Working Paper 92-35
September 1992

División de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624 9849

ON THE ESTIMATION OF THE INFLUENCE CURVE

Antonio Cuevas and Juan Romo*

Abstract

We prove the asymptotic validity of bootstrap confidence bands for the influence curve from its usual estimator (the sensitive curve). The proof is based on the use of Gill's (1989) *generalized delta method* for Hadamard differentiable operators. The scope and applicability of this result are also discussed.

Key words:
Influence curve, sensitivity curve, bootstrap confidence bands, Hadamard differentiability.

*Cuevas, Departamento de Matemáticas, Universidad Autónoma de Madrid, Spain; Romo, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Spain.

1. Introduction and background

It is well-known that, in many cases of practical interest, the estimators can be considered as restrictions of functionals defined on the space $\mathcal{F}$ of distribution functions. In fact, this idea goes back to the origins of mathematical statistics since it is implicit in the early notion of consistency proposed by Fisher. In precise terms, let $T_n = T_n(X_1, \ldots, X_n)$ be (for all $n = 1, 2, \ldots$) an estimator taking values in $\Re$, defined on random samples $X_1, \ldots, X_n$ from a univariate distribution. The sequence $\{T_n\}$ is said to be generated by a functional $T : \mathcal{F}_0 \subset \mathcal{F} \longmapsto \Re$, if for all n and for each sample $X_1, \ldots, X_n$, we have $T_n(X_1, \ldots, X_n) = T(F_n)$, where F_n is the empirical distribution associated with $X_1, \ldots, X_n$. Many usual estimators fulfil this condition; this is the case, for instance, of M- and L-estimators [see, e. g. Huber (1981)]. By $\mathcal{F}_n$ we will represent the set of empirical distributions of order n in $\mathcal{F}$, that is, the set of discrete probability measures in $\mathcal{F}$ whose atoms have probabilities equal to $1/n$ or to a multiple of $1/n$. Obviously, the domain $\mathcal{F}_0$ of T has to include $\mathcal{F}_n$ for all $n \in \mathcal{N}$.

In this setting, a natural idea is to use the differentiability properties of the functional T in order to get statistical results for the sequence $\{T_n\}$. The works of von Mises (1947) and Kallianpur and Rao (1945) are pioneering contributions on this topic but, in fact, the use of differentiation techniques only became really popular in the late sixties coinciding with the rapid development of the robustness theory. An important example is the so-called *influence function*, $T'(F; x)$ (of a functional T at a distribution $F \in \mathcal{F}$), which is nothing but the *partial derivative* of T along the *direction* corresponding to the degenerate distribution δ_x (for each x), that is, $T'(F; x) = \lim_{\epsilon \to 0^+} [T((1-\epsilon)F + \epsilon\delta_x) - T(F)]/\epsilon$ [see Hampel (1974), Hampel *et al.* (1987)]. If we assume that the sequence $\{T_n\}$of estimators generated by T is consistent, in probability under G (for each G), to $T(G)$ then $T'(F; x)$ represents (for small values of ϵ) the approximate value of *asymptotic bias* introduced by a contamination of type $(1-\epsilon)F + \epsilon\delta_x$ at the distribution F. Some quantitative measures of robustness (*gross-error sensitivity, local-shift sensitivity, rejection points*) are also defined from the influence curve.

However, in order to get a deeper insight into the meaning of the influence function, we need to impose (on T) further differentiability assumptions, stronger than the mere existence of $T'(F; x)$. This situation is similar to that of the classical analysis for functions $f : \Re^p \longrightarrow \Re$; the true significance of the gradient ∇f (which is the analog of the influence function) arises when we assume that f is differentiable since, in this case, ∇f defines the *best linear local approximant* of f.

The general concept of differentiability, for operators or functionals, is inspired on the same idea: let $\mathcal{G}$ and $\mathcal{D}$ be normed spaces and let $V : \mathcal{G} \longrightarrow \mathcal{D}$ be an operator. We will say that V is differentiable at $G \in \mathcal{G}$, with respect to a collection $\mathcal{S}$ of subsets of $\mathcal{G}$, if *there exists a linear and continuous map* $DV(G;.) : \mathcal{G} \longrightarrow \mathcal{D}$ (which we will call *the differential of* V *at* G) such that for Δ in some neighbourhood of zero,

$$V(G+\Delta) = V(G) + DV(G;\Delta) + R(G+\Delta),$$

where the remainder R satisfies

$$\lim_{t\to 0} \frac{R(G+t\Delta)}{t} = 0,$$

uniformly in $\Delta \in S$, for every $S \in \mathcal{S}$.

The most interesting particular cases correspond to the following choices of $\mathcal{S}$: $\mathcal{S}$ = all singletons of $\mathcal{G}$, $\mathcal{S}$ = all compact subsets of $\mathcal{G}$, and $\mathcal{S}$ = all bounded subsets of $\mathcal{G}$. They lead, respectively, to the concepts of *Gâteaux*, *Hadamard* (or *compact*) and *Fréchet* differentiability.

The application of these concepts (borrowed from the functional analysis) to statistical functionals $T : \mathcal{F} \longrightarrow \Re$, presents an obvious hurdle: $\mathcal{F}$ is not a normed space. A simple device to overcome this difficulty is embedding $\mathcal{F}$ in the space $\mathcal{G} \equiv \{\lambda(F-H) : F, H \in \mathcal{F}, \lambda \in \mathcal{R}\}$, endowed with the supremum norm.

The statistical functionals can be often extended in a natural way to the space $\mathcal{G}$ (or appropriate subspaces of it). In such cases the use of the above notions of differentiability is a very useful tool which allows to consider the influence function from a different perspective. Moreover, if the functional T is Fréchet (or Hadamard) differentiable at F and the differential can be expressed in the form

$$DT(F;\Delta) = \int_{-\infty}^{\infty} \Psi(x) d\Delta(x),$$

then it is not difficult to prove [see Boos and Serfling (1980)] that $\Psi(x)$ coincides with the influence curve, and the sequence $\{T_n\}$ of estimators generated by T is asymptotically normal with asymptotic variance

$$\int_{-\infty}^{\infty} T'(F;x)^2 dF(x).$$

This is, perhaps, the most important point in connection with the influence curve: under standard conditions *the asymptotic variance can be expressed in terms of* $T'(F;x)$. In particular, the estimates of the influence curve are potentially useful in the estimation of the asymptotic variance [see Presedo (1991)].

The choice between Hadamard or Fréchet differential in each particular application is usually guided by technical considerations. In general terms, Fréchet differential is more natural and easier to handle. Some applications can be found in Kallianpur and Rao (1955), Boos and Serfling (1980), Clarke (1986), Parr (1985) and Arcones and Giné (1992). Nevertheless, the compact differentiation has, in principle, a broader applicability since it imposes a weaker (less restrictive) condition; it is, in fact, the weakest notion of differential which is still manageable in the sense of fulfilling the chain rule. For applications, see Fernholz (1983), Esty *et al.* (1985) and Gill (1989).

In this paper we use Hadamard differential to prove (in Section 2 below) the validity of bootstrap confidence bands for the standard estimator of the influence curve. The basic tools used in the proof are the results on bootstrap of empirical processes [see Giné and Zinn (1990)] and the *generalized delta method* established by Gill (1989). Section 3 contains some final remarks.

2. BOOTSTRAP CONFIDENCE BANDS FOR THE INFLUENCE CURVE

We consider now the problem of estimating the influence curve $T'(F;x)$ from a random sample $X_1,\ldots,X_n$ of F. Three estimators have been considered in the literature: the *sensitivity curve*, the *empirical influence curve*, and the *jackknife approximation* [see Hampel *et al.* (1987), p. 92]. The first one is perhaps the most popular: it is defined by

$$SC_n(x) = \frac{T((1-\frac{1}{n})F_{n-1} + \frac{1}{n}\delta_x) - T(F_{n-1})}{1/n}.$$

Curiously enough, the asymptotic properties of this estimate have received little attention; maybe the reason is that the influence curve is often used for descriptive aims, in order to get a general idea of the behavior of the sequence

T_n associated with T. However, if we want to use $T'(F;x)$ for quantitative purposes (calculation of the asymptotic variance or the gross-error sensitivity, for instance), we need to have precise statements of consistency for the estimates of the influence curve.

Our target here is to obtain bootstrap confidence bands for $T'(F;x)$ from the estimator $SC_n(x)$. To be more concrete, we want to calculate (at least approximately) the sampling distribution of the statistic

$$D_n = \sup_x \sqrt{n} \mid SC_n(x) - T'(F;x) \mid .$$

We will use bootstrap methodology, that is, we will approximate the distribution of D_n under F by that of its bootstrap version

$$D_n^* = \sup_x \sqrt{n} \mid SC_n^*(x) - SC_n(x) \mid ,$$

under F_n, where $SC_n^*(x)$ denotes the sensitivity curve $SC_n(x)$ calculated from the bootstrap sample $X_1^*, \ldots, X_{n-1}^*$ (whose empirical distribution is represented by F_{n-1}^*), which is drawn by resampling from the original data $X_1, \ldots, X_{n-1}$.

The validity of such an approximation is established in Theorem 1 below. First, we introduce an auxiliary family of functionals defined by

$$\tau_x(G;t) = T[(1-t)G + t\delta_x], \qquad t \geq 0.$$

Observe that

$$T'(F;x) = \frac{d}{dt}\tau_x(G;t) \mid_{t=0}$$

In the sequel, the derivative $\frac{d}{dt}\tau_x(G;t)$ will be denoted by $\tau_x'(G;t)$. Thus, $T'(F;x) = \tau_x'(G;0)$.

THEOREM 1. Let $D(\bar{\mathcal{R}})(\equiv D[-\infty + \infty])$ be the space of *cadlag* (i.e. right continuous with left-hand side limits) functions endowed with the $\| \cdot \|_\infty$ (essential supremum) norm. Let $T : \mathcal{F} \to \mathcal{R}$ be a statistical functional with associated influence functional $T'(F;x)$. Assume that for each $F \in \mathcal{F}$, we have

(i) there exist constants $C \in \Re$ and $\eta > \frac{1}{2}$ such that for all empirical distribution F_n close enough (in the weak topology) to F ,

$$| \tau_x'(F_n;t) - \tau_x'(F_n;0) | \leq Ct^\eta, \qquad \textit{for all} \;\; 0 < t \leq \frac{1}{n} \qquad a.s.$$

(ii) the influence function $T'(F;\cdot)$ belongs to $D(\bar{\mathcal{R}})$ (in particular, it is bounded),

(iii) the influence functional $T'(F;\cdot)$ can be extended to the vector space $\mathcal{G}$ (the linear span of F, defined above) and the transformation (from $\mathcal{G}$ to $D(\bar{\mathcal{R}})$) : $H \mapsto T'(H;\cdot)$ is continuously Hadamard differentiable.

Then, the statistic D_n *can be bootstrapped*, in the sense that its bootstrap version D_n^* converges weakly a.s. to the same limit as D_n does.

PROOF. As a previous step in the proof, we study the asymptotic behavior of $T'(F_n;\cdot)$ (a natural estimator of the influence curve). From Doob-Donsker's theorem,

$$\sqrt{n}(F_n - F) \longrightarrow B^0(F),$$

weakly in $D(\bar{\mathcal{R}})$, where B^0 is the Brownian bridge on [0,1] considered as a random element [see, e.g. Pollard (1984, p. 97)]. The $\delta - method$ [as in Gill (1989)] gives that

$$\sqrt{n}(T'(F_n;\cdot) - T'(F;\cdot)) \longrightarrow_w DT'(F;\cdot)B^0(F), \tag{1}$$

where $DT'(F;\cdot)$ is the Hadamard differential of T'. From the general result by Giné and Zinn (1990) on bootstrap of empirical measures, we have that

$$\sqrt{n}(F_n^* - F_n) \longrightarrow_w B^0(F), \qquad a.s.$$

in $D(\bar{\mathcal{R}})$ and, again from Gill (1989) and hypothesis (iii), it follows that

$$\sqrt{n}(T'(F_n^*;\cdot) - T'(F_n;\cdot)) \longrightarrow_w DT'(F;\cdot)B^0(F). \qquad a.s. \tag{2}$$

Note that (1) and (2) mean that the bootstrap works for the "plug-in" estimator $T'(F_n;\cdot)$ of the influence curve. The analogous result for the sensitivity curve will be established if we prove that

$$\sqrt{n}(SC_n(\cdot) - T'(F;\cdot)) \longrightarrow_w DT'(F;\cdot)B^0(F) \tag{3}$$

and

$$\sqrt{n}(SC_n^*(\cdot) - SC_n(\cdot)) \longrightarrow_w DT'(F;\cdot)B^0(F) \qquad a.s. \tag{4}$$

Since

$$\sqrt{n}(SC_n(x) - T'(F;x)) = \sqrt{n}\left(\frac{T((1-\frac{1}{n})F_{n-1} + \frac{1}{n}\delta_x) - T(F_{n-1})}{1/n} - T'(F;x)\right), \tag{5}$$

by applying the Mean Value Theorem, (5) equals to

$$\sqrt{n}(\tau_x'(F_{n-1};t_n) - \tau_x'(F;0)) =$$

$$= \sqrt{n}(\tau_x'(F_{n-1};0) - \tau_x'(F;0)) + \sqrt{n}(\tau_x'(F_{n-1};t_n) - \tau_x'(F_{n-1};0)),$$

for some $t_n \in (0, \frac{1}{n})$. As seen before, by Gill's δ-method, the first term in the sum tends to $DT'(F;\cdot)B^0(F)$ and by hypothesis (i), the second one tends to zero; so (3) follows. To get (4), it is enough to show that

$$\| \sqrt{n}(SC_n^*(\cdot) - SC_n(\cdot)), \sqrt{n}(T'(F_n^*;\cdot) - T'(F_n;\cdot)) \|_\infty \longrightarrow_{a.s.} 0 \qquad a.s. \tag{6}$$

Again, by applying the Mean Value Theorem in (6), the left hand side equals to

$$\| (\sqrt{n}(\tau_\cdot'(F_n^*;t_n^*) - \tau_\cdot'(F_n;t_n)) - \sqrt{n}(\tau_\cdot'(F_n^*;0) - \tau_\cdot'(F_n;0)) \|_\infty \leq$$

$$\leq \sqrt{n} \| (\tau_\cdot'(F_n^*;t_n^*) - \tau_\cdot'(F_n^*;0)) \|_\infty + \sqrt{n} \| \tau_\cdot'(F_n;t_n) - \tau_\cdot'(F_n;0) \|_\infty \leq$$

$$\leq C(t_n^{*\eta} + t_n^\eta) \longrightarrow_{a.s.} 0 \qquad a.s. \tag{7}$$

Now, from (2) and (7), we get that

$$\sqrt{n}(SC_n^*(\cdot) - SC_n(\cdot)) \longrightarrow_w DT'(F;\cdot)B^0(F) \qquad a.s. \tag{8}$$

Since $\| \cdot \|_\infty$ is continuous with respect to its own topology, by using the Continuous Mapping Theorem [see, e.g. Pollard (1984), p.44], we conclude from (3) and (8) that D_n and D_n^* converge weakly a.s. to the same limit and the result follows.

Some remarks:

(a) Hypothesis (i) is a not very restrictive regularity assumption. It is fulfilled, for instance, by the sample mean $\bar{X}$ as well as by any other statistic defined as a smooth function of $\bar{X}$.

(b) Assumptions (ii) and (iii) hold for the important class of L-estimators, whose influence curve has the expression [see, e.g. Huber (1981), p. 57]

$$T'(F;x) = \int_{-\infty}^{x} h'(y)m(F(y))dy - \int_{-\infty}^{\infty}(1-F(y))h'(y)m(F(y))dy. \quad (9)$$

Indeed, assuming that h' is integrable and m is continuously differentiable, one can check out that the expression for this influence curve given above is just a composition of continuously differentiable operators: observe that the transformation $F \longmapsto \int_{-\infty}^{x} h'(y)F(y)dy$ is linear and continuous and, hence, differentiable. Also, $F \mapsto m(F(\cdot))$ is a differentiable map.

As for the M-estimators, the form of their influence function is, under some regularity assumptions [see Deniau *et al* (1977)]

$$T'(F;x) = \frac{\Psi(x;T(F))}{-\int \frac{\partial}{\partial\theta}\Psi(y;\theta)\mid_{\theta=T(F)} dF(y)}. \quad (10)$$

Hypotheses (i)-(iii) in the theorem could be checked for particular choices of Ψ. However, it doesn't seem straightforward to give a simple general condition on Ψ ensuring the validity of these hypotheses.

(c) In those cases where the influence function $T'(F;.)$ is not bounded, Theorem 1 still holds in order to provide confidence bands on compact intervals $[-M,M]$. It would suffice to replace $\mathcal{D}(\bar{\mathcal{R}})$ by the corresponding space $\mathcal{D}[-M,M]$.

3. Final remarks

(a) From a methodological point of view, Theorem 1 presents the interesting feature of providing a relatively simple application of Gill's (1989) generalized delta method. Although, in principle, this method applies to estimators of type $T(F_n)$ (defined as restrictions of operators), the proof of Theorem 1 shows that the applicability of this technique can sometimes be extended to more general cases by using standard differentiability arguments.

(b) In the proof of Theorem 1 arises, as an auxiliary tool, the *plug-in* estimator $T'(F_n; x)$. It is worth mentioning that this estimator could be of practical interest in those situations where the functional form of the influence function $T'(F; x)$ is known in advance. This is the case, for example, of the M- and L-estimators mentioned above [see expressions (9) and (10)]. While the proof of Theorem 1 suggests that the plug-in estimator and the sensitivity curve are asymptotically equivalent, a higher finite-sample efficiency is to be expected for the first one. In any case, the detailed comparative study of the different estimators for the influence curve seems to be an interesting open problem. In particular, as indicated in the introduction, the influence curve is closely related with the asymptotic variance: so, every different estimator of the influence curve provides an estimator for the asymptotic variance. In particular, note that, as an application of Theorem 1, bootstrap confidence intervals could be obtained for the asymptotic variance $\sigma_T^2(F) = \int_{-\infty}^{\infty} T'(F; x)^2 dF(x)$.

REFERENCES

ARCONES, M.A. and GINE, E. (1992). On the bootstrap of M-estimators and other statistical functionals. *Proc. IMS Bootstrap Conference, East Lansing 1990.*

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.

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

DENIAU, C., OPPENHEIM, G. and VIANO, C. (1977). Courbe d'influence et sensitivité. *Asterisque* **43-44**, 151-187.

ESTY, W., GILLETTE, R., HAMILTON, M. and TAYLOR, D. (1985). Asymptotic distribution theory of statistical functionals: the compact derivative approach for robust estimators. *Ann. Math. Statist.* **37 A**, 109-129.

FERNHOLZ, L. T. (1983). *von Mises Calculus for Statistical Functionals.* Springer-Verlag, New York.

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part I) *Scand. J. Statist.*, **16**, 97-128.

GINE, E. and ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Prob.* **18**, 851-869.

HAMPEL, F.R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69**, 383-393.

HAMPEL, F.R., RONCHETTI, E.M., ROUSSEEUW, P.J. and STAHEL, W.A. (1987). *Robust Statistics. The Approach Based on Influence Functions.* Wiley, New York.

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

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

MISES, R. von (1947). On the asymptotic distributions of differentiable statistical functions. *Ann. Math. Statist.* **18**, 309-348.

PARR, W. C. (1985). The bootstrap: some sample theory and connections with robustness. *Stat. Prob. Letters*, **3**, 97-100.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

PRESEDO, M.A. (1991). Un análisis mediante la función de influencia de la estimación no paramétrica de las funciones de regresión. Ph. D. dissertation, Univ. of Santiago de Compostela.