International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 6 (Special Issue), 2007, pp 305-314

Heuristics of Influence Function

M. Nasser
Dept. of Statistics
University of Rajshahi
Rajshahi-6205
Bangladesh
E-mail: nasser_nomi@yahoo.com

[Received February 13, 2005; Revised May 15, 2007; Accepted July 29, 2007]

Abstract

This article discusses heuristics of influence functions, an important concept in robust statistics, with some new counter-examples. It re-emphasises the caution raised by Davies (1993, 1994 and 1998) about using heuristics of influence function thoughtlessly. It is suggested that the influence functions of more smooth functionals have more desirable properties. Finally some examples of more smooth functionals including a new class of uniformly Fréchet differentiable L- location functional is presented.

Keywords and Phrases: Heuristics of influence function, (Strongly) Fréchet differentiability and Uniform Fréchet differentiability.

AMS Classification: 62G35.

1 Introduction

Hampel in his Ph.D. thesis (1968) developed three concepts: qualitative robustness (also Π- robustness), breakdown point and influence function to assess robustness in estimation and thus raised rigorousness in robust estimation to a satisfactory level. Hampel discussed and elaborated qualitative robustness at the outset of his thesis, breakdown point and influence function in the latter part. Though his seminal article on qualitative robustness (1971) was published three years before his mostly quoted article on the 'influence function' (1974) (originally termed as 'influence curve'), the latter has become most popular among statistical community. He and his co-researchers used several heuristics (not theorems) of influence function and developed a new approach to Robust Statistics. In this article I have critically examined the heuristics

and tried to raise caution against indiscriminate use of those heuristics, and thereby argued for the need to seek more smooth functionals in the line of Davies (1993, 1998). In his seminal article in 1974 Hampel first gave the definition for the particular case (both sample space and functional range space are R or subsets of R) and then for the general case (sample space X, a polish space and functional range space, R^k). We could generalize the concept even further by assuming functional range spaces at least a topological vector space, V and X, a general sample space. But considering objectives of the article we define influence function of statistical as follows:
Definition1.Let X be the sample space and T be a R^k-valued mapping from a subset of the probability measures,$D_T(X)$, a finitely full and convex subset of $S_p(X)$, the set all probability measures on X. Let $F \in D_T(X)$ and δ_x denote the atomic probability measure concentrated at any given point, x. Then the vector-valued influence function of T at F (here F is a measure) is defined pointwise by

$$IF_{T,F}(x) = \lim_{\epsilon \to 0} \frac{\mathrm{T}[(1-\epsilon)\mathrm{F} + \epsilon\delta_{\mathrm{x}}] - \mathrm{T}(\mathrm{F})}{\epsilon} \qquad (1)$$

Though for a particular T, the influence is generally considered as a function of x and$F$, later, for brevity, it is denoted by $IF(x)$. 'The IF is mainly a heuristic tool, with an intuitive interpretation $'$ (Hampel et al., 1986, p-83). It can be intuitively interpreted as a suitably normed asymptotic influence of outliers on the value of an estimate or test statistic,$T(F_n)$. It is a local robustness property. Various characteristics of an influence function are used to develop various concepts such as Gross Error Sensitivity (GES), γ^* (supremum of $IF(x)$ w.r.t. x for fixed F) and maximum-bias curve (graph of GES vs F), Local -Shift Sensitivity (LSS) (supremum of slope of $IF(x)$), λ^*, Rejection Point, ρ^* (related to the upper limit of the range outside which the influence function vanishes) etc to delineate definite but different aspects of local robustness properties. As important by- products of the attempt to quantify the effect of outlier on the estimators, the concept of Change of Variance Function (CVF) has been developed from $IF(x)$ to plot asymptotic variance vs F. While using the heuristics of influence function we should keep in mind that the heuristics of influence function are heuristics, not theorems. But tendencies to use them as theorems are not rare in literature.
The article is arranged as follows. Section 2 presents the first heuristic of influence function with a new result whereas section 1 introduces the topic. Section 3 and 4 present second and third heuristics respectively with some new counter-examples. The concluding section illustrates some strongly and uniformly Fréchet differentiable functionals. The article emphasizes that heuristics of influence functions, which are important for judging robustness of the associated functionals but require careful scrutiny at every step. The new results are termed as propositions.

2 First Heuristic

2.1 Introduction. Let T have a bounded influence function $IF(x)$ at F and

$$\gamma^* = \sup_{x \in R} |IF(x)| < \infty \tag{2}$$

According to Davies (1993) the first heuristic conclusion for a bounded $IF(x)$ is that we expect for ϵ sufficiently small,

$$\sup_{x \in R} |T[(1-\epsilon)F + \epsilon\delta_x] - T(F)| \leq 2\epsilon\gamma^* \tag{3}$$

i.e. bias caused by a small amount ϵ of point contamination is bounded by ϵ. But Hampel, the first proponent of influence formulated it in two other forms in his Ph.D thesis and without sufficient arguments claimed that the following inequalities roughly hold,

$$\sup_{G \in S_p X} |T[(1-\epsilon)F + \epsilon G] - T(F)| \leq \epsilon\gamma^* \tag{4}$$

where G is arbitrary, and H belongs to $B_\delta(F)$, a ball of radius δ, generated by total variation metric,

$$\sup_{H \in B_\delta(F)} |T(H) - T(F)| \leq 2\epsilon\gamma^* \tag{5}$$

2.2 Counter-examples and Comments. Davies (1993) showed that the above mentioned Hampel's claim is not always true giving an practical counter-example, the case of the middle of the shortest functional T_{MSH} (defined as $\frac{(a+b)}{2}$ where a and b are such as to minimize $(a-b)$ subject to $P_r([a,b]) \geq 1/2$) at $N(0,1)$. Now comes generally the question $-$ what conditions guarantee the inequalities 4 and 5.
It can be proved easily that if $\lim_{\epsilon \to 0} \frac{\mathrm{T[(1-\epsilon)F+\epsilon\delta_x]-T(F)}}{\epsilon} - IF(x) = 0$ uniformly in x, then $\ni \delta > 0$ such that for all ϵ, $0 < \epsilon < \delta$ and some K, $0 < K \leq 1$

$$\sup_{x \in R} |T[(1-\epsilon)F + \epsilon\delta_x] - T(F)| \leq (1+K)\epsilon\gamma^* \tag{6}$$

$$\Rightarrow \sup_{x \in R} |T[(1-\epsilon)F + \epsilon\delta_x] - T(F)| \leq \epsilon\gamma^* + O(\epsilon) \tag{7}$$

Inequalities 6 and 7 are equally true for Banach -valued functionals. Huber (1977) showed that if T is Fréchet differentiable at F and $G_\epsilon = (1-\epsilon)F + \epsilon H, 0 < \epsilon < 1$, then

$$\sup_{H \in S_p X} |T(G_\epsilon) - T(F)| \leq \epsilon\gamma^* + o(\epsilon) \tag{8}$$

8 holds true for Banach -valued functional T. Though it is an improvement on 7, Hampel's first claim (inequality 4) is not still satisfied. Since the most popular gross error model, $G_\epsilon = (1-\epsilon)\Phi + \epsilon H, 0 < \epsilon < \frac{1}{2}$, despite the generality of H, doesn't include some, perhaps the majority laws G_ϵ treated as normal to an acceptable approximation in practice, such as laws G on R with $G([0,\infty)) = 1$, laws discretized by rounding to finitely many decimal places, it is imperative for us to seek the condition that guarantees the existence of an inequality of the similar type over a weak neighborhood. The

following proposition shows an improvement on 8 in the desired line assuming strongly Fréchet differentiability of functionals with continuous influence function of bounded variation. Several examples of functionals meeting this requirement are in section 5.

Proposition 2.1 If T is strongly Fréchet differentiable at F w.r.t. Kolmogorov norm with continuous influence function $\Psi(x)$ of bounded variation, then

i) 8 holds.

ii) There exists $\epsilon, 0 < \epsilon < 1$ and $M \geq \|\Psi(x)\|_v > 0$ s.t.

$$\forall H, G \in \{G | \|G - F\| < \delta\} \Rightarrow |T(H) - T(G)| \leq \epsilon M \tag{9}$$

Here $2\delta = \epsilon$

Proof. We prove ii. We know, strong Fréchet differentiability of T at $F \Leftrightarrow T$is Fréchet differentiable at F and $\forall \rho > 0 \ni \delta > 0$ s.t. $r(H) = T(H) - T(F) - \int \Psi(x) dH(x)$ is ρ-Lipschitzean $\forall H \in \{G | \|G - F\| < \delta\}$;
that is $\forall H, G \in \{G | \|G - F\| < \delta\} \Rightarrow |r(H) - r(G)| \leq \rho \|H - G\|$. Again as$|r(H) - r(G)| = |T(H) - T(G) - \int \Psi(x) d(H - G)|$ and $|\int \Psi(x) d(H(x) - G(x))| \leq \|\Psi\|_v \|H - G\|$, we have
$\|G - F\| < \delta \Rightarrow |T(H) - T(G)| \leq \rho \|H - G\| + |\Psi\|_v \|H - G\| \leq (\rho + |\Psi\|_v) \|H - G\|$
Assuming $M = (\rho + |\Psi\|_v)$ and $2\delta = \epsilon$, we get the desired result. (Proved)

It is to be noted that i) since ρ is arbitrary, we could assume $M = (|\Psi\|_v)$ and
ii) if F is continuous, $\{G | \|G - F\| < \delta\}$ always contains a weak neighborhood.

3 Second Heuristic

3.1 Introduction. If T is well defined at all empirical distributionsF_n derived from independently and identically distributed random variables $X_1, X_2, \cdots X_n$ with common distribution F, then we expect (Hampel et al., 1986, p-85)

$$n^{\frac{1}{2}} [T(F_n) - T(F)] = n^{-\frac{1}{2}} \sum IF(X_i) + o_p(1) \tag{10}$$

In particular, $n^{\frac{1}{2}} [T(F_n) - T(F)] \Rightarrow N(0, \sigma^2)$ where

$$\sigma^2(F) = \int IF(x)^2 dF \tag{11}$$

3.2 Counter-examples and Comments. We should be cautious about the fact that even existence of inequalities 2-7 does not imply the inequality 10 without additional assumption. Let F be continuous and $T_c(F)$ = size of the largest atom. Then it

is obvious, $IF(x) = 1$ for all x and it satisfies 2 and 3, but $T_c(F_n) = \frac{1}{\mathrm{n}}\forall n$. So $n^{\frac{1}{2}}[T_c(F_n) - T_c(F)] = n^{-\frac{1}{2}}$ and $n^{-\frac{1}{2}}\sum IF(X_i) = n^{\frac{1}{2}}$. Therefore, 10 and 11 are not satisfied. Hampel et al. (1986, P-85) claimed; '2.1.8 (our 11) gives the right answer in all practical cases we know of'. Davies (1993, pp 1856-1857) showed that the middle of the shortest functional T_{MSH} at $N(0,1)$ does not satisfy the heuristic despite existence of its influence function, where as Rousseeuw and Leroy (1987) argued for non-existence of the influence function due to its cube root convergence. The weakest set of sufficient conditions for existence the validity of 10 is available in Reed (1976).
Let $T : E \to F$. Both E and F are normed spaces and
a) T is Hadamard differentiable at $F$with differential $dT(F)$,
b) $F_n \in E$ and $n^{\frac{1}{2}}(F_n - F)$ is tight,
Then

$$[T(F_n) - T(F)] = n^{-1}\sum IF(X_i) + o_p(n^{-\frac{1}{2}})$$

The following point should be noted here, 1) Since Hadamard differentiability of T at F implies continuity of T at F, $T(F_n) \overset{a.e.}{\to} T(F)$ if$F_n \overset{a.e.}{\to} F$ in the topology of the domain, and that implies $E_F(IF(X) = 0$ by the weak law of large numbers.
But more stringent Fréchet differentiability begets extra benefit in the expansion 10 as shown by Bednarski (1993). The expansion is valid if F_n comes from G_n belonging to shrinking nbd, $N = \{G_n : \|G_n - F\|_\infty \leq \frac{\epsilon}{\mathrm{n}^{\frac{1}{2}}}\}$,

$$[T(F_n) - T(F)] = n^{-1}\sum IF(X_i) + o_p(\|F_n - F\|_\infty) \tag{12}$$

So it is natural that uniformly Fréchet differentiable functionals as well as strongly Fréchet differentiable functionals enjoy more properties in this regard.

4 Third Heuristic

As like any function of population, $IF(x)$'s relevance to a particular sample can only be approximate. So there is an imperative to search for finite version of $IF(x)$ which enjoy similar asymptotic properties but more immediate relation to the actual samples. We can call them estimators of $IF(x)$. Several types of these estimators with different properties and use are available in literature (Andrews et al, 1972; Mallows, 1975; Hampel et al, 1986; Efron, 1992, Devison and Hinkly, 1997 etc.). Mainly they are used with the following objectives:

i) To have a first-hand idea of the effect of outliers on the statistic $T(F_n)$.

ii) To measure influence of individual observation on $T(F_n)$.

iii) To estimate asymptotic variance $\sigma^2(F) = \int IF(x)^2 dF < \infty$ of $T(F_n)$.

Among all these estimators of $IF(x)$, empirical influence function, sensitivity curves and jackknife influence functions are important ones for both theoretical and practical considerations. For details see Nasser and Alam (2006). Generally it is expected that the (stylized) sensitivity curve (defined in the following paragraph) of $T(F_n)$ tends to $IF(x)$ for all x. Davies (1993) disproves it for the same functional in the same article. Mallows (1975) mentioned, consistency property holds for all these estimators without proofs. Some counter-examples including a new one are also given in the next subsection.

4.1 Sensitivity Curves

Let us suppose

i) $x_1, x_2, \cdots, x_{n-1}$ be a sample, x be added to the sample and $Tn(.)$ be an estimator defined on a sample of size n. Let $I^1(x) = T_n(x_1, x_2, \cdots, x_{n-1}, x)$

ii) $x_1, x_2, \cdots, x_n$ be a sample and x_n be replaced by x, then $I^2(x) = T_n(x_1, x_2, \cdots, x_{n-1}, x)$.

$I^1(x)$ simply represents the value of the estimators when a new observation x is added to the sample (addition-corruption), while the $I^2(x)$upholds the value of the estimator after replacing a observation, say x_n by x (replacement corruption). Sensitivity curves, first introduced by Tukey (1970-1971), are nothing but translated and scaled versions of $I^1(x)$ and $I^2(x)$. We take the case of addition- corruption that is translated and scaled versions of $I^1(x)$, Sensitivity curve is defined as
$SC_n(x) = n[T_n(x_1, x_2, \cdots, x_{n-1}, x) - T_{n-1}(x_1, x_2, \cdots, x_{n-1})]$
When $T_{n-1}(x_1, x_2, \ldots, x_{n-1}) = T(F_{n-1})$

$$\begin{aligned} SC_n(x)) &= \frac{I^1(x) - T(F_{n-1})}{\frac{1}{n}} \\ &= \frac{T\left[(1-\frac{1}{n})F_{n-1} + \frac{1}{n}\delta_x\right] - T(F_{n-1})}{\frac{1}{n}} \\ &= g(\frac{1}{n}, x; F_{n-1}) \end{aligned}$$

where $g(\epsilon, x, F) = \frac{T[(1-\epsilon)F + \epsilon\delta_x] - T(F)}{\epsilon}$
Stylized sensitivity curve is based on an artificial sample of size $n-1$ instead a real sample of same size.

$$\begin{aligned} SC_n(x)) &= \frac{T\left[(1-\frac{1}{n})F_{n-1} + \frac{1}{n}\delta_x\right] - T(F_{n-1})}{\frac{1}{n}} \\ &= n\left(T\left[(1-\frac{1}{n})F_{n-1} + \frac{1}{n}\delta_x\right] - T(F_{n-1})\right) \end{aligned}$$

where $\quad F_{n-1} = \frac{1}{n-1}\sum_{j=1}^{n-1}\delta_{x_j}x_j = F^{-1}(j/n)$

4.2 Consistency of Sensitivity Curve

Taking the same functional in the counterexample in 3.2 we can easily show that sensitivity curve is not always a consistent estimator of $IF(x)$. Since $T_c[(1-\frac{1}{n})F_{n-1}+\frac{1}{n}\delta_x] = \frac{1}{n}$ a.e. and $T_c(F_{n-1}) = \frac{1}{n-1}$, it follows that $SC_n(x) = \frac{1}{n-1}$ a.e. but $IF(x) = 1$. Davies (1993) showed that for T_{MSH} at $N(0,1)\lim_{n\to\infty} SC_n(x) = 4IF(x)$ for stylized sensitivity curve. Croux (1998) demonstrated inconsistency of sensitivity of sample median but strong and uniform consistency for trimmed mean and smooth M-estimator. So to have consistent estimators we need extra conditions. It is meaningful to find sufficient conditions that guarantee consistency, specially strong and uniform consistency.
Now we mention two cases of functionals for which $SC_n(x)$ is a consistent estimators of $IF(x)$. Since in the first one consistency is strong, the related functional is better from the point under consideration.

Proposition 4.1 Let T be Gâteaux differentiable over neighborhood of F with continuous derivative and V, the vector space generated by $D_T(X)$ be topologized in such a way (Kolmogorov norm or Kuiper metric or weak topology are appropriate for this) that $\delta_x - F_n \overset{a.e.}{\to} \delta_x - F$. Then $SC_n(x) \overset{a.e.}{\to}= IF(x)$

Proof. Proof is straightforward.
The result implies that uniformly Fréchet differentiability of T over a neighborhood of F with respect to Kolmogorov norm is a desirable property.

Proposition 4.2 Let $\langle X, A\rangle$ be a measurable space. G be a probability measure on X, $a()$ be a random variable defined on X and $g : R \to R$ be a continuously differentiable function.

Let $g(\int a(x)dG(x)) = T(G)$ and $\int a(x)dG(x) = H(G)$. Then

i) $IF_H(x) = a(x) - H(G),\ SC_n^H = a(x) - H(G_n)$

ii) $IF_T(x) = g^{(1)}(a(x) - H(G)),\ SC_n^H = g^{(1)}(a(x) - H(G_n))$

iii) $SC_n^H(x) \overset{P}{\to} IF_H(x),\ SC_n^T(x) \overset{P}{\to} IF_T(x)$

Proof. Proof is straightforward.
The class of functionals that satisfy this proposition includes moments, variance, correlation and other important functionals.

5 Examples of Uniformly Fréchet Differentiable Functional

Now we place some examples of functionals which satisfy our conditions in the propositions.

Strong Fréchet diffferentiability of L-Functionals(L-estimator). Let us consider the functional, where $T(G) = \int xJ(G(x))dG(x) = \int G^{-1}(t)J(t)dt$ where $J(t)$ is a function defined on $[0,1]$. $T(F_n)$ is called L-estimator of $T(F)$. Following the line Boos (1979)who showed (theorem 1) under

a) $J(u)$ bounded and continuous a.e. Lebesgue and a.e F^{-1}

b) $J(u)$ vanishes for $u < \alpha$ and $u > \beta$ when $0 < \alpha < \beta < 1$. $T(G)$ has a Fréchet differential a w.r.to $\|\|_\infty$

$dT(F)(G-F) = \int_{-\infty}^{\infty}\Big(G(y) - F(y)\Big)J(F(y))dy$ i.e. its influence function

$dT(F)(\delta_x - F) = \int_{-\infty}^{\infty}(\delta_x(y) - F(y)J(F(y))dy$

Parr (1985) demonstrated that T is strongly Fréchet differentiable at under the same set of assumption. It is easy to show that $\psi_F(x)$ is continuous and of bounded variation. So it satisfies our requirement. Examples;

i) The trimmed mean with $J(t) = \frac{(\alpha_1 < t < 1-\alpha_2)}{1-\alpha_1-\alpha_2}$, obviously satisfies b. If $F$is such that $F^{-1}(\alpha_1)$ and $F^{-1}(1-\alpha_2)$ are uniquely determined, then a is satisfied.

ii) Let us cite another class of $J(t)$ which satisfies a,b and some other interesting properties,$J(t) = k(t-\alpha)^p(\beta-t)^p, 0 < \alpha < t < \beta < 1, p > 0$ and k is the normalized constant.Then the L-functional $T(G)$ can be shown to be uniformly differentiable on $\{G|\|G-F\|_\infty < \delta\}$ for all F where $\delta \leq \min\{\alpha, 1-\beta\}$ (Nasser, 2ooo).

M-Location and Scale Functionals. Davies in his illuminating article (1998) deduced two construction of locally uniformly linearizable high breakdown location and scale functionals. It can be shown that functionals in both case are Lipschitz continuous of order 1 (i.e. uniformly continuous) with respect to Kolmogorov norm (Kuiper metric) over a neighborhood $N_\delta(F)$ for all $W(\eta)$ (using Davies' notation).

M and Minimum Cramer-von Mises estimators of Location Functionals. Parr (1985) also presentd a M-location and a minimum Cramer-von-Mises location functional which meet our requirements.

Cramer -von Mises Test Statistic. Shao (1993) showed the following functional is strongly Fréchet differentiable at F with respect to $\|\|_\infty$. Let F_0 be a specified hypothetical distribution and $T(G) = \int |G(x) - F_0(x)|^2 dF_0(x)$ $T(F_n)$ is Cramer -von

Mises test statistic for the test problem. $H_0 : F = F_0$ vs $H_1 : H_0 : F \neq F_0$ Since here $\psi_F(y) = 2\int\{\delta_y(x) - F(x)\}\{F(x) - F_0(x)\}dF_0(x)$, (if F_0 is continuous) we have a) $\psi_F(y)$ is of bounded variation. b) $\psi_F(y)$ is continuous. Then $T(G)$ is strongly Fréchet differentiable at F and satisfies our required conditions.

References

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). *Robust Estimates of Location:Survey and Advances*, Princeton University Press, Priceton, N.J.

Bednarski, T. (1993). Fréchet differentiability of statistical functionals and implications to robust statistics, In: Morgenthaler, S, Ronchetti, E. and Stahel, W.A., (Eds.), *New Directions in Statistical Data Analysis and Robustness*, Besel, Birkhauser Verlag, 25-34.

Boos, D.D. (1979). A differential for L-statistics, *Annal. Statist.* 7, 955-959.

Croux, C. (1998). Limit behaviour of the empirical influence function of the median, *Statist.Probab. Lett.* 37, 331-340.

Davies, P.L. (1993). Aspects of robust regression, *Annal. Statist.* 21, 1843-1899.

Davies, P. L. (1994). Desirable properties, breakdown and efficiency in the linear regression model statistics *Probability Letters* 19, 361-370.

Davies, P.L. (1998). On locally uniformly linearizable high breakdown location and scale functionals, *A*nnal. Statist. 26, 1103-1125.

Davison, A.C and Hinkley, D.V. (1997): *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge CB2 2RU, U.K.

Efron, B. (1992). Jackknife-after-boostrapping standard errors and influence functions, *J.Royal Statist. Soc.* B 54, 83-127.

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. Stat.*, 42, 1887-1896.

Hampel, F.R. (1974). The influence curve and its role in robust estimation, *J. Amer. Statist.Assoc.* 62, 1179-1186.

Hampel, F.R., Rousseeuw, P.J., Ronchtti, E.M. and Stahel, W.A., (1986). *Robust Statistic: The Approach Based on Influence Functon,* Wiley, New York.

Huber, J. (1977). *Robust Statistical Procedures*, Regional Conference Series in Applied Mathematical No. 27, Soc. Industries. Appl. Math., Philadelphia, Penn.

Mallows, C.L. (1975). *On Some Topics in Robustness*, Technical Memorandum. Bell Telephone Laboratories, Murray Hill, N.J.

Nasser, M (2000). A new class of uniformly Fréchet differentials L-functions (L-estimator), *Proceedings of The seventh National Statistical conference on Statistics in the New Millennium*, Dhaka, 135-140.

Nasser, M and Alam, M.(2006). Estimators of influence functions Communications in Statistics-Theory and Methods, 35, 21-32.

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

Reeds, J. A. (1976). *On the definition of von Mises functionals*, Ph. D. thesis, Dept of Statistics Harvard University, Cambridge, Mass.

Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*, John Wiley and Sons, New York.

Shao, J. (1993). Differentiability of statistical functionals and consistency of the Jackknife, *Annal. Statist.* 21, 61-75.

Tukey, J.W. ((1970-71) 1977). *Exploratory Data Analysis (1970-71: preliminary edition).* Addison-Wesley, Reading. Mass.