Nonparametric Notions of Multivariate "Scatter Measure" and "More Scattered" Based on Statistical Depth Functions

Yijun Zuo[1]
Arizona State University

and

Robert Serfling[2]
University of Texas at Dallas

November, 1998 (revised June, 1999)

[1]Department of Mathematics, Arizona State University, Tempe, Arizona 85287-1804, USA.
[2]Programs in Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 78083-0688, USA. Supported by NSF Grant DMS-9705209.

Abstract

Nonparametric notions of multivariate "scatter measure" and "more scattered", based on statistical depth functions, are investigated. In particular, notions of "more scattered" based on the "halfspace" depth function are shown to generalize versions introduced by Bickel and Lehmann (1976, 1979) in the univariate case and by Eaton (1982) and Oja (1983) in the multivariate case. Scatter measures are also discussed, with emphasis on those based on the halfspace depth. Basic desirable properties established for the previous versions of "more scattered" are shown to carry over to the depth-based notions as well, in both the univariate and multivariate cases. Further, some properties unique to the depth-based notions are established.

AMS 1991 Subject Classification: Primary 62H05; Secondary 62G05.

Key words and phrases: Scatter measure; more scattered; statistical depth functions; multivariate; nonparametric; statistical inference; halfspace depth.

1 Introduction

Statistical depth functions (defined below) have become increasingly used to develop center-outward orderings of the points of a multivariate distribution as well as of sample values. Such orderings provide a foundation for new nonparametric methods in multivariate data analysis and inference. Roughly speaking, a "location measure" or "center" is given by a point of maximal depth, one distribution is defined to be "more scattered" than another if its depth-trimmed regions have less volume than corresponding ones of the other distribution, and a "scatter measure" is a numerical functional that partially orders distributions in accord with a given "more scattered" relation. It is attractive, of course, that nonparametric location and scatter measures adopted in practice be coherently related to each other through some common defining principle. Depth functions provide such an approach.

Here we focus on "scatter" and formally study depth-based notions of "scatter measure" and "more scattered" showing that particular depth-based versions generalize other univariate and multivariate notions in the literature while retaining favorable properties. This complements work of Liu, Parelius, and Singh (1997), who develop some practical depth-based data-analytic methodology for treating multivariate scatter.

Following a conventional approach, corresponding to a given notion of "more scattered" we define an associated "scatter measure" as follows. Denote by $\boldsymbol{\mathcal{F}}$ a class of distributions on $\mathbf{R}^d$ and by $\mathbf{R}^+$ the nonnegative real numbers. For convenience, a distribution on $\mathbf{R}^d$ and its corresponding probability measure on Borel sets will be denoted by the same symbol.

Definition 1.1 Corresponding to a given notion of "more scattered", we define a nonparametric multivariate *scatter measure* in $\mathbf{R}^d$ as any functional $\varphi : \boldsymbol{\mathcal{F}} \to \mathbf{R}^+$ satisfying

(i) $\varphi(P) \geq \varphi(Q)$, for any P, Q in $\boldsymbol{\mathcal{F}}$ such that P is more scattered than Q;

(ii) $\varphi(AX + b) = |\det(A)|\varphi(X)$, for any $d \times d$ nonsingular matrix A and vector b in $\mathbf{R}^d$.

A formal definition of "more scattered" (equivalently, "more dispersed", "more spread out", "less concentrated") based on arbitrary depth functions is provided in Section 2. There also, in connection with the particular case of "halfspace depth" (defined below), we illustrate the use of Definition 1.1 to define "scatter measure" and discuss these measures relative to competing approaches.

In Section 3, confining attention to halfspace depth, we show that corresponding notions of "scatter measure" and "more scattered" generalize leading previous notions in the literature, i.e., univariate notions of Bickel and Lehmann (BL) (1976, 1979) and multivariate notions of Eaton (1982) and Oja (1983). In Section 4, desirable properties and characterizations established for previous notions of "more scattered" are shown to hold as well for depth-based notions, in both the univariate and multivariate cases. Also, some properties unique to the depth-based notions are established.

We conclude the present section with a brief discussion of statistical depth functions. For a given distribution P on $\mathbf{R}^d$, any function $D(\cdot, P)$ on $\mathbf{R}^d$ which provides a P-based center-outward ordering of points in $\mathbf{R}^d$ may be regarded (potentially, at least) as a *depth* *function.* For unimodal distributions, regions of high depth correspond to more central regions. A leading example, indeed the earliest, was introduced by Tukey (1975), has received considerable study and popularity, and will serve as the primary special case considered in the present paper. Namely, with respect to a probability distribution P on $\mathbf{R}^d$, the *halfspace depth* of a point x is the minimum probability mass carried by a closed halfspace containing x. That is,

$$HD(x, P) = \inf\{P(H) : H \text{ closed halfspace, } x \in H\}.$$

For comparison we mention a popular competitor introduced by Liu (1990). Namely, the *simplicial depth* of a point x with respect to P is the probability that x belongs to a random d-dimensional simplex. That is,

$$SD(x, P) = P(x \in S[X_1, \ldots, X_{d+1}]),$$

where $X_1, \ldots, X_{d+1}$ is a random sample from P and $S[x_1, \ldots, x_{d+1}]$ denotes the d-dimensional simplex with vertices $x_1, \ldots, x_{d+1}$.

Here we confine attention to *affine invariant* depth functions, that is, those satisfying

$$D(x, F_X) = D(Ax + b, F_{AX+b}),$$

for any $F \in \boldsymbol{\mathcal{F}}$, nonsingular matrix A, and vector $b \in \mathbf{R}^d$, where F_ξ denotes the distribution of a given random variable ξ. The halfspace and simplicial depth functions are readily seen to be affine invariant.

Further properties, such as "maximality at point of symmetry (if any)", "monotonicity outward from point of symmetry (if any)", and "vanishing at infinity", are sometimes required of depth functions (see Zuo and Serfling (1998a)). All three of these are satisfied, for example, by the halfspace depth. We shall not, however, require these to hold in the present development.

Finally, we note that depth functions may be described conveniently in terms of their contours and regions bounded by the contours. For a given depth function $D(\cdot, P)$ and for $\alpha > 0$, we call

$$D^\alpha(P) = \{x \in \mathbf{R}^d : D(x, P) \geq \alpha\}$$

the corresponding *α-trimmed region.* Its boundary $\partial D^\alpha(P)$ is called the corresponding *α-contour.* Depth contours are useful for viewing multivariate distributions geometrically and for exploring the shape of multivariate datasets. Further, the notion of trimmed regions will be basic to our definition of "more scattered" in the next section.

In particular, for the *halfspace* depth, the relevant range of α is $(0, 1)$, and it is straightforward to establish

$$D^\alpha(P) = \bigcap\{H : H \text{ closed halfspace, } P(H) \geq 1 - \alpha\}. \tag{1}$$

That is, here the α-trimmed region is the intersection of all closed halfspaces of probability measure $1 - \alpha$ or greater, and we note that *convexity* thus follows. (See [28, Theorem 2.11] for a similar result for a particular class of depth functions containing the halfspace depth.)

For recent results covering various depth functions, their properties, convergence of their sample versions, convergence of their contours, and their applications to multivariate location, scale, skewness, regression, etc., see, for example, [9], [11], [12], [13], [14], [15], [18], [20], [21], [23], [25], [28], [29], and [30].

2 Depth-based notions of "scatter measure" and "more scattered"

Let $\Delta(\cdot)$ be a *volume* function in $\mathbf{R}^d$. For simplicity, we take $\Delta(S)$ = Lebesgue measure of S, for Lebesgue measurable $S \subset \mathbf{R}^d$.

Definition 2.1 Let $D(\cdot,\cdot)$ be a statistical depth functional on $\mathbf{R}^d \times \boldsymbol{\mathcal{F}}$. A distribution P in $\boldsymbol{\mathcal{F}}$ is said to be *more scattered* than distribution Q in $\boldsymbol{\mathcal{F}}$, denoted "$P \overset{sc}{\geq} Q$", if $\Delta(D^\alpha(P)) \geq \Delta(D^\alpha(Q))$ for each $\alpha > 0$.

Thus P is "more scattered" than Q if, for any $\alpha > 0$, the α-trimmed region based on $D(\cdot, P)$ has greater volume than that based on $D(\cdot, Q)$. A corresponding notion of scatter measure is then available via Definition 1.1, as seen in the following example.

Example 2.1 Fix $\alpha > 0$ and define, based on a given depth functional $D(\cdot,\cdot)$,

$$\varphi^\alpha(P) = \Delta(D^\alpha(P)),\ P \in \boldsymbol{\mathcal{F}}.$$

As shown in Theorem 2.1 below, φ^α is a nonparametric multivariate scatter measure. In particular, for *halfspace* depth, convexity of the regions $D^\alpha(P)$ gives $\varphi^\alpha(P)$ strong intuitive appeal. □

Theorem 2.1 *For any fixed $\alpha > 0$, the functional φ^α is a nonparametric multivariate scatter measure in the sense of Definitions 1.1 and 2.1.*

PROOF. Fix $\alpha > 0$. By Definition 2.1, it is easy to see that φ^α satisfies condition (i) of Definition 1.1. Further, condition (ii) of Definition 1.1 follows by Theorem 2.20 of Rudin (1987) or Theorem 12.2 of Billingsley (1986). □

Given a scatter measure $\varphi(\cdot)$, its value $\varphi(P)$ for a given distribution P may be estimated by the sample analogue $\hat{\varphi} = \varphi(\widehat{P}_n)$, where $\widehat{P}_n$ denotes the usual empirical measure on $\mathbf{R}^d$ given by placing mass $1/n$ at each observation in a random sample $X_1, \ldots, X_n$ from P.

Example 2.2 Continuing Example 2.1, we estimate $\varphi^\alpha(P)$ by $\hat{\varphi}^\alpha = \varphi^\alpha(\widehat{P}_n) = \Delta(D^\alpha(\widehat{P}_n))$. Note that the asymptotic behavior of such estimators is derived from that of the regions $D^\alpha(\widehat{P}_n)$ in $\mathbf{R}^d$. For such results in the case of halfspace depth, see Donoho and Gasko (1992), Nolan (1992), and Massé and Theodorescu (1994), in the case of arbitrary depth functions for elliptical distributions, He and Wang (1997), and in the case of arbitrary depth functions for general distributions, Zuo and Serfling (1998b).

The regions $D^\alpha(\widehat{P}_n)$ represent d-dimensional analogues of the α-trimmed sample upon which many univariate methods are based. It should be noted, however, that the halfspace-based α-trimmed regions for $d > 1$ do not retain all interpretations of the univariate case. For example, in the case $d = 1$, the region $D^{1/4}(\widehat{P}_n)$ comprises the middle half of the data and accordingly $\varphi^{1/4}(\widehat{P}_n)$ is the interquartile range, a long-established nonparametric spread measure which, further, satisfies the criteria of [5]. For $d > 1$, however, $D^{1/4}(\widehat{P}_n)$ does not necessarily represent the middle half of the data. For some data sets the maximum halfspace depth can be as low as $1/(d+1)$ (see [9]), in which case $D^{1/4}(\widehat{P}_n)$ is empty for $d > 4$. Consequently, for $d > 1$, the "middle half" of the data is given by $D^\alpha(\widehat{P}_n)$ for a *data-dependent* choice of α, as developed by Rousseeuw and Ruts (1997) in their "bagplot", a multivariate version of Tukey's boxplot.

One appeal of estimators based on trimmed samples is their typical favorable breakdown point α, whereas M-estimators and related other competitors have breakdown points $\leq 1/(d+1)$, thus unfortunately decreasing with higher dimension. It should be noted, however, that some particular trimming approaches can have unfavorable properties despite their good breakdown points. For example, coordinatewise trimming is not affine invariant, eliminating only outliers for which at least one coordinate is extreme. And convex hull peeling is prone to eliminate not only outliers, but, excessively, non-outliers. For further discussion and perspective, see Donoho and Gasko (1992), Nolan (1992), Barnett and Lewis (1995), pp. 277–279, Liu, Parelius, and Singh (1997), and Zuo and Serfling (1998a). $\square$

3 Correspondences with previous notions of "more scattered"

Here the notion of "more scattered" based on *halfspace depth* is shown to generalize leading previous notions. This broadens the class of pairs of distributions (P, Q) for which P and Q may be meaningfully compared with respect to "scatter". In particular, we consider four previous notions: two of BL (1976, 1979) for *symmetric* and *arbitrary univariate* distributions, respectively, and two of Eaton (1982) and Oja (1983), respectively, for *arbitrary multivariate* distributions.

In discussing the univariate versions, we denote by $X_1 \overset{st}{\leq} X_2$ that X_1 is *stochastically smaller than* X_2, i.e., that $F_{X_1}(x) \geq F_{X_2}(x)$ for every x, and we define the usual quantile function by $F^{-1}(p) = \inf\{x : F(x) \geq p\}$, $0 < p < 1$.

The following one-dimensional notion of "more scattered" for symmetric distributions, essentially equivalent to the "peakedness" ordering of Birnbaum (1948), was introduced by BL (1976).

D1: In $\mathbf{R}$, for X and Y symmetric about μ and ν, respectively, we say that X is *more scattered about* μ *than* Y *about* ν if

$$|X - \mu| \overset{st}{\geq} |Y - \nu|.$$

An analogue for arbitrary distributions in $\mathbf{R}$ was given in BL (1979).

D2: In $\mathbf{R}$, for arbitrary random variables X and Y, we say that X *is more scattered than* Y, denoted $F_X \overset{sc}{\geq} F_Y$, if

$$F_X^{-1}(v) - F_X^{-1}(u) \geq F_Y^{-1}(v) - F_Y^{-1}(u), \ \forall\, 0 < u < v < 1.$$

The essence of D2 was introduced by Brown and Tukey (1946). Note that D1 and D2 do not coincide for symmetric distributions. Eaton (1982) extended D1 to higher dimension:

D3: P is *more scattered about* μ *than* Q *about* ν if

$$P(C+\mu) \leq Q(C+\nu), \ \forall \text{ convex } C \in \mathbf{R}^d \text{ with } C = -C.$$

Oja (1983) likewise extended D2:

D4: P is *more scattered than* Q, $P \overset{sc}{\geq} Q$, if for X having distribution P, and some ϕ such that

$$\Delta(S[\phi(x_1), \phi(x_2), \ldots, \phi(x_{d+1})] \leq \Delta(S[x_1, x_2, \ldots, x_{d+1}]),$$

for all $x_1, x_2, \ldots, x_{d+1}$ in $\mathbf{R}^d$, the random vector $\phi(X)$ has distribution Q.

We will show that the notions D1–D4, and that given by Definition 2.1 with *halfspace* depth, are interrelated as follows:

- For *symmetric univariate* distributions,
  - (a) D3 is equivalent to D1. Thus D3 generalizes D1. (See Lemma 3.1 and Remarks 3.1.)
  - (b) Definition 2.1 with halfspace depth generalizes D1 and D3. (See Theorem 3.1 and Remarks 3.2.)
- For *strictly increasing univariate* cdf's, D4 is equivalent to D2. (See Remarks 3.3.)
- For *centrally symmetric multivariate* distributions, Definition 2.1 with halfspace depth generalizes D3. (See Theorem 3.2 and Remarks 3.4.)
- For *arbitrary multivariate* distributions, Definition 2.1 with halfspace depth generalizes D4. (See Theorem 3.3 and Remarks 3.4.)

We now provide the details, beginning with some useful characterizations of D1.

Lemma 3.1 *In* $\mathbf{R}$, *let* X *and* Y *have distributions symmetric about* μ *and* ν, *respectively. Then the following statements are equivalent:*

(1) X *is more scattered about* μ *than* Y *about* ν *in the sense of D1;*

(2) $P(|X-\mu| \le a) \le P(|Y-\nu| \le a), \quad \forall\, a > 0;$

(3) $t\,(F_{X-\mu}(t) - F_{Y-\nu}(t)) \le 0, \quad \forall\, t \in \mathbf{R};$

(4) $(s - \frac{1}{2})\left(F^{-1}_{X-\mu}(s) - F^{-1}_{Y-\nu}(s)\right) \ge 0, \quad \forall\, s \in (0,1).$

Proof. (a) (1) $\Leftrightarrow$ (2). This follows in a straightforward fashion from the definition of "stochastically smaller".

(b) (2) $\Leftrightarrow$ (3). By symmetry of $X-\mu$ and $Y-\nu$, (2) holds iff $F_{X-\mu}(t) \le F_{Y-\nu}(t)$ for $t > 0$ and $F_{X-\mu}(t) \ge F_{Y-\nu}(t)$ for $t < 0$, that is, iff (3) holds.

(c) (3) $\Leftrightarrow$ (4). This follows easily from the symmetry of $X-\mu$ and $Y-\nu$. □

Remarks 3.1 Clearly, by the above result, for *symmetric* distributions in $\mathbf{R}$, D1 is equivalent to D3. Thus D3 indeed generalizes D1. □

Theorem 3.1 *For symmetric distributions in* $\mathbf{R}$*, Definition 2.1 with halfspace depth is equivalent to D1 (and D3).*

Proof. Let X have cdf F_X symmetric about μ and Y have cdf F_Y symmetric about ν. For simplicity, let these distributions be continuous; extension to the general case is straightforward.

For halfspace depth and $d=1$, it is readily checked that for any continuous cdf G,

$$\Delta(D^\alpha(G)) = |G^{-1}(1-\alpha) - G^{-1}(\alpha)|, \quad 0 < \alpha < 1.$$

Applying symmetry, we thus have

$$\Delta(D^\alpha(F_X)) = 2|F^{-1}_{X-\mu}(\alpha)|, \quad 0 < \alpha < 1,$$

and

$$\Delta(D^\alpha(F_Y)) = 2|F^{-1}_{Y-\nu}(\alpha)|, \quad 0 < \alpha < 1,$$

whence

$$\Delta(D^\alpha(F_X)) - \Delta(D^\alpha(F_Y)) = \begin{cases} -2[F^{-1}_{X-\mu}(\alpha) - F^{-1}_{Y-\nu}(\alpha)], & 0 < \alpha \le 1/2, \\ +2[F^{-1}_{X-\mu}(\alpha) - F^{-1}_{Y-\nu}(\alpha)], & 1/2 \le \alpha < 1. \end{cases}$$

Referring to condition (4) of Lemma 3.1, we see that Definition 2.1 with halfspace depth is equivalent to D1. □

Remarks 3.2 By Theorem 3.1, Definition 2.1 is indeed a generalization of BL (1976). □

Remarks 3.3 For strictly increasing distributions in $\mathbf{R}$, D4 is equivalent to D2; see Theorem 1 of BL (1979) for the proof. Also, D4 implies D1, but the converse fails since the two "more scattered" notions of BL (1976, 1979) are inconsistent. □

In extending the preceding comparisons to $\mathbf{R}^d$, we use the following standard nonparametric notion of multivariate symmetry. A random vector X in $\mathbf{R}^d$ (or its distribution P) is *centrally symmetric* about θ if $X - \theta \stackrel{d}{=} \theta - X$, where "$\stackrel{d}{=}$" denotes "equal in distribution", or, equivalently, if $P(S) = P(S^{(\theta)})$ for any Borel set S and its reflection $S^{(\theta)}$ about θ. In the following result, D3 is exhibited as a special case of Definition 2.1.

Theorem 3.2 *Let P and Q be centrally symmetric about the origin in $\mathbf{R}^d$. Take as depth function the halfspace depth. Then $P \stackrel{sc}{\geq} Q$ in the sense of D3 implies $P \stackrel{sc}{\geq} Q$ in the sense of Definition 2.1.*

PROOF. Suppose, for some $\alpha > 0$, that there exists a point $x \in \mathbf{R}^d$ such that $D(x, Q) = \alpha$ and $D(x, P) < \alpha$. Then there exists a closed halfspace H_x with x on the boundary such that $P(H_x) < \alpha$ and $P(H_x^-) < \alpha$, where H_x^- denotes the reflection of H_x about the origin. Applying D3, we have

$$Q(\mathbf{R}^d - (H_x \cup H_{-x})) \geq P(\mathbf{R}^d - (H_x \cup H_{-x})) > 1 - 2\alpha,$$

which implies that $Q(H_x) = Q(H_{-x}) < \alpha$, contradicting the assumption that $D(x, Q) = \alpha$. That is, $D(x, P) \geq D(x, Q)$, for any $x \in \mathbf{R}^d$, whence $D^\alpha(Q) \subset D^\alpha(P)$. □

Theorem 3.3 *Take as depth function the halfspace depth. Suppose that $P \stackrel{sc}{\geq} Q$ in the sense of D4 with ϕ an affine transformation. Then $P \stackrel{sc}{\geq} Q$ in the sense of Definition 2.1.*

PROOF. (a) $\mathbf{R}^1$ case. It is clear that $\Delta(D^\alpha(P)) = \Delta([x_1, x_2])$ for some x_1 and x_2. By the affine invariance property of statistical depth functions,

$$\begin{aligned} \Delta(D^\alpha(Q)) &= \Delta(\{y \in \mathbf{R}^1 \mid D(y, Q) \geq \alpha\}) \\ &= \Delta(\{y = \phi(x) \in \mathbf{R}^1 \mid D(x; P) \geq \alpha\}) \\ &= \Delta([\phi(x_1), \phi(x_2)]). \end{aligned}$$

It follows from D4 that $\Delta(D^\alpha(P)) \geq \Delta(D^\alpha(Q))$, that is, $P \stackrel{sc}{\geq} Q$ in the sense of Definition 2.1.

(b) $\mathbf{R}^d$ $(d > 1)$ case. By 1, $D^\alpha(P)$ is convex and compact and thus can be approximated by d-dimensional simplices, whence

$$\Delta(D^\alpha(P)) = \cup_{i=1}^\infty \Delta(S_i)$$

for a sequence of d-dimensional simplices $\{S_i\}$. The affine invariance of statistical depth functions again implies that

$$\begin{aligned} \Delta(D^\alpha(Q)) &= \Delta(\{y \in \mathbf{R}^d \mid D(y, Q) \geq \alpha\}) \\ &= \Delta(\{y = \phi(x) \in \mathbf{R}^d \mid D(x; P) \geq \alpha\}) \\ &= \cup_{i=1}^\infty \Delta(\phi(S_i)). \end{aligned}$$

Now since $\Delta(S[\phi(x_1), \ldots, \phi(x_{d+1})]) \leq \Delta(S[x_1, \ldots, x_{d+1}])$ for $x_1, \ldots, x_{d+1} \in \mathbf{R}^d$, we have $\Delta(D^\alpha(P)) \geq \Delta(D^\alpha(Q))$, proving that $P \stackrel{sc}{\geq} Q$ in the sense of Definition 2.1. □

Remarks 3.4 Theorems 3.2 and 3.3 imply that Definition 2.1 for "more scattered" is weaker (i.e., more general) than those of Eaton (1982) and Oja (1983), under typical conditions. □

4 Properties of depth-based notions of "more scattered"

Here we investigate properties related to the depth-based notion of "more scattered" given by Definition 2.1.

Immediately we have that the relation $\overset{sc}{\geq}$ is reflexive, transitive and antisymmetric. These are properties that should be possessed by any reasonable notion of "more scattered".

Further, the ordering $\overset{sc}{\geq}$ is preserved under affine transformation. That is, the affine invariance property we are assuming for statistical depth functions yields

Theorem 4.1 *If $X \overset{sc}{\geq} Y$, then $AX + b \overset{sc}{\geq} AY + b$ for any $d \times d$ nonsingular matrix A and vector b in $\mathbf{R}^d$.*

Thus, for example, a comparison of two clouds of data points with respect to scatter, is not altered by a change of coordinate system.

For simplicity, in the remainder of this section we confine attention to the special case of *halfspace* depth. Analogous results may undoubtedly be established for some other depth functions.

The following result provides a simple sufficient condition for two distributions in $\mathbf{R}^d$ to be comparable by the halfspace-based notion of "more scattered".

Theorem 4.2 *Let P and Q be distributions in $\boldsymbol{\mathcal{F}}$. If $P(H) \geq Q(H)$ for every closed halfspace H in $\mathbf{R}^d$ with $P(H) \leq 1/2$, then $P \overset{sc}{\geq} Q$ in the sense of Definition 2.1 with halfspace depth.*

PROOF. Similar to that of Theorem 3.2. □

Remarks 4.1 If P and Q are symmetric about the origin in $\mathbf{R}$, then the condition in the above theorem is also necessary. It is not clear, however, whether this holds true in $\mathbf{R}^d$ for $d > 1$. □

Theorem 4.2 yields the following result of BL (1976) on scatter comparison of univariate distributions having monotone likelihood ratio.

Corollary 4.1 *Suppose that P and Q are symmetric about the origin in $\mathbf{R}$, with densities f and g respectively. If $f(x)/g(x)$ is increasing for $x > 0$, then $P \overset{sc}{\geq} Q$ in the sense of Definition 2.1 with halfspace depth.*

Proof. Since $f(x)/g(x)$ is increasing $\forall x > 0$, by symmetry of P and Q, for any closed halfspace $H = [x, \infty)$ such that $P(H) \leq 1/2$, we have $P(H) = P([x, \infty)) \geq Q([x, \infty)) = Q(H)$. A similar result holds for any closed halfspace $H = (-\infty, y]$. We now apply Theorem 4.2. □

Let us now consider mixtures of distributions. In particular, we may ask whether contamination of a given distribution by another with greater scatter yields a distribution more scattered than the uncontaminated distribution.

Theorem 4.3 *Suppose that $P_1 \overset{sc}{\geq} P_2$ in the sense of Definition 2.1 with halfspace depth, and that $D^\alpha(P_1) \supset D^\alpha(P_2)$ for all $\alpha > 0$. Then $(1-\theta)P_1 + \theta P_2 \overset{sc}{\geq} P_2$ in the same sense, for any $\theta \in (0,1)$.*

Proof. Put $P_\theta = (1-\theta)P_1 + \theta P_2$. We show that

$$(*) \qquad D^\alpha(P_\theta) \supset D^\alpha(P_2),$$

for each $\alpha > 0$. Let $x \in D^\alpha(P_2)$ be given and assume that $D(x, P_2) = \beta \geq \alpha$. Suppose that $D(x, P_\theta) < \alpha$. Then there exists a closed halfspace H_x with x on its boundary such that $P_\theta(H_x) < \alpha$. Now $P_\theta(H_x) = (1-\theta)P_1(H_x) + \theta P_2(H_x)$. Since $D^\alpha(P_1) \supset D^\alpha(P_2)$, $P_1(H_x) \geq \alpha$. It follows that $\theta P_2(H_x) < \alpha - (1-\theta)P_1(H_x) \leq \theta\alpha$. Thus $P_2(H_x) < \alpha$, contradicting the assumption that $D(x, P_2) = \beta \geq \alpha$. Hence $x \in D^\alpha(P_2)$ implies $x \in D^\alpha(P_\theta)$ for any $\alpha > 0$. Thus $(*)$ holds, proving the result. □

In particular, for multivariate normal models we have

Corollary 4.2 *In the sense of Definition 2.1 with halfspace depth, a d-variate standard normal distribution $N_d(\mathbf{0}, \mathbf{I})$ contaminated with another d-variate normal distribution $N_d(\mathbf{0}, \sigma^2\mathbf{I})$ with $\sigma > 1$ (Tukey model in $\mathbf{R}^d$) is more scattered than the uncontaminated d-variate standard normal distribution $N_d(\mathbf{0}, \mathbf{I})$.*

For the univariate case, Theorem 4.3 together with Theorem 3.1 yields the following result of BL (1976).

Corollary 4.3 *In $\mathbf{R}$, if F and G are symmetric about the origin, and G is more scattered than F, then the mixture $H_\theta = (1-\theta)F + \theta G$ is more scattered than F, for $0 < \theta < 1$.*

It is convenient and intuitively appealing when scatter comparisons of two distributions reduce to comparison of their covariance matrices. The following results show this to be true in the case of elliptical distributions and in particular multivariate normal distributions. A random vector X in $\mathbf{R}^d$ is said to be *elliptically distributed* if its density is of the form

$$f(x) = c|\Sigma|^{-1/2} h((x-\mu)'\Sigma^{-1}(x-\mu)),$$

denoted $X \sim E_d(h; \mu, \Sigma)$.

Theorem 4.4 *Suppose that $X \sim E_d(h; u, \Sigma_1)$ and $Y \sim E_d(h; v, \Sigma_2)$. Then $Y \overset{sc}{\geq} X$ in the sense of Definition 2.1 with halfspace depth if and only if $\Sigma_2 - \Sigma_1 \geq 0$ (positive semidefinite).*

Proof. One can show (see Zuo and Serfling (1998b)) in this case that

$$\begin{aligned} D^\alpha(X) &= \{x \in \mathbf{R}^d \,|\, D(x; X) \geq \alpha\} \\ &= \{x \in \mathbf{R}^d \,|\, (x-u)'\Sigma_1^{-1}(x-u) \leq r_\alpha^2\}, \end{aligned}$$

for any $\alpha > 0$ and some corresponding $r_\alpha > 0$ in $\mathbf{R}$. A similar result holds for $D^\alpha(Y)$. It is not difficult to see that $\Sigma_1^{-1} - \Sigma_2^{-1} \geq 0$ if and only if $\Delta(D^\alpha(Y)) \geq \Delta(D^\alpha(X))$. Since $\Sigma_2 - \Sigma_1 \geq 0$ if and only if $\Sigma_1^{-1} - \Sigma_2^{-1} \geq 0$, it follows that $Y \overset{sc}{\geq} X$ in the sense of Definition 2.1 if and only if $\Sigma_2 - \Sigma_1 \geq 0$. □

For the multivariate normal case we obtain

Corollary 4.4 *Consider independent random vectors $X_i \sim N_d(u_i, \Phi_i)$, $1 \leq i \leq n$, and independent random vectors $Y_i \sim N_d(v_i, \Psi_i)$, $1 \leq i \leq n$. Then $\sum_{i=1}^n (B_i \Psi_i B_i' - A_i \Phi_i A_i') \geq 0$ if and only if $\sum_i^n B_i Y_i \overset{sc}{\geq} \sum_i^n A_i X_i$ in the sense of Definition 2.1 with halfspace depth.*

Remarks 4.2 Corollary 4.4 implies, for example, that the sample mean $\overline{X}$ from a multivariate normal random vector X is more scattered than the sample mean $\overline{Y}$ from another multivariate normal random vector Y if and only if X is more scattered than Y, or equivalently if and only if $\mathrm{cov}(X) - \mathrm{cov}(Y) \geq 0$.

The above remark mentions a special case in which the ordering $\overset{sc}{\geq}$ is preserved under the taking of convolutions of random variables in $\mathbf{R}^d$. We now look at this more generally. Our first result treats the case $d = 1$.

Theorem 4.5 *Let $d = 1$. Assume that*

1^o *X_1 and X_2 are independent with respective distributions F_1 and F_2, and Y_1 and Y_2 are independent with respective distributions G_1 and G_2;*

2^o *F_i and G_i are symmetric about the origin in $\mathbf{R}$ $(i = 1, 2)$;*

3^o *F_1 and G_2 have unimodal densities;*

4^o *Y_i is more scattered than X_i, for $i = 1, 2$, in the sense of Definition 2.1 with halfspace depth.*

Then $Y_1 + Y_2$ is more scattered than $X_1 + X_2$ in the same sense.

PROOF. For any constant $c > 0$, since X_1 and X_2 are independent and symmetric, we have

$$\begin{aligned}
P(|X_1 + X_2| < c) &= 2\int_0^\infty (F_1(x+c) - F_1(x-c))\, dF_2(x) \\
&= 2\left((F_1(x+c) - F_1(x-c))F_2(x)|_0^\infty\right) \\
&\quad -2\int_0^\infty F_2(x)(f_1(x+c) - f_1(x-c))\, dx.
\end{aligned}$$

The unimodality of F_1 implies that $f_1(x+c) - f_1(x-c) \leq 0$ for any $x > 0$, and 4^o and Lemma 3.1 imply that $F_2(x) \geq G_2(x)$ for any $x > 0$. The symmetry of F_2 and G_2 about the origin implies that $F_2(0) = G_2(0) = 1/2$. Thus

$$\begin{aligned}
P(|X_1 + X_2| < c) &= 2\int_0^\infty (F_1(x+c) - F_1(x-c))\, dF_2(x) \\
&\geq 2\int_0^\infty (F_1(x+c) - F_1(x-c))\, dG_2(x) \\
&= 2\left((F_1(x+c) - F_1(x-c))G_2(x)|_0^\infty\right) \\
&\quad -2\left(\int_c^\infty G_2(x-c)\, dF_1(x) - \int_{-c}^\infty G_2(x+c) d\, F_1(x)\right) \\
&= 2\left((\frac{1}{2} - F_1(c)) - \int_0^\infty (G_2(x-c) - G_2(x+c))\, dF_1(x)\right) \\
&\quad +2\left(\int_0^c (G_2(x-c) + G_2(-x+c))\, dF_1(x)\right) \\
&= 2\left((\frac{1}{2} - F_1(c)) - \int_0^\infty (G_2(x-c) - G_2(x+c))\, dF_1(x)\right) \\
&\quad +2\left((F_1(c) - 1/2)\right) \\
&= 2\int_0^\infty (G_2(x+c) - G_2(x-c))\, dF_1(x).
\end{aligned}$$

Repeating the argument used above, we have

$$\begin{aligned}
P(|X_1 + X_2| < c) &\geq 2\int_0^\infty (G_2(x+c) - G_2(x-c))\, dF_1(x) \\
&\geq 2\int_0^\infty (G_2(x+c) - G_2(x-c))\, dG_1(x) \\
&= P(|Y_1 + Y_2| < c).
\end{aligned}$$

The result now follows immediately from Lemma 3.1 and Theorem 3.1. □

Remarks 4.3 (1) Theorem 4.5 generalizes Theorem 1 of BL (1976), using a similar method of proof but dropping their simplifying assumption of independence of X_1 and Y_2.

(2) Theorem 4.5 also generalizes the main result of Birnbaum (1948), where continuity of Y_1 and X_2 was required. □

The next result treats general dimension d.

Theorem 4.6 *Let X be a random vector in $\mathbf{R}^d$ with density $f(x)$ satisfying*

1^o *$f(x) = f(-x)$;*

2^o *$\{x | f(x) \geq u\}$ is convex for any $u > 0$.*

Let Y be a random vector independently distributed and symmetric about the origin in $\mathbf{R}^d$. Then $X+Y \overset{sc}{\geq} bX+aY$ in the sense of Definition 2.1 with halfspace depth, for $0 \leq a \leq b \leq 1$.

Proof. It suffices to show that $X + Y \overset{sc}{\geq} X + aY$ in the sense of Definition 2.1 for all $a \in [0, 1]$. We show that

$$(*) \qquad D^\alpha(X + Y) \supset D^\alpha(X + aY),$$

for all $\alpha > 0$. Suppose that $x \in D^\alpha(X + aY)$ and put $D(x; X + aY) = \beta \geq \alpha$. Suppose that $x \notin D^\alpha(X + Y)$, that is, $D(x; X + Y) < \alpha$. Then there exists a closed halfspace H_x with x on its boundary such that $\alpha > P(X + Y \in H_x)$. On the other hand, it is not difficult to show that

$$\begin{aligned} P(X + Y \in H_x) &= \int_{\mathbf{R}^d} P(X \in H_x - t)\, dF_Y(t) \\ &= \int_{H_0^{(x)}} (P(X \in H_x + t) + P(X \in -H_x + t))\, dF_Y(t), \end{aligned}$$

where $H_0^{(x)}$ is the closed halfspace with the origin on its boundary which is hyperparallel to the boundary of H_x, and contains H_x. By Theorem 1 of Anderson (1955) (also see Mudholker (1966) and Anderson (1996)), we have that

$$\begin{aligned} P(X + Y \in H_x) &= \int_{H_0^{(x)}} (P(X \in H_x + t) + P(X \in -H_x + t))\, dF_Y(t) \\ &\geq \int_{H_0^{(x)}} (P(X \in H_x + at) + P(X \in -H_x + at))\, dF_Y(t) \\ &= \int_{R^d} P(X \in H_x - at)\, dF_Y(t) \\ &= P(X + aY \in H_x). \end{aligned}$$

Hence $\alpha > P(X + aY \in H_x)$, contradicting the assumption that $D(x; X + aY) = \beta \geq \alpha$. Thus $x \in D^\alpha(X + aY)$ implies that $x \in D^\alpha(X + Y)$, for any $\alpha > 0$, that is, $(*)$ holds. The result now follows immediately from Definition 2.1. □

Finally, we examine preservation of "more scattered" under convolutions within the important class of spherically symmetric distributions. A random vector X has *spherically symmetric* distribution if $X \overset{d}{=} TX$ for any orthogonal matrix T (see Muirhead (1982), p. 32).

Theorem 4.7 *Let X_i and $Y_i, i = 1,2,$ in $\mathbf{R}^d$ be independent with spherically symmetric distributions F_i and G_i, respectively. Suppose*

1^o $Y_i \overset{sc}{\geq} X_i$, for $i = 1,2$, in the sense of Definition 2.1 with halfspace depth;

2^o F_1 and G_2 have unimodal densities.

Then $Y_1 + Y_2 \overset{sc}{\geq} X_1 + X_2$ in the same sense.

Proof. For any $x \in D^\alpha(X_1 + X_2)$, $\alpha > 0$, suppose that $D(x; X_1 + X_2) = \beta \geq \alpha$. We claim that $x \in D^\alpha(Y_1 + Y_2)$, that is, $D(x; Y_1 + Y_2) \geq \alpha$. If not, then $D(x; Y_1 + Y_2) < \alpha$ and there exists a closed halfspace H_x with x on its boundary such that $I \equiv P(Y_1 + Y_2 \in H_x) < \alpha$. On the other hand, applying an argument similar to that for Theorem 4.6, and utilizing spherical symmetry, we have

$$\begin{aligned}
I &= \int_{\mathbf{R}^d} P(Y_2 \in H_x - t)\, dG_1(t) \\
&= \int_{H_0^{(x)}} (P(Y_2 \in t + H_x) + P(Y_2 \in t - H_x))\, dG_1(t) \\
&= 2^{d-1} \cdot \int_{\mathbf{R}_+^d} (P(Y_2 \in t + H_x) + P(Y_2 \in t - H_x))\, dG_1(t) \\
&= 2^{d-1} \cdot \int_0^\infty \cdots \int_0^\infty \gamma(Y_2; t, H_x)\; d\left(\frac{\partial^{d-1} G_1(t)}{\partial t_2 \cdots \partial t_d}\right)\, dt_2 \cdots dt_d,
\end{aligned}$$

where G is the distribution function of Y_1, $\mathbf{R}_+^d$ is the first quadrant of $\mathbf{R}^d$, $t = (t_1, \ldots, t_d)'$, and $\gamma(Y_2; t, H_x) = P(Y_2 \in t + H_x) + P(Y_2 \in t - H_x)$. By Theorem 1 of Anderson (1955), the latter function of t is an increasing function of t_1. Since Y_1 is more scattered than X_1, then by Lemma 3.1 and spherical symmetry

$$\frac{\partial^{d-1} G_1(t)}{\partial t_2 \cdots \partial t_d} \leq \frac{\partial^{d-1} F_1(t)}{\partial t_2 \cdots \partial t_d},$$

for any $t_1 \geq 0$ $(t_i \geq 0, i = 2, \ldots, d)$. Integration by parts now yields

$$\begin{aligned}
I &= 2^{d-1} \int_0^\infty \cdots \int_0^\infty \gamma(Y_2; t, H_x)\; d\left(\frac{\partial^{d-1} G_1(t)}{\partial t_2 \cdots \partial t_d}\right)\, dt_2 \cdots dt_d, \\
&\geq 2^{d-1} \int_0^\infty \cdots \int_0^\infty \gamma(Y_2; t, H_x)\; d\left(\frac{\partial^{d-1} F_1(t)}{\partial t_2 \cdots \partial t_d}\right)\, dt_2 \cdots dt_d, \\
&= P(X_1 + Y_2 \in H_x).
\end{aligned}$$

By similar steps, we thus obtain

$$P(Y_1 + Y_2 \in H_x) \geq P(X_1 + Y_2 \in H_x) \geq P(X_1 + X_2 \in H_x).$$

Hence $\alpha > P(X_1 + X_2 \in H_x)$, contradicting the assumption that $D(x; X_1 + X_2) \geq \alpha$. Therefore, $x \in D^\alpha(X_1 + X_2)$ implies $x \in D^\alpha(Y_1 + Y_2)$, for any $\alpha > 0$, proving that $Y_1 + Y_2 \overset{sc}{\geq} X_1 + X_2$ in the sense of Definition 2.1. □

Remarks 4.4 Theorem 4.7 and Corollary 4.4 are established under the assumptions that the underlying distributions are spherically distributed or multivariate normally distributed. We conclude with an open question: Under what weaker assumption(s) on distributions does an analogue of Theorem 4.5 hold in the multivariate setting?

5 Acknowledgment

The authors greatly appreciate two referees' helpful remarks, which led to several important corrections and improvements.

References

[1] Anderson, T. W. (1955). The integral of a symmetric unimodal function. *Proc. Amer. Math. Soc.* **6** 170–176.

[2] Anderson, T. W. (1996). Some inequalities for symmetric convex sets with applications. *Ann. Statist.* **24** 753–762.

[3] Barnett, V., and Lewis, T. (1995). *Outliers in Statistical Data*, 3rd edition. John Wiley & Sons, New York.

[4] Bickel, P. J., and Lehmann, E. L. (1976). Descriptive statistics for nonparametric models. III. Dispersion. *Ann. Statist.* **4** 1139–1158.

[5] Bickel, P. J., and Lehmann, E. L. (1979). Descriptive statistics for nonparametric models. IV. Spread. In J. Jurečková, ed., *Contributions to Statistics. Hàjek Memorial Volume*, pp. 33–40. Academia, Prague.

[6] Billingsley, P. (1986). *Probability and Measure*, 2nd edition. John Wiley & Sons, New York.

[7] Birnbaum, Z. W. (1948). On random variables with comparable peakedness. *Ann. Math. Statist.* **37** 1593–1601.

[8] Brown, G., and Tukey, J. W. (1946). Some distributions of sample means. *Ann. Math. Statist.* **7** 1–12.

[9] Donoho, D. L., and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827.

[10] Eaton, M. L. (1982). A review of selected topics in multivariate probability inequalities. *Ann. Statist.* **10** 11–43.

[11] He, X., and Wang, G. (1997). Convergence of depth contours for multivariate data sets. *Ann. Statist.* **25** 495–504.

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

[13] Liu, R. Y., Parelius, J. M., and Singh, K. (1997). Multivariate Analysis by data depth: descriptive statistics, graphics and inference. To appear in *Ann. Statist.* **27**.

[14] Liu, R. Y., and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88** 252–260.

[15] Massé, J.-C., and Theorodescu, R. (1994). Halfplane trimming for bivariate distributions. *J. Multiv. Analysis* **48** 188–202.

[16] Mudholkar, G. (1966). The integral of an invariant unimodal function over an invariant convex set – an inequality and applications. *Proc. Amer. Math. Soc.* **17** 1327–1333.

[17] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, New York.

[18] Nolan, D. (1992). Asymptotics for multivariate trimming. *Stoch. Proc. Appl.* **42** 157–169.

[19] Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statist. and Prob. Letters* **1** 327–333.

[20] Rousseeuw, P. J., and Hubert, M. (1999). Regression depth (with discussion). *J. Amer. Statist. Assoc.* **94** 388–433.

[21] Rousseeuw, P. J., and Ruts, I. (1996). Bivariate location depth. *Appl. Statist.* **45** 516–526.

[22] Rousseeuw, P. J., and Ruts, I. (1997). The bagplot: a bivariate box-and-whiskers plot. Technical Report, University of Antwerp.

[23] Rousseeuw, P. J., and Struyf, A. (1998). Computing location depth and regression depth in higher dimensions. *Statist. and Comput.* **8** 193–203.

[24] Rudin, W. (1987). Real and Complex Analysis. McGraw-Hill, New York.

[25] Ruts, I., and Rousseeuw, P. J. (1996). Computing depth contours of bivariate point clouds. *Comput. Statist. & Data Analysis* **23** 153–168.

[26] Tukey, J. W. (1975). Mathematics and picturing data. *Proc. Int. Congr. Math. Vancouver 1974* **2** 523–531.

[27] Yeh, A. B., and Singh, K. (1997). Balanced confidence regions based on Tukey's depth and the bootstrap. *J. Roy. Statist. Soc. B* **59** 639–652.

[28] Zuo, Y., and Serfling, R. (1998a). General notions of statistical depth function and some related convergence results. Submitted.

[29] Zuo, Y., and Serfling, R. (1998b). Structural properties and convergence results for contours of sample statistical depth functions. Submitted.

[30] Zuo, Y., and Serfling, R. (1999). On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *J. Statist. Plann. Inf.* to appear.