Partial Influence Functions

Ana M. Pires and João A. Branco

Department of Mathematics and Applied Mathematics Centre, Instituto Superior Técnico, Technical University of Lisbon, Av. Rovisco Pais, 1049-001 Lisboa, Portugal
E-mail: Ana.Pires@math.ist.utl.pt

Abstract:

In this paper we extend the definition of the influence function to functionals of more than one distribution, that is for estimators depending on more than one sample, such as the pooled variance, the pooled covariance matrix and the linear discriminant analysis coefficients. In this case the appropriate designation should be "partial influence functions", following the analogy with derivatives and partial derivatives. Some useful results are derived, such as an asymptotic variance formula. These results are then applied to several estimators of the Mahalanobis distance between two populations and the linear discriminant function coefficients.

Key Words: asymptotic variance, influence function, linear discriminant function, Mahalanobis distance, robust estimators

1. INTRODUCTION

The idea of influence function is recent (Hampel, 1974) but nowadays it is a well established statistical tool. It is fundamental for robustness studies (Hampel, Ronchetti, Rousseeuw and Stahel, 1986) and it is becoming increasingly important in the area of diagnostics through one of its several empirical versions (Cook and Weisberg, 1982; Critchley, 1985; Critchley and Vitiello, 1991; He and Simpson, 1992; Fung, 1996; Lu, Ko and Chang, 1997).

Although influence functions have already been used for estimators that depend on more than one sample (Campbell, 1978; Radhakrishan and Kshirsagar, 1981; Radhakrishan, 1983; Critchley and Vitiello, 1991; Fung, 1996; Rousseeuw and Ronchetti, 1981, Hampel *et al.*, 1986) we have felt that an explicit definition is still missing and that this definition will allow all the potential of the influence function to be disclosed and adequately explored. In Section 2 this definition and the derived formulae are presented. Following the analogy with derivatives and partial derivatives the designation “partial influence functions” is introduced (note, however, that this use of the term “partial” differs from that of Rieder, 1994). In order to simplify the presentation we have considered only the two-sample case. The results, however, can be immediately extended to the k-sample case, $k > 2$. For the same reason the proofs were removed to the Appendix. Section 3 discusses, as examples, the Mahalanobis distance between two populations and the coefficients of the linear discriminant function. The conclusions are drawn in Section 4.

2. DEFINITION AND PROPERTIES

We will consider only estimators which are functionals. Let $\mathcal{X}$ be a sample space contained in $\mathbb{R}^m$. Let $\mathcal{F}$ be the set of all finite signed measures on $\mathcal{X}$ and let $\boldsymbol{T} = \boldsymbol{T}(F_1, F_2)$ be a functional with domain $\mathcal{D}(\boldsymbol{T}) \subset \mathcal{F} \times \mathcal{F}$ taking values in $\mathbb{R}^p$ (m is the dimension of the sample space and p is the dimension of the parameter space). We assume that $\mathcal{D}(\boldsymbol{T})$ is a convex set containing more than one element. In the following $\Delta_{\boldsymbol{x}}$ denotes the probability measure which puts mass 1 at the point $\boldsymbol{x} \in \mathcal{X}$.

Definition 2.1. The partial influence functions of the functional $\boldsymbol{T}$ at $(F_1, F_2) \in \mathcal{D}(\boldsymbol{T})$, with relation to F_1 and F_2, respectively, are given by:

$$IF_1(\boldsymbol{x}; \boldsymbol{T}, F_1, F_2) = \lim_{\varepsilon \to 0} \frac{\boldsymbol{T}[(1-\varepsilon)F_1 + \varepsilon \Delta_{\boldsymbol{x}}, F_2] - \boldsymbol{T}(F_1, F_2)}{\varepsilon},$$

$$IF_2(\boldsymbol{x}; \boldsymbol{T}, F_1, F_2) = \lim_{\varepsilon \to 0} \frac{\boldsymbol{T}[F_1, (1-\varepsilon)F_2 + \varepsilon \Delta_{\boldsymbol{x}}] - \boldsymbol{T}(F_1, F_2)}{\varepsilon},$$

in those $\boldsymbol{x} \in \mathcal{X}$ where each limit exists.

IF_i is thus a function $\mathcal{X} \subset \mathbb{R}^m \longrightarrow \mathbb{R}^p$. By analogy with the one-sample case, the heuristic interpretation is that each component of $IF_i(\boldsymbol{x}; \boldsymbol{T}, F_1, F_2)$ measures, approximately, n_i times the change on the corresponding component of $\boldsymbol{T}$ caused by an additional observation in $\boldsymbol{x}$, in the i-th sample, when $\boldsymbol{T}$ is applied to a large combined sample of (n_1, n_2) observations. Or, equivalently, $(n_1 + n_2)w_i$ times the same change, where $w_i = n_i/(n_1 + n_2)$.

The idea of partial influence function appears in Hampel *et al.* (1986) (although this designation is not used) but it deserves only light consideration within the definition of the influence function for two-sample tests.

The partial influence functions are closely related (as in the one-sample case) with the (partial) Gâteaux differentiability.

Definition 2.2. A functional $\boldsymbol{T}(F_1, F_2)$ is said to be partially Gâteaux able at $(F_1, F_2) \in \mathcal{D}(\boldsymbol{T})$, with relation to F_1, if there is a function $a_{11}(\boldsymbol{x})$:

$\mathcal{X} \longrightarrow \mathbb{R}^p$ such that for all $(G_1, F_2) \in \mathcal{D}(\boldsymbol{T})$ it holds that

$$\lim_{\varepsilon \to 0} \frac{\boldsymbol{T}[(1-\varepsilon)F_1 + \varepsilon G_1, F_2] - \boldsymbol{T}(F_1, F_2)}{\varepsilon} = \int a_{11}(\boldsymbol{x}) dG_1(\boldsymbol{x})$$

and it is said to be partially Gateaux differentiable at (F_1, F_2), with relation to F_2, if there is a function $a_{12}(\boldsymbol{x}) : \mathcal{X} \longrightarrow \mathbb{R}^p$ such that for all $(F_1, G_2) \in \mathcal{D}(\boldsymbol{T})$ it holds that

$$\lim_{\varepsilon \to 0} \frac{\boldsymbol{T}[F_1, (1-\varepsilon)F_2 + \varepsilon G_2] - \boldsymbol{T}(F_1, F_2)}{\varepsilon} = \int a_{12}(\boldsymbol{x}) dG_2(\boldsymbol{x}),$$

$a_{11}(\boldsymbol{x})$ and $a_{12}(\boldsymbol{x})$ are called the first (partial) kernel functions.

Comparing Definitions 1 and 2 we see that if $\boldsymbol{T}(F_1, F_2)$ is partially Gâteaux differentiable at (F_1, F_2), with relation to F_1 and F_2, and if both $(\Delta_{\boldsymbol{x}}, F_2)$ and $(F_1, \Delta_{\boldsymbol{x}})$ belong to $\mathcal{D}(\boldsymbol{T})$ then

$$a_{1i}(\boldsymbol{x}) \equiv IF_i(\boldsymbol{x}; \boldsymbol{T}, F_1, F_2), \quad i = 1, 2,$$

and

$$\int a_{1i}(\boldsymbol{x}) dF_i(\boldsymbol{x}) = \int IF_i(\boldsymbol{x}; \boldsymbol{T}, F_1, F_2) dF_i(\boldsymbol{x}) = 0, \quad i = 1, 2. \tag{1}$$

The next theorem gives the first order Taylor formula for the functional $\boldsymbol{T}$ evaluated at a distribution (G_1, G_2) "near" (F_1, F_2), also known as the first order von Mises expansion.

Theorem 2.1. *Given $(F_1, F_2), (G_1, G_2) \in \mathcal{D}(\boldsymbol{T})$ if the following regularity conditions are verified*

(R1)there is $\delta > 0$ such that $\boldsymbol{T}$ is partially Gâteaux differentiable at $((1-t_1)F_1 + t_1 G_1, (1-t_2)F_2 + t_2 G_2)$, for all $0 \leq t_1, t_2 < 1$ and $|t_1 - t_2| < \delta$, with relation to F_1 and F_2;

(R2)$IF_i(\boldsymbol{x}; \boldsymbol{T}, (1-t)F_1 + tG_1, (1-t)F_2 + tG_2)$, $i = 1, 2$, are continuous functions of t;

(R3)The functions

$$g_i(t) = \frac{\partial}{\partial t} \int IF_i(\boldsymbol{x}; \boldsymbol{T}, (1-t)F_1 + tG_1, (1-t)F_2 + tG_2) d(G_i - F_i)(\boldsymbol{x}), \quad i = 1, 2$$

are well defined for all $0 < t < 1$;

(R4)$IF_i(\boldsymbol{x}; \boldsymbol{T}, (1-t-\Delta t)F_1 + (t+\Delta t)G_1, (1-t-\Delta t)F_2 + (t+\Delta t)G_2)$, is dominated as $\Delta t \to 0$ by an integrable function relatively to $d(G_i - F_i)(\boldsymbol{x})$, $i = 1, 2$.

then

$$\begin{aligned}\boldsymbol{T}(G_1, G_2) &= \boldsymbol{T}(F_1, F_2) + \sum_{i=1}^{2} \int IF_i(\boldsymbol{x}; \boldsymbol{T}, F_1, F_2) dG_i(\boldsymbol{x}) + \\ &+ R\left((G_1, G_2) - (F_1, F_2)\right)\end{aligned}$$

and there is $0 < c < 1$ *such that*

$$R\left((G_1, G_2) - (F_1, F_2)\right) = \frac{g_1(c) + g_2(c)}{2}.$$

The proof is given in the Appendix. An important result is obtained when the theorem is applied to the empirical distribution (F_{n_1}, F_{n_2}):

$$\begin{aligned}\boldsymbol{T}(F_{n_1}, F_{n_2}) &= \boldsymbol{T}(F_1, F_2) + \sum_{i=1}^{2} \int IF_i(\boldsymbol{x}; \boldsymbol{T}, F_1, F_2) dF_{n_i}(\boldsymbol{x}) + \\ &+ \mathrm{R}\left((F_{n_1}, F_{n_2} - (F_1, F_2)\right) = \\ &= \boldsymbol{T}(F_1, F_2) + \sum_{i=1}^{2} \sum_{j=1}^{n_i} \frac{IF_i(\boldsymbol{x}_{ij}; \boldsymbol{T}, F_1, F_2)}{n_i} + \\ &+ \mathrm{R}\left((F_{n_1}, F_{n_2}) - (F_1, F_2)\right).\end{aligned}$$

Then, if $\boldsymbol{x}_{ij} \overset{\text{iid}}{\sim} F_i$, $i = 1, 2$, and $\mathrm{R}\left((F_{n_1}, F_{n_2}) - (F_1, F_2)\right) \xrightarrow{P} 0$, by the Central Limit Theorem we can conclude that

$$\sqrt{n_1 + n_2}\left(\boldsymbol{T}(F_{n_1}, F_{n_2}) - \boldsymbol{T}(F_1, F_2)\right)$$

is asymptotically normal with null mean vector and covariance matrix $(p \times p)$ given by

$$\begin{aligned}V(\boldsymbol{T}, F_1, F_2) &= \lim_{n_1+n_2 \to \infty, n_1/n_2 = w} (n_1 + n_2) \mathrm{var}\, \boldsymbol{T}(F_{n_1}, F_{n_2}) = \\ &= (n_1 + n_2) \frac{n_1 \mathrm{var}\, IF_1}{n_1^2} + (n_1 + n_2) \frac{n_2 \mathrm{var}\, IF_2}{n_2^2} = \\ &= \frac{1}{w_1} V_1(\boldsymbol{T}, F_1, F_2) + \frac{1}{w_2} V_2(\boldsymbol{T}, F_1, F_2) \qquad (2)\end{aligned}$$

where $w_i = n_i/(n_1 + n_2)$, $w = n_1/n_2 = w_1/w_2$ remains fixed, and

$$V_i(\boldsymbol{T}, F_1, F_2) = \int IF_i(\boldsymbol{x}; \boldsymbol{T}, F_1, F_2) IF_i(\boldsymbol{x}; \boldsymbol{T}, F_1, F_2)^T dF_i(\boldsymbol{x}) \quad (i = 1, 2). \qquad (3)$$

A development that has points of contact with the one presented here is the expansion for multisample based U-statistics which can be found in Sen (1981, section 3.6). In the case of an estimator that can, simultaneously, be represented by a differentiable functional and a U-statistic, both approaches are equivalent. However, there are cases for which this theory becomes extremely difficult (such as some of the robust estimators considered in the examples in Section 3, or the projection-pursuit estimators referred in Section 4, or the multi-sample estimators for a common covariance matrix proposed by Hawkins and McLachlan, 1997, and He and Fung, 2000). Moreover, the influence function has a role to play as a diagnostic tool which must not be forgotten.

The asymptotic variance formula here derived has also strong connections with the delta method and the jackknife procedure, like in the one-sample case.

3. EXAMPLES

3.1. Mahalanobis distance

An important parameter in multivariate analysis is the squared Mahalanobis distance (Δ^2) between two populations, or groups, with distributions F_1 and F_2, expectations $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively, and common non-singular covariance matrix $\boldsymbol{\Sigma}$: $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. A natural estimator of this parameter is

$$D^2_{\boldsymbol{T},\boldsymbol{S}}(F_{n_1}, F_{n_2}) = (\boldsymbol{T}(F_{n_1}) - \boldsymbol{T}(F_{n_2}))^T \boldsymbol{S}_c^{-1}(F_{n_1}, F_{n_2}) (\boldsymbol{T}(F_{n_1}) - \boldsymbol{T}(F_{n_2})),$$

where $\boldsymbol{T}(F_{n_1})$ and $\boldsymbol{T}(F_{n_2})$ are multivariate estimators of location for each population and $\boldsymbol{S}_c(F_{n_1}, F_{n_2})$ is an estimator of the common multivariate scatter, generally of the form $w_1\boldsymbol{S}(F_{n_1}) + w_2\boldsymbol{S}(F_{n_2})$, with $\boldsymbol{S}(F_{n_i})$ being a multivariate estimator of scatter for population i. If all these estimators are equivalent to a functional and Fisher consistent then $D^2_{\boldsymbol{T},\boldsymbol{S}}(F_1, F_2) = \Delta^2$. The partial influence functions of the functional D^2 can be easily determined by the usual differentiation rules.

Noticing that Δ^2 is an affine invariant parameter and that D^2 is an affine invariant estimator of Δ^2 if $\boldsymbol{T}$ and $\boldsymbol{S}$ are affine equivariant (which is usually the case), it is sufficient to study the influence functions at "central" model distributions F_1^* and F_2^*. We will consider these "central" distributions to be those — within the family of F_1 and F_2 — having location parameters $\boldsymbol{\mu}_1^* = (\Delta/2, 0, \ldots, 0)^T$ and $\boldsymbol{\mu}_2^* = (-\Delta/2, 0, \ldots, 0)^T$, respectively, and common covariance matrix $\boldsymbol{\Sigma}^* = \boldsymbol{I}$. There is an affine (non-singular) transformation, $\boldsymbol{x}_i^* = \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{b}$, such that if $\boldsymbol{x}_i \sim F_i$ then $\boldsymbol{x}_i^* \sim F_i^*$. It is easy to verify that $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}^{-1/2}$ and $\boldsymbol{b} = -\boldsymbol{A}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2$, where $\boldsymbol{\Sigma}^{-1/2}$ is the unique symmetric matrix such that $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{U}$ is a non-unique orthogonal matrix such that its first row is $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1/2}/\Delta$. Because of the invariance of D^2 we then have

$$IF_i(\boldsymbol{x}; D^2_{\boldsymbol{T},\boldsymbol{S}}, F_1, F_2) = IF_i(\boldsymbol{x}^*; D^2_{\boldsymbol{T},\boldsymbol{S}}, F_1^*, F_2^*) = \\ = 2(-1)^{i+1}\Delta IF(\boldsymbol{x}^*; T_1, F_i^*) - w_i\Delta^2 IF(\boldsymbol{x}^*; S_{11}, F_i^*). \tag{4}$$

For the classical estimators, that is, the sample mean, $\boldsymbol{T}(F_{n_i}) = \bar{\boldsymbol{x}}_i = \sum_j \boldsymbol{x}_{ij}/n_i$, and the sample pooled covariance matrix (not adjusted),

$$\boldsymbol{S}_c(F_{n_1}, F_{n_2}) = w_1\boldsymbol{S}_1 + w_2\boldsymbol{S}_2,$$

with $\boldsymbol{S}_i = \sum_j(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)^T/n_i$, it is well known that

$$IF(\boldsymbol{x}^*; T_1, F_i^*) = x_1^* - \mu_{i1}^* \quad \text{and} \quad IF(\boldsymbol{x}^*; S_{11}, F_i^*) = (x_1^* - \mu_{i1}^*)^2 - 1.$$

The non-robustness of $D^2_{\bar{\boldsymbol{x}},\boldsymbol{S}}$ comes from the fact that these influence functions are unbounded. However, it should be noted that, because the large values of $IF_i(\boldsymbol{x}^*; D^2_{\bar{\boldsymbol{x}},\boldsymbol{S}}, F_1^*, F_2^*)$ are negative and $D^2 > 0$, the adverse effect of outliers is that of reducing the estimate towards zero, an effect usually called “implosion”, in opposition to “explosion”, which occurs when the influence function goes to $+\infty$.

If $\boldsymbol{T}$ and $\boldsymbol{S}$ are B-robust estimators, that is, if they have bounded influence functions, so will be $D^2_{\boldsymbol{T},\boldsymbol{S}}$, and the behavior of its partial influence functions can easily be studied using (4).

Figure 1 shows tridimensional plots of the first partial influence function for $D^2_{\boldsymbol{T},\boldsymbol{S}}$, using several estimators at the “central” (bivariate) distributions with $\Delta = 2$ ($\boldsymbol{\mu}_1 = (1,0)^T$, $\boldsymbol{\mu}_2 = (-1,0)^T$, $\boldsymbol{\Sigma} = \boldsymbol{I}$) and $w_1 = w_2 = 1/2$. The second partial influence function can be obtained by a symmetry to the plane $x_1^* = 0$.

The first pair of estimators used is the classical, for which this influence function does not depend on the form of F_i but only on their first and second moments. The second pair considered is formed by the Minimum Covariance Determinant (MCD) estimators of multivariate location and scatter (Rousseeuw, 1985) whose influence function have recently been explicitly derived (Croux and Haesbroeck, 1999). A third pair is the one step reweighted MCD estimators (denoted MCD^1) also described in Croux and Haesbroeck (1999) and influence functions given in Lopuhaä (1998). Finally we have considered the S-estimators (Rousseeuw and Yohai, 1984) whose influence function can be found in Lopuhaä (1989). For the three robust estimators a breakdown point of 25% was imposed and the plots shown are for underlying normal distributions.

All the robust estimators show their bounded influence. In the MCD and MCD^1 cases the influence function is redescending but not to zero, so outliers may introduce some bias. Due to very steep jumps some erratic behaviour, leading to a high variance is expected (worse for MCD than for MCD[1]). For S-estimators the influence is very smooth (a smaller variance is antecipated) but not redescending, therefore, the bias introduced by large contamination, though bounded, may be large.

Given the partial influence functions, asymptotic variances for $D^2_{\boldsymbol{T},\boldsymbol{S}}$ are easy to obtain using formulas (2) and (3) from Section 2. (The necessary regularity conditions and asymptotic normality have already been proved, in the references cited above, for all the estimators $\boldsymbol{T}$ and $\boldsymbol{S}$ considered).

$$\begin{aligned} V_i(D^2_{\boldsymbol{T},\boldsymbol{S}}, F_1, F_2) &= 4\Delta^2 V(T_1, F_i^*) + w_i^2 \Delta^4 V(S_{11}, F_i^*) - \\ &\quad -4(-1)^{i+1} w_i \Delta^3 Cov(T_1, S_{11}, F_i^*). \end{aligned}$$

If F_i^* are of the same form $V(T_1, F_i^*)$ and $V(S_{11}, F_i^*)$ will not depend on i, and if furthermore they are both elliptically symmetric the last term vanishes, for the estimators under consideration, because the location and scatter components are asymptotically independent. Therefore

$$\begin{aligned} V(D^2_{\boldsymbol{T},\boldsymbol{S}}, F_1, F_2) &= 4\Delta^2 V(T_1, F_i^*)\left(\frac{1}{w_1} + \frac{1}{w_2}\right) + \\ &\quad + \Delta^4 V(S_{11}, F_i^*)\left(\frac{w_1^2}{w_1} + \frac{w_2^2}{w_2}\right) = \\ &= 4\Delta^2\left(\alpha(w_1w_2)^{-1} + \beta\Delta^2\right), \end{aligned} \tag{5}$$

where $\alpha = V(T_1, F_i^*)$ and $\beta = V(S_{11}, F_i^*)/4$. If $(w_1w_2)^{-1}$ is large, that is, if the dimensions of the two samples are very disparate the variance of the location estimator dominates, whereas if Δ^2 is much larger than $(w_1w_2)^{-1}$ it is the variance of the variance estimator which dominates.

Table 1 gives the values of α and β under several elliptical distributions for the estimators previously described.

All the distributions have their parameters scaled in order to have the identity as covariance matrix. This means for instance that the $t_{\nu(\boldsymbol{\mu},\boldsymbol{I})}$ distribution (with ν degrees of freedom, $\nu > 2$) has density given by

$$g(\boldsymbol{x}) = \frac{\Gamma\left(\frac{m+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi^{m/2}(\nu-2)^{m/2}}\left[1 + \frac{(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{x}-\boldsymbol{\mu})}{\nu-2}\right]^{-\frac{m+\nu}{2}}$$

and that the symmetric contaminated normal ($\mathrm{SCN}_{(\boldsymbol{\mu},\boldsymbol{I})}$) distribution, that is $(1-\varepsilon)\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{I}/a) + \varepsilon\mathcal{N}(\boldsymbol{\mu}, c\boldsymbol{I}/a)$ with density given by

$$g(\boldsymbol{x}) = (1-\varepsilon)\left(\frac{a}{2\pi}\right)^{p/2}\exp\left[-\frac{a(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{x}-\boldsymbol{\mu})}{2}\right] +$$

$$+\varepsilon\left(\frac{a}{2\pi c}\right)^{p/2}\exp\left[-\frac{a(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{x}-\boldsymbol{\mu})}{2c}\right],$$

where $a = 1 + (c-1)\varepsilon$. Under these circumstances we have, for the classical estimators, $V(\bar{x}_1, F_i^*) = 1$ (at all the distributions considered), $V(S_{11}, F_i^*) = 2$ (at the normal distribution), $V(S_{11}, F_i^*) = (2\nu-2)/(\nu-4)$ (at the t_ν distribution, with $\nu > 4$), and $V(S_{11}, F_i^*) = 3[1+(c^2-1)\varepsilon]/a^2-1$ (at the SCN distribution).

For all the robust estimators the variances in Table 1 were evaluated by integration of the corresponding squared influence functions. All the integrals were computed accurately using symbolic computation (in Mathematica). Numerical methods were used only for determination of some specific constants required by each robust estimator.

The results in Table 1 are not surprising and are in close agreement with results obtained by Croux and Haesbroeck (1999) (note, however, that these authors did not consider the distribution SCN) .Overall, S-estimators emerge as the best compromise.

3.2. Linear discriminant function coefficients

Another important parameter in multivariate analysis is the vector ($\boldsymbol{\alpha}$) of the coefficients of Fisher's linear discriminant function between two populations (under the conditions described in the first paragraph of Subsection 3.1). This vector is usually written as $\boldsymbol{\alpha} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, but any non-negative scalar multiple of it will serve the same purposes of discrimination between the two populations. In particular, the normalized form $\boldsymbol{\alpha}_N = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|$. Natural estimators of these parameters are obtained by plugging in the defining formulae estimators of $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}_i$, leading to

$$\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}}(F_{n_1}, F_{n_2}) = \boldsymbol{S}_c^{-1}(F_{n_1}, F_{n_2})\left(\boldsymbol{T}(F_{n_1}) - \boldsymbol{T}(F_{n_2})\right)$$

and

$$\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}(F_{n_1}, F_{n_2}) = \hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}}(F_{n_1}, F_{n_2}) / \left\|\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}}(F_{n_1}, F_{n_2})\right\|.$$

If the location and scatter estimators are Fisher consistent then

$$\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}}(F_1, F_2) = \boldsymbol{\alpha}.$$

For $\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}(F_1, F_2)$ to be Fisher consistent it is sufficient that $\boldsymbol{T}$ is consistent and $\boldsymbol{S}$ is consistent up to a constant ($\boldsymbol{S}_c(F_1, F_2) \propto \boldsymbol{\Sigma}$). This situation is common for robust estimators.

The partial influence functions of $\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}}$, $\left\|\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}}\right\|$ and $\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}$ follow easily ($\boldsymbol{F} = (F_1, F_2)$):

$$IF_i(\boldsymbol{x}; \hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}}, \boldsymbol{F}) = -w_i \boldsymbol{\Sigma}^{-1} IF(\boldsymbol{x}; \boldsymbol{S}, F_i) \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) +$$

$$+(-1)^{i+1}\boldsymbol{\Sigma}^{-1}IF(\boldsymbol{x};\boldsymbol{T},F_i)$$

$$IF_i(\boldsymbol{x};\|\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}}\|,\boldsymbol{F}) = \frac{\boldsymbol{\alpha}^T IF_i(\boldsymbol{x};\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}},\boldsymbol{F})}{\|\boldsymbol{\alpha}\|}$$

$$IF_i(\boldsymbol{x};\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}},\boldsymbol{F}) = \left(\boldsymbol{I} - \boldsymbol{\alpha}_N\boldsymbol{\alpha}_N^T\right)\frac{IF_i(\boldsymbol{x};\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}},\boldsymbol{F})}{\|\boldsymbol{\alpha}\|} \tag{6}$$

$\boldsymbol{\alpha}$ is equivariant for affine transformations, that is, if $\boldsymbol{x}_i^* = \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{b}$, with distribution F_i^*, $i = 1,2$, then $\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}}(F_1^*,F_2^*) = \boldsymbol{\alpha}^* = (\boldsymbol{A}^{-1})^T\boldsymbol{\alpha}$. As in the previous subsection $\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}}$ is an affine equivariant estimator of the parameter $\boldsymbol{\alpha}$ if $\boldsymbol{T}$ and $\boldsymbol{S}$ are affine equivariant. In that case it is also sufficient to study the influence functions (and the asymptotic variances) at "central" distributions F_1^* and F_2^*. We can use the same F_1^* and F_2^*, for which $\boldsymbol{\alpha}^* = (\Delta,0,\ldots,0)^T$, $\|\boldsymbol{\alpha}^*\| = \Delta$ and $\boldsymbol{\alpha}_N^* = (1,0,\ldots,0)^T$. We then have

$$IF_i(\boldsymbol{x};\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}},F_1,F_2) = \boldsymbol{A}^T IF_i(\boldsymbol{x}^*;\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}},F_1^*,F_2^*) \tag{7}$$

and

$$IF_i(\boldsymbol{x}^*;\hat{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{S}},F_1^*,F_2^*) = -w_i\Delta IF(\boldsymbol{x}^*;\boldsymbol{S}_1,F_i^*) + (-1)^{i+1}IF(\boldsymbol{x}^*;\boldsymbol{T},F_i^*),$$

where $\boldsymbol{S}_1$ is the first column of $\boldsymbol{S}$. A similar relation concerning $\hat{\boldsymbol{\alpha}}_N$ is not so clear but it can also be established. In other words there is a matrix $\boldsymbol{B}$ such that

$$IF_i(\boldsymbol{x};\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}},F_1,F_2) = \boldsymbol{B}^T IF_i(\boldsymbol{x}^*;\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}},F_1^*,F_2^*),$$

or, using (6) and (7), such that

$$\left(\boldsymbol{I} - \boldsymbol{\alpha}_N\boldsymbol{\alpha}_N^T\right)\boldsymbol{A}^T\frac{\Delta}{\|\boldsymbol{\alpha}\|} = \boldsymbol{B}^T\left(\boldsymbol{I} - \boldsymbol{\alpha}_N^*\boldsymbol{\alpha}_N^{*T}\right).$$

First note that $\boldsymbol{I} - \boldsymbol{\alpha}_N\boldsymbol{\alpha}_N^T$ and $\boldsymbol{I} - \boldsymbol{\alpha}_N^*\boldsymbol{\alpha}_N^{*T}$ are symmetric idempotent matrices with rank $m-1$. Therefore they have $m-1$ unit eigenvalues and one null. Also $\boldsymbol{I} - \boldsymbol{\alpha}_N^*\boldsymbol{\alpha}_N^{*T} = \mathrm{diag}(0,1,\ldots,1)$. So we can write

$$\boldsymbol{I} - \boldsymbol{\alpha}_N\boldsymbol{\alpha}_N^T = \boldsymbol{V}\left(\boldsymbol{I} - \boldsymbol{\alpha}_N^*\boldsymbol{\alpha}_N^{*T}\right)\boldsymbol{V}^T,$$

where $\boldsymbol{V}$ is an eigenvector matrix of $\boldsymbol{I} - \boldsymbol{\alpha}_N\boldsymbol{\alpha}_N^T$ but with the first and last columns interchanged. Finally noting that the effect of right (left) multiplication of any matrix by $\boldsymbol{I} - \boldsymbol{\alpha}_N^* \boldsymbol{\alpha}_N^{*T}$ is to put the null vector into its first column (row), we conclude that $\boldsymbol{B}$ is not unique (its first row is undetermined) and that a solution is

$$\boldsymbol{B} = \frac{\Delta}{\|\boldsymbol{\alpha}\|} \boldsymbol{A}\boldsymbol{V}\left(\boldsymbol{I} - \boldsymbol{\alpha}_N^* \boldsymbol{\alpha}_N^{*T}\right)\boldsymbol{V}^T = \frac{\Delta}{\|\boldsymbol{\alpha}\|}\boldsymbol{A}\left(\boldsymbol{I} - \boldsymbol{\alpha}_N \boldsymbol{\alpha}_N^T\right).$$

An interesting aspect arising from the above discussion is that the influence functions of the first component of $\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}$ at (F_1^*, F_2^*) are identically null, and that, in general

$$\boldsymbol{\alpha}_N^T IF_i(\boldsymbol{x}; \hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}, F_1, F_2) \equiv 0.$$

This does not mean that an observation at a point $\boldsymbol{x}$ (or $\boldsymbol{x}^*$) does not change the corresponding estimate, it only means that the change is not $O(n_i^{-1})$ but $o(n_i^{-1})$.

The relevant changes are thus orthogonal to de discriminant direction and it is sufficient to inspect the influence functions for the second component of $\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}$ at the "central" distributions:

$$IF_i(\boldsymbol{x}^*; \hat{\alpha}_{N2,\boldsymbol{T},\boldsymbol{S}}, F_1^*, F_2^*) = -w_i IF(\boldsymbol{x}^*; S_{12}, F_i^*) + (-1)^{i+1}\frac{IF(\boldsymbol{x}^*; T_2, F_i^*)}{\Delta},$$

which for the classical estimators becomes

$$IF_i(\boldsymbol{x}^*; \hat{\alpha}_{N2,\hat{\boldsymbol{x}},\boldsymbol{S}}, F_1^*, F_2^*) = -w_i(x_1^* - \mu_{i1}^*)x_2^* + (-1)^{i+1}\frac{x_2^*}{\Delta}.$$

Figure 2 shows tridimensional plots of the first partial influence function for the second component of the normalized discriminant vector using several estimators at the "central" bivariate distribution with $\Delta = 2$ and $w_1 = w_2 = 1/2$. The pairs of estimators and their specific characteristics are the same used in the previous subsection. The second partial influence function can be obtained by a symmetry to the point $(x_1^*, x_2^*) = (0, 0)$.

Again the boundedness of the influence function of the robust estimators is observed. The remarks concerning their smoothness are similar to those made about Figure 1. A difference is, however, observed in what concerns the redescending behaviour, since now the influence redescends to zero for all the robust estimators, that is, they completely reject outliers. This may be a clear advantage for S-estimators and a strong point for its use if a discriminant estimator is desired.

Because of the affine equivariance the asymptotic variances of $\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}$ is

$$V(\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}, F_1, F_2) = \boldsymbol{B}^T V(\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}, F_1^*, F_2^*)\boldsymbol{B}.$$

By the reasons stated before expression (5) in Subsection 3.1 and using again formulas (2) and (3) in Section 2, we further have

$$V(\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}, F_1^*, F_2^*) = \operatorname{diag}\left[0; \gamma + \frac{\alpha(w_1w_2)^{-1}}{\Delta^2}; \ldots; \gamma + \frac{\alpha(w_1w_2)^{-1}}{\Delta^2}\right],$$

where $\alpha = V(T_1, F_i^*)$, $\beta = V(S_{11}, F_i^*)/4$ and $\gamma = V(S_{12}, F_i^*)$. For large Δ^2 this variance is dominated by the variance of the covariance estimator.

It is not surprising that the asymptotic covariance matrix of $\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}$ is singular, since for the normalized vector one of the components is a function of the others and we are dealing with first order expansions (this is also a direct consequence of the remarks made about the influence functions of $\hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}$). Again this does not mean that the variance of $\hat{\alpha}_{1,N,\boldsymbol{T},\boldsymbol{S}}$ is zero for a sample of (F_1^*, F_2^*) it only means that the variance is not $O\left((n_1+n_2)^{-1}\right)$ but $o\left((n_1+n_2)^{-1}\right)$.

Table 2 gives the values of γ for the cases and conditions considered in Table 1. For the classical estimators we have: $V(S_{12}, F_i^*) = 1$ (at the normal distribution), $V(S_{12}, F_i^*) = (\nu-2)/(\nu-4)$ (at the t_ν distribution, with $\nu > 4$) and $V(S_{12}, F_i^*) = [1 + (c^2-1)\varepsilon]/a^2$ (at the SCN distribution). The remarks that could made about Table 2 are very similar to those made about Table 1 in the previous subsection.

The results given in this section are useful to obtain partial influence functions and asymptotic variances for the estimator of the linear discriminant function at a given new observation, $\boldsymbol{y}$: $\hat{L}(\boldsymbol{y}) = \hat{\boldsymbol{\alpha}}_N^T\boldsymbol{y} - \hat{\boldsymbol{\alpha}}_N^T(\hat{\boldsymbol{\mu}}_1 + \hat{\boldsymbol{\mu}}_2)/2$ (assuming equal *a priori* probabilities and misclassification costs). For the partial influence functions of $\hat{L}(\boldsymbol{y})$ we have

$$IF_i(\boldsymbol{x}; \hat{L}(\boldsymbol{y}), F_1, F_2) = \\ = IF_i(\boldsymbol{x}; \hat{\boldsymbol{\alpha}}_{N,\boldsymbol{T},\boldsymbol{S}}, F_1, F_2)^T \left(\boldsymbol{y} - \frac{\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2}{2}\right) - \boldsymbol{\alpha}_N^T \frac{IF(\boldsymbol{x}; \hat{\boldsymbol{\mu}}, F_i)}{2},$$

and the same kind of exercise could be done for any differentiable function of $\hat{L}(\boldsymbol{y})$, like *a posteriori* probabilities of group membership. However, the influence of the classification rule, which assigns $\boldsymbol{y}$ to one of the two groups, does not exist since the classification rule is a two valued non differentiable function of $\hat{L}(\boldsymbol{y})$.

It would be of interest to obtain asymptotic results for the estimators considered by Hawkins and McLchlan (1997) and He and Fung (2000) but this is behind the scope of this paper.

Additional aspects which are of interest considering are the finite sample behaviour of the estimators. It is known, for instance, that the convergence to asymptotic behaviour is slower for MCD than for S-estimators (Rocke and Woodruff, 1997).

4. CONCLUSIONS

In this paper it is shown how the influence function can be generalized and used to study the asymptotic behaviour of estimators depending on samples from more than one population.

In the examples presented the asymptotic variance could easily be obtained by the delta method because of the functional dependence of the multivariate estimators of location and scatter, but for other types of estimators, like projection pursuit estimators, there is no such dependence and this is a possible approach for looking at their asymptotic properties. Projection pursuit estimators of linear discriminant coefficients are described for instance in Van Ness and Yang (1998), following a suggestion in Huber (1985). In some cases an estimator of the squared Mahalanobis distance (Δ^2) can also be obtained: if the projection index is based on Fisher's separation criterion (linear combination of the variables which maximizes the ratio of the between to the within groups variability) then the maximum of the index is an estimate of Δ^2.

One of the important conclusions arising from these examples is the superiority of the S-estimators relatively to the MCD and MCD^1 estimators, especially for the linear discriminant coefficients.

APPENDIX

Proof of Theorem 1:
Consider the auxiliary function $\boldsymbol{\psi} : \mathbb{R} \longrightarrow \mathbb{R}^p$,

$$\boldsymbol{\psi}(t) = \boldsymbol{T}[(1-t)F_1 + tG_1, (1-t)F_2 + tG_2],$$

then $\boldsymbol{\psi}(1) = \boldsymbol{T}(G_1, G_2)$ and $\boldsymbol{\psi}(0) = \boldsymbol{T}(F_1, F_2)$. If all components of $\boldsymbol{\psi}(t)$ are second order differentiable it holds that

$$\boldsymbol{\psi}(1) = \boldsymbol{\psi}(0) + \boldsymbol{\psi}'(0) + \text{Remainder} \quad \exists_{0<c<1} : \text{ Rest} = \frac{\psi''(c)}{2}. \quad \text{(A.1)}$$

and there is $0 < c < 1$ such that

$$\text{Remainder} = \frac{\boldsymbol{\psi}''(c)}{2}.$$

To evaluate $\boldsymbol{\psi}'(t)$, $0 \leq t < 1$ we use the definition

$$\boldsymbol{\psi}'(t) =$$
$$= \lim_{\Delta t \to 0} \left\{ \frac{\boldsymbol{T}[(1-t-\Delta t)F_1 + (t+\Delta t)G_1, (1-t-\Delta t)F_2 + (t+\Delta t)G_2]}{\Delta t} - \right.$$

$$
\begin{aligned}
& - \left.\frac{\boldsymbol{T}[(1-t)F_1+tG_1,(1-t)F_2+tG_2]}{\Delta t}\right\} = \\
= & \lim_{\Delta t\to 0}\left\{\frac{\boldsymbol{T}[(1-t-\Delta t)F_1+(t+\Delta t)G_1,(1-t-\Delta t)F_2+(t+\Delta t)G_2]}{\Delta t}-\right.\\
& - \left.\frac{\boldsymbol{T}[(1-t)F_1+tG_1,(1-t-\Delta t)F_2+(t+\Delta t)G_2]}{\Delta t}\right\} + \\
& + \lim_{\Delta t\to 0}\left\{\frac{\boldsymbol{T}[(1-t)F_1+tG_1,(1-t-\Delta t)F_2+(t+\Delta t)G_2]}{\Delta t}-\right.\\
& - \left.\frac{\boldsymbol{T}[(1-t)F_1+tG_1,(1-t)F_2+tG_2]}{\Delta t}\right\} = \\
= & \ \boldsymbol{\psi}_1'(t)+\boldsymbol{\psi}_2'(t).
\end{aligned}
$$

If we write

$$
\begin{aligned}
(1-t-\Delta t)F_i+(t+\Delta t)G_i &= \left(1-\frac{\Delta t}{1-t}\right)[(1-t)F_i+tG_i]+\frac{\Delta t}{1-t}G_i \\
&= H_i, \qquad \text{(A.2)}
\end{aligned}
$$

with $i=2$, we can conclude by the partial Gâteaux differentiability that

$$
\begin{aligned}
\boldsymbol{\psi}_2'(t) &= \lim_{\Delta t\to 0}\left\{\frac{1}{1-t}\times\right.\\
&\times\left.\frac{\boldsymbol{T}\left[(1-t)F_1+tG_1,H_2\right]-\boldsymbol{T}[(1-t)F_1+tG_1,(1-t)F_2+tG_2]}{\Delta t/(1-t)}\right\}= \\
&= \frac{1}{1-t}\int IF_2(\boldsymbol{x};\boldsymbol{T},(1-t)F_1+tG_1,(1-t)F_2+tG_2)dG_2(\boldsymbol{x})
\end{aligned}
$$

but as $G_2=(1-t)F_2+tG_2+(1-t)(G_2-F_2)$ by (1) we have

$$
\boldsymbol{\psi}_2'(t)=\int IF_2(\boldsymbol{x};\boldsymbol{T},(1-t)F_1+tG_1,(1-t)F_2+tG_2)d(G_2-F_2)(\boldsymbol{x}).
$$

On the other hand noting that the partial differentiability in each direction is an ordinary differentiability we can use the mean value theorem. Then, using (A.2),

$$
\boldsymbol{\psi}_1'(t)=\lim_{\Delta t\to 0}\frac{\int IF_1(\boldsymbol{x};\boldsymbol{T},H_\theta,H_2)d(H_1-(1-t)F_1-tG_1)(\boldsymbol{x})}{\Delta t},
$$

with $H_\theta=(1-\theta)H_1+\theta[(1-t)F_1+tG_1]$, $0<\theta<1$. But

$$
H_1-(1-t)F_1-tG_1=\Delta t(G_1-F_1),
$$

therefore

$$\boldsymbol{\psi}_1'(t) = \lim_{\Delta t \to 0} \int IF_1(\boldsymbol{x}; \boldsymbol{T}, H_\theta, (1-t-\Delta t)F_2 + (t+\Delta t)G_2)d(G_1 - F_1)(\boldsymbol{x}).$$

Because $H_\theta = [1-t-(1-\theta)\Delta t]F_1 + [t+(1-\theta)\Delta t]G_1$, the integrand is a continuous function of t (by regularity condition R2) and it is possible to interchange the limit and the integration (by regularity condition R4) we finally conclude that

$$\boldsymbol{\psi}_1'(t) = \int IF_1(\boldsymbol{x}; \boldsymbol{T}, (1-t)F_1 + tG_1, (1-t)F_2 + tG_2)d(G_1 - F_1)(\boldsymbol{x}).$$

Then

$$\boldsymbol{\psi}'(0) = \sum_{i=1}^{2} \int IF_i(\boldsymbol{x}; \boldsymbol{T}, F_1, F_2)dG_i(\boldsymbol{x})$$

and

$$\boldsymbol{\psi}''(t) = \sum_{i=1}^{2} \frac{\partial}{\partial t} \int IF_i(\boldsymbol{x}; \boldsymbol{T}, (1-t)F_1 + tG_1, (1-t)F_2 + tG_2)d(G_i - F_i)(\boldsymbol{x}),$$

which by substitution into (A.1) completes the proof.

ACKNOWLEDGMENT

This work was partially supported by the project Praxis/PCEX/P/MAT/41/96. We thank two anonymous referees for their suggestions regarding the contents of the paper.

REFERENCES

1. N. A. Campbell, The influence function as an aid in outlier detection in multivariate analysis, *Applied Statistics* **27** (1978), 251–258.
2. R. D. Cook and S. Weisberg, "Residuals and Influence in Regression," Chapman and Hall, New York, 1982.
3. F. Critchley, Influence in principal components analysis, *Biometrika* **72** (1985), 627–636.
4. F. Critchley and C. Vitiello, The influence of observations on misclassification probability estimates in linear discriminant analysis, *Biometrika* **78** (1991), 677–690.
5. C. Croux and G. Haesbroeck, Influence function and efficiency of the minimum covariance determinant scatter matrix estimator, *Journal of Multivariate Analysis* **71** (1999), 161–190.
6. W. K. Fung, The influence of observations on misclassification probability in multiple discriminant analysis, *Communications in Statistics – Theory and Methods* **25** (1996), 1917–1930.

7. F. R. Hampel, The influence curve and its role in robust estimation, *Journal of the American Statistical Association* **69** (1974), 383–393.

8. 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.

9. D. Hawkins and G. J. McLachlan, High-breakdown discriminant analysis, *Journal of the American Statistical Association* **92** (1997), 136–143.

10. X. He and W. K. Fung, High-breakdown estimation for multiple populations with applications to discriminant analysis, *Journal of Multivariate Analysis* **72** (2000), 151–162.

11. X. He and D. G. Simpson, Robust direction estimation, *The Annals of Statistics* **20** (2000), 351–369.

12. P. J. Huber, Projection pursuit (with discussion), *Annals of Statistics* **13** (1985), 435–475.

13. H. P. Lopuhaä, On the relation between S-estimators and M-estimators of multivariate location and covariance, *The Annals of Statistics* **17** (1989), 1662–1683.

14. H. P. Lopuhaä, Asymptotics of reweighted estimators of multivariate location and scatter, Technical Report 98-09, Department of Technical Mathematics and Informatics, TUDelft, 1998.

15. J. Lu, D. Ko and T. Chang, The standardized influence matrix and its applications, *Journal of the American Statistical Association* **92** (1997), 1572–1580.

16. R. Radhakrishnan, Influence functions for certain parameters in discriminant analysis, *Metron*, **30** (1983), 183–194.

17. R. Radhakrishnan and A. M. Kshirsagar, Influence functions for certain parameters in multivariate analysis, *Communications in Statistics – Theory and Methods*, **10** (1981), 515–529.

18. H. Rieder. “Robust Asymptotic Statistics,” Springer-Verlag, New York, 1994.

19. D. M. Rocke and D. L. Woodruff, Robust estimation of multivariate location and shape, *Journal of Statistical Planning and Inference*, **57** (1997), 245–255.

20. P. Rousseeuw, Multivariate estimation with high breakdown point, in “Mathematical Statistics and Applications,” Vol. B, (W. Grossman, G. Pflug, I. Vincze and W. Wertz, Eds.), pp.283–297, Reidel, Dordrecht, 1985.

21. P. Rousseeuw and E. Ronchetti, Influence curves of general statistics, *Journal of Computational and Applied Mathematics* **7** (1981), 161–166.

22. P. Rousseeuw and V. J. Yohai, Robust regression by means of S-estimators, in “Robust and Non-linear Time Series Analysis,” pp.256–272, Lecture Notes in Statistics **26**, Springer, New York, 1984.

23. P. K. Sen, “Sequential Nonparametrics,” Wiley, New York, 1981.

24. J. W. Van Ness and J. J. Yang, Robust discriminant analysis: training data breakdown point, *Journal of Statistical Planning and Inference* **67** (1998), 67–83.

TABLE 1.

Values of α and β for several distributions, estimators and number of variables

		$p=2$		$p=3$		$p=5$		$p=10$	
Dist.	Est.	α	β	α	β	α	β	α	β
Φ	$\bar{\boldsymbol{X}}, \boldsymbol{S}$	1.000	0.500	1.000	0.500	1.000	0.500	1.000	0.500
	MCD	2.478	1.907	2.145	1.666	1.883	1.366	1.674	1.090
	MCD^1	1.139	0.836	1.107	0.740	1.082	0.661	1.061	0.600
	S	1.097	0.556	1.051	0.532	1.025	0.515	1.010	0.506
t_5	$\bar{\boldsymbol{X}}, \boldsymbol{S}$	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
	MCD	1.225	1.646	1.077	1.473	0.963	1.269	0.878	1.089
	MCD^1	0.865	0.986	0.869	0.949	0.874	0.932	0.879	0.932
	S	0.789	0.833	0.771	0.815	0.762	0.801	0.758	0.794
t_8	$\bar{\boldsymbol{X}}, \boldsymbol{S}$	1.000	0.875	1.000	0.875	1.000	0.875	1.000	0.875
	MCD	1.607	1.694	1.394	1.493	1.227	1.254	1.097	1.041
	MCD^1	0.963	0.892	0.956	0.831	0.952	0.789	0.950	0.765
	S	0.923	0.716	0.895	0.693	0.880	0.677	0.874	0.669
t_{15}	$\bar{\boldsymbol{X}}, \boldsymbol{S}$	1.000	0.636	1.000	0.636	1.000	0.636	1.000	0.636
	MCD	1.965	1.768	1.695	1.545	1.478	1.275	1.303	1.031
	MCD^1	1.038	0.850	1.021	0.772	1.007	0.712	0.997	0.670
	S	1.012	0.636	0.976	0.611	0.956	0.594	0.946	0.586
SCN	$\bar{\boldsymbol{X}}, \boldsymbol{S}$	1.000	1.833	1.000	1.833	1.000	1.833	1.000	1.833
	MCD	1.259	1.621	1.086	1.408	0.959	1.169	0.872	0.968
	MCD^1	0.780	1.144	0.805	1.268	0.831	1.408	0.858	1.564
	S	0.684	0.844	0.663	0.867	0.649	0.896	0.637	0.922

TABLE 2.

Values of γ for several distributions, estimators and number of variables

Dist.	Est.	$p = 2$	$p = 3$	$p = 5$	$p = 10$
Φ	$\bar{\boldsymbol{X}}, \boldsymbol{S}$	1.000	1.000	1.000	1.000
	MCD	6.127	4.296	3.087	2.285
	MCD^1	1.360	1.199	1.115	1.071
	S	1.177	1.082	1.034	1.012
t_5	$\bar{\boldsymbol{X}}, \boldsymbol{S}$	3.000	3.000	3.000	3.000
	MCD	4.562	3.337	2.531	2.010
	MCD^1	1.434	1.404	1.421	1.460
	S	1.458	1.346	1.282	1.252
t_8	$\bar{\boldsymbol{X}}, \boldsymbol{S}$	1.500	1.500	1.500	1.500
	MCD	4.973	3.556	2.624	2.017
	MCD^1	1.352	1.279	1.259	1.265
	S	1.351	1.242	1.184	1.160
t_{15}	$\bar{\boldsymbol{X}}, \boldsymbol{S}$	1.182	1.182	1.182	1.182
	MCD	5.418	3.821	2.767	2.072
	MCD^1	1.330	1.219	1.171	1.154
	S	1.268	1.164	1.111	1.090
SCN	$\bar{\boldsymbol{X}}, \boldsymbol{S}$	2.778	2.778	2.778	2.778
	MCD	4.786	3.402	2.526	1.983
	MCD^1	1.673	1.809	2.001	2.240
	S	1.333	1.273	1.237	1.198

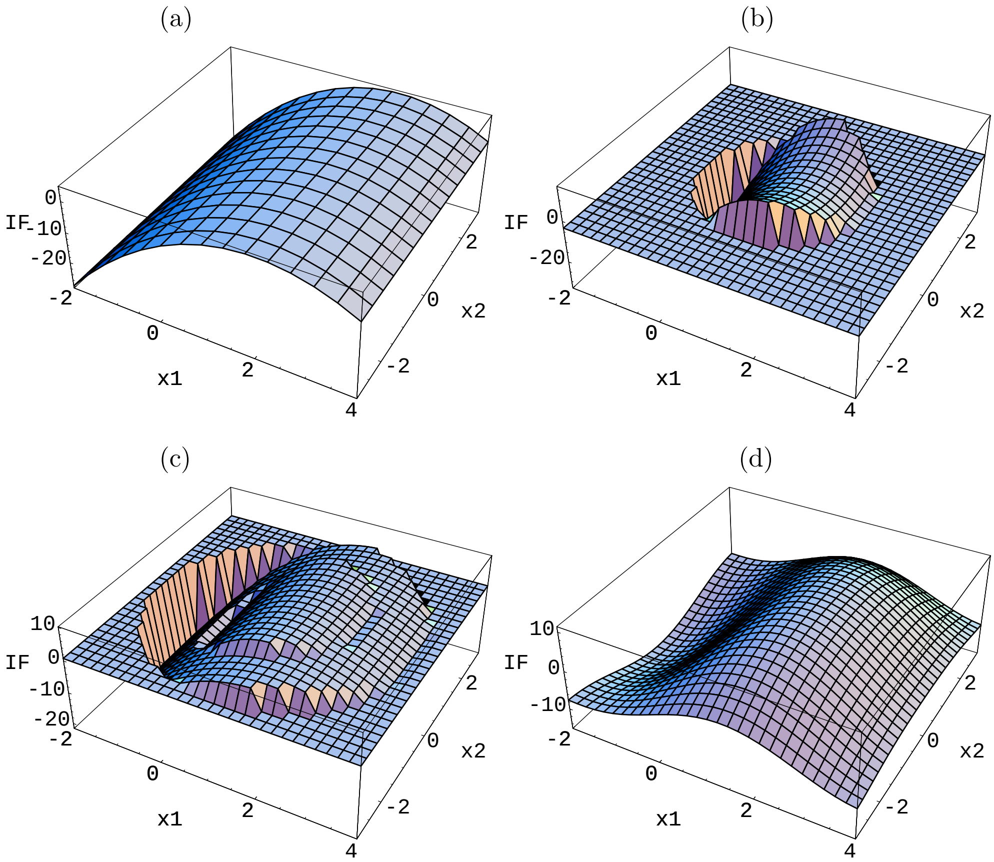

FIG. 1. First partial influence function of several estimators of Δ^2: (a) classical estimator; (b) with MCD; (c) with MCD^1; (d) with S-estimators.

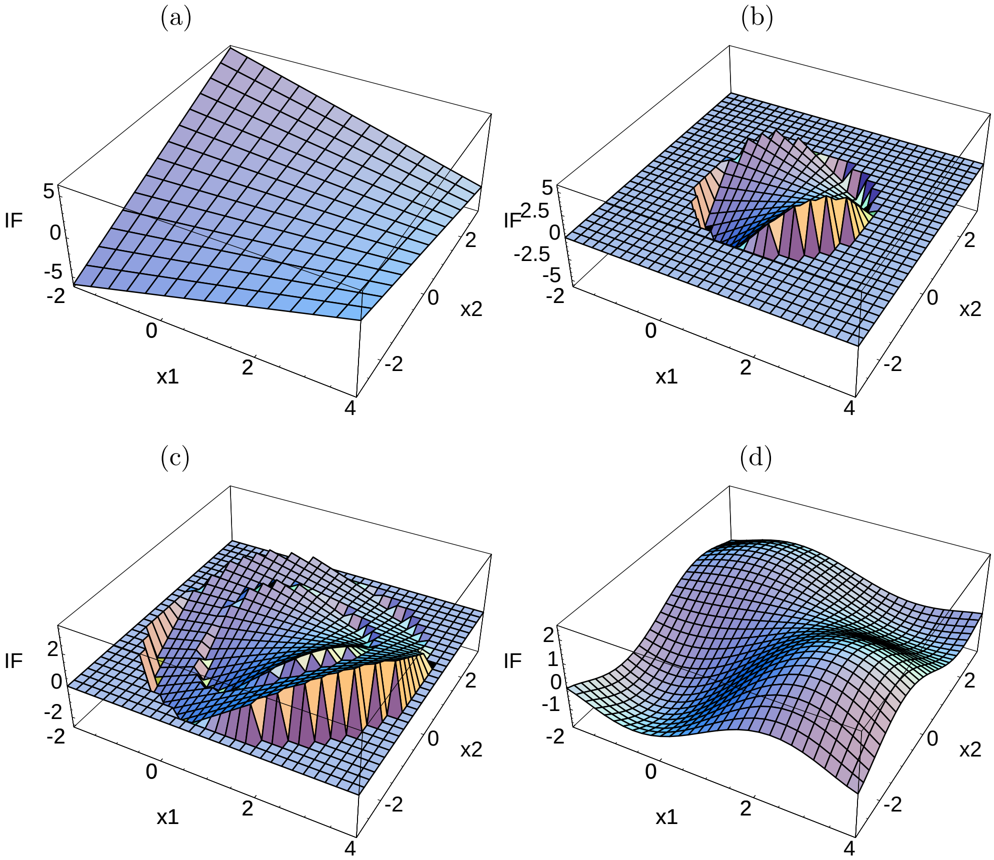

FIG. 2. First partial influence function for several estimators of the second component of the normalized discriminant vector: (a) classical estimator; (b) with MCD; (c) with MCD^1; (d) with S-estimators.