation
	F	[68]Dudley	p-variation norms
	F	[42]Field & Smith	A weighted maximum likelihood estimator
	CFGO	[69]Rieder	Asymptotic theory
1995	C	[70]Ren & Sen	Multivariate Hadamard Differentiability
	CFG	[32]Shao and Tu	Jackknife and Bootstrap
1997	C	[70]Ren & Gu	Regression and censored data using M-estimates
	C	[71]Fernholz	Target estimation
1998	F	[6]Davies	location and scale functionals
1999	C	[43]Bednarski & Clarke	Adaptive trimmed likelihood estimator
2000	F	[8]Clarke,Gamble & Bednarski	β -trimmed mean

Key Approach :
VM : Von Mises; F : Fréchet; G : Gâteaux; C : Compact or hadamard; O ; Other

References

1 D D Boos *Ann Statist* **7** (1979) 955
2 D D Boos and R J Serfling *Ann Statist* **8** (1980) 618
3(a) B R Clarke *Ann Statist* **11** (1983) 1196
(b) B R Clarke *Probab Th Rel Fields* **73**(1986) 197
4 T Bednarski, B R Clarke and W Kolkiewicz *J Austral Math Soc (Series A)* **50** (1991) 88
5(a) T Bednarski *Asymptotic Statistics, Proc Fifth Prague Symp* (Eds. P Mandl and Husková) (1993) p 49
(b) T Bednarski *Scand J Statist* **20** (1993) 213
(c) T Bednarski *New Directions in Statistical Data Analysis and Robustness* (Eds : S Morgenthaler, E Ronchetti and W A Stahel) Basel Birkhäuser Verlag (1993) p 25
6(a) P L Davies *J Mult Anal* **40** (1992) 311
(b) P L Davies *Ann Statist* **21** (1993) 1843
(c) P L Davies *Ann Statist* **26** (1998) 1103
7 T Bednarski and B R Clarke *Discussiones Mathematicae : Algebra and Stochastic Methods* **18** (1998) 155
8 B R Clarke, D K Gamble and T Bednarski *Austral and New Zea J Statist* **42** (2000) (to *appear*)
9 B R Clarke PhD Thesis (1980) Australian National University Canberra
10 G Kallianpur and C R Rao *Sankhya* **15** (1955) 331
11 M Fréchet *Ann Ecole Norm Sup* **42** (1925) 293
12 R M Dudley *Ann Probab* **20** (1992) 1968
13 C R Rao *Sankhya* **18** (1957) 139
14 G Kallianpur *Sankhya, A* **23** (1963) 149
15 B R Clarke *Probab Math Statist* **11** (1991) 149
16 R von Mises *Ann Math Statist* **18** (1947) 309
17 J A Reeds Ph D Thesis (1976) Harvard University Cambridge MA
18 L T Fernholz *Von Mises Calculus for Statistical Functionals* Lecture Notes in Statistics Springer Verlag New York **19** (1983)
19 R D Gill *Scand J Statist* **16** (1989) 97
20 V I Averbukh and O G Smolyanov *Russian Math Surveys* **22** (1967) 201

21 M J Fabian *Gâteaux Differentiability of Convex Functions and Topology: Weak Asplund Spaces* Wiley New York (1997)

22(a) P J Huber *Robust Statistics* Wiley New York (1981) p 37

(b) F R Hampel, E M Ronchetti, P J Rousseeuw and W A Stahel *Robust Statistics: The Approach Based on Influence Functions* Wiley New York (1986) p 102

23(a) P J Huber *Ann Math Statist* **35** (1964) 73

(b) P J Huber *Proc of the Fifth Berkeley Symp on Math Statist and Prob* **1**(1967) 221

24 F R Hampel *J Amer Statist Assoc* **62** (1974) 1197

25 A E Beaton and J W Tukey *Technometrics* **16** (1974) 147

26 P J Huber *Robust Statistical Procedures* SIAM Philadelphia (1977) (1996)

27 F R Hampel PhD Thesis (1968) University of California Berkeley

28(a) D Boos *J Amer Statist Assoc* **76**(1981) 663;

(b) W C Parr and T de Wet *Comm Statist Theory Methods A* **10** (1981) 1149

(c) C R Heathcote and M J Silvapulle *Biometrika* **68** (1981) 501

(d) T P Hettmansperger, I Hueter and J Hüsler *J Statist Planning and Inf* **41** (1994) 291

(e) Ö Öztürk and T P Hettmansperger *Biometrika* **84** (1997) 283

29(a) B R Clarke *Statist and Prob Letters* **7**(1989) 275

(b) B R Clarke and C R Heathcote *Ann Inst Statist Math* **46** (1994) 83

30(a) C C Heesterman and R D Gill *Statist Neerland* **46** (1992) 165; (b) A van der Vaart *Scand J Statist* **18** (1991) 63

31(a) J Shao *J Amer Statist Assoc* **84** (1989) 114

(b) J Shao *Statist and Prob Letters* **8** (1989) 397

(c) (c) J Shao *Ann Statist* **21** (1993) 61

32 J Shao and D Tu *The Jackknife and Bootstrap* Springer Series in Statistics Springer Verlag New York (1995)

33 B R Clarke, R K Milne and G F Yeo *Scand J Statist* **20** (1993) 133

34 R J Serfling *Approximation Theorems of Mathematical Statistics* Wiley New York (1980)

35 N Reid *Ann Statist* **9** (1981) 78

36(a) T Bednarski and S Zontek *Ann Statisti* **24** (1996) 1493;

(b) R M Huggins *Austral J Statist* **35** (1993) 43

37 T Bednarski and B R Clarke *Austral J Statist* **35** (1993) 141

38(a) M H Quenouille *Biometrika* **43** (1956) 353

(b) J W Tukey *Ann Math Statist* **29** (1958) 614

39(a) W C Parr *J R Statist Soc B* **47** (1985) 56

(b) J Shao and C F J Wu *Ann Statist* **17** (1989) 1176

40 R G Staudte and S J Sheather *Robust Estimation and Testing* Wiley New York (1990) p79

41 B Efron *Ann Statist* **7**(1979) 1

42 C Field and B Smith *Int Statist Rev* **62**(1994) 405

43 T Bednarski and B R Clarke *Mathematics and Statistics Murdoch University Research Report* (1999) Western Australia

44(a) P J Bicke,l C A J Klaassen, Y Ritov and J A Wellner *Efficient and Adaptive Estimation for Semiparametric Models* The John Hopkins University Press Maryland (1993); (b) L Le Cam *Théorie Asymptotique de la Décision Statistique* Les Presses de l'Université de Montréal (1969); (c) A van der Vaart *Scand J Statist* **18** (1991) 63; (d) A van der Vaart *Ann Statist* **19** (1991) 204

45(a) H Jeffreys *Theory of Probability* Clarendon Press Oxford (1967) p 214

(b) H Jeffreys *The Earth: Its Origin, History and Physical Constitution* Cambridge University Press (1970) p 71

46 R S Anderssen and J R Cleary *Physics of the Earth and Planetary Interiors* **23** (1980) 207

47 M Shimshoni *Geophys J R Astron Soc* **13** (1967) 471

48 J R Cleary and R A W Hadden *Nature* **240** (1972) 549

49(a) V A Hasselblad *Technometrics* **8** (1966) 431

(b) N E Day *Biometrika* **56** (1969) 463

(c) D W Jr Hosmer *Communications in Statist* **1** (1973) 217

(d) E B Fowlkes *J Amer Statist Assoc* **74** (1979) 561

(e) R A Redner and H F Walker *SIAM Rev* **26** (1984) 195

(f) W A Woodward W C Parr W R Schucany and H Lindsey *J Amer Statist Assoc* **79** (1984) 590

50 P Rabinowitz *Numerical Methods for Non-Linear Algebraic Equations* Gordon and Breach London (1970)

51 A H Welsh *Aspects of Statistical Inference* Wiley New York (1996)

52 A R Fillipova *Theor Probab Appl* **7**(1962) 24

53(a) R Beran *Ann Statist* **5** (1977) 445

(b) R Beran *Ann Statist* **10** (1982) 212

54 P Papantoni-Kazakos and R M Gray *Ann Probab* **9** (1979) 989

55 M Pollak *Commun Statist-Theor Meth A* **8** (1979) 509

56 P J Bickel and D A Freedman *Ann Statist* **9** 1196

57 R D Cook and S Weisberg *Residuals and Influence in Regression* Chapman and Hall New York (1982)

58 C S Withers *Ann Statist* **11** (1983) 577

59(a) W C Parr *Statist Prob Lett* **1** (1983) 177

(b) W C Parr *Statist Prob Lett* **3** (1985) 97

60 H Künsch *Ann Statist* **12** (1984) 843

61(a) W Esty, R Gillette, M Hamilton and D Taylor *Ann Inst Statist Math* **37** (1985) 109

(b) D C Taylor *Ann Inst Statist Math* **37** (1985) 131

62 S S Yang *Statistica Neerlandica* **39** (1985) 375

63 D Pollard (1985) *Econometric Theory* **1** (1985) 295

64 R D Martin and V J Yohai *Ann Statist* **14** (1986) 781

65 K Lohse *Statistics and Decisions* **5**(1987)

66 P K Sen *Ann Statist* **16** (1988) 450

67 B Cheng *Statist and Probab Letters* **15** (1992) 339

68 R M Dudley *Ann Statist* **22** (1994) 1

69 H Rieder *Robust Asymptotic Statistics* Springer Verlag New York (1994)

70(a) J Ren and P K Sen *J Mult Anal* **55** (1995) 14

(b) J Ren and M Gu *Ann Statist* **25** (1997) 2638

71 L T Fernholz L_t *Statistical Procedures and Related Topics* (Ed. Y Dodge) IMS Lecture Notes-Monograph Series **31** (1997) 363