Journal of Multivariate Analysis **77**, 138–161 (2001)
doi:10.1006/jmva.2000.1929, available online at http://www.idealibrary.com on IDEAL®

Influence Function of Halfspace Depth

Mario Romanazzi

University of Venice, Venice, Italy
E-mail: romanaz@unive.it

Received May 18, 1999

The sensitivity of halfspace depth values and contours to perturbations of the underlying distribution is investigated. The influence function of the halfspace depth of any point $x \in \mathbb{R}^p$ is bounded and discontinuous; it is constant and positive when the perturbing observation z is placed in any optimal halfspace and it is constant and negative when z is placed in any non-optimal halfspace. When the optimal [illegible] all the depth regions but only the outer region can be arbitrarily expanded. To obtain the same effect on the inner regions the size of the perturbation is required to be not less than the depth orders. Numerical illustrations of the results are given. © 2001 Academic Press

AMS 1991 subject classifications: 62G35, 62H12.

Key words and phrases: halfspace depth; halfspace region; influence function; simplicial depth.

1. INTRODUCTION

A depth function is a device to measure the degree of centralness of points in p-dimensional space with respect to a probability distribution F or an observed data set. Two specific definitions are Tukey's halfspace and Liu's simplicial depth [9, 22]. These depth functions do not assume a particular class of distributions (e.g., elliptically symmetric) and so they provide a distribution-free location ordering of multivariate data. This property is exploited in several applications. For example, the set of deepest points (with respect to halfspace or simplicial depth) has the meaning of median region and the center of gravity of this set can be used as a location parameter or estimator instead of the classical centroid (Donoho and Gasko [4], Rousseeuw and Ruts [16]). More generally, the contours of constant depth are interpreted as generalized quantile surfaces (Massé and Theodorescu [10]). Romanazzi [13] and Rousseeuw *et al.* [18] plot selected contours of halfspace depth to graphically display location, spread and shape of a bivariate sample and to reveal outliers.

0047-259X/01 $35.00

In the mentioned applications it is often claimed that the inner contours of halfspace depth are resistant to substantial contamination of the underlying distribution. This follows in part from the results of Donoho and Gasko [4] who derive the breakdown point of the deepest point and other location estimators based on halfspace depth. The breakdown point of the simplicial median is investigated by Chen [2]. In the present paper we use the influence function to study the sensitivity to local perturbations of the ordering induced by halfspace depth. Notations and definitions are established in Section 2. The theoretical (population) influence function is derived in Section 3 and a corresponding sample version is derived in Section 4. Here the leave-one-out effects on the depth contours are investigated and a sample version of the influence function of simplicial depth is given. Section 5 contains some numerical illustrations and Section 6 is devoted to an overall discussion of the results.

2. HALFSPACE AND SIMPLICIAL DEPTH

Let F be a probability distribution on p-dimensional Euclidean space $\mathbb{R}^p$. For any point $x \in \mathbb{R}^p$ we denote with $HS(x)$ and $\mathscr{H}_x$, respectively, any closed halfspace including x and the set of all such halfspaces. We also write $P_F(HS(x)) \equiv P_F(z \in HS(x))$ for the probability, under F, of the halfspace $HS(x)$. Tukey's halfspace depth is the functional of the probability distribution F defined as follows.

Definition 1. The halfspace depth of $x \in \mathbb{R}^p$ with respect to F is

$$d_{HS}(x; F) = \inf_{\mathscr{H}_x} P_F(HS(x)).$$

Any closed halfspace $HS(x)$ satisfying $P_F(HS(x)) = d_{HS}(x; F)$ will be called "optimal" halfspace.

Definition 2. Fix a value d of halfspace depth, $0 < d \leqslant \bar{d} = \sup_{x \in \mathbb{R}^p} d_{HS}(x; F)$. The depth region of level d is the set

$$D_{HS}(d; F) = \{x \in \mathbb{R}^p : d_{HS}(x; F) \geqslant d\}.$$

Each set $D_{HS}(d; F)$ is the intersection of all closed halfspaces with probability greater than $1 - d$. This implies that $\{D_{HS}(d; F), 0 < d \leqslant \bar{d}\}$ is a family of nested convex bodies of $\mathbb{R}^p$.

Consider a sample $X_1, ..., X_n$ of n independent and identical distributed (iid) observations from F and denote with $\hat{F}_n$ the corresponding empirical distribution. The sample halfspace depth $d_{HS}(x; \hat{F}_n)$, obtained by substituting $\hat{F}_n$ for F in Definition 1, is proportional to the minimum number of sample observations contained in any closed halfspace including x. Donoho and Gasko [4, p. 1817] show that, for any distribution F, $d_{HS}(x; \hat{F}_n) \to d_{HS}(x; F)$, almost surely, as $n \to \infty$. The sample depth region $D_{HS}(d; \hat{F}_n)$ is the subset of $\mathbb{R}^p$ whose elements are all the points with sample depth at least equal to d. Its properties are investigated by Donoho and Gasko [4], Massé and Theodorescu [10] and He and Wang [7].

Like halfspace depth, the simplicial depth is a functional of F, and its definition is based on the geometric notion of simplex. Let $X_1, ..., X_{p+1}$ be $p+1$ iid observations from F and let $S(X_1, ..., X_{p+1})$ be the random open simplex with vertices at $X_1, ..., X_{p+1}$.

Definition 3. The simplicial depth of $x \in \mathbb{R}^p$ with respect to F is

$$d_S(x; F) = P_F(S(X_1, ..., X_{p+1}) : x \in S(X_1, ..., X_{p+1})).$$

A sample version of $d_S(x; F)$ is obtained by considering all possible subsamples (sampling without replacement) of $p+1$ observations $X_{i_1}, ..., X_{i_{p+1}}$ from the random sample $X_1, ..., X_n$. The sample simplicial depth $d_S(x; \hat{F}_n)$ is equal to the number of the simplices including x, divided by $\binom{n}{p+1}$, the number of possible simplices. If F is absolutely continuous with bounded density, then $\sup_{x \in \mathbb{R}^p} |d_S(x; \hat{F}_n) - d_S(x; F)| \to 0$, almost surely, as $n \to \infty$ [9, Theorem 5]. Simplicial depth regions, defined in exactly the same way as for halfspace depth, could be envisaged, but they are not convex subsets of $\mathbb{R}^p$ [7, Example 3.2].

3. INFLUENCE FUNCTION

Let $\tilde{F} = (1-\varepsilon) F + \varepsilon\delta_z$, where δ_z is the point-mass distribution concentrated at z. Here F is interpreted as the reference theoretical distribution contaminated by a perturbation located at z. If $T(F)$ is a functional defined for all distributions in a suitable class $\mathscr{F}$, then the influence function of T at F is [6]

$$I(z; T(F)) = \lim_{\varepsilon \to 0^+} \frac{T((1-\varepsilon) F + \varepsilon\delta_z) - T(F)}{\varepsilon},$$

provided this limit exists for every $z \in \mathbb{R}^p$.

3.1. *Influence function of halfspace depth.* Let x be a fixed point of $\mathbb{R}^p$ and let $d_{HS}(x; F)$ be the value of the halfspace depth function at x. For any

$z \in \mathbb{R}^p$, $z \neq x$, we can partition the set $\mathscr{H}_x$ of closed halfspaces including x into two mutually exclusive subsets by putting

$$\mathscr{H}_{x,z} = \{HS(x) \in \mathscr{H}_x : z \in HS(x)\},$$
$$\mathscr{H}_{x,\bar{z}} = \{HS(x) \in \mathscr{H}_x : z \notin HS(x)\}.$$

Clearly, $\mathscr{H}_{x,z}$ is the set of all closed halfspaces including both x and z, whereas $\mathscr{H}_{x,\bar{z}} = \mathscr{H}_x \cap \overline{\mathscr{H}_{x,z}}$ is the set of all closed halfspaces including x but not z. Then $\mathscr{H}_{x,z} \cap \mathscr{H}_{x,\bar{z}} = \emptyset$ and $\mathscr{H}_{x,z} \cup \mathscr{H}_{x,\bar{z}} = \mathscr{H}_x$. Putting

$$d_{HS}^{(z)}(x; F) = \inf_{\mathscr{H}_{x,z}} P_F(HS(x)),$$
$$d_{HS}^{(\bar{z})}(x; F) = \inf_{\mathscr{H}_{x,\bar{z}}} P_F(HS(x)),$$

it follows $d_{HS}(x; F) = \min\{d_{HS}^{(z)}(x; F), d_{HS}^{(\bar{z})}(x; F)\}$. Now, the probability of any halfspace $HS(x)$ with respect to the perturbed distribution $\tilde{F}$ is

$$P_{\tilde{F}}(HS(x)) = \begin{cases} (1-\varepsilon)\, P_F(HS(x)) + \varepsilon, & \text{if} \quad HS(x) \in \mathscr{H}_{x,z}, \\ (1-\varepsilon)\, P_F(HS(x)), & \text{if} \quad HS(x) \in \mathscr{H}_{x,\bar{z}}. \end{cases}$$

To obtain the expression of $d_{HS}(x; \tilde{F})$, and the influence function, it is convenient to distinguish three cases.

1. If $d_{HS}^{(\bar{z})}(x; F) \leqslant d_{HS}^{(z)}(x; F)$, then $d_{HS}(x; F) = d_{HS}^{(\bar{z})}(x; F)$ and

$$d_{HS}(x; \tilde{F}) = (1-\varepsilon)\, d_{HS}^{(\bar{z})}(x; F) = (1-\varepsilon)\, d_{HS}(x; F).$$

Using the definition of influence function it follows

$$I(z; d_{HS}(x; F)) = \lim_{\varepsilon \to 0^+} \frac{(1-\varepsilon)\, d_{HS}^{(\bar{z})}(x; F) - d_{HS}^{(\bar{z})}(x; F)}{\varepsilon}$$
$$= -d_{HS}^{(\bar{z})}(x; F) = -d_{HS}(x; F).$$

2. If $d_{HS}^{(\bar{z})}(x; F) > d_{HS}^{(z)}(x; F)$ and $(1-\varepsilon)\, d_{HS}^{(\bar{z})}(x; F) \geqslant (1-\varepsilon)\, d_{HS}^{(z)}(x; F) + \varepsilon$, then $d_{HS}(x; F) = d_{HS}^{(z)}(x; F)$ and

$$d_{HS}(x; \tilde{F}) = (1-\varepsilon)\, d_{HS}^{(z)}(x; F) + \varepsilon = (1-\varepsilon)\, d_{HS}(x; F) + \varepsilon.$$

Hence

$$IF(z; d_{HS}(x; F)) = \lim_{\varepsilon \to 0^+} \frac{(1-\varepsilon)\, d_{HS}^{(z)}(x; F) + \varepsilon - d_{HS}^{(z)}(x; F)}{\varepsilon}$$
$$= 1 - d_{HS}^{(z)}(x; F) = 1 - d_{HS}(x; F).$$

3. Finally, if $d_{HS}^{(\bar{z})}(x; F) > d_{HS}^{(z)}(x; F)$ and $(1-\varepsilon)\, d_{HS}^{(\bar{z})}(x; F) < (1-\varepsilon)\, d_{HS}^{(z)}(x; F) + \varepsilon$, then $d_{HS}(x; F) = d_{HS}^{(z)}(x; F)$ and

$$d_{HS}(x; \tilde{F}) = (1-\varepsilon)\, d_{HS}^{(\bar{z})}(x; F).$$

Moreover, when the size of the perturbation ε tends to 0, there must exist $\varepsilon_0 > 0$ satisfying

$$(1-\varepsilon)\, d_{HS}^{(\bar{z})}(x; F) \geqslant (1-\varepsilon)\, d_{HS}^{(z)}(x; F) + \varepsilon$$

for all $\varepsilon < \varepsilon_0$. This implies

$$IF(z; d_{HS}(x; F)) = 1 - d_{HS}(x; F) = 1 - d_{HS}^{(z)}(x; F),$$

as in the second case.

The overall expression of the influence function turns out to be

$$IF(z; d_{HS}(x; F)) = \begin{cases} -d_{HS}(x; F), & \text{if} \quad d_{HS}^{(\bar{z})}(x; F) < d_{HS}^{(z)}(x; F); \\ 1 - d_{HS}(x; F), & \text{otherwise.} \end{cases}$$

If $z = x$, then $\mathscr{H}_{x,z} = \mathscr{H}_x$, $\mathscr{H}_{x,\bar{z}} = \varnothing$, $P_{\tilde{F}}(HS(x)) = (1-\varepsilon)\, P_F(HS(x)) + \varepsilon$ for any closed halfspace including x, which implies $d_{HS}(x; \tilde{F}) = (1-\varepsilon)\, d_{HS}(x; F) + \varepsilon$. Hence $IF(x; d_{HS}(x; F)) = 1 - d_{HS}(x; F)$.

3.2. *Properties of the influence function.*

1. Since $d_{HS}(x; F)$ is a probability, $IF(z; d_{HS}(x; F))$ is bounded and, for all $z \in \mathbb{R}^p$, $-1 \leqslant IF(z; d_{HS}(x; F)) \leqslant 1$.

2. Putting an additional infinitesimal mass at a point z in an optimal halfspace has a positive influence on $d_{HS}(x; F)$, that is, the degree of centralness of x increases. Conversely, if z does not belong to an optimal halfspace, the perturbation has a negative influence, and the degree of centralness of x decreases. More specifically, since

$$d_{HS}(x; \tilde{F}) = \begin{cases} d_{HS}(x; F) - \varepsilon d_{HS}(x; F), & \text{if} \quad d_{HS}^{(\bar{z})}(x; F) < d_{HS}^{(z)}(x; F); \\ d_{HS}(x; F) + \varepsilon(1 - d_{HS}(x; F)), & \text{otherwise,} \end{cases}$$

the halfspace depth of x can be arbitrarily augmented (reduced) by concentrating enough mass on a point z in an optimal (non-optimal) halfspace.

3. $IF(z; d_{HS}(x; F))$ is a step function. For all z belonging to optimal halfspaces, $IF(z; d_{HS}(x; F))$ is constant and equal to $1 - d_{HS}(x; F)$, whereas

for z belonging to non-optimal halfspaces, $IF(z; d_{HS}(x; F))$ is constant and equal to $-d_{HS}(x; F)$. It follows that the influence function is discontinuous at the boundary between optimal and non-optimal halfspaces. The jump is equal to 1, irrespective of the value of $d_{HS}(x; F)$. This implies that, in general, the local-shift sensitivity of $d_{HS}(x; F)$ is infinite. A notable exception are the deepest points because in this case the union of the optimal halfspaces is $\mathbb{R}^p$ [17, Proposition 12], hence $IF(z; d_{HS}(x; F))$ is everywhere continuous and the local-shift sensitivity is equal to zero. Appendix 1 shows some pictures of the influence function for a bivariate Pareto distribution.

4. According to intuition, the depth measure of peripheral points is more exposed to contamination of the theoretical distribution than the depth measure of central points. Consider, for example, the uniform distribution with positive density on the points of the unit square $\{x=(x_1, x_2)^T \in \mathbb{R}^2 : 0 \leqslant x_i \leqslant 1, i=1, 2\}$. The deepest point is the centroid $x=(1/2, 1/2)^T$ and $d_{HS}((1/2, 1/2)^T; F)=1/2$. Since the union of the optimal halfspaces is $\mathbb{R}^2$, $IF(z; d_{HS}((1/2, 1/2)^T; F))$ is constant and equal to 1/2 for all $z \in \mathbb{R}^2$ and $\sup_{z \in \mathbb{R}^2} |IF(z; d_{HS}((1/2, 1/2)^T; F))| = 1/2$. Moreover, for any point on the boundary of the square, $d_{HS}(x; F)=0$ and $\sup_{z \in \mathbb{R}^2} |IF(z; d_{HS}(x; F))|=1$. This confirms, though in an indirect way, the robustness of statistical functionals using inner contours of halfspace depth.

5. For any fixed point $x \in \mathbb{R}^p$, $IF(z; d_{HS}(x; F))$ defines a two-values random number whose distribution depends on F. By definition of halfspace depth, the probability p_x of this random number being equal to $1-d_{HS}(x; F)$ is the total probability of points z belonging to the optimal halfspaces, that is, $p_x = \int_{A_x} dF(z)$, where $A_x = \{z \in \mathbb{R}^p : z \text{ belongs to optimal halfspaces}\}$. Then

$$E_F(IF(z; d_{HS}(x; F))) = (1-d_{HS}(x; F))\, p_x - d_{HS}(x; F)(1-p_x)$$
$$= p_x - d_{HS}(x; F),$$

and

$$Var_F(IF(z; d_{HS}(x; F))) = (1-d_{HS}(x; F))^2\, p_x + d^2_{HS}(x; F)(1-p_x)$$
$$- (p_x - d_{HS}(x; F))^2$$
$$= p_x(1-p_x).$$

If the optimal halfspace is unique, then $p_x = d_{HS}(x; F)$ and in such a case $E_F(IF(z; d_{HS}(x; F)))=0$, $Var_F(IF(z; d_{HS}(x; F))) = d_{HS}(x; F)(1-d_{HS}(x; F))$.

6. For probability distributions F satisfying suitable regularity conditions (e.g., [21, Chap. 6]), if the optimal halfspace associated to x is unique, we can write the von Mises expansion

$$d_{HS}(x; \hat{F}_n) - d_{HS}(x; F) = \frac{1}{n} \sum_{i=1}^{n} IF(X_i; d_{HS}(x; F)) + \text{remainder term}.$$

Therefore the sample depth $d_{HS}(x; \hat{F}_n)$ could be asymptotically Normal, with mean $d_{HS}(x; F)$ and variance $d_{HS}(x; F)(1 - d_{HS}(x; F))/n$. The remainder term is investigated in Appendix 2. This result suggests that the distribution of sample depth could be more dispersed for points with halfspace depth near to 1/2.

4. SAMPLE INFLUENCE

When the available information is a random sample of n iid observations X_i from F, several estimators of the theoretical influence function can be used (e.g., [3]). In this work we consider the sample influence values

$$SI(X_i; T(F)) = -(n-1)(T(\hat{F}^{(i)}_{n-1}) - T(\hat{F}_n)),$$

where $\hat{F}^{(i)}_{n-1}$ is the empirical distribution of the subsample of size $n-1$ obtained by discarding the ith observation X_i. Up to the normalization factor $-(n-1)$, the SI values give the leave-one-out effect of each data point on the estimate $T(\hat{F}_n)$.

4.1. *Halfspace depth.* Fix $x \in \mathbb{R}^p$ and let $d_{HS}(x; \hat{F}_n)$ be the sample halfspace depth of x. Identifying z with the i-th sample observation X_i and assuming $X_i \neq x$, the derivation of the sample influence values of $d_{HS}(x; F)$ closely parallels Section 3.1. Here $\mathscr{H}_{x, X_i}$ is the set of closed halfspaces including both x and X_i, whereas $\mathscr{H}_{x, \bar{X}_i}$ is the set of closed halfspaces including x, but not X_i. Moreover, we denote with $d^{(X_i)}_{HS}(x; \hat{F}_n) \equiv l(x)/n$ and $d^{(\bar{X}_i)}_{HS}(x; \hat{F}_n) \equiv m(x)/n$ the sample depth of x conditional on halfspaces including and not including X_i, respectively.

If $m(x) < l(x)$, then $d_{HS}(x; \hat{F}_n) = m(x)/n$ and $d_{HS}(x; \hat{F}^{(i)}_{n-1}) = m(x)/(n-1)$ because $d^{(\bar{X}_i)}_{HS}(x; \hat{F}^{(i)}_{n-1}) = m(x)/(n-1) \leqslant d^{(X_i)}_{HS}(x; \hat{F}^{(i)}_{n-1}) \equiv (l(x)-1)/(n-1)$. This implies

$$SI(X_i; d_{HS}(x; F)) = -(n-1)\left(\frac{m(x)}{n-1} - \frac{m(x)}{n}\right)$$

$$= -\frac{m(x)}{n} = -d_{HS}(x; \hat{F}_n).$$

If $m(x) \geqslant l(x)$, then $d_{HS}(x; \hat{F}_n) = l(x)/n$, $d_{HS}(x; \hat{F}^{(i)}_{n-1}) = (l(x)-1)/(n-1)$ and

$$SI(X_i; d_{HS}(x; F)) = -(n-1)\left(\frac{l(x)-1}{n-1} - \frac{l(x)}{n}\right)$$

$$= 1 - \frac{l(x)}{n} = 1 - d_{HS}(x; \hat{F}_n).$$

In the practical applications, interest often centers on the values $SI(X_i; d_{HS}(X_j; F))$, $j = 1, ..., n$, $j \neq i$, which describe the effect of deleting the i-th observation on the depth measure of the remaining observations. As in Section 3.2, the variation of the halfspace depth of X_j is positive or negative according to whether X_i belongs to optimal halfspaces or not. Note that $SI(X_i; d_{HS}(X_i; F)) = 1 - d_{HS}(X_i; \hat{F}_n) > 0$.

4.2. *Halfspace regions.* When the ith sample observation is discarded, the (absolute) depth $nd_{HS}(x; \hat{F}_n)$ either remains unchanged or is reduced by one. This suggests a regular pattern of variation of the perturbed halfspace regions $D_{HS}(d; \hat{F}^{(i)}_{n-1})$ with respect to the unperturbed regions $D_{HS}(d; \hat{F}_n)$. A few properties are described in Proposition 1.

For $x \in \mathbb{R}^p$, the empirical depth $d_{HS}(x; \hat{F}_n)$ is a rational number k/n, $k \in \{0, 1, ..., k^*_n\}$. For data sets in general position $\lceil n/(p+1) \rceil \leqslant k^*_n \leqslant \lceil n/2 \rceil$ [4, Proposition 2.3], where $\lceil n/2 \rceil$ is the smallest integer not less than $n/2$.

PROPOSITION 1. *Suppose the observation X_i, $i \in \{1, ..., n\}$, with halfspace depth $d_{HS}(X_i; \hat{F}_n) = k/n$, is discarded from the sample. For $h \in \{1, ..., k^*_n\}$, the depth regions $D_{HS}(h/(n-1); \hat{F}^{(i)}_{n-1})$ satisfy*

(a) $D_{HS}(h/(n-1); \hat{F}^{(i)}_{n-1}) \subseteq D_{HS}(h/n; \hat{F}_n)$ *and* $D_{HS}(h/(n-1); \hat{F}^{(i)}_{n-1}) = D_{HS}(h/n; \hat{F}_n)$ *if* $h < k$;

(b) $D_{HS}((h+1)/n; \hat{F}_n) \subseteq D_{HS}(h/(n-1); \hat{F}^{(i)}_{n-1}) \subseteq D_{HS}(h/n; \hat{F}_n)$.

Proof. (a) Let $\mathscr{H}(j)$, $j \in \{n-h+1, n-h+2, ..., n\}$ be the family of closed halfspaces containing j sample points chosen from $\{X_l, l = 1, ..., n\}$ and let $\mathscr{K}(j)$, $j \in \{n-h, n-h+1, ..., n-1\}$ be the family of closed halfspaces containing j sample points chosen from $\{X_l, l = 1, ..., n, l \neq i\}$. Then $D_{HS}(h/n; \hat{F}_n)$ is the intersection of all closed halfspaces belonging to the sets $\mathscr{H}(n-h+1)$, ..., $\mathscr{H}(n)$ and $D_{HS}(h/(n-1); \hat{F}^{(i)}_{n-1})$ is the intersection of the halfspaces belonging to the sets $\mathscr{K}(n-h)$, ..., $\mathscr{K}(n-1)$. Consider a closed halfspace $HS(j)$ belonging to $\mathscr{H}(j)$. If $X_i \in HS(j)$, then $HS(j) \in \mathscr{K}(j-1)$, otherwise $HS(j) \in \mathscr{K}(j)$. Moreover, any closed halfspace $HS(n-h) \in \mathscr{K}(n-h)$ which does not include X_i can not belong to any set $\mathscr{H}(j)$. Thus $D_{HS}(h/(n-1); \hat{F}^{(i)}_{n-1}) \subseteq D_{HS}(h/n; \hat{F}_n)$.

If $h<k$, the discarded observation X_i must be an inner point of the closed halfspaces containing $n-h+1$ sample observations. Otherwise there would exist a closed halfspace with h sample observations and containing X_i and the halfspace depth of X_i with respect to $\hat{F}_n$ would be equal to $h/n<k/n$. Then, for $j \in \{n-h+1, n-h+2, ..., n\}$, $\mathscr{H}(j)=\mathscr{K}(j-1)$ which implies $D_{HS}(h/(n-1); \hat{F}^{(i)}_{n-1}) = D_{HS}(h/n; \hat{F}_n)$.

(b) We have only to show that $D_{HS}((h+1)/n; \hat{F}_n) \subseteq D_{HS}(h/(n-1); \hat{F}^{(i)}_{n-1})$. Note that $D_{HS}((h+1)/n; \hat{F}_n)$ is the intersection of the closed halfspaces belonging to $\mathscr{H}(j)$, $j \in \{n-h, n-h+1, ..., n\}$. Let $HS(j)$ be any closed halfspace belonging to $\mathscr{K}(j)$. If $X_i \in HS(j)$, then $HS(J) \in \mathscr{H}(j+1)$, otherwise $HS(j) \in \mathscr{H}(j)$. Further, any closed halfspace $HS(n-h) \in \mathscr{H}(n-h)$ which includes X_i can not belong to any set $\mathscr{K}(j)$. ▮

By interchanging the rôles of $\hat{F}_n$ and $\hat{F}^{(i)}_{n-1}$ in Proposition 1, it is easy to describe what happens when a new observation X_{n+1} is added to the sample. We denote with $\hat{F}^{(X_{n+1})}_{n+1}$ the distribution function of the augmented sample $\{X_1, ..., X_n, X_{n+1}\}$.

PROPOSITION 2. *Let* $d_{HS}(X_{n+1}; \hat{F}^{(X_{n+1})}_{n+1}) = k/(n+1)$, $k \in \{1, ..., k^*_{n+1}\}$. *For* $h \in \{1, ..., k^*_{n+1}\}$, *the depth regions* $D_{HS}(h/(n+1); \hat{F}^{(X_{n+1})}_{n+1})$ *satisfy*

(a) $D_{HS}(h/n); \hat{F}_n) \subseteq D_{HS}(h/(n+1); \hat{F}^{(X_{n+1})}_{n+1})$ *and* $D_{HS}(h/n; \hat{F}_n) = D_{HS}(h/(n+1); \hat{F}^{(X_{n+1})}_{n+1})$ *if* $h<k$;

(b) $D_{HS}(h/n; \hat{F}_n) \subseteq D_{HS}(h/(n+1); \hat{F}^{(X_{n+1})}_{n+1}) \subseteq D_{HS}((h-1)/n; \hat{F}_n)$, *where* $D_{HS}(0; \hat{F}_n) \equiv \mathbb{R}^p$.

According to Proposition 2, a single new observation X_{n+1} outside the convex hull of $\{X_1, ..., X_n\}$ may alter all the depth regions $D_{HS}(h/n; \hat{F}_n)$, but only $D_{HS}(1/n; \hat{F}_n)$ can be arbitrarily expanded by letting $\max_{1 \leqslant j \leqslant p} |X_{n+1,j}| \to \infty$. In general, to arbitrarily expand all the depth regions up to order h, it would be necessary to add at least h new observations at X_{n+1} and let $\max_{1 \leqslant j \leqslant p} |X_{n+1,j}| \to \infty$ (compare with [4, Lemma 3.1]). In this sense deeper regions are more resistant to sample contamination than more external regions. Propositions 1 and 2 also suggest to take $\overline{D_{HS}((h-1)/n; \hat{F}_n) \cap D_{HS}((h+1)/n; \hat{F}_n)}$ as an empirical uncertainty region in the estimation of the boundary of $D_{HS}(h/n; F)$. The simplest way to measure the discrepancy between $D_{HS}(h/n; \hat{F}_n)$ and the perturbed region $D_{HS}(h/(n-1); \hat{F}^{(i)}_{n-1})$ (or $D_{HS}(h/(n+1); \hat{F}^{(X_{n+1})}_{n+1})$) is to take the (hyper)volume of $\overline{D_{HS}(h/(n-1); \hat{F}^{(i)}_{n-1}) \cap D_{HS}(h/n; \hat{F}_n)}$. Since the depth regions are nested convex bodies in $\mathbb{R}^p$, it could also be appropriate to use Hausdorff distance [8, p. 231].

4.3. *Simplicial depth.* Suppose that the n sample observations $X_1, ..., X_n$ are in general position. Form the set of $\binom{n}{p+1}$ open $(p+1)$-simplices corresponding to all possible subsets of $p+1$ data points and consider the subset $\mathscr{S}_{n,x}$ of the simplices having $x \in \mathbb{R}^p$ as an inner point. For any $i \in \{1, ..., n\}$ we can partition $\mathscr{S}_{n,x}$ into two complementary subsets: the subset $\mathscr{S}^{(X_i)}_{n,x}$ of the simplices having X_i as a vertex, and the subset $\mathscr{S}^{(\bar{X}_i)}_{n,x}$ of the simplices whose vertices are different from X_i. Let $s_n(x)$, $s_n^{(X_i)}(x)$ and $s_n^{(\bar{X}_i)}(x)$ be the cardinalities of $\mathscr{S}_{n,x}$, $\mathscr{S}^{(X_i)}_{n,x}$ and $\mathscr{S}^{(\bar{X}_i)}_{n,x}$, respectively. Then $s_n(x) = s_n^{(X_i)}(x) + s_n^{(\bar{X}_i)}(x)$, $d_S(x; \hat{F}_n) = s_n(x)/\binom{n}{p+1}$ and $d_S(x; \hat{F}^{(i)}_{n-1}) = s_n^{(\bar{X}_i)}(x)/\binom{n-1}{p+1} = (s_n(x) - s_n^{(X_i)}(x))/\binom{n-1}{p+1}$. Note that

$$d_S(x; \hat{F}_n) = \frac{\binom{n-1}{p+1}}{\binom{n}{p+1}} \frac{s_n^{(\bar{X}_i)}(x)}{\binom{n-1}{p+1}} + \frac{\binom{n-1}{p}}{\binom{n}{p+1}} \frac{s_n^{(X_i)}(x)}{\binom{n-1}{p}}$$

$$= \left(1 - \frac{p+1}{n}\right) d_S^{(\bar{X}_i)}(x; \hat{F}_n) + \frac{p+1}{n} d_S^{(X_i)}(x; \hat{F}_n),$$

where $d_S^{(\bar{X}_i)}(x; \hat{F}_n) = s_n^{(\bar{X}_i)}(x)/\binom{n-1}{p+1}$ is the simplicial depth of x conditional on simplices belonging to $\mathscr{S}^{(\bar{X}_i)}_{n,x}$ and $d_S^{(X_i)}(x; \hat{F}_n) = s_n^{(X_i)}(x)/\binom{n-1}{p}$ is the simplicial depth conditional on simplices belonging to $\mathscr{S}^{(X_i)}_{n,x}$. Thus the sample simplicial depth of x can be interpreted as a weighted mean of $d_S^{(\bar{X}_i)}(x; \hat{F}_n)$ and $d_S^{(X_i)}(x; \hat{F}_n)$. Now, $SI(X_i; d_S(x; F)) = -(n-1)(d_S(x; \hat{F}^{(i)}_{n-1}) - d_S(x; \hat{F}_n))$, and since $d_S(x; \hat{F}^{(i)}_{n-1}) - d_S(x; \hat{F}_n) = ((p+1)/n)(d_S^{(\bar{X}_i)}(x; \hat{F}_n) - d_S^{(X_i)}(x; \hat{F}_n))$, it follows

$$SI(X_i; d_S(x; F)) = \frac{n-1}{n}(p+1)(d_S^{(X_i)}(x; \hat{F}_n) - d_S^{(\bar{X}_i)}(x; \hat{F}_n)).$$

The SI values depend essentially on the difference of the conditional depths, therefore $|SI(X_i; d_S(x; F))| \leqslant (1 - 1/n)(p+1)$ and $SI(X_i; d_S(x; F)) = 0$ iff $d_S^{(X_i)}(x; \hat{F}_n) = d_S^{(\bar{X}_i)}(x; \hat{F}_n)$.We expect $d_S^{(X_i)}(x; \hat{F}_n) > d_S^{(\bar{X}_i)}(x; \hat{F}_n)$, hence positive SI values, for observations X_i close to x. Moreover, the factor $p+1$ suggests that simplicial depth can be more affected by sample contaminations as the dimensionality grows higher.

5. NUMERICAL ILLUSTRATIONS

5.1. *Star-cluster data set.* In this illustration we use the star-cluster data set (Rousseeuw and Leroy [14]) where the values of X: logarithm of the effective surface temperature and Y: logarithm of the light intensity are recorded for $n = 47$ stars. A remarkable property is the presence of four units far removed from the main group (stars no. 11, 20, 30 and 34, see the scatter plot in Fig. 1) which can be thought of as a contamination of the true distribution.

The sample values of halfspace and simplicial depth were computed by the DEPTH routine [15] and the results are given in Table I. The maximum value of halfspace depth among the sample observations is 18/47 (star no. 28 with coordinates 4.42, 4.90) and the maximum value of simplicial depth is 5221/16215 (stars no. 33 and 38, both with coordinates 4.45, 5.22). The discrepancy between these two results could suggest that simplicial depth is less resistant to sample contamination than halfspace depth.

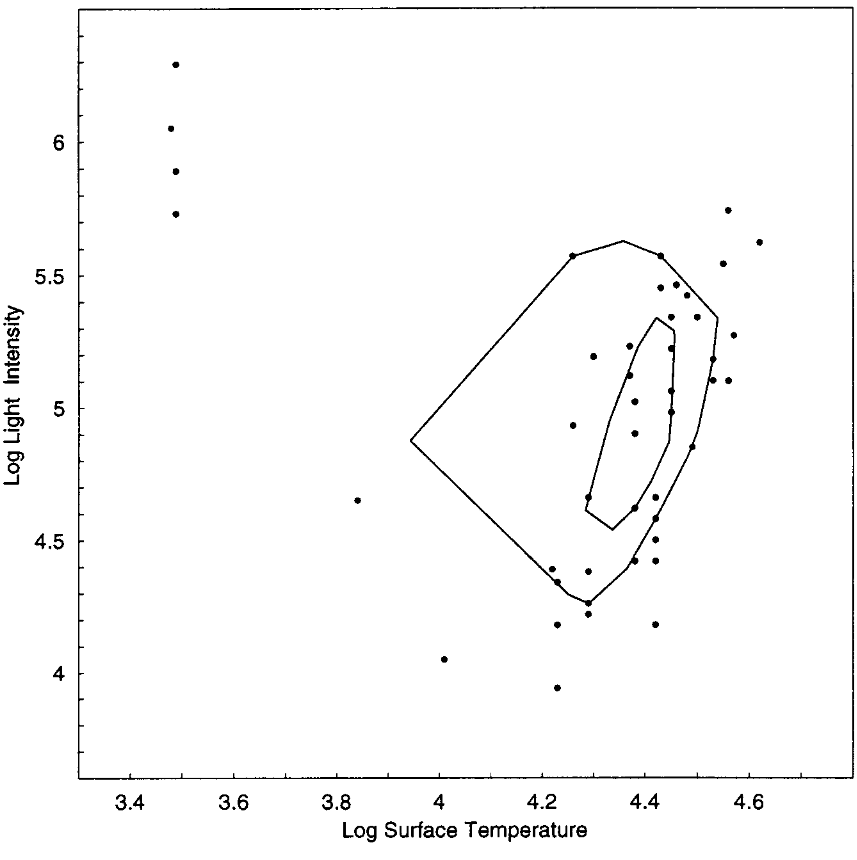


FIG. 1. Star data. Scatter plot with halfspace depth contours.

TABLE I

Star Data: Halfspace and Simplicial Depth (H-depth, $47 \times$ halfspace depth; S-depth, $\binom{47}{3} \times$ simplicial depth)

Star no.	*H-Depth*	*S-Depth*	*Star no.*	*H-Depth*	*S-Depth*
1	10	3629	25	17	4922
2	2	2025	26	8	3127
3	8	3281	27	12	3768
4	2	2025	28	18	4972
5	7	3303	29	7	2221
6	6	2355	30	1	1035
7	2	1079	31	4	1843
8	2	1121	32	1	1035
9	5	1903	33	15	5221
10	13	4291	34	1	1035
11	1	1035	35	6	1961
12	8	2571	36	1	1035
13	6	2709	37	4	1877
14	1	1035	38	15	5221
15	5	2052	39	5	2173
16	5	2549	40	5	1713
17	1	1035	41	12	3795
18	1	1035	42	13	4362
19	3	1380	43	7	2820
20	2	1122	44	11	3583
21	8	2884	45	4	1619
22	3	1533	46	12	3917
23	2	1601	47	3	2081
24	4	1469			

The depth contours of levels 5/47 anf 12/47, computed by the ISODEPTH routine [20], are also shown in Fig. 1. The shape of the outer region is perturbed by the four outlying stars, whereas the inner region is scarcely affected.

To illustrate the behaviour of the influence function, first we consider the point $x_A = (4.2, 5.8)^T$ and calculate the SI values with respect to halfspace and simplicial depth for each sample observation X_i. Now, $d_{HS}(x_A; \hat{F}_n) = 2/47$ and $d_S(x_A; \hat{F}_n) = 213/\binom{47}{3}$. Moreover,

$$SI(X_i; d_{HS}(x_A; F)) = \begin{cases} -2/47, & i \notin \{30, 34\}, \\ 45/47, & i \in \{30, 34\}, \end{cases}$$

in perfect agreement with the expected results, because X_{30} and X_{34} are the two sample observations included in the optimal halfplane. On the other hand,

$$SI(X_i;\, d_S(x_A; F)) = \begin{cases} -0.026, & i \notin \{2, 4, 30, 34, 36\}, \\ 0.095, & i = 36, \\ 0.216, & i = 2,\, i = 4, \\ 0.222, & i = 30, \\ 0.344, & i = 34, \end{cases}$$

which confirms the positive influence of sample points located in a neighbourhood of x_A. It is clear that the influence function of simplicial depth gives a more detailed description than the influence function of halfspace depth. Note, in particular, the inversion of signs for observations no. 2, 4, and 36.

Next, consider the point $x_B = (4.4, 5.05)^T$ which is located in the median region with respect to halfspace depth. The depth values are $d_{HS}(x_B; \hat{F}_n) = 19/47$, $d_S(x_B; \hat{F}_n) = 4189/\binom{47}{3}$ and

$$SI(X_i;\, d_{HS}(x_B; F)) = \begin{cases} -19/47, & i \in \mathscr{I}, \\ 28/47, & i \notin \mathscr{I}, \end{cases}$$

where $\mathscr{I} = \{15, 16, 17, 18, 21, 22, 23, 26, 28, 31, 41, 47\}$. The stem-and-leaf display of the SI values of $d_S(x_B; F)$ is given in Table II. Now, the number

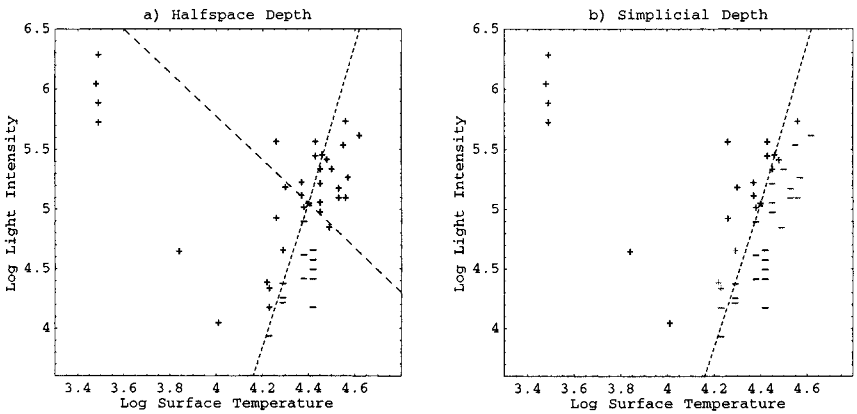


FIG. 2. Star data. Scatter plot and SI values (darker points have higher values of $|SI|$; "$-$", negative SI; "$+$", positive SI; "*" marks the position of x_B). (a) Halfspace depth, (b) simplicial depth.

TABLE II

Star Data: Stem-and-Leaf Display of $SI(X_i; d_S(x_B; F))$, $i = 1, ..., 47$

-1l	
-1h	10, 10, 10, 10, 10, 10, 10
-0l	92, 92, 92, 74, 74, 56, 56, 50
-0h	25, 25, 25, 19, 19, 13, 13, 13, 13
0l	05, 05, 05, 41, 41, 41, 41
0h	66, 66, 66, 72, 72, 72, 72, 72, 72, 72
1l	02, 02, 02, 08, 44, 44
1h	

of sample observations belonging to any optimal halfplane is equal to 19. Nevertheless, the sample influence function of $d_{HS}(x_B; F)$ is positive for 35 observations out of 47 because x_B has several optimal halfplanes, each containing 19 observations (Fig. 2a shows two such halfplanes). From Table II, the behaviour of the influence function of simplicial depth appears different. The stem-and-leaf suggests an almost symmetric distribution, centered at zero. Moreover, the scatter plot in Fig. 2b reveals that data points with positive SI values for simplicial depth tend to concentrate in one of the two optimal halfplanes for halfspace depth.

To observe the effects on the depth contours of addition/deletion of single units, we considered the depth region $D_{HS}(5/47; \hat{F}_n)$ and the perturbed regions $D_{HS}(5/46; \hat{F}^{(9)}_{n-1})$, obtained by discarding unit no. 9, and

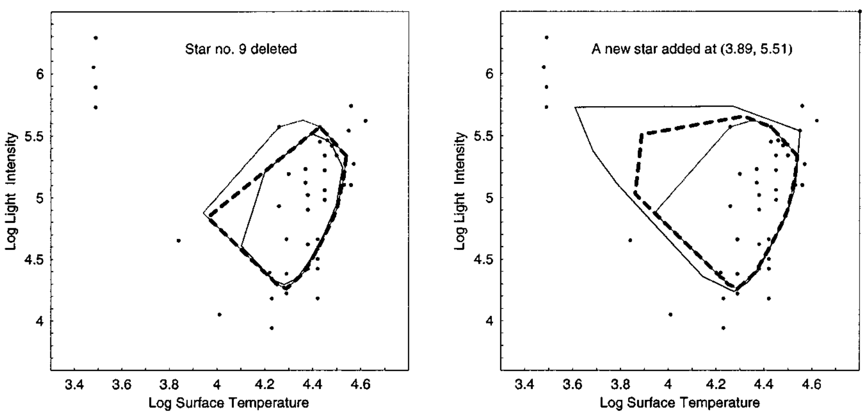


FIG. 3. Star data. Perturbation of the halfspace region of level 5/47 by deletion/addition of a single unit (solid lines, contours of level 6/47, 5/47, and 4/47; dashed lines, perturbed contours).

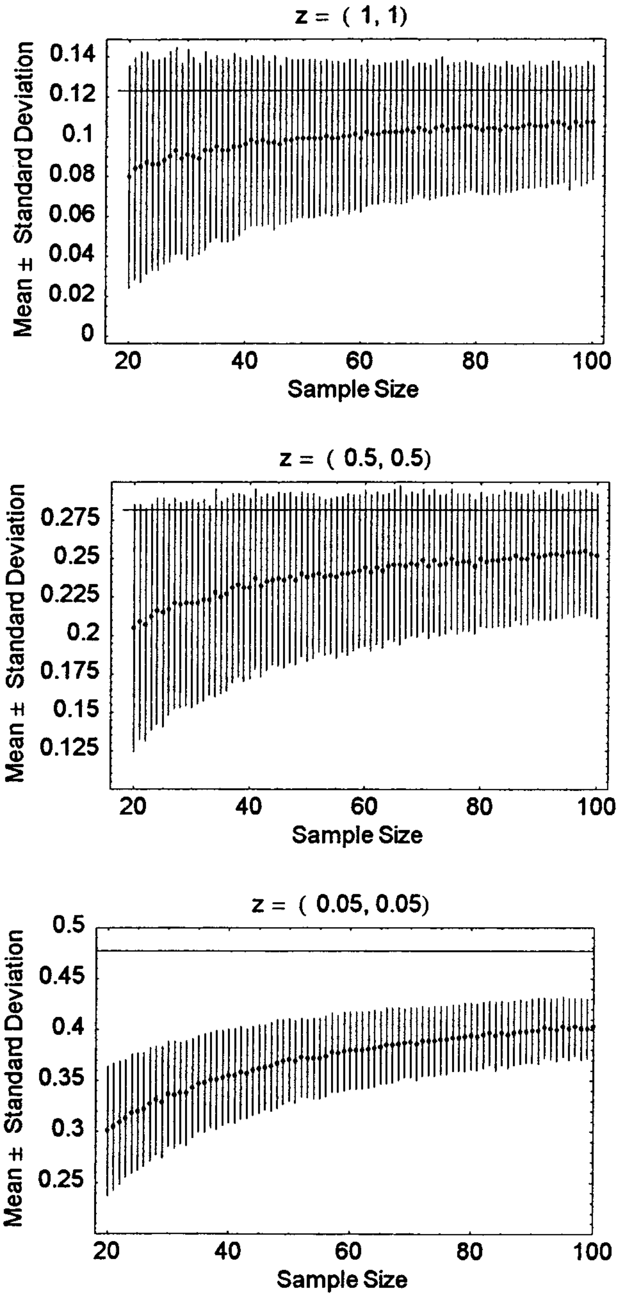


FIG. 4. Artificial normal data. Mean and standard deviation over 1000 replications of $d_{HS}(z; \hat{F}_n)$ (horizontal line: population depth).

$D_{HS}(5/48; \hat{F}^{(z)}_{n+1})$, obtained by augmenting the sample with $z=(3.89, 5.51)^T$. According to Propositions 1 and 2, the boundaries of the perturbed regions (dashed lines in Fig. 3) are included in the set $D_{HS}(4/47; \hat{F}_n) \cap \overline{D_{HS}(6/47; \hat{F}_n)}$.

5.2. *Simulation results.* To investigate the properties of the distribution of the sample halfspace depth, we performed the following Monte Carlo experiment. For each value of $n \in \{20, 21, ..., 100\}$, 1000 pseudo-random samples of size n were drawn from the bivariate Normal distribution centered at $(0, 0)^T$ and with covariance matrix $\Sigma = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$. The empirical distribution of $d_{HS}(z; \hat{F}_n)$ over the 1000 replications was examined for $z_1=(1, 1)^T$, $z_2=(0.5, 0.5)^T$ and $z_3=(0.05, 0.05)^T$. The population values of halfspace depth are $d_{HS}(z_1; F) = \Phi(-2/\sqrt{3}) \simeq 0.123$, $d_{HS}(z_2; F) = \Phi(-1/\sqrt{3}) \simeq 0.282$ and $d_{HS}(z_3; F) = \Phi(-1/(10\sqrt{3})) \simeq 0.476$, where $\Phi(.)$ is the cumulative distribution function of the standard Normal distribution.

The most important finding is that the average of $d_{HS}(z; \hat{F}_n)$ over the 1000 replications is smaller than the theoretical value $d_{HS}(z; F)$ for each value of n and all z. This suggests that $d_{HS}(z; \hat{F}_n)$ tends to underestimate $d_{HS}(z; F)$, in particular when the population depth is near to 1/2 (we investigate this point from the theoretical point of view in Appendix 2). As Fig. 4 shows, the bias is a decreasing function of n, but it is still important for sample sizes in the range $80-100$. As a consequence, the empirical coverage of the asymptotic confidence interval $d_{HS}(x; \hat{F}_n) \mp k_{1-\alpha/2} \sqrt{d_{HS}(x; \hat{F}_n)(1-d_{HS}(x; \hat{F}_n))/n}$ turned out to be always smaller than the nominal value, with wide fluctuations over n. In this case there is no evidence of a systematic improvement as n grows higher, maybe because not only the bias but also the standard error is a decreasing function of n. A little better results were obtained with the variance-stabilizing transformation $d_{HS}(x; \hat{F}_n) \rightarrow \arcsin \sqrt{d_{HS}(x; \hat{F}_n)}$.

The histograms of the values of $d_{HS}(x; \hat{F}_n)$ (not shown in this work) suggest a positively skewed distribution with a tendency to become almost symmetric as n grows higher. However, if $d_{HS}(x; F)$ is near to 1/2, the distribution is negatively skewed and the degree of asymmetry remains important even for n as high as $80-100$.

6. DISCUSSION

The sensitivity analysis of halfspace depth values and regions gives new insight on the behaviour of these statistics. Central values and regions are more stable, whereas outer values and regions are markedly affected by sample contaminations. To arbitrarily expand a depth region of (absolute) level k, one must place k new observations at a sufficiently far point z, outside the convex hull of the sample. In a sense, the breakdown point of the k-level depth region is $k/(n+k)$. This behaviour compares favourably with the classical parametric measure of centrality, Mahalanobis' (squared) distance $d_M^2(x; F) = (x-\mu)^T \Sigma^{-1}(x-\mu)$, whose influence function

$$IF(z; d_M^2(x; F)) = d_M^2(x; F) - 2(z-\mu)^T \Sigma^{-1}(x-\mu) \\ - ((x-\mu)^T \Sigma^{-1}(z-\mu))^2$$

is unbounded.

Distributional properties of statistics arising from depth measures are investigated in several papers. Eddy [5] considers the joint distribution of the order statistics of $\{u^T X_i, i, 1, ..., n\}$ and $\{v^T X_i, i, 1, ..., n\}$ for arbitrary p-vectors u and v. Nolan [11] proves that the radius function of a depth region converges weakly to a Gaussian process. In the same paper she describes the limit distribution of the direction normal to any optimal halfspace. More recently, Nolan [12] gives the asymptotic distribution of the deepest point in the bivariate case and Bai and He [1] extend her result to the general p-dimensional case. In the present paper we concentrate on the empirical depth of a fixed point x. Under random sampling from F, the number of sample observations included in any halfspace $HS(x)$ has a Binomial distribution with parameters n, the sample size, and p_F, the probability of $HS(x)$. Hence $d_{HS}(x; \hat{F}_n)$ is the minimum of a family of correlated Binomial distributions. Rousseeuw and Struyf [19] prove that the search for the minimum can be confined to a finite number of distributions. Under suitable conditions on F, the von Mises expansion described in Section 3 implies that $d_{HS}(x; \hat{F}_n)$ is asymptotically Normal, with mean $d_{HS}(x; F)$ and standard deviation $\sqrt{d_{HS}(x; F)(1-d_{HS}(x; F))/n}$. This result suggests the naive confidence interval $d_{HS}(x; \hat{F}_n) \mp k_{1-\alpha/2} \sqrt{d_{HS}(x; \hat{F}_n)(1-d_{HS}(x; \hat{F}_n))/n}$ for the population depth $d_{HS}(x; F)$. However, the Monte Carlo experiment described in Section 5 shows some discrepancies between the finite-sample and the asymptotic distribution. In particular, there is a strong evidence, supported by theoretical results of Appendix 2, that $d_{HS}(x; \hat{F}_n)$ is a biased estimator.

The evaluation of the sampling variability of the halfspace regions is a more trying task. Our suggestion is to take the set $D_{HS}((h-1)/n; \hat{F}_n) \cap D_{HS}((h+1)/n; \hat{F}_n)$ as an uncertainty region for the estimate of the boundary of $D_{HS}(h/n; F)$, but this proposal needs further research and empirical checking.

APPENDIX 1

The bivariate Pareto distributions are a parametric family indexed by a positive parameter α. The density function

$$f(x;\alpha)=\begin{cases}\alpha(\alpha+1)(x_1+x_2-1)^{-(\alpha+2)}, & x_1>1\cap x_2>1,\\ 0, & \text{elsewhere,}\end{cases}$$

is decreasing along any ray from $(1,1)^T$ and is constant on the segments $x_1+x_2=c$, $c>0$, $x_1>1\cap x_2>1$. The halfspace depth function of the Pareto distribution with $\alpha=1$ is

$$d_{HS}(x;F)=\begin{cases}d_1(x)=4(x_2-1)/(x_1+x_2-1)^2,\\ \qquad x_1>x_2\cap x_1>2,\\ d_2(x)=4(x_1-1)/(x_1+x_2-1)^2,\\ \qquad x_1\leqslant x_2\cap x_2>2,\\ d_3(x)=4(x_1-1)(x_2-1)/(x_1+x_2-1)^2,\\ \qquad 1<x_1\leqslant 2\cap 1<x_1\leqslant 2.\end{cases}$$

The deepest point is $x_A=(2,2)^T$, coincident with the coordinate-wise median and $d_{HS}(x_A;F)=4/9<1/2$. Note that the cumulative distribution function is equal to 1/2 at $\tilde{x}=((7+\sqrt{17})/4)(1,1)^T$ and $d_{HS}(\tilde{x};F)\simeq 0.342$.

The contours of $d_{HS}(x;F)$ are formed by three arcs: the first and the second are arcs of parabolae deriving from $d_i(x)=d$, $i\in\{1,2\}$ (d is a constant value of depth, $0\leqslant d\leqslant 4/9$); the third is an arc of hyperbola obtained from $d_3(x)=d$. They are symmetric with respect to the line $x_1-x_2=0$ but not with respect to the orthogonal line $x_1+x_2=4$. The positive orientation of the contours is coherent with the slopes of the regression lines $E(X_j\,|\,x_i)=1+x_i$, $i,j\in\{1,2\}$, $i\neq j$. The contours corresponding to $d=0.01$, 0.10, 0.25 are shown in Fig. 5 together with a 3D-plot of the depth function.

Now we discuss the influence function of selected points.

The simplest case is the deepest point x_A. It is easy to check that three optimal halfplanes are

$$x_1-2x_2+2\leqslant 0,\qquad x_1+x_2-4\leqslant 0,\qquad 2x_1-x_2-2\geqslant 0,$$

whose union is $\mathbb{R}^2$ (note that, according to [17], Proposition 12, a ray basis can be derived from the boundary lines of these halfplanes). Thus, for all $z\in\mathbb{R}^2$, $IF(z;d_{HS}(x_A;F))=5/9$, a constant (non-random) number: in a sense, any perturbation can only rise the depth of x_A. As stated in Proposition 12 of Rousseeuw and Ruts [17], this behaviour is typical of deepest points of absolutely continuous distributions.

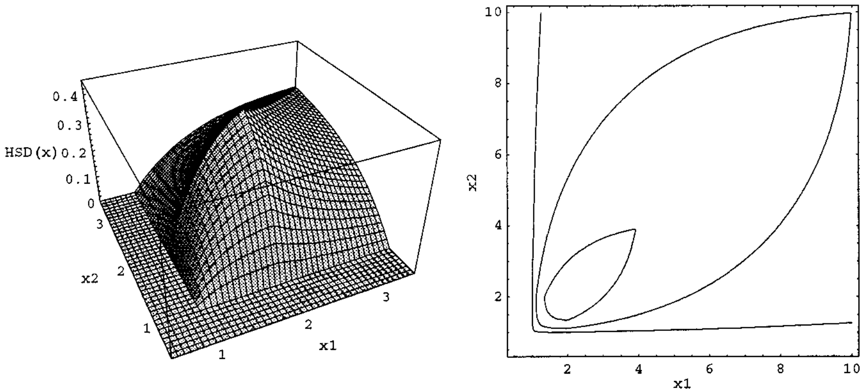


FIG. 5. 3D-plot and contour plot of the halfspace depth function of the bivariate Pareto distribution with parameter $\alpha = 1$.

The second example is $x_B = (1, 1)^T$, whose depth is obviously zero. The optimal region is the union of the halfplanes

$$x_1 \leqslant 1, \qquad x_2 \leqslant 1.$$

It follows that the influence function is

$$IF(z; d_{HS}(x_B; F)) = \begin{cases} 0, & z_1 > 1 \cap z_2 > 1, \\ 1, & \text{elsewhere.} \end{cases}$$

This random number is a.e. (with respect to F) equal to zero.

The point $x_C = (3, 3)^T$ with depth $d_{HS}(x_C; F) = 8/25$ has two optimal halfplanes

$$4x_1 - x_2 - 9 \geqslant 0, \qquad x_1 - 4x_2 + 9 \leqslant 0,$$

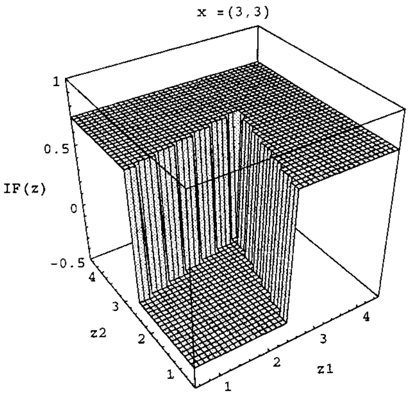
and the influence function is

$$IF(z; d_{HS}(x_C; F)) = \begin{cases} -8/25, & 4z_1 - z_2 - 9 < 0 \cap z_1 - 4z_2 + 9 > 0, \\ 17/25, & \text{elsewhere.} \end{cases}$$

Finally, for $x_D = (3, 5)^T$ with $d_{HS}(x_C; F) = 16/49$, the optimal halfplane is $3x_1 - 4x_2 + 11 \leqslant 0$ and the influence function is

$$IF(z; d_{HS}(x_D; F)) = \begin{cases} -16/49, & 3z_1 - 4z_2 + 11 > 0, \\ 33/49, & \text{elsewhere.} \end{cases}$$

The plots of the influence functions of x_C and x_D are given in Fig. 6.



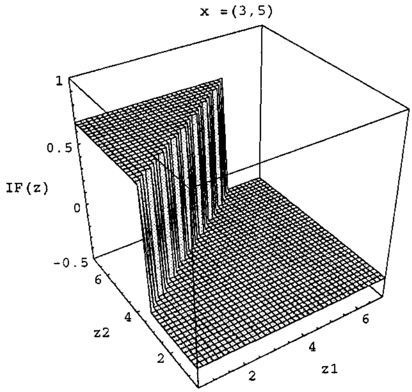


FIG. 6. 3D-plots of $IF(z; d_{HS}(x_C; F))$ and $IF(z; d_{HS}(x_D; F))$; F is the bivariate Pareto distribution with parameter $\alpha = 1$.

APPENDIX 2

For $x \in \mathbb{R}^p$ let $\psi(z; HS(x))$ be the indicator function of any closed halfspace $HS(x)$ including x:

$$\psi(z; HS(x)) = \begin{cases} 1, & z \in HS(x), \\ 0, & z \notin HS(x). \end{cases}$$

The halfspace depth of x is the functional

$$d_{HS}(x; F) = \inf_{\mathscr{H}_x} \int \psi(z; HS(x))\, dF(z)$$

$$= \inf_{\mathscr{H}_x} E_F(\psi(z; HS(x))),$$

and, if the optimal halfspace $HS^*(x; F) \equiv HS_x^*$ is unique, the influence function of $d_{HS}(x; F)$ can be written as

$$IF(z; d_{HS}(x; F)) = \psi(z; HS_x^*) - d_{HS}(x; F)$$

$$= \psi(z; HS_x^*) - \int \psi(z; HS_x^*)\, dF(z).$$

The first-order von Mises approximation of a functional $T(F)$ at the empirical distribution function $\hat{F}_n$ is (e.g., [21], Chapter 6)

$$T(\hat{F}_n) = T(F) + T_1(F; \hat{F}_n - F) + R_1(F; \hat{F}_n - F),$$

where

$$T_1(F; \hat{F}_n - F) = \int IF(z; T(F))\, d(\hat{F}_n - F)(z),$$

and

$$R_1(F; \hat{F}_n - F) = T(\hat{F}_n) - T(F) - T_1(F; \hat{F}_n - F).$$

For the halfspace depth functional, the linear term is

$$\begin{aligned} T_1(F; \hat{F}_n - F) &= \int (\psi(z; HS^*_x) - d_{HS}(x; F))\, d(\hat{F}_n - F)(z) \\ &= \int (\psi(z; HS^*_x) - d_{HS}(x; F))\, d\hat{F}_n(z) \\ &= \frac{1}{n} \sum_{i=1}^{n} \psi(X_i; HS^*_x) - E_F(\psi(z; HS^*_x)) \\ &= P_{\hat{F}_n}(HS^*_x) - P_F(HS^*_x), \end{aligned}$$

where $\{X_1, ..., X_n\}$ is a set of n iid observations from F. Thus, the von Mises expansion approximates the sampling error $d_{HS}(x; \hat{F}_n) - d_{HS}(x; F)$ by the difference between the empirical measure and the theoretical P_F-measure of the same P_F-optimal halfspace HS^*_x. Of course, $nP_{\hat{F}_n}(HS^*_x) = \sum_{i=1}^{n} \psi(X_i; HS^*_x)$ has the Binomial distribution $Bi(n; d_{HS}(x; F))$.

For a given n, let $HS^*(x; \hat{F}_n) \equiv HS^*_{n,x}$ be the $P_{\hat{F}_n}$-optimal halfspace (i.e., the halfspace including the minimum number of sample observations) and let $\psi(z; HS^*_{n,x})$ be the corresponding indicator function. The remainder term is

$$\begin{aligned} R_1(F; \hat{F}_n - F) &= E_{\hat{F}_n}(\psi(z; HS^*_{n,x}) - \psi(z; HS^*_x)) \\ &= \frac{1}{n} \sum_{i=1}^{n} (\psi(X_i; HS^*_{n,x}) - \psi(X_i; HS^*_x)) \\ &= P_{\hat{F}_n}(HS^*_{n,x}) - P_{\hat{F}_n}(HS^*_x), \end{aligned}$$

that is, the difference between the empirical measure of the $P_{\hat{F}_n}$-optimal halfspace and the empirical measure of the P_F-optimal halfspace. The very definition of optimal halfspace implies $P_{\hat{F}_n}(HS^*_{n,x}) \leqslant P_{\hat{F}_n}(HS^*_x)$, thus $R_1(F; \hat{F}_n - F) \leqslant 0$ for any n. This proves that $d_{HS}(x; \hat{F}_n)$ is a biased estimator of $d_{HS}(x; F)$.

We have already shown that, if the P_F-optimal halfspace is unique, then

$$E_F(IF(z; d_{HS}(x; F))) = 0,$$

and

$$Var_F(IF(z; d_{HS}(x; F))) = d_{HS}(x; F)(1 - d_{HS}(x; F)) > 0,$$

provided that $d_{HS}(x; F) > 0$. A central limit theorem argument would then follow from $\sqrt{n}\, R_1(F; \hat{F}_n - F) \to^P 0$, i.e.,

$$\sqrt{n}\,(P_{\hat{F}_n}(HS^*_{n,x}) - P_{\hat{F}_n}(HS^*_x)) \to^P 0.$$

Since $-\sqrt{n}\, R_1(F; \hat{F}_n - F) = \sqrt{n}\,(P_{\hat{F}_n}(HS^*_x) - P_{\hat{F}_n}(HS^*_{n,x}))$ is a non-negative random variable, by Chebycev's inequality

$$\begin{aligned} 0 &\leqslant P\{\sqrt{n}\,(P_{\hat{F}_n}(HS^*_x) - P_{\hat{F}_n}(HS^*_{n,x})) \geqslant \varepsilon\} \\ &\leqslant E\{\sqrt{n}\,(P_{\hat{F}_n}(HS^*_x) - P_{\hat{F}_n}(HS^*_{n,x}))\}/\varepsilon \\ &= \sqrt{n}\,\{d_{HS}(x; F) - E(P_{\hat{F}_n}(HS^*_{n,x}))\}/\varepsilon. \end{aligned}$$

Now, $d_{HS}(x; \hat{F}_n) \equiv P_{\hat{F}_n}(HS^*_{n,x})$ is a consistent estimator of $d_{HS}(x; F)$, therefore, under suitable regularity conditions on the theoretical distribution F,

$$E(P_{\hat{F}_n}(HS^*_{n,x})) = d_{HS}(x; F) - \beta_n^2 + o_P(n^{-\gamma}),$$

where $\beta_n^2 = O_P(n^{-\gamma})$ is the finite-sample bias of $d_{HS}(x; \hat{F}_n)$. It follows that a sufficient condition for $\sqrt{n}\, R_1(F; \hat{F}_n - F) \to^P 0$ is $\gamma > 1/2$.

A more precise description of β_n^2 can be given. Let u^*_x be the unit vector perpendicular to the boundary of the P_F-optimal halfspace $HS^*_x = \{z \in \mathbb{R}^p : u^{*T}_x z \leqslant u^{*T}_x x\}$ and let $u^*_{n,x}$ be the unit vector perpendicular to the boundary of the $P_{\hat{F}_n}$-optimal halfspace $HS^*_{n,x} = \{z \in \mathbb{R}^p : u^{*T}_{n,x} z \leqslant u^{*T}_{n,x} x\}$. Under uniqueness of the P_F-optimal halfspace, Nolan [11, p. 162] shows that $u^*_{n,x}$ is a consistent estimator of u^*_x, and suggests the following representation of $u^*_{n,x}$

$$u^*_{n,x} = v_n + \sqrt{1 - v_n^T v_n}\, u^*_x,$$

where v_n is a random vector in a neighbourhood of the null vector 0_p, orthogonal to u^*_x. An important result for the present analysis is that $v_n^T v_n = O_P(-2/3)$ [11, Theorem 2, p. 166] because, up to a positive constant depending on x, $\beta_n^2 = E(v_n^T v_n)$. The regularity conditions which must be satisfied are (i) F has a continuous density, (ii) twice differentiability of $P_F(HS_{n,z})$ at $v_n = 0_p$ for z near x, and (iii) optimality of the vector $u^*_{n,x}$, i.e., $P_{\hat{F}_n}(z \in \mathbb{R}^p : u^{*T}_{n,x} z \leqslant u^{*T}_{n,x} x) \leqslant P_{\hat{F}_n}(z \in \mathbb{R}^p : u^T z \leqslant u^T x)$.

REFERENCES

1. Z.-D. Bai and X. He, Asymptotic distributions of the maximal depth estimators for regression and multivariate location, *Ann. Statist.* **27** (1999), 1616–1637.
2. Z. Chen, Bounds for the breakdown point of the simplicial median, *J. Multivariate Anal.* **55** (1995), 1–13.
3. F. Critchley, Influence in principal component analysis, *Biometrika* **72** (1985), 627–636.
4. D. L. Donoho and M. Gasko, Breakdown properties of location estimates based on halfspace depth and projected outlyingness, *Ann. Statist.* **20** (1992), 1803–1827.
5. W. F. Eddy, Ordering of multivariate data, *in* "Computer Science and Statistics: Proceedings of the 16th Symposium on the Interface" (L. Billard, Ed.), pp. 25–30, North-Holland, Amsterdam, 1985.
6. F. Hampel, The influence curve and its role in robust estimation, *J. Amer. Statist. Assoc.* **69** (1974), 383–393.
7. X. He and G. Wang, Convergence of depth contours for multivariate datasets, *Ann. Statist.* **25** (1997), 495–504.
8. P. J. Kelly and M. L. Weiss, "Geometry and Convexity," Wiley, New York, 1979.
9. R. Y. Liu, On a notion of data depth based on random simplices, *Ann. Statist.* **18** (1990), 405–414.
10. J.-C. Massé and R. Theodorescu, Halfplane trimming for bivariate distributions, *J. Multivariate Anal.* **48** (1994), 188–202.
11. D. Nolan, Asymptotics for multivariate trimming, *Stochastic Process. Appl.* **42** (1992), 157–169.
12. D. Nolan, On min-max majority and deepest points, *Statist. Probab. Lett.* **43** (1999), 325–333.
13. M. Romanazzi, A schematic plot for bivariate data based on halfspace depth, *Student* **2** (1997), 149–158.
14. P. J. Rousseeuw and A. M. Leroy, "Robust Regression and Outlier Detection," Wiley, New York, 1987.
15. P. J. Rousseeuw and I. Ruts, Algorithm AS 307: Bivariate location depth, *J. Roy. Statist. Soc. Ser. C* **45** (1996), 516–526.
16. P. J. Rousseeuw and I. Ruts, Constructing the bivariate Tukey median, *Statist. Sinica* **8** (1998), 827–839.
17. P. J. Rousseeuw and I. Ruts, The depth function of a population distribution, *Metrika* **49** (1999), 213–244.
18. P. J. Rousseeuw, I. Ruts, and J. W. Tukey, The bagplot: a bivariate boxplot, *Amer. Statist.* **53** (1999), 382–387.
19. P. J. Rousseeuw and A. Struyf, Computing location depth and regression depth in higher dimensions, *Statist. Comput.* **8** (1998), 193–203.

20. I. Ruts and P. J. Rousseeuw, Computing depth contours of bivariate point clouds, *Comput. Statist. Data Anal.* **23** (1996), 153–168.
21. R. J. Serfling, "Approximation Theorems of Mathematical Statistics," Wiley, New York, 1980.
22. J. W. Tukey, Mathematics and the picturing of data, *Proc. Internat. Congr. Math.* **2** (1975), 523–531.

Printed in Belgium