Robustness by generalized influence functions: a new approach

Matteo Fini[1], Davide La Torre[1]

[1] Department of Economics, Business and Statistics, University of Milan
(e-mail: matteo.fini@unimi.it, davide.latorre@unimi.it)

ABSTRACT: The Influence Function (IF) was introduced by Hampel in 1968. On this definition are based many stability and robustness results of statistical estimators. In this paper we extend it to more general settings.

KEYWORDS: Robustness, Influence Functions, Set Valued Analysis

1 The Influence Function (IF): the definition

The classical notion of Influence Function was introduced by Hampel in 1968 (see the references) and it is based on the classical notion of Gateaux derivative of a functional on a vector space. Let $(X_1,\ldots,X_n)$ be a set of random variables i.i.d. and a $(x_1,\ldots,x_n)$ be a sample of observations. The observations belong to the classical sample space X. $\mathcal{F}(X)$ is the set of all probability measures defined on the classical probability space (X,B). $\theta \in \Theta$ is the unknown parameter of the parameter space, F_θ is the classical distribution function and f_θ the density function. The IF is used to analyze the behavior of an estimator when a new observation x is introduced or one observation slides from the supposed distribution. The classical definition of IF (see Hampel for details) is

$$IF(x;T,F) = \lim_{t \downarrow 0} \frac{T((1-t)F + t\delta_x) - T(F)}{t} \tag{1}$$

with $x \in X$ (if the previous limit exists), where T is the estimator, F is the distribution function and δ_x is the Dirac measure at the point x. What we wish to study here is how to generalize this concept to a more general setting. There are examples in which this definition fails; these are the cases in which the functional is not Gateaux differentiable and the previous limit doesn't exist. To do this we will use the notion of Dini derivative of a vector function which is the topic of the next section.

2 The first order Dini directional derivative

Many papers in literature deal with extensions of classical derivatives for nonsmooth vector functions $f : X \to Y$, where X and Y are real vector spaces. Here we are interested in the notion of Dini because it is the easiest and simplest among all definitions. The underlying idea is to regard the set of all cluster points of first order incremental ratio.

Definition 2.1. *Let $f : X \to Y$ be a given function and $x_0 \in X$. Dini generalized derivative f'_D at x_0 in the direction $d \in X$ is*

$$f'_D(x_0;d) = \left\{ l : l = \lim_{k \to +\infty} \frac{f(x_0 + t_k d) - f(x_0)}{t_k}, t_k \downarrow 0 \right\}. \tag{2}$$

We first observe that $f'_D(x_0;d) \subset Y$ and so $f'(x;d) : X \rightrightarrows Y$ for all $x \in X$. Obviously this set is nonempty when locally lipschitz functions are considered and $\dim Y$ is finite.

3 A notion of Generalized Influence Functions (GIF)

The previous definition of Dini derivative leads to consider a generalized influence function (GIF) which is a multifunction. This is what we pay if we want to avoid the request that the previous limit exists at each $x \in X$. Consider the following definition

$$IF(x;T,F) = \left\{ l : l = \lim_{t_n \downarrow 0} \frac{T((1-t_n)F + t_n \delta_x) - T(F)}{t_n} \right\} \tag{3}$$

and so

$$IF(x,T,F) : X \rightrightarrows \mathbb{R} \tag{4}$$

For this definition we will prove a generalized Taylor expansion and some results concerning stability and robustness of statistical estimators.

References

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J, Stahel, W.A. 1984, *Robust statistica. The approach based on influence functions*, Wiley.

Hampel, F.R. 1968, *Contributions to the theory of robust estimation*, Ph.D. thesis. Berkley.

Hampel, F.R. 1971, *A general qualitative definition of robustness*, Ann. Math. Statist. 42.

Hampel, F.R. 1974, *The influence curve and its role in robust estimation*, J. Am. Statist. Assoc. 69.

Pires, A.M., Branco, J.A. 2002, *Partial influence functions*, J. Multivariate Anal, 83.