Influence Functions for a General Class of Depth-Based Generalized Quantile Functions

Jin Wang[1]

Northern Arizona University

and

Robert Serfling[2]

University of Texas at Dallas

June 2005

Final preprint version for *Journal of Multivariate Analysis*, 2005, to appear

[1]Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, Arizona 86011-5717. Email: jin.wang@nau.edu.

[2]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Email: serfling@utdallas.edu. Website: www.utdallas.edu/~serfling.

Support by National Science Foundation Grant DMS-0103698 is gratefully acknowledged.

Abstract

Given a multivariate probability distribution F, a corresponding *depth function* orders points according to their "centrality" in the distribution F. One useful role of depth functions is to generate two-dimensional curves for convenient and practical description of particular features of a multivariate distribution, such as dispersion and kurtosis. Here the robustness of sample versions of such curves is explored via the influence function approach applied to the relevant functionals, using structural representations of the curves as generalized quantile functions. In particular, for a general class of so-called Type D depth functions including the well-known Tukey or halfspace depth, we obtain influence functions for the depth function itself, the depth distribution function, the depth quantile function, and corresponding depth-based generalized quantile functions. Robustness behavior similar to the usual univariate quantiles is found and quantified: the influence functions are of step function form with finite gross error sensitivity but infinite local shift sensitivity. Applications to a "scale" curve, a Lorenz curve for "tailweight", and a "kurtosis" curve are treated. Graphical illustrations are provided for the influence functions of the scale and kurtosis curves in the case of the bivariate standard normal distribution and the halfspace depth function.

AMS 2000 Subject Classification: Primary 62G35 Secondary 62H05.

Key words and phrases: robustness; multivariate analysis; nonparametric methods; generalized quantiles; dispersion; kurtosis.

1 Introduction

A methodology gaining increasing application in nonparametric multivariate analysis is the use of *depth functions*. These provide a way of ordering points according to a notion of "centrality", or equivalently "outlyingness", where typically the point of maximal depth represents a reasonable notion of multivariate median and usually agrees with notions of center defined by symmetry considerations when such are applicable. See Liu, Parelius and Singh [10], Zuo and Serfling [25], and Mosler [12] for broad treatments of depth functions and depth-based methods and Serfling [18] for a recent overview. Since depth functions are basic to the development of the present paper, brief background is provided in an Appendix.

One leading role of depth functions is to generate *multivariate quantile functions*, via the contours of equal depth or equivalently equal outlyingness, and corresponding quantile-based nonparametric descriptive measures for dispersion, skewness, and kurtosis – see Serfling [17]. Another important use of depth functions – introduced by Liu, Parelius and Singh [10] and, indeed, the focus of the present paper – is to generate convenient, *one-dimensional* sample curves designed to provide visual display of particular features or characteristics of higher-dimensional distributions. In particular, they discuss and illustrate depth-based one-dimensional curves for visualization of *scale* and *kurtosis* features of multivariate distribution and comment "it is the very simplicity of such objects which make them powerful as a general tool for the practicing statistician".

With data we examine the *sample versions* of such "feature curves", as we might term them, analogously to the use of univariate empirical distribution functions. These, however, have the added appeal of being smooth curves instead of step functions. Some partial results on their distribution theory may be found in [15], [16]. In the present paper, our purpose is to characterize the *robustness* of any such depth-based sample "feature curve", via the *influence function approach* applied to the corresponding population curve. For example, for a particular depth-based sample kurtosis curve, we shall obtain bounded influence and thus finite gross error sensitivity, in contrast with the unbounded influence functions of moment-based kurtosis functionals.

Let us now formulate technically the setting and objectives of this paper. For a given multivariate distribution, Einmahl and Mason [2] define corresponding *generalized quantile functions*, curves designed to summarize in convenient two-dimensional plots certain features of the given multivariate distribution. Specifically, given a probability distribution F on $\mathbb{R}^d$, a class $\boldsymbol{\mathcal{A}}$ of Borel sets in $\mathbb{R}^d$, and a real-valued set function $\lambda(A)$ defined over $A \in \boldsymbol{\mathcal{A}}$, they define an associated "generalized quantile function" by

$$U_F(p) = \inf\{\lambda(A) : F(A) \geq p,\ A \in \boldsymbol{\mathcal{A}}\},\ 0 < p < 1. \tag{1}$$

For $d = 1$, $\boldsymbol{\mathcal{A}}$ the class of halflines, and $\lambda((-\infty, x]) = x$, we obtain the usual univariate quantile function. As shown in [15], the above-discussed depth-based sample "feature curves" of [10] may be conveniently represented for theoretical purposes as sample versions of depth-based generalized quantile functions as given by (1). Consequently, our goal may be expressed as finding the influence function of a generalized quantile function that is defined in terms of a depth function.

Assume given a *depth function* $D(\boldsymbol{x}, F)$ providing an F-based ordering of the points $\boldsymbol{x}$ in $\mathbb{R}^d$ according to their "centrality" in the distribution F. Denoting the central regions or level sets of $D(\cdot, F)$ by $I(\alpha, D, F) = \{\boldsymbol{x} : D(\boldsymbol{x}, F) \geq \alpha\}$, $\alpha > 0$, we define $C_{F,D}(p)$ to be the central region having probability weight p and introduce

CONDITION A.

(i) $F(\cdot)$ and $D(\cdot, F)$ are continuous functions, and

(ii) $\lambda(I(\alpha, D, F))$ is a decreasing function of α.

It follows easily [15] that $C_{F,D}(p) = I(F^{-1}_{D(\boldsymbol{X},\,F)}(1-p), D, F)$, and that, with $\boldsymbol{\mathcal{A}}$ given by $\{C_{F,D}(p),\ 0 < p < 1\}$,

$$U_F(p) = \lambda(C_{F,D}(p)) = \lambda(I(F^{-1}_{D(\boldsymbol{X},\,F)}(1-p), D, F)),\ 0 < p < 1. \tag{2}$$

For fixed F, the curve given by (2) for $0 < p < 1$ is a convenient two-dimensional plot and can describe some key feature of the multivariate F. Particular choices of $\lambda(\cdot)$ yield, for example, a *scale* curve for dispersion [10] [15] [22] and a *Lorenz curve* for tailweight [10] [15]. Further, a transform of the scale curve yields a *kurtosis curve* [23].

To explore the *robustness* of sample versions of such depth-based generalized quantile functions, we consider for each fixed p the functional of F given by

$$T_p(F) = \lambda(C_{F,D}(p))$$

and apply the *influence function* approach [6], which characterizes the limiting effect on the functional when F undergoes a small perturbation of the form

$$F_{\boldsymbol{y},\,\varepsilon} = (1-\varepsilon)F + \varepsilon\delta_{\boldsymbol{y}},$$

where $\delta_{\boldsymbol{y}}$ denotes the cdf placing mass 1 at $\boldsymbol{y}$, $\boldsymbol{y} \in \mathbb{R}^d$. In general, the influence function (IF) of a functional T on distributions F on $\mathbb{R}^d$ is defined at each $\boldsymbol{y} \in \mathbb{R}^d$ and choice of F as

$$IF(\boldsymbol{y}, T, F) = \lim_{\varepsilon \downarrow 0} \frac{T(F_{\boldsymbol{y},\,\varepsilon}) - T(F)}{\varepsilon}.$$

For F_n the sample distribution function based on a sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from F, we have under mild regularity conditions the approximation

$$T(F_n) - T(F) \doteq \frac{1}{n} \sum_{i=1}^{n} IF(\boldsymbol{X}_i, T, F), \tag{3}$$

which indicates the contribution of each observation (especially an outlier) to the estimation error and yields as well the asymptotic variance of $T(F_n)$: $\int IF(\boldsymbol{x}, T, F)^2\, dF(\boldsymbol{x})$. See [7], [14] and [6] for details.

There are a variety of possible choices for $D(\cdot, \cdot)$, of course. In the present development we confine attention to the case of *Type D depth functions*, defined in [25] as those of form

$$D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) = \inf\{F(C) : \boldsymbol{x} \in C \in \boldsymbol{\mathcal{C}}\}, \tag{4}$$

where $\boldsymbol{\mathcal{C}}$ is a specified class of closed subsets of $\mathbb{R}^d$. (See the Appendix for further discussion.) Further, we assume that D and $\boldsymbol{\mathcal{C}}$ satisfy

Condition B.

(i) If $C \in \boldsymbol{\mathcal{C}}$, then $\overline{C^c} \in \boldsymbol{\mathcal{C}}$, and

(ii) $\max_{\boldsymbol{x}} D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) < 1$,

where A^c denotes the complement, and $\overline{A}$ the closure, of a set A.

Specifically, then, our primary goal in the present paper is to obtain the IF of a generalized quantile functional $T_p(F) = \lambda(C_{F,D(\cdot,\,F,\boldsymbol{\mathcal{C}})}(p))$ based on a Type D depth function with F, D, and $\boldsymbol{\mathcal{C}}$ satisfying Conditions A and B. As an intermediate step, we also obtain the IF of the Type D depth functions themselves, a result significant in its own right, because depth functions are used for outlier identification and any reasonable outlier identifier should *itself* be robust against outliers.

Specifically, in Section 2 we derive the IF's for

(a) the *depth function* $D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})$ for any fixed $\boldsymbol{x}$,

(b) the *depth distribution function* $F_{D(\boldsymbol{X},\,F,\boldsymbol{\mathcal{C}})}(z)$ for any fixed real z, and

(c) the *depth quantile function* $F^{-1}_{D(\boldsymbol{X},\,F,\boldsymbol{\mathcal{C}})}(p)$ for any fixed $p \in (0, 1)$.

The IF for (b) is used in getting that for (c), which, as a matter of interest, can also be represented as a generalized quantile function (see [15]). In turn, the IF for (c) is used to obtain our target IF, that for $T_p(F)$. This result is derived in Section 3.

Let us also summarize our results qualitatively. Robustness behavior similar to the usual univariate quantiles is found and quantified: *the influence functions for (a), (b), and (c) are step functions in form, with finite gross error sensitivity* [6], [14] *but infinite local shift sensitivity*. This yields similar behavior for the IF of $T_p(F)$.

We conclude this introduction with illustrative preliminary discussion of two important generalized quantile curves, i.e., functionals $T_p(F)$ designed to measure *scale* and *kurtosis*, respectively. These will be revisited with further details in Sections 3.1 and 3.3, along with a *tailweight* example in Section 3.2.

Example *Scale and kurtosis curves, for elliptically symmetric F and halfspace depth.* Let F have a density of form

$$f(\boldsymbol{x}) = |\boldsymbol{\Sigma}|^{-1/2} h((\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})),\ \boldsymbol{x} \in \mathbb{R}^d,$$

for a continuous and positive scalar function $h(\cdot)$ and a positive definite matrix $\boldsymbol{\Sigma}$. Let D be the halfspace depth, a well-known special case of Type D depth functions (see Appendix). Then the central regions turn out to be nested ellipsoids of form

$$\{\boldsymbol{x} \in \mathbb{R}^d : (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \le c\}.$$

Scale. With $\lambda(A)$ the *volume* of A, the above general functional $T_p(F)$ becomes simply the volume of the central region having probability weight p, $0 < p < 1$. This defines a particular generalized quantile function called the *scale curve* [10], for it quantifies the expansion of the (nested) central regions with increasing probability weight p. Detailed treatment is found in [8], [10], and [22]. Corollary 3.1 below augments these previous studies by presenting the IF of such a scale curve (pointwise) in the case of Type D depth. In particular, for the example at hand, with F_R the cdf of the squared Mahalanobis distance of $\boldsymbol{X}$ from $\boldsymbol{\mu}$,

$$R = (\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}),$$

we have

$$\mathrm{IF}(\mathrm{scale}, \boldsymbol{y}) = \frac{p - \mathbf{1}\{(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) \le F_R^{-1}(p)\}}{|\boldsymbol{\Sigma}|^{-1/2} h(F_R^{-1}(p))},$$

which is positive or negative according as $\boldsymbol{y}$ is outside or inside the ellipsoid

$$\{\boldsymbol{x} \in \mathbb{R}^d : (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \le F_R^{-1}(p)\}.$$

Specializing to F *bivariate standard normal*, we have $f(\boldsymbol{x}) = (1/2\pi)\exp\{-\boldsymbol{x}'\boldsymbol{x}/2\}$ and $h(r) = (1/2\pi)\exp\{-r/2\}$, and $R = \boldsymbol{X}'\boldsymbol{X}$ has the chi-squared distribution with 2 degrees of freedom, i.e., is exponential with scale parameter 2. Thus $F_R(r) = 1 - \exp\{-r/2\}$, $F_R^{-1}(p) = -2\ln(1-p)$, and $h(F_R^{-1}(p)) = (1-p)/2\pi$, and the above IF reduces to

$$\mathrm{IF}(\mathrm{scale}, \boldsymbol{y}) = \begin{cases} -2\pi, & \boldsymbol{y}'\boldsymbol{y} \le -2\ln(1-p) \\ 2\pi p/(1-p), & \boldsymbol{y}'\boldsymbol{y} \ge -2\ln(1-p). \end{cases}$$

This shows that, as might be expected for estimation of scale, serious "inliers" cause underestimation and serious "outliers" cause overestimation. The features of this IF are illustrated in Figure 1 and will be seen in Corollary 3.1 to hold under more general conditions.

Kurtosis. For nonparametric description of a distribution, the natural next step after treating location, spread, symmetry and skewness is to characterize *kurtosis*. The classical univariate (moment-based) notion of kurtosis, the standardized fourth central moment, has been construed as simply a discriminator between heavy peakedness and heavy tails, but is more properly understood as a measure concerning the structure of the distribution in the region between, and linking, the center and the tails. The boundary between the center and tails represents the so-called "shoulders" of the distribution. In these picturesque terms, classical univariate kurtosis measures *dispersion of probability mass away from the shoulders*, toward either the center or the tails or both. Thus peakedness, kurtosis and tailweight are distinct, although interrelated, features of a distribution. See [23] for detailed discussion.

For a distribution in $\mathbb{R}^d$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the classical univariate kurtosis is generalized by Mardia [11] to the fourth moment of the Mahalanobis distance of $\boldsymbol{X}$ from $\boldsymbol{\mu}$. This may be seen to measure the dispersion of $\boldsymbol{X}$ about the points on the ellipsoid $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = d$, interpreting this surface as the "shoulders" of the distribution. Higher kurtosis arises when probability mass is diminished near the shoulders and greater either near $\boldsymbol{\mu}$ (greater peakedness) or in the tails (greater tailweight) or both. Such a measure does not, however, indicate the *shape* of the distribution in this region. See [23] for further discussion.

Alternative kurtosis measures have been introduced which are *quantile-based*. These complement the moment-based types by characterizing the shape of the distribution within affine equivalence. The univariate case was treated by Groeneveld and Meeden [5] and a depth-based extension to the multivariate case has recently been introduced by Wang and Serfling [23], where detailed treatment is found. Such measures may be represented as transforms of the volume function or equivalently of the above-discussed scale curve:

$$k_{F,D}(p) = \frac{V_{F,D}(\frac{1}{2}-\frac{p}{2}) + V_{F,D}(\frac{1}{2}+\frac{p}{2}) - 2\,V_{F,D}(\frac{1}{2})}{V_{F,D}(\frac{1}{2}+\frac{p}{2}) - V_{F,D}(\frac{1}{2}-\frac{p}{2})}.$$

For this notion of kurtosis, the "shoulders" of F are given by the contour of the central region of probability 1/2, i.e., the "interquartile region" $C_{F,D}(\frac{1}{2})$. The quantity $k_{F,D}(p)$ measures the relative volumetric difference between equiprobable regions just without and just within the shoulders, in the tail and central parts, respectively. The IF of $k_{F,D}(p)$ under general conditions is given in Section 3.3. Again specializing to the case of elliptically symmetric F

and halfspace depth, we obtain

$$
\begin{aligned}
&\mathrm{IF}(\boldsymbol{y}, k_{\cdot,D}(p), F)\\
&= \frac{2\Gamma(d/2+1)}{\pi^{d/2}[(F_R^{-1}(\frac{1}{2}+\frac{p}{2}))^{d/2}-(F_R^{-1}(\frac{1}{2}-\frac{p}{2}))^{d/2}]^2} \times \\
&\quad \Big\{[(F_R^{-1}(\tfrac{1}{2}))^{d/2}-(F_R^{-1}(\tfrac{1}{2}-\tfrac{p}{2}))^{d/2}]\cdot \frac{(\frac{1}{2}+\frac{p}{2})-\mathbf{1}\{(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\le F_R^{-1}(\frac{1}{2}+\frac{p}{2})\}}{h(F_R^{-1}(\frac{1}{2}+\frac{p}{2}))}\\
&\quad +[(F_R^{-1}(\tfrac{1}{2}+\tfrac{p}{2}))^{d/2}-(F_R^{-1}(\tfrac{1}{2}))^{d/2}]\cdot \frac{(\frac{1}{2}-\frac{p}{2})-\mathbf{1}\{(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\le F_R^{-1}(\frac{1}{2}-\frac{p}{2})\}}{h(F_R^{-1}(\frac{1}{2}-\frac{p}{2}))}\\
&\quad -[(F_R^{-1}(\tfrac{1}{2}+\tfrac{p}{2}))^{d/2}-(F_R^{-1}(\tfrac{1}{2}-\tfrac{p}{2}))^{d/2}]\cdot \frac{\frac{1}{2}-\mathbf{1}\{(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\le F_R^{-1}(\frac{1}{2})\}}{h(F_R^{-1}(\frac{1}{2}))}\Big\}.
\end{aligned}
$$

We see that the IF of $k_{F,D}(p)$ can be either positive or negative for contamination at $\boldsymbol{y}$ within the $(\frac{1}{2}-\frac{p}{2})$th central region, jumps by a positive amount as $\boldsymbol{y}$ moves out of this region, jumps again by a positive amount as $\boldsymbol{y}$ moves out of the $\frac{1}{2}$th central region, and finally again by a positive amount as $\boldsymbol{y}$ moves out of the $(\frac{1}{2}+\frac{p}{2})$th central region. Outside this region, under some conditions on F, the IF is positive. (See Groeneveld [4] for discussion in the univariate case.) In any case, the IF is bounded and thus has finite gross error sensitivity, in contrast with the unbounded IF's of moment-based kurtosis measures.

For the case of F bivariate standard normal, the above IF becomes

$$
\begin{aligned}
&\mathrm{IF}(\text{kurtosis}, \boldsymbol{y})\\
&= \begin{cases}
0, & 0 \le \boldsymbol{y}'\boldsymbol{y} \le -2\ln((1+p)/2)\\
-4w(p)[\ln(1-p)]/(1+p), & -2\ln((1+p)/2) \le \boldsymbol{y}'\boldsymbol{y} \le -2\ln(1/2)\\
-4w(p)[\ln(1+p)-p(1+p)^{-1}\ln(1-p), & -2\ln(1/2) \le \boldsymbol{y}'\boldsymbol{y} \le -2\ln((1-p)/2)\\
4pw(p)[(1-p)^{-1}\ln(1+p)+(1+p)^{-1}\ln(1-p), & \boldsymbol{y}'\boldsymbol{y} \ge -2\ln((1-p)/2),
\end{cases}
\end{aligned}
$$

where $w(p) = [\ln((1+p)/(1-p))]^2$. See Figure 2 for illustration. □

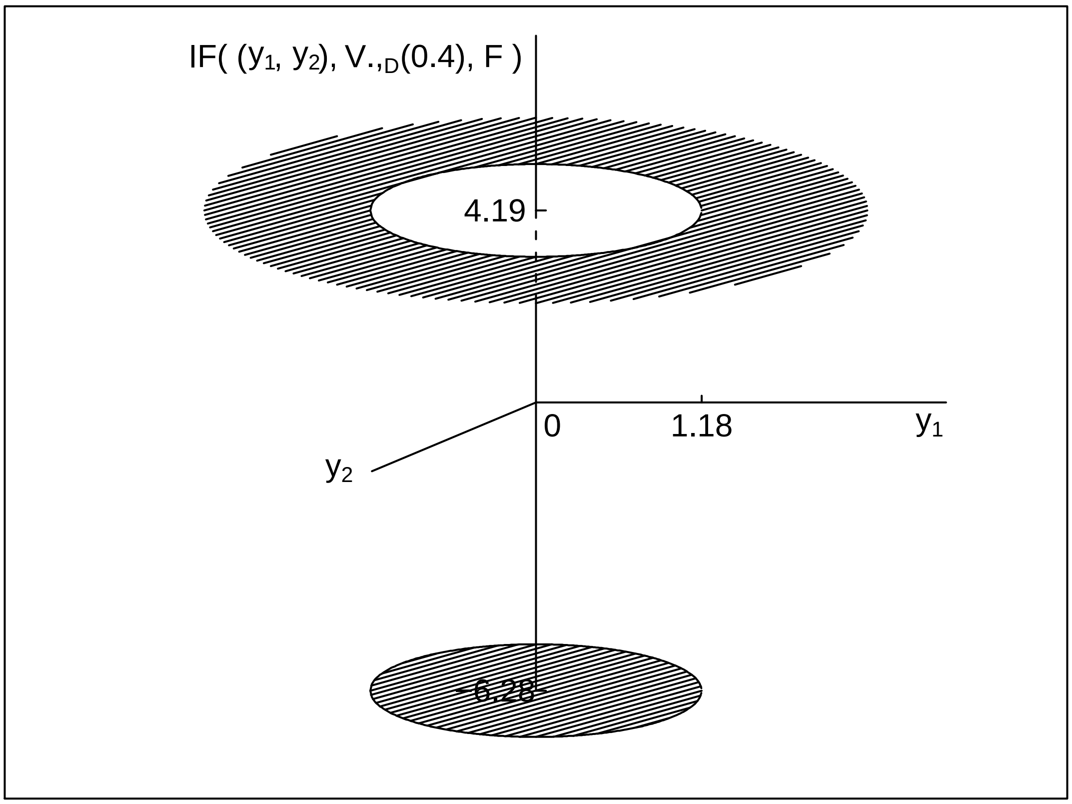

Figure 1: The IF of the scale curve for F bivariate standard normal.

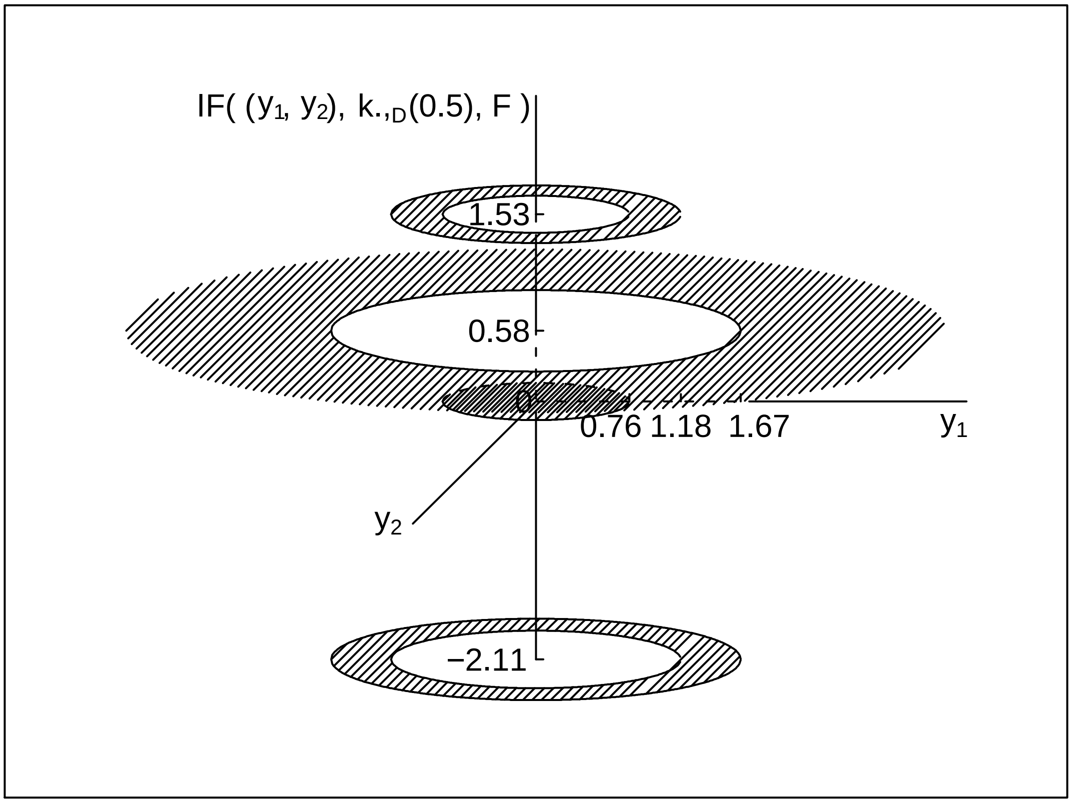

Figure 2: The IF of the kurtosis curve for F bivariate standard normal.

2 The IF's of Type D depth, distribution, and quantile functions

Consider now a Type D depth function as given by (4) and satisfying Condition B. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, defining

$$\begin{aligned} D^{(\boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) &= \inf\{F(C) : \boldsymbol{x} \in C \in \boldsymbol{\mathcal{C}},\ \boldsymbol{y} \in C\}, \\ D^{(\sim \boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) &= \inf\{F(C) : \boldsymbol{x} \in C \in \boldsymbol{\mathcal{C}},\ \boldsymbol{y} \notin C\}, \end{aligned}$$

with the convention that $\inf \emptyset = \infty$, we have

$$D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) = \min\{D^{(\boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}),\ D^{(\sim \boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})\}.$$

That is, for a Type D depth function, the depth of any point $\boldsymbol{x}$ may be represented, for any choice of another point $\boldsymbol{y}$, as the minimum of the depths of the point $\boldsymbol{x}$ taken with respect to the subclasses of sets C which either contain or do not contain the point $\boldsymbol{y}$, respectively.

The IF's to be derived all will involve in their formulas the sets

$$S_{\boldsymbol{y}} = \{\boldsymbol{x} : D^{(\boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) < D^{(\sim \boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})\},\ \boldsymbol{y} \in \mathbb{R}^d.$$

Under typical smoothness conditions on F and $D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})$, there exists a unique deepest point as center, say $\boldsymbol{M}$, and for each $\boldsymbol{x} \neq \boldsymbol{M}$ a unique "optimal" set $C_{\boldsymbol{x}}$ for which $D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})$ $= F(C_{\boldsymbol{x}})$. In this case, $\boldsymbol{x} \in S_{\boldsymbol{y}}$ if and only if $\boldsymbol{y}$ belongs to the optimal set for $\boldsymbol{x}$. For $\boldsymbol{x} = \boldsymbol{M}$, however, there typically are multiple optimal sets, in which case, for every $\boldsymbol{y} \neq \boldsymbol{M}$, some contain $\boldsymbol{y}$ and some do not, and $\boldsymbol{M} \notin S_{\boldsymbol{y}}$. Further aspects of $S_{\boldsymbol{y}}$ are provided by the following result.

Lemma 2.1 *For $\boldsymbol{x} \in S_{\boldsymbol{y}}$, $D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) \geq D(\boldsymbol{y}, F, \boldsymbol{\mathcal{C}})$. Equivalently,*

$$S_{\boldsymbol{y}} \subset I(D(\boldsymbol{y}, F, \boldsymbol{\mathcal{C}}), D(\cdot, F, \boldsymbol{\mathcal{C}}), F).$$

Proof. For $\boldsymbol{x} \in S_{\boldsymbol{y}}$,

$$\begin{aligned} D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) &= \min\{D^{(\boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}),\ D^{(\sim \boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})\} \\ &= D^{(\boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) \\ &= \inf\{F(C) : \boldsymbol{x} \in C \in \boldsymbol{\mathcal{C}},\ \boldsymbol{y} \in C\} \\ &\geq \inf\{F(C) : \boldsymbol{y} \in C \in \boldsymbol{\mathcal{C}}\} \\ &= D(\boldsymbol{y}, F, \boldsymbol{\mathcal{C}}), \end{aligned}$$

establishing the stated inequality. The inclusion is equally straightforward. ■

The converse is not true. For example, we have $D(\boldsymbol{M}, F, \boldsymbol{\mathcal{C}}) \geq D(\boldsymbol{y}, F, \boldsymbol{\mathcal{C}})$, each $\boldsymbol{y}$, but, as noted above, $\boldsymbol{M} \notin S_{\boldsymbol{y}}$ for $\boldsymbol{y} \neq \boldsymbol{M}$.

2.1 The IF of $D(\boldsymbol{y}, F, \boldsymbol{\mathcal{C}})$

The IF for the halfspace depth has been treated in an excellent study by Romanazzi [13]. Theorem 2.1 below covers Type D depth functions in general and shows that the influence upon $D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})$ due to perturbation of F by contamination at $\boldsymbol{y}$ takes one of two values according as $\boldsymbol{x} \in S_{\boldsymbol{y}}$ or $\boldsymbol{x} \notin S_{\boldsymbol{y}}$, in the first case positively incrementing the depth at $\boldsymbol{x}$ and in the second case negatively incrementing it. In other words, contamination at $\boldsymbol{y}$ causes the centrality of $\boldsymbol{x}$ to increase or decrease, according to whether or not $S_{\boldsymbol{y}}$ contains $\boldsymbol{x}$. The boundedness of this IF keeps the gross error sensitivity [6], [14] of the functional $D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})$ finite, as desired. Due to the step function structure of the IF, however, the local shift sensitivity [6], [14] of $D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})$ is infinite, except in the case of a unique deepest point $\boldsymbol{M}$, for which, as seen above, $\boldsymbol{y} \neq \boldsymbol{M}$ implies $\boldsymbol{M} \notin S_{\boldsymbol{y}}$, so that the IF of $D(\boldsymbol{M}, F, \boldsymbol{\mathcal{C}})$ assumes the constant value $-D(\boldsymbol{M}, F, \boldsymbol{\mathcal{C}})$ for all $\boldsymbol{y} \in \mathbb{R}^d$, making the local shift sensitivity zero.

Theorem 2.1 *The influence function of $D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})$ is, for $\boldsymbol{y} \in \mathbb{R}^d$,*

$$IF(\boldsymbol{y}, D(\boldsymbol{x}, \cdot, \boldsymbol{\mathcal{C}}), F) = \mathbf{1}\{\boldsymbol{x} \in S_{\boldsymbol{y}}\} - D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}).$$

Proof. It is easily checked that

$$\begin{aligned} D(\boldsymbol{x}, F_{\boldsymbol{y},\varepsilon}, \boldsymbol{\mathcal{C}}) &= \min\{D^{(\boldsymbol{y})}(\boldsymbol{x}, F_{\boldsymbol{y},\varepsilon}, \boldsymbol{\mathcal{C}}),\ D^{(\sim \boldsymbol{y})}(\boldsymbol{x}, F_{\boldsymbol{y},\varepsilon}, \boldsymbol{\mathcal{C}})\} \\ &= \min\{(1-\epsilon)D^{(\boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) + \varepsilon,\ (1-\epsilon)D^{(\sim \boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})\}. \end{aligned}$$

Then, for $\boldsymbol{x} \notin S_{\boldsymbol{y}}$, we have $D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) = D^{(\sim \boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})$ and hence

$$D(\boldsymbol{x}, F_{\boldsymbol{y},\varepsilon}, \boldsymbol{\mathcal{C}}) = (1-\epsilon)D^{(\sim \boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) = (1-\epsilon)D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}),$$

yielding

$$IF(\boldsymbol{y}, D(\boldsymbol{x}, \cdot, \boldsymbol{\mathcal{C}}), F) = \lim_{\varepsilon \downarrow 0} \frac{D(\boldsymbol{x}, F_{\boldsymbol{y},\varepsilon}, \boldsymbol{\mathcal{C}}) - D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})}{\varepsilon} = -D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}).$$

On the other hand, for $\boldsymbol{x} \in S_{\boldsymbol{y}}$ we have $D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) = D^{(\boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})$ and, for ε sufficiently small,

$$(1-\epsilon)D^{(\sim \boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) > (1-\epsilon)D^{(\boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) + \varepsilon,$$

whence

$$D(\boldsymbol{x}, F_{\boldsymbol{y},\varepsilon}, \boldsymbol{\mathcal{C}}) = (1-\epsilon)D^{(\boldsymbol{y})}(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) + \varepsilon = (1-\epsilon)D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) + \varepsilon,$$

yielding

$$IF(\boldsymbol{y}, D(\boldsymbol{x}, \cdot, \boldsymbol{\mathcal{C}}), F) = 1 - D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})$$

and completing the proof. ■

2.2 The IF of $F_{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}})}(z)$

Denoting by $F_{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}})}$ the distribution function of the random depth $D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}})$, we find the IF of $F_{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}})}(z)$ for any fixed real z.

Theorem 2.2 *If $F_{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}})}$ is continuous with density $f_{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}})}$, then the influence function of $F_{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}})}(z)$ is, for $\boldsymbol{y} \in \mathbb{R}^d$,*

$$
\begin{aligned}
&IF(\boldsymbol{y}, F_{D(\boldsymbol{X},\cdot,\boldsymbol{\mathcal{C}})}(z), F) \\
&\quad = \begin{cases} f_{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}})}(z)\cdot z - F_{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}})}(z), & D(\boldsymbol{y},F,\boldsymbol{\mathcal{C}}) \geq z, \\ f_{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}})}(z)\cdot z - F_{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}})}(z) + 1 - f_{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}})|\boldsymbol{X}\in S_{\boldsymbol{y}}}(z)\cdot P(\boldsymbol{X}\in S_{\boldsymbol{y}}), & D(\boldsymbol{y},F,\boldsymbol{\mathcal{C}}) < z. \end{cases}
\end{aligned}
$$

Proof. For all $\varepsilon > 0$ sufficiently small, and with P and $P_{\boldsymbol{y},\varepsilon}$ the probability measures corresponding to F and $F_{\boldsymbol{y},\varepsilon}$ respectively, we have

$$
\begin{aligned}
F_{D(\boldsymbol{X},F_{\boldsymbol{y},\varepsilon},\boldsymbol{\mathcal{C}})}(z) &= P_{\boldsymbol{y},\varepsilon}\{D(\boldsymbol{X},F_{\boldsymbol{y},\varepsilon},\boldsymbol{\mathcal{C}}) \leq z\} \\
&= (1-\varepsilon)P\{D(\boldsymbol{X},F_{\boldsymbol{y},\varepsilon},\boldsymbol{\mathcal{C}}) \leq z\} + \varepsilon\,\mathbf{1}\{D(\boldsymbol{y},F_{\boldsymbol{y},\varepsilon},\boldsymbol{\mathcal{C}}) \leq z\} \\
&= (1-\varepsilon)\big[P\{D(\boldsymbol{X},F_{\boldsymbol{y},\varepsilon},\boldsymbol{\mathcal{C}}) \leq z \text{ and } \boldsymbol{X}\in S_{\boldsymbol{y}}\} \\
&\quad + P\{D(\boldsymbol{X},F_{\boldsymbol{y},\varepsilon},\boldsymbol{\mathcal{C}}) \leq z \text{ and } \boldsymbol{X}\notin S_{\boldsymbol{y}}\}\big] + \varepsilon\,\mathbf{1}\{D(\boldsymbol{y},F_{\boldsymbol{y},\varepsilon},\boldsymbol{\mathcal{C}}) \leq z\} \\
&= (1-\varepsilon)\big[P\{(1-\varepsilon)D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}}) + \varepsilon \leq z \text{ and } \boldsymbol{X}\in S_{\boldsymbol{y}}\} \\
&\quad + P\{(1-\varepsilon)D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}}) \leq z \text{ and } \boldsymbol{X}\notin S_{\boldsymbol{y}}\}\big] + \varepsilon\,\mathbf{1}\{D(\boldsymbol{y},F_{\boldsymbol{y},\varepsilon},\boldsymbol{\mathcal{C}}) \leq z\} \\
&= (1-\varepsilon)\left[P\left\{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}}) \leq \frac{z-\varepsilon}{1-\varepsilon} \text{ and } \boldsymbol{X}\in S_{\boldsymbol{y}}\right\}\right. \\
&\quad \left. + P\left\{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}}) \leq \frac{z}{1-\varepsilon} \text{ and } \boldsymbol{X}\notin S_{\boldsymbol{y}}\right\}\right] + \varepsilon\,\mathbf{1}\{D(\boldsymbol{y},F_{\boldsymbol{y},\varepsilon},\boldsymbol{\mathcal{C}}) \leq z\}.
\end{aligned}
$$

(i) Suppose that $D(\boldsymbol{y},F,\boldsymbol{\mathcal{C}}) > z$. Then for ε sufficiently small we have

$$
\left\{\boldsymbol{x} : D(\boldsymbol{x},F,\boldsymbol{\mathcal{C}}) \leq \frac{z-\varepsilon}{1-\varepsilon} \text{ and } \boldsymbol{x}\in S_{\boldsymbol{y}}\right\} = \emptyset,
$$

$$
\left\{\boldsymbol{x} : D(\boldsymbol{x},F,\boldsymbol{\mathcal{C}}) \leq \frac{z}{1-\varepsilon} \text{ and } \boldsymbol{x}\notin S_{\boldsymbol{y}}\right\} = \left\{\boldsymbol{x} : D(\boldsymbol{x},F,\boldsymbol{\mathcal{C}}) \leq \frac{z}{1-\varepsilon}\right\}, \text{ and}
$$

$$
\boldsymbol{y} \notin \{\boldsymbol{x} : D(\boldsymbol{x},F_{\boldsymbol{y},\varepsilon},\boldsymbol{\mathcal{C}}) \leq z\}.
$$

Thus

$$F_{D(\boldsymbol{X},\,F_{\boldsymbol{y},\,\varepsilon},\,\boldsymbol{\mathcal{C}})}(z) = (1-\varepsilon)\,P\left\{D(\boldsymbol{X},F,\boldsymbol{\mathcal{C}}) \le \frac{z}{1-\varepsilon}\right\} = (1-\varepsilon)\,F_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}\left(\frac{z}{1-\varepsilon}\right),$$

and we readily obtain

$$IF(\boldsymbol{y}, F_{D(\boldsymbol{X},\,\cdot,\,\boldsymbol{\mathcal{C}})}(z), F) = f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(z)\cdot z - F_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(z).$$

(ii) Next suppose that $D(\boldsymbol{y},F,\boldsymbol{\mathcal{C}}) = z$. Then, since $z \le \max_{\boldsymbol{x}} D(\boldsymbol{x},F,\boldsymbol{\mathcal{C}}) < 1$,

$$\frac{z-\varepsilon}{1-\varepsilon} < z$$

and it follows that

$$\left\{\boldsymbol{x} : D(\boldsymbol{x},F,\boldsymbol{\mathcal{C}}) \le \frac{z-\varepsilon}{1-\varepsilon} \text{ and } \boldsymbol{x}\in S_{\boldsymbol{y}}\right\} = \emptyset,$$

and, for ε sufficiently small, that

$$\boldsymbol{y} \notin \{\boldsymbol{x} : D(\boldsymbol{x}, F_{\boldsymbol{y},\,\varepsilon}, \boldsymbol{\mathcal{C}}) \le z\}.$$

Then, by steps similar to those in (i), we obtain

$$\begin{aligned}
F_{D(\boldsymbol{X},\,F_{\boldsymbol{y},\,\varepsilon},\,\boldsymbol{\mathcal{C}})}(z) &= (1-\varepsilon)F_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})|\boldsymbol{X}\,\notin\,S_{\boldsymbol{y}}}\left(\frac{z}{1-\varepsilon}\right)\cdot P(\boldsymbol{X}\notin S_{\boldsymbol{y}}),\\
F_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(z) &= F_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})|\boldsymbol{X}\,\notin\,S_{\boldsymbol{y}}}(z)\cdot P(\boldsymbol{X}\notin S_{\boldsymbol{y}}),\\
f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(z) &= f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})|\boldsymbol{X}\,\notin\,S_{\boldsymbol{y}}}(z)\cdot P(\boldsymbol{X}\notin S_{\boldsymbol{y}}).
\end{aligned}$$

which used together yield

$$\begin{aligned}
IF(\boldsymbol{y}, F_{D(\boldsymbol{X},\,\cdot,\,\boldsymbol{\mathcal{C}})}(z), F) &= [f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})|\boldsymbol{X}\,\notin\,S_{\boldsymbol{y}}}(z)\cdot z - F_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})|\boldsymbol{X}\,\notin\,S_{\boldsymbol{y}}}(z)]\cdot P(\boldsymbol{X}\notin S_{\boldsymbol{y}})\\
&= f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(z)\cdot z - F_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(z).
\end{aligned}$$

(iii) Finally, suppose that $D(\boldsymbol{y},F,\boldsymbol{\mathcal{C}}) < z$. Then $D(\boldsymbol{y},F_{\boldsymbol{y},\,\varepsilon},\boldsymbol{\mathcal{C}}) \le z$ for ε sufficiently small, and by steps similar to the preceding, we arrive at

$$\begin{aligned}
&IF(\boldsymbol{y}, F_{D(\boldsymbol{X},\,\cdot,\,\boldsymbol{\mathcal{C}})}(z), F)\\
&= f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(z)\cdot z - F_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(z) + 1 - f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})|\boldsymbol{X}\,\in\,S_{\boldsymbol{y}}}(z)\cdot P(\boldsymbol{X}\in S_{\boldsymbol{y}}).
\end{aligned}$$

Combining (i), (ii) and (iii), the proof is complete. ■

2.3 The IF of $F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p)$

It is now relatively straightforward to obtain the IF of the depth quantile function $F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p)$ for any fixed $p \in (0,1)$.

Theorem 2.3 *If $f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p)) > 0$, then the influence function of $F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p)$ is, for $\boldsymbol{y} \in \mathbb{R}^d$,*

$$IF(\boldsymbol{y}, F^{-1}_{D(\boldsymbol{X},\,\cdot,\,\boldsymbol{\mathcal{C}})}(p), F)$$

$$= \begin{cases} -F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p) + \frac{p}{f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p))}, & \\ \qquad\qquad \text{for } D(\boldsymbol{y}, F, \boldsymbol{\mathcal{C}}) \geq F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p), & \\ -F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p) + \frac{p-1+f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})|\boldsymbol{X}\,\in\,S_{\boldsymbol{y}}}\,(F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p))\cdot P(\boldsymbol{X} \in S_{\boldsymbol{y}})}{f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p))}, & \\ \qquad\qquad \text{for } D(\boldsymbol{y}, F, \boldsymbol{\mathcal{C}}) < F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p). & \end{cases}$$

Proof. Using a standard implicit function theorem approach with

$$F_{D(\boldsymbol{X},\,F_{\boldsymbol{y},\,\varepsilon},\,\boldsymbol{\mathcal{C}})}(F^{-1}_{D(\boldsymbol{X},\,F_{\boldsymbol{y},\,\varepsilon},\,\boldsymbol{\mathcal{C}})}(p)) = p,$$

we have

$$\frac{d}{d\varepsilon} F_{D(\boldsymbol{X},\,F_{\boldsymbol{y},\,\varepsilon},\,\boldsymbol{\mathcal{C}})}(F^{-1}_{D(\boldsymbol{X},\,F_{\boldsymbol{y},\,\varepsilon},\,\boldsymbol{\mathcal{C}})}(p))\Big|_{\varepsilon=0} = 0,$$

which yields in straightforward fashion, by carrying out the differentiation,

$$IF(\boldsymbol{y}, F^{-1}_{D(\boldsymbol{X},\,\cdot,\,\boldsymbol{\mathcal{C}})}(p), F) = -\frac{IF(\boldsymbol{y}, F_{D(\boldsymbol{X},\,\cdot,\,\boldsymbol{\mathcal{C}})}(z), F)\Big|_{z=F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p)}}{f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(p))}.$$

Now apply Theorem 2.2. ■

3 The IF of a Type D depth-based generalized quantile functional

We are now prepared to investigate the IF of the general functional

$$T_p(F) = \lambda(C_{F,D}(p)) = \lambda(I(F^{-1}_{D(\boldsymbol{X},\,F)}(1-p), D, F)),$$

assuming Conditions A and B and a further condition on λ:

Condition C.

$\lambda(\cdot)$ is finitely additive: for A and B disjoint, $\lambda(A \cup B) = \lambda(A) + \lambda(B)$.

In this case, for $\alpha > 0$ and $\boldsymbol{y} \in \mathbb{R}^d$, defining

$$\begin{aligned}\lambda^{(\boldsymbol{y})}(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F) &= \lambda(\{\boldsymbol{x} : D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) \geq \alpha \text{ and } \boldsymbol{x} \in S_{\boldsymbol{y}}\}) \\ \lambda^{(\sim \boldsymbol{y})}(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F) &= \lambda(\{\boldsymbol{x} : D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) \geq \alpha \text{ and } \boldsymbol{x} \notin S_{\boldsymbol{y}}\}),\end{aligned}$$

we have

$$\lambda(I(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F)) = \lambda^{(\boldsymbol{y})}(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F) + \lambda^{(\sim \boldsymbol{y})}(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F).$$

Theorem 3.1 *If Conditions A, B and C hold,* $\lambda^{(\boldsymbol{y})}(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F)$ *and* $\lambda^{(\sim \boldsymbol{y})}(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F)$ *are differentiable functions of* α, *and* $f_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}})}(F^{-1}_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}})}(1-p)) > 0$, *then the IF of* $T_p(F)$ *is, for* $\boldsymbol{y} \in \mathbb{R}^d$,

$$IF(\boldsymbol{y}, T_p(\cdot), F)$$

$$= \begin{cases} \frac{d}{d\alpha}\lambda(I(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F))\Big|_{\alpha = F^{-1}_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}})}(1-p)} \times \frac{1-p}{f_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}})}(F^{-1}_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}})}(1-p))}, \\ \qquad\qquad \textit{for } D(\boldsymbol{y}, F, \boldsymbol{\mathcal{C}}) \geq F^{-1}_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}})}(1-p), \\ \\ \frac{d}{d\alpha}\lambda(I(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F))\Big|_{\alpha = F^{-1}_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}})}(1-p)} \times \frac{-p + f_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}}) | \boldsymbol{X} \in S_{\boldsymbol{y}}}(F^{-1}_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}})}(1-p)) \cdot P(\boldsymbol{X} \in S_{\boldsymbol{y}})}{f_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}})}(F^{-1}_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}})}(1-p))} \\ \qquad\qquad - \frac{d}{d\alpha}\lambda^{(\boldsymbol{y})}(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F)\Big|_{\alpha = F^{-1}_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}})}(1-p)}, \\ \qquad\qquad \textit{for } D(\boldsymbol{y}, F, \boldsymbol{\mathcal{C}}) < F^{-1}_{D(\boldsymbol{X}, F, \boldsymbol{\mathcal{C}})}(1-p). \end{cases}$$

Proof. For ε sufficiently small, and by steps similar to those used in Section 2,

$$\begin{aligned} & T_p(F_{\boldsymbol{y}, \varepsilon}) \\ &= \lambda(C_{F_{\boldsymbol{y}, \varepsilon},\, D(\cdot, F_{\boldsymbol{y}, \varepsilon}, \boldsymbol{\mathcal{C}})}(p)) \\ &= \lambda(I(F^{-1}_{D(\boldsymbol{X}, F_{\boldsymbol{y}, \varepsilon}, \boldsymbol{\mathcal{C}})}(1-p), D(\cdot, F, \boldsymbol{\mathcal{C}}), F_{\boldsymbol{y}, \varepsilon})) \\ &= \lambda(\{\boldsymbol{x} : D(\boldsymbol{x}, F_{\boldsymbol{y}, \varepsilon}, \boldsymbol{\mathcal{C}}) \geq F^{-1}_{D(\boldsymbol{X}, F_{\boldsymbol{y}, \varepsilon}, \boldsymbol{\mathcal{C}})}(1-p)\} \\ &= \lambda^{(\boldsymbol{y})}\left(\frac{F^{-1}_{D(\boldsymbol{X}, F_{\boldsymbol{y}, \varepsilon}, \boldsymbol{\mathcal{C}})}(1-p) - \varepsilon}{1-\varepsilon}, D(\cdot, F, \boldsymbol{\mathcal{C}}), F\right) + \lambda^{(\sim \boldsymbol{y})}\left(\frac{F^{-1}_{D(\boldsymbol{X}, F_{\boldsymbol{y}, \varepsilon}, \boldsymbol{\mathcal{C}})}(1-p)}{1-\varepsilon}, D(\cdot, F, \boldsymbol{\mathcal{C}}), F\right). \end{aligned}$$

Thus

$$
\begin{aligned}
IF(\boldsymbol{y}, T_p(\cdot), F) &= IF(\boldsymbol{y}, \lambda(C_{\cdot,\ D(\cdot,\cdot,\boldsymbol{\mathcal{C}})}(p), F)) \\
&= \frac{d}{d\varepsilon}\lambda(C_{F_{\boldsymbol{y},\varepsilon},\ D(\cdot,\ F_{\boldsymbol{y},\varepsilon},\boldsymbol{\mathcal{C}})}\ (p))\Big|_{\varepsilon=0} \\
&= \frac{d}{d\alpha}\lambda(I(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F))\Big|_{\alpha=F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(1-p)} \times \left[IF(\boldsymbol{y}, F^{-1}_{D(\boldsymbol{X},\,\cdot,\,\boldsymbol{\mathcal{C}})}(1-p), F) + F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(1-p)\right] \\
&\qquad - \frac{d}{d\alpha}\lambda^{(\boldsymbol{y})}(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F)\Big|_{\alpha=F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(1-p)}.
\end{aligned}
$$

In the case $\alpha \le D(\boldsymbol{y}, F, \boldsymbol{\mathcal{C}})$, we have $\{\boldsymbol{x} : D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) \ge \alpha \text{ and } \boldsymbol{x} \in S_{\boldsymbol{y}}\} = S_{\boldsymbol{y}}$, so that

$$
\lambda^{(\boldsymbol{y})}(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F) = \lambda(\{\boldsymbol{x} : D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}}) \ge \alpha \text{ and } \boldsymbol{x} \in S_{\boldsymbol{y}}\}) = \lambda(S_{\boldsymbol{y}})
$$

is a constant function of α, therefore satisfying

$$
\frac{d}{d\alpha}\lambda^{(\boldsymbol{y})}(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F) = 0.
$$

It is now straightforward to complete the proof using Theorem 2.3. ■

The preceding result shows that the influence upon $T_p(F)$ due to perturbation of F by contamination at $\boldsymbol{y}$ takes one of two values according as $\boldsymbol{y}$ lies within or without the pth central region. Thus this functional has finite gross error sensitivity but infinite local shift sensitivity. These features reflect familiar aspects of the IF of the usual univariate quantile function $F^{-1}(p)$, such as that the influence of a contaminating observation depends only on whether it lies above or below a certain threshold but not on its particular value, that is, equivalently, only on whether its depth lies above or below a certain threshold but not on its particular depth value.

Application of Theorem 3.1 in several contexts is illustrated in the following subsections.

3.1 The IF of the scale curve based on Type D depth

In the Example of Section 1, we introduced the "scale curve" and illustrated its IF in a special case. Here we obtain the general result as an application of Theorem 3.1. For convenience, let us introduce some reasonable conditions under which the IF assumes a simpler form.

Condition D.

(i) $D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})$ is strictly decreasing in $\boldsymbol{x}$ in the support of F along any ray starting from the point of maximal depth,

(ii) $D(\boldsymbol{x}, F, \boldsymbol{\mathcal{C}})$ is differentiable in $\boldsymbol{x}$, and

(iii) For any $\alpha \geq 0$, the density $f(\boldsymbol{x})$ is constant over the α depth contour,

$$\partial I(\alpha, D(\cdot, F, \boldsymbol{\mathcal{C}}), F) = D^{-1}(\alpha, F, \boldsymbol{\mathcal{C}}).$$

We then have

Corollary 3.1 *Assume the conditions of Theorem 3.1 along with Conditions D. Then, for any p such that $f(D^{-1}(F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(1-p), F, \boldsymbol{\mathcal{C}})) > 0$, we have, for $\boldsymbol{y} \in \mathbb{R}^d$,*

$$IF(\boldsymbol{y}, V_{\cdot,D}(p), F) = \frac{p - \mathbf{1}\{D(\boldsymbol{y}, F, \boldsymbol{\mathcal{C}}) \geq F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(1-p)\}}{f(D^{-1}(F^{-1}_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(1-p), F, \boldsymbol{\mathcal{C}}))}. \tag{5}$$

Proof. For $\lambda(\cdot)$ the volume function, using transformation to polar coordinates it is straightforward to show (or see [22] for a similar result and proof) that

$$f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})|\boldsymbol{X}\,\in\,S_{\boldsymbol{y}}}(z) \cdot P(\boldsymbol{X} \in S_{\boldsymbol{y}}) = -f(D^{-1}(z, F, \boldsymbol{\mathcal{C}})) \times \frac{d}{dz}\lambda^{(\boldsymbol{y})}(z, D(\cdot, F, \boldsymbol{\mathcal{C}}), F)$$

and

$$f_{D(\boldsymbol{X},\,F,\,\boldsymbol{\mathcal{C}})}(z) = -f(D^{-1}(z, F, \boldsymbol{\mathcal{C}})) \times \frac{d}{dz}\lambda(I(z, D(\cdot, F, \boldsymbol{\mathcal{C}}), F)).$$

The result now follows immediately. ∎

We see that the IF of $V_{F,D}(p)$ is a two-valued step function with a jump on the boundary of the pth central region, from a negative value inside to a positive value outside, reflecting that serious "inliers" cause underestimation and serious "outliers" overestimation. Corollary 3.1 quantifies this effect. We note that the IF conforms to the role of the depth function in defining contours that demark degrees of "centrality" or "outlyingness".

3.2 The IF of a Lorenz curve based on Type D depth

Following Gastwirth [3], the *Lorenz curve* of a positive univariate random variable Y having cdf G is defined as

$$L_G(p) = \frac{1}{E(Y)} \int_0^p G^{-1}(t)\, dt = \frac{1}{E(Y)} \int_0^{G^{-1}(p)} y\, dG(y),\ \ 0 \leq p \leq 1,$$

which we note is scale-free due to normalization by $E(Y)$. As a method to characterize *tailweight* for multivariate distributions, Liu, Parelius and Singh [10] introduce a depth-based approach utilizing the Lorenz curve of the univariate depth distribution F_D. As discussed in [15], the depth-based Lorenz curve $L_{F_D}(p)$ may be represented as a generalized quantile function. That is, with $\boldsymbol{\mathcal{A}}$ given by the *outer* regions of $D(\cdot, F)$ and

$$\lambda(A) = \frac{1}{ED(X,F)} \int_A D(x,F)\,dF, \;\; A \in \boldsymbol{\mathcal{A}},$$

we obtain for $0 < p < 1$

$$T_p(F) = \lambda(C_{F,D}(p)^c) = \frac{1}{ED(X,F)} \int_0^{F_D^{-1}(p)} y\,dF_D(y) = \frac{1}{ED(X,F)} \int_0^p F_D^{-1}(p)\,dp = L_{F_D}(p).$$

For the case of Type D depth, the IF of this functional can be obtained from Theorem 3.1. We omit the details.

3.3 The IF of a kurtosis curve based on Type D depth

For the case of Type D depth, the IF of this kurtosis measure is derived via "influence function calculus" in conjunction with the IF of the scale curve, obtaining

$$\begin{aligned}
&\mathrm{IF}(\boldsymbol{y}, k_{\cdot,D}(p), F) \\
&= \frac{2}{[V_{F,D}(\frac{1}{2}+\frac{p}{2}) - V_{F,D}(\frac{1}{2}-\frac{p}{2})]^2} \times \Big\{ [V_{F,D}(\tfrac{1}{2}) - V_{F,D}(\tfrac{1}{2}-\tfrac{p}{2})] \cdot \mathrm{IF}(\boldsymbol{y}, V_{\cdot,D}(\tfrac{1}{2}+\tfrac{p}{2}), F) \\
&\qquad + [V_{F,D}(\tfrac{1}{2}+\tfrac{p}{2}) - V_{F,D}(\tfrac{1}{2})] \cdot \mathrm{IF}(\boldsymbol{y}, V_{\cdot,D}(\tfrac{1}{2}-\tfrac{p}{2}), F) \\
&\qquad - [V_{F,D}(\tfrac{1}{2}+\tfrac{p}{2}) - V_{F,D}(\tfrac{1}{2}-\tfrac{p}{2})] \cdot \mathrm{IF}(\boldsymbol{y}, V_{\cdot,D}(\tfrac{1}{2}), F) \Big\}.
\end{aligned}$$

For Type D depth, we apply results for the scale curve IF to see that the IF of $k_{F,D}(p)$ is a step function with jumps at three contours, one defining the interquartile region or "shoulders", and the other two demarking the annuli of equal probability $p/2$ within and without the shoulders.

4 Acknowledgments

Constructive suggestions by an Associate Editor and a referee are very much appreciated and have greatly improved the paper. Support by National Science Foundation Grant DMS-0103698 is also gratefully acknowledged.

References

[1] Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estinates based on halfspace depth and projected outlyingness. *Annals of Statistics* **20** 1803–1827.

[2] Einmahl, J. H. J. and Mason, D. M. Mason (1992). Generalized quantile processes. *Annals of Statistics* **20** 1062–1078.

[3] Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica* **39** 1037–1039.

[4] Groeneveld, R. A. (1998). A class of quantile measures for kurtosis. *The American Statistician* **51** 325–329.

[5] Groeneveld, R. A. and Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician* **33** 391–399.

[6] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

[7] Huber, P. J. (1981). *Robust Statistics.* Wiley, New York.

[8] Liu, R. Y. (1990). On a notion of data depth based on random simplices. *Annals of Statistics* **18** 405–414.

[9] Liu, R. Y. (1992). Data depth and multivariate rank tests. In L_1-*Statistics and Related Methods* (Y. Dodge, ed.), pp. 279–294, North-Holland, Amsterdam.

[10] Liu, R. Y., Parelius, J. M. and Singh, K. Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Annals of Statistics* **27** 783–858.

[11] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57** 519–530.

[12] Mosler, K. (2002). *Multivariate Dispersion, Central Regions and Depth.* Springer, New York.

[13] Romanazzi, M. (2001). Influence function of halfspace depth. *Journal of Multivariate Analysis* **77** 138–161 (with correction note, *ibid.* **81** (2002) 187–188).

[14] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

[15] Serfling, R. (2002a) Generalized quantile processes based on multivariate depth functions, with applications in nonparametric multivariate analysis. *Journal of Multivariate Analysis* **83** (2002), 232–247.

[16] Serfling, R. (2002b). A depth function and a scale curve based on spatial quantiles. In *Statistical Data Analysis Based On the L_1-Norm and Related Methods* (Y. Dodge, ed.), pp. 25-38. Birkhaüser.

[17] Serfling, R. (2004). Nonparametric multivariate descriptive measures based on spatial quantiles. *Journal of Statistical Planning and Inference* **123** 259–278

[18] Serfling, R. (2005). Depth and quantile functions in nonparametric multivariate analysis. In *Proceedings of May 2003 DIMACS Workshop on Data Depth: Robust Multivariate Analysis, Computational Geometry, and Applications* (R. Liu, R. Serfling, D. Souvaine, Y. Vardi, eds.), American Mathematical Society DIMACS Book Series, 2005, to appear.

[19] Small, C. G. (1987). Measures of centrality for multivariate and directional distributions. *Canadian Journal of Statistics* **15** 31–39.

[20] Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver 1974* (R. D. James, ed.), **2** 523–531.

[21] Vardi, Y. and Zhang, C.-H. (2000). The multivariate L_1-median and associated data depth. *Proceedings of National Academy of Science USA* **97** 1423–1426.

[22] Wang, J. and Serfling, R. (2006). On scale curves for nonparametric description of dispersion. In *Proceedings of May 2003 DIMACS Workshop on Data Depth: Robust Multivariate Analysis, Computational Geometry, and Applications* (R. Liu, R. Serfling, D. Souvaine, Y. Vardi, eds.), American Mathematical Society DIMACS Book Series, to appear.

[23] Wang, J. and Serfling, R. (2005). Nonparametric multivariate kurtosis and tailweight measures. *Journal of Nonparametric Statistics*, in press.

[24] Zuo, Y. (2003). Projection-based depth functions and associated medians. *Annals of Statistics* **31** 1460–1490.

[25] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Annals of Statistics* **28** 461–482.

[26] Zuo, Y. and Serfling, R. (2000). Structural properties and convergence results for contours of sample statistical depth functions. *Annals of Statistics* **28** 483–499.

5 Appendix: Brief background on depth functions

In passing from the real line to higher dimensional Euclidean space $\mathbb{R}^d$, an effective way to compensate for the lack of a natural order for $d > 1$ is to orient to a center. *Depth functions* provide center-outward orderings in $\mathbb{R}^d$, with depth decreasing with increasing distance from the center. This supports the intended interpretation of the depth of a point $\boldsymbol{x}$ as measuring its "centrality" (or, inversely, "outlyingness"). Since outlyingness is a *globally* oriented feature, depth functions differ in role from probability density functions, which describe *local* structure at points $\boldsymbol{x}$. That is, the contours of equal probability density have interpretability merely in a *local* sense: they characterize the amount of probability mass in a neighborhood of a point. Using depth functions, on the other hand, we may organize points into contours of equal outlyingness. This is basic to a proper generalization of univariate quantile-based inference to the multivariate case.

Although there are earlier antecedents, the first depth function was the *halfspace depth* introduced by Tukey (1975) and popularized by Donoho and Gasko (1992). This is given by Type D depth functions defined in Section 1 with $\boldsymbol{\mathcal{C}}$ the class of halfspaces in $\mathbb{R}^d$. This generates a corresponding affine equivariant notion of median as center. (In the univariate case with F continuous, the halfspace depth of x is simply $\min\{F(x), 1 - F(x)\}$.) For $\boldsymbol{\mathcal{C}}$ the class of sets of the form $\{\boldsymbol{x} \in \mathbb{R}^d : (-1)^\nu \pi_i(\boldsymbol{x}) > a\}$, where $i = 1, \ldots, d$, $\nu = 0$ or 1, and π_i projects $\boldsymbol{x}$ to its ith coordinate, we obtain a depth function which generates as center the vector of coordinatewise medians. For a general treatment of Type D depth functions in an equivalent form as "index functions", along with further examples, see Small [19]. For $\boldsymbol{\mathcal{C}}$ a Vapnik-Červonenkis class, the almost sure uniform convergence of sample Type D depth functions to their population counterparts is established in [26, Thm. B2].

Important other examples of depth function include the *simplicial depth* of Liu (1990), the *spatial depth* of Vardi and Zhang (2000), and the *projection depth* introduced by Liu (1992) and popularized by Zuo (2003). General overview papers have been cited in Section 1.