**CONTRIBUTIONS TO THE THEORY AND APPLICATIONS OF STATISTICAL DEPTH FUNCTIONS**

APPROVED BY SUPERVISORY COMMITTEE:

Robert Serfling, Chair

Larry Ammann

John Van Ness

Michael Baron

To Chuangping and her parents,

to my mother, and to the memory of my father

# CONTRIBUTIONS TO THE THEORY AND APPLICATIONS OF STATISTICAL DEPTH FUNCTIONS

by

Yijun Zuo, M.S., M.S., M.S., M.S.

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN MATHEMATICAL SCIENCES

THE UNIVERSITY OF TEXAS AT DALLAS

May, 1998

## ACKNOWLEDGMENTS

First of all, I sincerely thank my advisor, Professor Robert Serfling, for his excellent guidance and constant encouragement. Without his continual support this dissertation would never have been completed.

Special thanks are due to Dr. Georgia L. Thompson for her insightful comments and suggestions which have greatly improved the dissertation.

My thanks also go to Professors Larry Ammann, John Van Ness, and Michael Baron for their careful reading of this dissertation and useful suggestions.

I thank Dr. John Wiorkowski, Dr. Raimund Ober, Dr. Janos Turi, Dr. Tiberiu Constantinescu, Dr. Ali Hooshyar and the rest of the faculty and staff in the Programs in Mathematical Sciences for their support during my five years of studies at The University of Texas at Dallas.

Finally, I express my deep gratitude to my wife, Chuangping Yang, for her constant encouragement and support.

March 31, 1998

**CONTRIBUTIONS TO THE THEORY AND APPLICATIONS OF STATISTICAL DEPTH FUNCTIONS**

Publication No._______

Yijun Zuo, Ph.D.

The University of Texas at Dallas, 1998

Supervising Professor: Robert Serfling

For years people have sought ways to order multivariate statistical data. Statistical depth functions are proving to be one of the most promising methods. They provide a center-outward ordering of multivariate observations and allow one to define reasonable analogues of univariate order statistics and rank statistics. The latter, in turn, lead to generalizations of classical univariate L-statistics and R-statistics in the multivariate setting. Consequently, statistical depth functions play key roles in multivariate nonparametric and robust statistical procedures (e. g. multivariate location estimation), multivariate outlier detection, multivariate discriminant analysis and classification, testing of multivariate symmetry, quality control and system reliability, etc.

In this dissertation, desirable properties that depth functions should possess are formulated. A general definition of "statistical depth function" is introduced, which unifies *ad hoc* definitions. Two existing well-known notions of depth function are examined and it is found that one performs well but that the other lacks some favorable properties of "statistical depth functions" and consequently should not be treated in general as a statistical depth function.

General structures for depth functions are also constructed and studied, and some new attractive statistical depth functions are introduced. Applications of statistical depth functions in multivariate nonparametric and robust statistical procedures are explored. In particular, statistical depth function concepts are applied to introduce multivariate quantile *contours*, multivariate *location* measures, and multivariate *scatter* measures, and the resulting entities are investigated in detail.

Closely related to multivariate data ordering is the notion of multivariate *symmetry*. In this dissertation, a new notion of symmetry, called "*halfspace symmetry*", is introduced. Characterizations of this notion of symmetry as well as of other existing notions of symmetry, and interrelationships among these notions of symmetry, are developed.

It turns out that halfspace symmetry is a reasonable symmetry assumption on multivariate distributions for the discussion of nonparametric location inference and related statistical procedures. This new notion of symmetry not only supports more general approximations to actual distributions in modeling but also plays important roles in discussion of statistical depth functions and multivariate location measures.

# TABLE OF CONTENTS

# Chapter 1

# INTRODUCTION

No statisticians would doubt today that univariate order statistics have not only theoretical interest but also practical significance. They have played key roles in many areas such as nonparametric statistical procedures (e.g., robust location estimation), outlier detection, discriminant analysis and classification, testing of symmetry, censored sampling, quality control, etc. Univariate order statistics and rank statistics have been extensively studied by many authors; see David (1981), Hettmansperger (1984), and Arnold, Balakrishnan and Nagaraja (1993).

The extension to *higher dimension* of areas of application and methodological advantages, or general properties, of univariate order statistics, however, turns out to be complicated because of the lack of any obvious and unambiguous means of fully ordering, or ranking multivariate observations. Despite this fact, statisticians have made substantial efforts in defining various types of higher dimensional analogues of univariate order concepts. For extensive treatment of the techniques for ordering multivariate observations, see Barnett (1976) and Reiss (1989). Among existing methods of ordering multivariate data, *depth functions* are proving to be one of the most promising.

The main idea of depth functions is to provide an ordering of all points from a "center" of a given multivariate distribution outward. Depth functions thus allow one to define reasonable analogues of univariate order statistics and rank statistics, which, in turn, lead to generalizations of classical univariate L-statistics and R-statistics in the multivariate setting. Consequently, depth functions play key roles in multivariate extensions of the statistical methods listed above.

Tukey (1974) put forward a "*halfspace*" depth function for points in $\mathbf{R}^d$ with respect to given multivariate distributions. Liu (1988, 1990) introduced a "*simplicial*" depth function in $\mathbf{R}^d$. Recent studies of data depth and applications include Donoho and Gasko (1992), Nolan (1992), Liu and Singh (1993), Massé and Theodorescu (1994), Rousseeuw and Hubert (1996), Ruts ans Rousseeuw (1996), Rousseeuw and Struyf (1997), Yeh and Singh (1997), Koshevoy and Mosler (1997), He and Wang (1997), Liu, Parelius and Singh (1997).

Although *ad hoc* depth functions have attracted considerable attention in the literature, some basic issues about depth functions have not been addressed. For example, (i) What desirable properties should depth functions have? (ii) Do existing depth functions possess all of the properties? (iii) If not, can one build more attractive depth functions?

In this dissertation, desirable properties that depth functions should possess are formulated. A general definition of "statistical depth function" is thus introduced, which unifies *ad hoc* definitions. Some existing well-known depth functions are examined. It is found, for example, that the halfspace depth function performs well but that the simplicial depth function, on the other hand, lacks some favorable properties of "statistical depth functions" and consequently should not be treated in full generality as a statistical depth function.

Some general structures for depth functions are also constructed and studied, and some new attractive statistical depth functions are introduced. Applications of statistical depth functions in multivariate nonparametric and robust statistical procedures are explored. In particular, statistical depth function concepts are applied to introduce multivariate quantile *contours*, multivariate *location* measures, and multivariate *scatter* measures, and the resulting entities are investigated in detail.

The work here provides general foundations for *center-outward ordering* of multivariate data based on depth functions and for development of analogues of univariate order statistics and rank statistics in the multivariate setting.

Closely related to multivariate data ordering is the notion of multivariate *symmetry*. Symmetry arises as major assumption on distributions in multivariate nonparametric location inference. It is understood, however, that actual distributions are typically not symmetric in any strict sense. Rather, symmetric distributions are used only as approximations to actual distributions in modeling. From this point of view, the weaker the assumption on symmetry, the more general the approximation that can be achieved. *Central symmetry*, as the weakest among traditional notions of symmetry, is the most typical assumption on distributions in multivariate nonparametric location inference. In 1990, Liu introduced a new notion of symmetry, called "*angular symmetry*", which is weaker than central symmetry. Since then, angular symmetry has received considerable attention; see Randles (1989), Liu (1990), Small (1990), Arcones, Chen and Giné (1994), Chen (1995) and Liu, Parelius and Singh (1997). However, characterizations of angular symmetry have not yet been studied formally.

Here, angular symmetry is examined and some useful characterizations are developed. Further, a new and even weaker notion of symmetry, called "*halfspace symmetry*", is introduced and studied. It turns out that halfspace symmetry is a reasonable symmetry assumption on multivariate distributions for the discussion of nonparametric location inference and related statistical procedures. This new notion not only supports more general approximations to actual distributions in modeling but also plays important roles in discussion of statistical depth functions and multivariate location measures.

The remainder of this dissertation is organized as follows. Chapter 2 is devoted to the discussion of *multivariate symmetry*. A new notion of symmetry, the halfspace symmetry, is introduced. Characterizations of this notion as well as other existing notions of symmetry, and interrelationships among these notions of symmetry, are developed.

Chapter 3 deals with *statistical depth functions*. Desirable properties of statistical depth functions are presented, and a general definition of statistical depth function is introduced. Two existing well-known notions of depth function, the *halfspace* depth function and the *simplicial* depth function, are examined. It turns out that the halfspace depth function performs well in general but that the simplicial depth function does not in some cases. General structures for statistical depth functions are introduced and studied in detail. The behavior of some sample depth functions is explored.

Chapter 4 concerns *multivariate depth contours* or *multivariate quantile contours*. Properties of depth contours, and the behavior of sample depth contours are thoroughly studied. Results are obtained which improve and generalize previous results in the literature.

Chapter 5 treats *depth-based multivariate nonparametric location measures*. A desirable property for such measures, called the "*center locating*" condition, is introduced as a new criterion. Depth-based multivariate medians, as multivariate nonparametric location measures, are introduced. The performance of these and some other multivariate nonparametric location measures is studied. It turns out that an existing well-known measure fails to satisfy the "center locating" condition and, consequently, should be used with caution in practice.

Chapter 6 discusses *depth-based multivariate nonparametric scatter measures*. Statistical depth functions are employed to introduce a notion of "*more scattered*" for comparison of one multivariate distribution with another. Relationships among this new notion and other existing notions of "more scattered", such as those of Bickel and Lehmann (1976) in $\mathbf{R}$, and of Eaton (1982) and Oja (1983) in $\mathbf{R}^d$ , are explored. It turns out that this notion is a generalization of that of Bickel and Lehmann (1976) in $\mathbf{R}$, and is more general than those of Eaton (1982) and Oja (1983) in $\mathbf{R}^d$ under some typical conditions. The properties of this depth-based notion are studied thoroughly. Finally, depth-based multivariate nonparametric scatter measures are defined, and some examples are presented and studied.

# Chapter 2

# SYMMETRY OF MULTIVARIATE DISTRIBUTIONS

## 2.1 Introduction

Symmetry of multivariate distributions plays so important a role in multivariate statistical inference, and in our discussions in later chapters, that we devote this chapter to its study.

*Spherical symmetry*, *elliptical symmetry* and *central symmetry* are the traditional notions of symmetry for statistical distributions. For definitions of these, see Muirhead (1982) and Eaton (1983). Spherical symmetry is the strongest notion among these three, whereas central symmetry is the weakest. Various characterizations of these symmetries have already been given in the literature. Other *ad hoc* notions of symmetry for statistical distributions exist; see Fang, Kotz and Ng (1990) for detailed discussion.

Liu (1988) introduced a new notion of symmetry, called *angular symmetry*, which is weaker than central symmetry. Since then it has received considerable attention; see Randles (1989), Liu (1990), Small (1990), Arcones, Chen and Giné (1994), Chen (1995) and Liu, Parelius and Singh (1997). However, formal characterizations of angular symmetry have not yet been studied carefully.

In this chapter, angular symmetry is examined and characterizations are developed.

Further, a new and even weaker notion of symmetry, called "*halfspace symmetry*", is introduced and studied. It turns out that halfspace symmetry is a reasonable assumption on multivariate distributions for the discussion of nonparametric multivariate location inference, including "multivariate medians", and of other multivariate statistical procedures. Since halfspace symmetry is the weakest among the above notions of symmetry, it provides the broadest foundation for approximating actual distributions, which are typically not symmetric in any strict sense, by "symmetric" ones.

## 2.2 Angular Symmetry

We begin with the notion of central symmetry of multivariate distributions.

**Definition 2.2.1** A random vector $X \in \mathbf{R}^d$ is *centrally symmetric* about a point $\theta \in \mathbf{R}^d$ if $X - \theta$ and $\theta - X$ are identically distributed.

It is often convenient to denote "$X$ has the same distribution as $Y$" by the notation "$X \stackrel{d}{=} Y$". Note that $\stackrel{d}{=}$ is an equivalence relation.

We first give characterizations of central symmetry of multivariate distributions. These will be used in later discussion of this section.

**Theorem 2.2.1** *The following statements are equivalent:*

(1) $X \in \mathbf{R}^d$ *is centrally symmetric about a point* $\theta \in \mathbf{R}^d$;

(2) $u'(X - \theta) \stackrel{d}{=} u'(\theta - X)$, *for any unit vector* $u$ *in* $\mathbf{R}^d$;

(3) $P(X - \theta \in H) = P(X - \theta \in -H)$, *for any closed halfspace* $H \in \mathbf{R}^d$.

PROOF: The equivalence of (2) and (3) is straightforward. The implication (1) $\Rightarrow$ (2) is immediate. Thus we need only show that (2) $\Rightarrow$ (1). Consider, for $t \in \mathbf{R}^d$, the characteristic functions $\psi_{X-\theta}(t)$ and $\psi_{\theta-X}(t)$ of $X-\theta$ and $\theta - X$. By (2) we deduce that

$$\psi_{X-\theta}(t) = E[e^{it'(X-\theta)}] = E[e^{it'(\theta-X)}] = \psi_{\theta-X}(t). \tag{2.1}$$

Thus, $X - \theta \stackrel{d}{=} \theta - X$. $\square$

Now we present the notion of angular symmetry of multivariate distributions, which was introduced by Liu (1988, 1990).

**Definition 2.2.2** A random vector $X \in \mathbf{R}^d$ is *angularly symmetric* about a point $\theta \in \mathbf{R}^d$ if and only if $(X-\theta)/\|X-\theta\|$ and $(\theta - X)/\|X-\theta\|$ are identically distributed, where $\|\cdot\|$ stands for Euclidean norm.

**Remarks 2.2.1** (1) It is easy to see that central symmetry implies angular symmetry, because $X \stackrel{d}{=} Y$ implies $f(X) \stackrel{d}{=} f(Y)$ for any measurable function $f$. Angular symmetry, however, does not necessarily imply central symmetry.
(2) The point (or the center) of angular symmetry of a random vector $X \in \mathbf{R}^d$, if it exists, is unique unless the distribution in $\mathbf{R}^d$ has all its probability mass on a line and its probability distribution on that line has more than one median. See Liu (1988, 1990) for proof, which is straightforward. We will assume in subsequent discussion that the point of angular symmetry (if any) of a multivariate distribution is unique unless stated otherwise. $\square$

**Definition 2.2.3** A set of $N$ vectors in Euclidean $\mathbf{R}^d$ is said to be *in general position* if every $d$-element subset is linearly independent, and a set of $N$ hyperplanes through the origin is said to be *in general position* if the corresponding set of normal vectors is in general position. The

$$2\sum_{i=0}^{d-1}\binom{N-1}{i} \tag{2.2}$$

regions generated by $N$ hyperplanes in general position through the origin are said to be (proper, nondegenerate) *convex cones* with the origin as vertices.

We now present characterizations of angular symmetry of multivariate distributions.

**Theorem 2.2.2** *The following statements are equivalent:*

(1) $X \in \mathbf{R}^d$ *is angularly symmetric about a point* $\theta \in \mathbf{R}^d$;

(2) $P(X - \theta \in C) = P(X - \theta \in -C)$, *for any circular cone* $C \in \mathbf{R}^d$ *with the origin as vertex*;

(3) $P(X - \theta \in C) = P(X - \theta \in -C)$, *for any convex cone* $C \in \mathbf{R}^d$ *with the origin as vertex;*

(4) $P(X - \theta \in H) = P(X - \theta \in -H)$, *for any closed halfspace* $H$ *with the origin on the boundary*;

(5) $P(u'(X - \theta) \geq 0) = P(u'(\theta - X) \geq 0)$, *for any unit vector* $u \in \mathbf{R}^d$.

PROOF: Assume, without loss of generality, that $\theta = 0$ and define $Y = \frac{X}{\|X\|}$ for $X \neq 0$ and $Y = 0$ for $X = 0$. Then, by definition, $X$ is angularly symmetric about the origin iff $Y$ is centrally symmetric about the origin.

(i) (1) $\Leftrightarrow$ (2). As just noted, $X$ is angularly symmetric about the origin iff $Y$ is centrally symmetric about the origin. By Theorem 2.2.1, $Y$ is centrally symmetric iff $P(Y \in H) = P(Y \in -H)$ for any closed halfspace $H$ in $\mathbf{R}^d$. Since $Y$ is distributed on the unit hypersphere $S^{d-1}$ and 0 in $\mathbf{R}^d$, $P(Y \in H) = P(Y \in -H)$ iff

$$P(Y \in H \cap S^{d-1}) = P(Y \in -H \cap S^{d-1}),$$

for any closed halfspace $H$ in $\mathbf{R}^d$. But $P(Y \in H \cap S^{d-1}) = P(Y \in -H \cap S^{d-1})$ for any closed halfspace $H$ in $\mathbf{R}^d$ iff $P(X \in C) = P(X \in -C)$ for any closed circular cone $C$ with the origin as vertex, since $Y \in H \cap S^{d-1}$ for some closed halfspace $H$ iff $X \in C$ for a corresponding closed circular cone $C$ with the origin as the vertex. Hence (1) $\Leftrightarrow$ (2).

(ii) (1) $\Leftrightarrow$ (3). Again we use the equivalence that $X$ is angularly symmetric about the origin iff $Y$ is centrally symmetric about the origin. Now $Y$ is centrally symmetric about the origin iff $P(Y \in B) = P(Y \in -B)$, for any Borel set $B$ in $\mathbf{R}^d$. Since $Y$ is distributed on the unit hypersphere $S^{d-1}$ and 0, $Y$ is centrally symmetric about the origin iff $P(Y \in B \cap S^{d-1}) = P(Y \in -B \cap S^{d-1})$ for any Borel set $B$ in $\mathbf{R}^d$. Since $X \in C$ for some convex cone $C$ with the origin as vertex iff $Y \in B \cap S^{d-1}$ for a corresponding1z Borel set $B$ in $\mathbf{R}^d$, $Y$ centrally symmetric about the origin implies that $P(X \in C) = P(X \in -C)$ for any closed convex cone $C$ with the origin as vertex, proving that (1) $\Rightarrow$ (3).

To prove (3) $\Rightarrow$ (1), note that if $P(X \in C) = P(X \in -C)$ for any closed convex cone $C$ with the origin as vertex, then

$$P(Y \in C \cap S^{d-1}) = P(Y \in -C \cap S^{d-1}).$$

Since any ellipsoid can be approximated by convex hulls, and since $P(A_n) \uparrow P(A)$ if $A_n \uparrow A$, we have that if

$$P(Y \in C \cap S^{d-1}) = P(Y \in -C \cap S^{d-1})$$

for any closed convex cone $C$ with the origin as the vertex, then $P(Y \in H \cap S^{d-1}) = P(Y \in -H \cap S^{d-1})$ for any closed halfspace $H \in \mathbf{R}^d$, that is, $Y$ is centrally symmetric about the origin, by Theorem 2.2.1. Hence (3) $\Rightarrow$ (1).

(iii) (1) $\Leftrightarrow$ (4). Now, from the proof of (1) $\Leftrightarrow$ (2), we have that $X$ is angularly symmetric about the origin iff

$$P(Y \in H \cap S^{d-1}) = P(Y \in -H \cap S^{d-1}),$$

for any closed halfspace $H \in \mathbf{R}^d$. Since $P(X \in H) = P(-X \in H)$ iff

$$P(Y \in H \cap S^{d-1}) = P(Y \in -H \cap S^{d-1}),$$

for any closed halfspace $H$ with the origin on the boundary, (1) $\Rightarrow$ (4).

To prove (4) $\Rightarrow$ (1), take $d = 2$ for the sake of simplicity. First we show that if $P(X \in H) = P(X \in -H)$ for any closed halfspace $H$ with the origin on the boundary, then

$$P(X \in H_1 \cap H_2) = P(X \in -H_1 \cap -H_2), \tag{2.3}$$

for any continuous or discrete random vector $X \in \mathbf{R}^d$ and any closed halfspaces $H_1$ and $H_2$ with the origin on their boundaries.

(a) Suppose that $X$ is continuous. In this case we have that

$$P(X \in H_1 \cap H_2) + P(X \in (H_1 \cap -H_2))$$

$$= \quad P(X \in -H_1 \cap -H_2) + P(X \in (H_1 \cap -H_2))$$

for any closed halfspaces $H_1$ and $H_2$ with the origin on the boundaries. Thus (2.3) holds.

(b) Suppose that $X$ is discrete. Assume (2.3) is violated. Then there exist closed halfspaces $H_1$ and $H_2$ with the origin on their boundaries and $\epsilon > 0$ such that

$$P(X \in H_1 \cap H_2) > P(X \in -H_1 \cap -H_2) + \epsilon. \tag{2.4}$$

Now it is not very difficult to see that there are closed halfspaces $H^3$ and $H^4$ through the origin such that $P(X \in H^3) = P(X \in H^4)$, and

$$\begin{aligned} P(X \in H^3) &\geq P(X \in H_1 \cap H_2) + P(X \in (H^1 \cap -H^2)^o) - \frac{\epsilon}{2}, \\ P(X \in H^4) &\leq P(\in -H_1 \cap -H_2) + P(X \in (H^1 \cap -H^2)^o) + \frac{\epsilon}{2}, \end{aligned}$$

where $A^o$ denotes the interior of set $A \subset \mathbf{R}^d$. The above two inequalities imply that $P(X \in H_1 \cap H_2) < P(X \in -H_1 \cap -H_2) + \epsilon$, contradicting (2.4). Hence (2.3) holds.

Now by (2.3), we have that

$$P(Y \in (H_1 \cap H_2) \cap S^1) = P(Y \in (-H_1 \cap -H_2) \cap S^1),$$

for any closed halfspace $H_1$ and $H_2$ with the origin on the boundary. Thus $P(Y \in C) = P(Y \in -C)$, for any closed arc $C$ on the unit circle. But the set of all closed arcs on the unit circle forms a $\pi$-*system* (see Billingsley (1986) p. 34 for the definition of $\pi$-system) and generates all Borel sets on the unit circle. By the unique extension theorem of $\pi$-system (Theorem 3.3 of Billingsley (1986)), $Y \stackrel{d}{=} -Y$ on the unit circle. Hence $X$ is angularly symmetric about the origin, proving that (4) $\Rightarrow$ (1).

(iv) (4) $\Leftrightarrow$ (5). This is straightforward. $\square$

For any random variable $Y \in \mathbf{R}$, define a *median* of its distribution to be a number $c$ such that

$$P(Y \le c) \ge \frac{1}{2}, \quad P(Y \ge c) \ge \frac{1}{2}.$$

Therefore, based on the Theorem 2.2.2, we obtain

**Corollary 2.2.1** *Suppose that the random vector $X \in \mathbf{R}^d$ is angularly symmetric about a point $\theta \in \mathbf{R}^d$. Then*

(1) $P(X \in H + \theta) = P(X \in -H + \theta) \ge \frac{1}{2}$ *for any closed halfspace $H$ with the origin on the boundary, and furthermore, if $X$ is continuous, then $P(X \in H + \theta) = \frac{1}{2}$;*

(2) $\mathrm{Med}(u'X) = u'\theta$ *for any unit vector $u \in \mathbf{R}^d$.*

PROOF: (1) By Theorem 2.2.2, we immediately have that $P(X \in H + \theta) = P(X \in -H + \theta)$ for any closed halfspace $H$ with the origin on the boundary. Since $P(X \in H + \theta) + P(X \in -H + \theta) \ge 1$, it follows that $P(X \in H + \theta) = P(X \in -H + \theta) \ge \frac{1}{2}$ and $P(X \in H + \theta) = \frac{1}{2}$ if $X$ is continuous.

(2) By Theorem 2.2.2, we obtain that $P(u'(X - \theta) \ge 0) = P(u'(\theta - X) \ge 0)$ for any unit vector $u \in \mathbf{R}^d$, which immediately implies that $\mathrm{Med}(u'X) = u'\theta$ for any unit vector $u \in \mathbf{R}^d$.

The proof is complete. $\square$

**Remark 2.2.1** The converse of (2) in Corollary 2.2.1 does not hold. That is, the condition that $\mathrm{Med}(u'X) = u'\theta$ for any unit vector $u \in \mathbf{R}^d$ does not necessarily imply that $X$ is angularly symmetric about $\theta$.

**Counterexamples**: (1) Let $P(X = 0) = \frac{1}{2}$, $P(X = -1) = \frac{1}{3}$, and $P(X = 1) = \frac{1}{6}$. Then $\mathrm{Med}(X) = 0$. By Theorem 2.2.2, however, it is easy to see that $X$ is not angularly symmetric about 0.

(2) Let $P(X = (0,0)) = \frac{2}{5}$, $P(X = (1, \pm 1)) = \frac{1}{5}$ and $P(X = (-1, \pm 1)) = \frac{1}{10}$. Then, it is not difficult to verify that $\mathrm{Med}(u'X) = u'(0,0)$ for any $u \in \mathbf{R}^2$. However, by Theorem 2.2.2, it is easy to see that $X$ is not angularly symmetric about (0,0). □

In the discrete case, we have a simple characterization of angular symmetry, which could be very useful in practice for the purpose of checking whether a given multivariate distribution is angularly symmetric.

**Theorem 2.2.3** *Suppose $X \in \mathbf{R}^d$ is discrete. Then the following statements are equivalent:*

(1) *$X \in \mathbf{R}^d$ is angularly symmetric about a point $\theta \in \mathbf{R}^d$;*

(2) *$P(X - \theta \in L) = P(X - \theta \in -L)$ for any ray $L$ passing through the origin.*

PROOF: (1) $\Rightarrow$ (2). By Theorem 2.2.2, we have that $P(X-\theta \in C) = P(X-\theta \in -C)$ for any circular cone $C$ with the origin as vertex. Consider the limit situation as the cone shrinks to a ray; we thus obtain (2). Hence (1) $\Rightarrow$ (2).

To show that (2) $\Rightarrow$ (1), note that (2) implies that $P(X - \theta \in C) = P(X - \theta \in -C)$ for any circular cone $C$ with the origin as vertex, since

$$P(X - \theta \in C) = \sum_i P(X \in l^i + \theta)$$

and

$$P(X - \theta \in -C) = \sum_i P(X \in -l^i + \theta)$$

for some rays $\{l^i\}$ (at most countably many) passing through the origin. By (2) of Theorem 2.2.2, we obtain that (2) $\Rightarrow$ (1). $\square$

### 2.3 Halfspace Symmetry

Remark 2.2.1 indicates that a point $\theta$, which is the median of the projected distribution of an underlying distribution in any direction (i.e., $\text{Med}(u'X) = u'\theta$ for a random vector $X$ and any unit vector $u$ in $\mathbf{R}^d$), is not necessarily the point of angular symmetry of the underlying distribution. On the other hand, it is clear that such a point $\theta$ is indeed a point of symmetry relative to an extended notion of symmetry defined as follows.

**Definition 2.3.1** A random vector $X \in \mathbf{R}^d$ is *halfspace symmetric* about a point $\theta \in \mathbf{R}^d$ if and only if $P(X \in H_\theta) \geq \frac{1}{2}$ for any closed halfspace $H_\theta$ with $\theta$ on the boundary.

**Remarks 2.3.1** (1) By Corollary 2.2.1, it is easy to see that angular symmetry implies halfspace symmetry. However, the converse does not hold. See the counterexamples in Remark 2.2.1 and also the following.

**Counterexample** Let $P(X = (0.2, 0.1)) = \frac{2}{5}$, $P(X = (1, \pm 1)) = \frac{1}{10}$ and $P(X = (-1, \pm 1)) = \frac{1}{5}$. Then it is easy to check that $X$ is halfspace symmetric about the point $(0.2, 0.1) \in \mathbf{R}^2$, but $X$ is not angularly symmetric about any point in $\mathbf{R}^2$.

(2) There are multivariate distributions which are not halfspace symmetric about any

point in $\mathbf{R}^d$. For example, let $P(X = (\pm 1, 0)) = \frac{1}{3}$ and $P(X = (0, 1)) = \frac{1}{3}$. Then it is easy to see that $X$ is not halfspace symmetric about any point in $\mathbf{R}^2$. $\square$

The uniqueness result about the point of angular symmetry of multivariate distributions also holds true for halfspace symmetry.

**Theorem 2.3.1** *The point* (*or the center*) *of halfspace symmetry $\theta$ of the distribution of a random vector $X \in \mathbf{R}^d$, if it exists, is unique unless the distribution of $X \in \mathbf{R}^d$ has all its probability mass on a line and its probability distribution on that line has more than one median.*

PROOF: Suppose that $\theta_1$ and $\theta_2$ are two points of symmetry of the halfspace symmetric distribution of a random vector $X$. Let $H_{\theta_1}$ be an arbitrary closed halfspace with $\theta_1$ on the boundary and containing $\theta_2$, and $H_{\theta_2}$ be the closed halfspace with $\theta_2$ on the boundary and containing $\theta_1$ such that its boundary is parallel to the boundary of $H_{\theta_1}$. Then, by the definition of halfspace symmetry, $P(X \in (H_{\theta_1} \cap H_{\theta_2})^o) = 0$. Thus $P(X \in \mathbf{R}^d) = P(X \in L_{\theta_1\theta_2})$, where $L_{\theta_1\theta_2}$ is the line in $\mathbf{R}^d$ passing through $\theta_1$ and $\theta_2$. Clearly, $\theta_1$ and $\theta_2$ are two medians of the distribution of $X$ on $L_{\theta_1\theta_2}$. The proof is complete. $\square$

We will assume in the subsequent discussion that the point of halfspace symmetry of multivariate distributions is unique (if any).

Denote a closed halfspace with $\theta$ on its boundary by $H_\theta$, and its reflection about $\theta$ by $\tilde{H}_\theta$. Similarly, denote a ray stemming from $\theta$ by $L_\theta$, and its reflection about $\theta$ by $\tilde{L}_\theta$. Now we present several characterizations of halfspace symmetry.

**Theorem 2.3.2** *The following statements are equivalent:*

(1) $X \in \mathbf{R}^d$ *is halfspace symmetric about a point* $\theta \in \mathbf{R}^d$;

(2) $\mathrm{Med}(u'X) = u'\theta$ *for any unit vector* $u \in \mathbf{R}^d$;

(3) $P(u'(X-\theta) \geq 0) \geq \frac{1}{2}$ *for any unit vector* $u \in \mathbf{R}^d$;

(4) $P(X \in (H_\theta^1 \cap H_\theta^2)^o) + P(X \in \pi^1 \cap \pi^2) \geq P(X \in (\tilde{H}_\theta^1 \cap \tilde{H}_\theta^2)^o)$ *for any closed halfspace* $H_\theta^i \subset \mathbf{R}^d$ *with* $\theta$ *on its boundary* $\pi^i$ $(i = 1, 2)$;

(5) $P(X \in (H_\theta^1 \cap H_\theta^2)) + P(X \in \pi^1 \cap \pi^2) \geq P(X \in (\tilde{H}_\theta^1 \cap \tilde{H}_\theta^2))$ *for any closed halfspace* $H_\theta^i \subset \mathbf{R}^d$ *with* $\theta \in \mathbf{R}^d$ *on its boundary* $\pi^i$ $(i = 1, 2)$.

PROOF: (i) (1) $\Leftrightarrow$ (3). $X$ is halfspace symmetric about $\theta$ iff $P(X \in H + \theta) \geq \frac{1}{2}$ for any closed halfspace $H \in \mathbf{R}^d$ with the origin on its boundary. But $X \in (H + \theta)$ for some closed halfspace $H \in \mathbf{R}^d$ with the origin on its boundary iff $u'(X - \theta) \geq 0$ for a corresponding unit vector $u \in \mathbf{R}^d$ which is a normal vector of $H$. Thus $X$ is halfspace symmetric about $\theta$ iff $P(u'(X-\theta) \geq 0) \geq \frac{1}{2}$ for any unit vector $u \in \mathbf{R}^d$. Thus (1) $\Leftrightarrow$ (3).

(ii) (2) $\Leftrightarrow$ (3). This is trivial.

(iii) (1) $\Leftrightarrow$ (4). Suppose that $X$ is halfspace symmetric about $\theta \in \mathbf{R}^d$. Let $H_\theta^i$ be a closed halfspace with $\theta$ on its boundary $\pi^i$ $(i = 1, 2)$. It is not difficult to see that there are closed halfspaces $H_\theta^3$ and $H_\theta^4$ such that

$$(H_\theta^1 \cap \pi^2) \cup (\tilde{H}_\theta^2 \cap \pi^1) \subset H_\theta^i \qquad \text{for } (i = 3, 4),$$

and for any $\epsilon > 0$

$$P(X \in H_\theta^4) \;\; \leq \;\; P(X \in (H_\theta^1 \cap H_\theta^2)^o) + P(X \in H_\theta^1 \cap \tilde{H}_\theta^2) + \frac{\epsilon}{2},$$

$$\begin{aligned} P(X \in (H_\theta^3)^o) &\geq P(X \in (\tilde{H}_\theta^1 \cap \tilde{H}_\theta^2)^o) + P(X \in H_\theta^1 \cap \tilde{H}_\theta^2) \\ &\quad - P(X \in \pi^1 \cap \pi^2) - \frac{\epsilon}{2}. \end{aligned}$$

Since $X$ is halfspace symmetric about $\theta \in \mathbf{R}^d$, we have

$$P(X \in (H_\theta^3)^o) \leq \frac{1}{2} \leq P(X \in H_\theta^4).$$

Thus

$$P(X \in (H_\theta^1 \cap H_\theta^2)^o) + P(X \in \pi^1 \cap \pi^2) \geq P(X \in (\tilde{H}_\theta^1 \cap \tilde{H}_\theta^2)^o) - \epsilon,$$

for any $\epsilon > 0$. Since $\epsilon$ is arbitrary, (1) $\Rightarrow$ (4) now follows.

To show (4) $\Rightarrow$ (1), take $H_\theta^1 = H_\theta^2$. Then

$$P(X \in (H_\theta^1)^o) + P(X \in \partial(H_\theta^1)) \geq P(X \in (\tilde{H}_\theta^1)^o).$$

Thus, $P(X \in H_\theta^1) \geq \frac{1}{2}$ for any closed halfspace $H_\theta^1$ with $\theta \in \mathbf{R}^d$ on its boundary, proving that $X$ is halfspace symmetric about $\theta$.

(iv) (1) $\Leftrightarrow$ (5). Similar to the proof of (1) $\Leftrightarrow$ (4). □

**Remarks 2.3.2** (1) Statement (2) of Theorem 2.3.2 indicates that it is reasonable to treat the point of halfspace symmetry as the "center" or the "multidimensional median" of any halfspace symmetric multivariate distribution.

(2) Thus halfspace symmetry, as a broader relaxation of the normality assumption on multivariate distributions than central symmetry and angular symmetry, is a reasonable symmetry assumption on multivariate distributions in nonparametric multivariate location inference and related statistical procedures.

(3) It is then desirable that any notion of multidimensional median resulting from a notion of multidimensional symmetry should agree with the point of halfspace symmetry when the underlying distributions are halfspace symmetric. □

When $d \leq 2$, Theorem 2.3.2 yields the following result, which could be useful in practice.

**Corollary 2.3.1** *For $d \leq 2$, $X \in \mathbf{R}^d$ is halfspace symmetric about a unique point $\theta \in \mathbf{R}^d$ if and only if*

$$P(X \in H_\theta^1 \cap H_\theta^2) + P(X = \theta) \geq P(X \in \tilde{H}_\theta^1 \cap \tilde{H}_\theta^2),$$

*for any closed halfspaces $H_\theta^1$ and $H_\theta^2$ in $\mathbf{R}^d$ with $\theta$ on their boundaries.*

PROOF: This follows in straightforward fashion from (5) of Theorem 2.3.2, because $\pi^1 \cap \pi^2$ in Theorem 2.3.2 now equals $\theta$. □

When $X$ is discrete, the following necessary conditions could be utilized to check halfspace symmetry of the underlying distribution.

**Theorem 2.3.3** *Suppose that $X \in \mathbf{R}^d$ is discrete and halfspace symmetric about a unique point $\theta \in \mathbf{R}^d$. Then*

(1) $P(X \in H_\theta^o) + P(X = \theta) \geq P\left(X \in (\tilde{H}_\theta)^o\right)$ *for any closed halfspace $H_\theta \subset \mathbf{R}^d$ with $\theta$ on the boundary;*

(2) $P(X \in L_\theta) + P(X = \theta) \geq P(X \in \tilde{L}_\theta)$ *for any ray $L_\theta \subset \mathbf{R}^d$.*

PROOF: The proof of the case $d = 1$ is trivial. Now we consider the case $d \geq 2$.

(1) Let $H_\theta \subset \mathbf{R}^d$ be a closed halfspace with $\theta \in \mathbf{R}^d$ on its boundary (hyperplane $\pi$).

Since $X$ is discrete, for any small $\epsilon > 0$, it is not difficult to see that there is a closed halfspace $H_\theta^1 \subset \mathbf{R}^d$ with a boundary hyperplane $\pi_1$ such that $P(X \in \pi \cap \pi_1) = P(X = \theta)$ and

$$P(X \in H_\theta^1) \leq P(X \in H_\theta^o) + P(X = \theta) + P(X \in (H_\theta^1 \cap \pi - \{\theta\})) + \epsilon,$$

and a closed halfspace $H_\theta^2 \subset \mathbf{R}^d$ with a boundary hyperplane $\pi_2$ such that $\pi \cap \pi_2 = \pi \cap \pi_1$ and

$$P(X \in (H_\theta^2)^o) \geq P(X \in (\tilde{H}_\theta)^o) + P(X \in (H_\theta^1 \cap \pi - \{\theta\})) - \epsilon.$$

Since $X$ is halfspace symmetric about $\theta$, $P(X \in H_\theta^1) \geq \frac{1}{2}$ and $P(X \in (H_\theta^2)^o) \leq \frac{1}{2}$, thus

$$P(X \in H_\theta^o) + P(X = \theta) \geq P(X \in (\tilde{H}_\theta)^o) - 2\epsilon.$$

Since the above inequality holds true for any sufficiently small $\epsilon$, therefore

$$P(X \in H_\theta^o) + P(X = \theta) \geq P(X \in (\tilde{H}_\theta)^o),$$

for any closed halfspace $H_\theta \subset \mathbf{R}^d$ with $\theta$ on the boundary.

(2) Let $L_\theta \subset \mathbf{R}^d$ be a ray stemming from $\theta$ with $\tilde{L}_\theta$ as its reflection about $\theta$. Since $X$ is discrete, it is not difficult to see that there is a closed halfspace $H_\theta \subset \mathbf{R}^d$ with a boundary (hyperplane $\pi$) containing the ray $L_\theta$ such that

$$P(X \in \pi) = P(X \in L_\theta).$$

For any $\epsilon > 0$, it is not difficult to see that there is a closed halfspace $H_\theta^1 \subset \mathbf{R}^d$ containing $L_\theta$ and with $\theta$ on its boundary such that

$$P(X \in H_\theta^1) \leq P(X \in L_\theta) + P(X \in H_\theta^o) + \epsilon,$$

and a closed halfspace $H_\theta^2 \subset \mathbf{R}^d$ containing $\tilde{L}_\theta$ and with $\theta$ on its boundary such that

$$P(X \in (H_\theta^2)^o) \geq P(X \in \tilde{L}_\theta) - P(X = \theta) + P(X \in H_\theta^o) - \epsilon.$$

Since $X$ is halfspace symmetric about $\theta$, $P(X \in H_\theta^1) \geq \frac{1}{2}$ and $P(X \in (H_\theta^2)^o) \leq \frac{1}{2}$, and thus

$$P(X \in L_\theta) + P(X = \theta) \geq P(X \in \tilde{L}_\theta) - 2\epsilon.$$

Hence

$$P(X \in L_\theta) + P(X = \theta) \geq P(X \in \tilde{L}_\theta),$$

for any ray $L_\theta \subset \mathbf{R}^d$ stemming from $\theta$. $\square$

Remarks 2.3.1 have shown that angular symmetry implies halfspace symmetry, and that the converse does not hold. A natural question now is: Under what conditions does halfspace symmetry imply angular symmetry? The answer is given in the following result.

**Theorem 2.3.4** *Suppose a random vector $X$ is halfspace symmetric about a point $\theta \in \mathbf{R}^d$, and either*

(1) *$X$ is continuous, or*

(2) *$X$ is discrete and $P(X = \theta) = 0$.*

*Then $X$ is angularly symmetric about $\theta$.*

PROOF: (1) By Definition 2.3.1 and the continuity of $X$, we have that

$$P(X \in H + \theta) = P(X \in -H + \theta) = \frac{1}{2},$$

for any closed halfspace $H$ with the origin on the boundary. Angular symmetry of $X$ now follows from Theorem 2.2.2.

(2) Suppose that $X$ is not angularly symmetric about the point $\theta \in \mathbf{R}^d$. By Theorem 2.2.3, there is a line $L$ through $\theta$ with rays $L_\theta^+$ and $L_\theta^-$ on either side of $\theta$, such that $P(X \in L_\theta^+) \neq P(X \in L_\theta^-)$. Assume, without loss of generality, that $P(X \in L_\theta^+) > P(X \in L_\theta^-) + \epsilon$ for some $\epsilon > 0$. Since $X$ is discrete, it is clear that there is a closed halfspace $H_\theta$ containing $L$ such that $P(X \in \partial(H_\theta)) = P(X \in L)$. Also, there exists $H_\theta^+ \subset \mathbf{R}^d$, a closed halfspace which just includes $L_\theta^+$ and excludes $L_\theta^-$, such that $P(X \in \partial(H_\theta^+)) = 0$ and

$$P(X \in H_\theta^+) > P(X \in (H_\theta)^o) + P(X \in L_\theta^+) - \frac{\epsilon}{2}.$$

Similarly, there exists $H_\theta^- \subset \mathbf{R}^d$, a closed halfspace which just includes $L_\theta^-$ and excludes $L_\theta^+$, such that $P(X \in \partial(H_\theta^-)) = 0$ and

$$P(X \in H_\theta^-) < P(X \in (H_\theta)^o) + P(X \in L_\theta^-) + \frac{\epsilon}{2}.$$

By Definition 2.3.1 and the properties of $H_\theta^+$ and $H_\theta^-$,

$$P(X \in H_\theta^+) = P(X \in H_\theta^-) = \frac{1}{2}.$$

Hence

$$\begin{aligned} P(X \in (H_\theta)^o) + P(X \in L_\theta^+) - \frac{\epsilon}{2} &< \frac{1}{2} \\ &< P(X \in (H_\theta)^o) + P(X \in L_\theta^-) + \frac{\epsilon}{2}, \end{aligned}$$

which contradicts the assumption about $P(X \in L_\theta^+)$ and $P(X \in L_\theta^-)$. Thus $P(X \in L_\theta^+) = P(X \in L_\theta^-)$ for any line $L$ through $\theta$ with rays $L_\theta^+$ and $L_\theta^-$ on either side of $\theta$. Angular symmetry of $X$ about $\theta$ now follows from Theorem 2.2.3. □

**Remark 2.3.1** Halfspace symmetry and angular symmetry thus coincide under some typical conditions. What, then, is the point of introducing the new notion, halfspace symmetry? The relevance of halfspace symmetry is based on the following points:

(1) Actual distributions are invariably discrete.

(2) In the discrete case, the center $\theta$ typically can be anticipated to carry some probability mass.

(3) Actual distributions typically are not symmetric in any sense. Halfspace symmetry, as the weakest among existing notions of symmetry, provides a more general foundation for approximating actual distributions by symmetric ones.

(4) Halfspace symmetry is useful in the performance evaluation of depth functionals and multivariate location measures, which will be introduced and discussed in later chapters.

□

## 2.4 Summary

In this chapter, a new notion of multivariate symmetry, halfspace symmetry, has been introduced. Characterizations of it as well as other existing notions of multivariate symmetry, and interrelationships among these notions have been studied. Halfspace symmetry not only supports more general approximations to actual distributions in modeling, but also plays important roles in discussions of later chapters.

# Chapter 3

# STATISTICAL DEPTH FUNCTIONS

## 3.1 Introduction

By assigning to each point $x \in \mathbf{R}^d$ a value based on a suitable unimodal function constructed with respect to a given underlying dataset, one can obtain a center-outward ordering of all points in $\mathbf{R}^d$. The value assigned to $x$ by this "depth function" is called the "depth" of $x$ with respect to the underlying dataset. Such a depth function allows one to define reasonable analogues of univariate order statistics and rank statistics, which then may lead to generalizations of classical univariate *L-statistics* and *R-statistics* in the multivariate setting. (See Serfling (1980) for discussion of classical *L-statistics* and *R-statistics*.) Consequently, depth functions can play key roles in multivariate nonparametric robust statistical procedures, multivariate outlier detection, testing of symmetry for multivariate distributions, etc..

Tukey (1974, 1977) introduced the first notion of statistical depth of a point in a multivariate dataset, as follows. In a one-dimensional dataset $X = \{X_1, X_2, \ldots, X_n\}$, the depth of a point $x$ is the minimum number of data points lying on either side of $x$. For a $d$-dimensional dataset, the depth of a point $x \in \mathbf{R}^d$ is the smallest depth of $x$ in any one-dimensional projection of the dataset. That is,

$$\begin{aligned} HD_n(x; P_n) &= \inf_{\|u\|=1} \sum_{i=1}^{n} \mathbf{I}\{u^T X_i \le u^T x\} \\ &= \inf_{H}\{P_n(H) \mid H \text{ is a closed halfspace} \\ &\qquad\qquad \text{with } x \text{ on the boundary}\}, \end{aligned} \tag{3.1}$$

where "$\mathbf{I}$" is an indicator function, $u$ is any unit vector in $\mathbf{R}^d$, $P_n$ is the empirical measure of the underlying probability measure $P$, defined as $P_n(H) \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\{X_i \in H\}$ and "*HD*" stands for "halfspace depth".

The population version of the halfspace depth (*HD*), with respect to a underlying probability measure $P$ on $\mathbf{R}^d$, can be defined as follows:

$$\begin{aligned} HD(x;\ P) &= \inf_{H}\{P(H) \mid H \text{ is a closed halfspace} \\ &\qquad \text{with } x \text{ on the boundary}\}, \quad x \in \mathbf{R}^d. \end{aligned} \tag{3.2}$$

Liu (1988, 1990) introduced a "simplicial" depth (*SD*) function on $\mathbf{R}^d$ with respect to a underlying probability measure $P$. Namely,

$$SD(x;\ P) = P(x \in S[X_1, \ldots, X_{d+1}]) \tag{3.3}$$

where $X_1, \ldots, X_{d+1}$ is a random sample from $P$ and $S[X_1, \ldots, X_{d+1}]$ is the $d$-dimensional simplex with vertices $X_1, \ldots, X_{d+1}$. The sample version of $SD(x; P)$ is defined as follows.

$$SD_n(x; P) = \frac{1}{\binom{n}{d+1}} \sum_{1 \le i_1 < \ldots < i_{d+1} \le n} \mathbf{I}(x \in S[X_{i_1}, \ldots, X_{i_{d+1}}]) \tag{3.4}$$

where $X_1, \ldots, X_n$ is a random sample from $P$. Note that, for each fixed $x$, $SD_n(x; P)$ is a *U-statistic*.

In this chapter, properties which a depth function should possess are first explored and then a general definition of statistical depth function is presented, which unifies various *ad hoc* definitions of data depth. Surprisingly, the simplicial depth of Liu (1988, 1990, 1993), which has received considerable study and attention in recent years, fails to satisfy some basic properties desired of statistical depth functions. A special type of depth function, called *E-depth* function, is introduced and examples are given. Depth "contours" and multivariate "quantile contours" are also defined and their properties and applications discussed.

### 3.2 Statistical Depth Functions

Before giving a general definition of statistical depth function, we first examine several properties that are desirable for depth functions. In the following we consider depth functions on $\mathbf{R}^d$ defined with respect to distributions that may be continuous or discrete.

- $(P1)$ **Affine Invariance** The depth of a point $x \in \mathbf{R}^d$ should not depend on the underlying coordinate system and the scales of the underlying measurements.

- $(P2)$ **Maximality at Center** For a distribution having a uniquely defined "center" (e.g., the point of symmetry with respect to some notion of symmetry, as discussed in Chapter 2), the depth function should attain maximum value at this center.

- ($P3$) **Monotonicity Relative to the Deepest Point** As a point $x \in \mathbf{R}^d$ moves away from the deepest point (the point at which the depth function attains maximum value, for a symmetric distribution, the center should be the deepest point) along with any fixed ray through the center, the depth function should decrease monotonically and approach zero as $\|x\|$ approaches infinity.

Besides the above three properties, we will also confine attention to depth functions that are nonnegative and bounded. Any functions possessing these properties may be regarded as a measure of depth of points with respect to the underlying distribution.

**Definition 3.2.1** *Let $P$ be a distribution function on $\mathbf{R}^d$. Let the mapping $D(\cdot\,; P)$ : $\mathbf{R}^d \to \mathbf{R}^1$ be bounded, nonnegative and satisfy ($P1$), ($P2$) and ($P3$). That is,*

*(1)* $D(x; P) = D(Ax + b; P_{AX+b})$.

*Further, if $P$ has a center $\theta$,*

*(2)* $D(\theta; P) = \sup_{x \in \mathbf{R}^d} D(x; P)$.

*Further, if (2) holds for any $\theta$, then*

*(3)* $D(x; P) \leq D(\theta + \alpha(x - \theta); P)$, *and* $D(x; P) \to 0$ *as* $\|x\| \to \infty$.

*Here $A$ is any $d \times d$ nonsingular matrix, $x$ and $b$ are any vectors in $\mathbf{R}^d$, $\alpha \in [0, 1]$, $X$ has distribution $P$, and $P_{AX+b}$ is the distribution function of $AX + b$. Then $D(\cdot\,; P)$ is said to be a statistical depth function with respect to $P$, i.e., $D(\cdot; P)$ measures the depth of $x$ with pespect to $P$, $x \in \mathbf{R}^d$.*

**Remark 3.2.1** A sample version of $D(x; P)$, denoted by $D_n(x) \equiv D(x; \hat{P}_n)$, may be defined by replacing $P$ by the empirical measure $\hat{P}_n$.

Let us now investigate whether Tukey's halfspace depth function *HD(x; P)* and Liu's simplicial depth function *SD(x; P)* are really statistical depth functions according to the above definition.

**Theorem 3.2.1** *The halfspace depth function HD(x; P) is a statistical depth function in the sense of Definition 3.2.1.*

PROOF: Clearly, *HD(x; P)* is bounded and nonnegative. We only need to check $(P1)$, $(P2)$ and $(P3)$.

(1). Affine Invariance. For any $d \times d$ nonsingular matrix $A$, vector $b$, and closed halfspace $H$,

$$\begin{aligned} X \in H &\iff AX + b \in AH + b, \\ x \in \partial H &\iff Ax + b \in \partial(AH + b), \end{aligned}$$

where $\partial H$ denotes the boundary of $H$.

Thus

$$\begin{aligned} HD(x; P_X) &= \inf_H \{P(X \in H) \mid H \text{ is a closed halfspace, } x \in \partial H\} \\ &= \inf_H \{P(AX + b \in AH + b) \mid \ AH + b \text{ is a closed halfspace,} \\ &\qquad Ax + b \in \partial(AH + b)\} \\ &= HD(Ax + b;\ P_{AX+b}). \end{aligned}$$

(2). Maximality at Center.

**(i)** Suppose that $P$ is *HS*-symmetric about a unique point $\theta \in \mathbf{R}^d$. Then by the definition of *HS*-symmetry, we have that

$$P(H_\theta) \geq \frac{1}{2}$$

for any closed halfspace $H$ with $\theta \in \partial H$ (such an $H$ is always denoted by $H_\theta$), and thus

$$HD(\theta; P) \geq \frac{1}{2}.$$

Now suppose that there is a point $x_0 \in \mathbf{R}^d$, $x_0 \neq \theta$, such that

$$HD(x_0; P) > \frac{1}{2}.$$

Then

$$P(H_{x_0}) > \frac{1}{2}$$

for any closed halfspace $H$ with $x_0 \in \partial H$. Hence, according to the definition of *HS*-symmetry, $P$ is also *HS*-symmetric about $x_0$, which contradicts the assumption that $P$ is *HS*-symmetric about a unique point $\theta \in \mathbf{R}^d$. Therefore, we have

$$HD(\theta;\ P) = \sup_{x \in \mathbf{R}^d} HD(x; P).$$

**(ii)** Suppose that $P$ is not *HS*-symmetric about any point. Then we simply define the center of the distribution to be the point where $D(x; P)$ attains maximum value.

(3). Monotonicity Relative to the Deepest Point.

**(i)** Suppose $\theta$ is the deepest point with respect to the underlying distribution, that is, $HD(\theta; P) = \sup_{x \in \mathbf{R}^d} HD(x; P)$. suppose $0 < \alpha < 1$. We are going to show that

$$HD(\alpha x; P) \leq HD(\theta + \alpha(x - \theta); P), \quad \forall \alpha \in (0, 1).$$

To compare $HD(x;P)$ and $HD(\theta+\alpha(x-\theta);P)$ we need only to consider the infimum over all closed halfspaces which do not contain $\theta$, since $HD(\theta;P) = \sup_{x\in\mathbf{R}^d} HD(x;P)$. For any $H_{\theta+\alpha(x-\theta)}$ (closed halfspace with $(\theta+\alpha(x-\theta)) \in \partial H$), by the separating hyperplane theorem there exists a closed halfspace $H_x$ such that

$$H_x \subset H_{\theta+\alpha(x-\theta)}.$$

Thus we have that

$$HD(x;P) \leq HD(\theta+\alpha(x-\theta);P), \quad \forall \alpha \in (0,1).$$

**(ii)** It is obvious that

$$P(\|X\| \geq \|x\|) \to 0 \qquad \text{as } \|x\| \to \infty,$$

and that for each $x$ there exists a closed halfspace $H_x$ such that

$$H_x \subset \{\|X\| \geq \|x\|\}$$

Thus

$$HD(x;P) \to 0 \qquad \text{as } \|x\| \to \infty.$$

We are done. $\square$

**Remark 3.2.2** For continuous angularly symmetric distributions, the simplicial depth function $SD(\cdot\,;P)$ is a statistical depth function in the sense of Definition 3.2.1. This is proved by Liu (1990). For discrete distributions, however, $SD(x;P)$ does not always possess the monotonicity property and also can fail to satisfy maximality at center property with respect to $HS$-symmetric distributions.

**Counterexamples.**

1. Take d = 1. Let $P(X=0)=\frac{1}{5}$, $P(X=\pm 1)=\frac{1}{5}$, $P(X=\pm 2)=\frac{1}{5}$. Then clearly X is centrally symmetric about 0, and

$$\begin{aligned} SD(\frac{1}{2};\ P) &= P(\frac{1}{2}\in \overline{X_1X_2}) \\ &= 2P(X_1\le \frac{1}{2}\le X_2) \\ &= 2P(X_1\le \frac{1}{2})P(X_2\ge \frac{1}{2}) \\ &= 2\cdot \frac{3}{5}\cdot \frac{2}{5} \\ &= \frac{12}{25}, \end{aligned}$$

where $X_1,\ X_2$ is a random sample from $X$. Similarly we have that

$$\begin{aligned} SD(1;\ P) &= P(1\in \overline{X_1X_2}) \\ &= 2P(X_1\le 1)P(X_2\ge 1) \\ &= 2\cdot \frac{4}{5}\cdot \frac{2}{5} \\ &= \frac{16}{25}. \end{aligned}$$

Since $SD(1;P) > SD(\frac{1}{2};P)$, the monotonicity property fails to hold.

2. Let d = 2. Let $P(X=(\pm 1,0))=\frac{1}{6}$, $P(X=(\pm 2,0))=\frac{1}{6}$, $P(X=(0,\pm 1))=\frac{1}{6}$. Then it is not difficult to see that $X$ is centrally symmetric about (0, 0), and

$$\begin{aligned} SD((1,0);P)\ -\ SD((\frac{1}{2},0);P) &= 3!\cdot 2\cdot \frac{1}{6}\cdot \frac{1}{6}\cdot \frac{1}{6} \\ &= \frac{1}{18} \\ &> 0, \end{aligned}$$

which also violates the monotonicity property.

3. Let d = 2. Let $P(X = \theta(0,0)) = \frac{19}{40}$, $P(X = A(-1,1)) = \frac{3}{40}$, $P(X = B(-1,-1)) = \frac{1}{40}$, $P(X = C(1,0)) = \frac{1}{40}$. Let $B\theta$ intersect $AC$ at $D$, $x$ be a point inside the triangle $\triangle A \theta D$, and $P(X = x) = \frac{16}{40}$. Then it is not difficult to verify based on the results established in Chapter 2 that $X$ is $HS$-symmetric about $\theta$, thus $\theta$ is the center of the distribution. However, we have

$$\begin{aligned} SD(x;P) - SD(\theta;P) &= \frac{3!}{40^3}(2 \times 16 \times 1 \times 3 - (3 \times 1 \times 19 + 1 \times 1 \times 19)) \\ &= \frac{3!}{40^3}(86 - 76) \\ &> 0, \end{aligned}$$

that is, the maximality at center property fails to hold. □

## 3.3 Some General Structures for Depth Functions

Four types of general structures that could be used to construct statistical depth functions with respect to any given distributions on $\mathbf{R}^d$ will be described here.

### 3.3.1 Type-A depth functions

Let $X_1, \ldots, X_r$ be independent observations on a distribution $F$. Let $h(x; X_1, \ldots, X_r)$ be any *bounded* and nonnegative functions which measures the relative position of $x$ with respect to the random sample $X_1, \ldots, X_r$. Then by taking the expectation of $h(x; X_1, \ldots, X_r)$ , one can measure the depth of $x$ with respect to the center or the deepest point of the underlying distribution. The resulting depth functions will be called *type-A depth functions*. Thus, the type-A depth functions may be represented as

$$D(x; F) \equiv E\,[\,h(x; X_1, \ldots, X_r)\,], \tag{3.5}$$

where $h(x; X_1, \ldots, X_r)$ is a function as described above.

**Theorem 3.3.1** *Suppose $\theta$ in $\mathbf{R}^d$ is the point of symmetry of distribution $F$ with respect to some notion of symmetry, as discussed in Chapter 2. Then the type-A depth functions $D(x; F)$ defined in (3.5) possess the maximality at center property if*

*(1) $h(x + b; x_1 + b, \ldots, x_r + b) = h(x; x_1, \ldots, x_r)$,*

*(2) $h(-x; -x_1, \ldots, -x_r) = h(x; x_1, \ldots, x_r)$,*

*(3) $h(x; x_1, \ldots, x_r)$ is concave in its argument $x$,*

*(4) there is a point $y$ in $\mathbf{R}^d$ such that*

$$y \in (\arg\sup_{x \in \mathbf{R}^d} E[h(x; X_1 - \theta, \ldots, X_r - \theta)]) \cap (\arg\sup_{x \in \mathbf{R}^d} E[h(x; \theta - X_1, \ldots, \theta - X_r)]),$$

*where $x$, $b$ and $x_1, \ldots, x_r$ are arbitrary vectors in $\mathbf{R}^d$, $X_1, \ldots, X_r$ is a random sample from $F$.*

PROOF: By (1) and (2) we have

$$\begin{aligned}
E\,[\,h(x; X_1 - \theta, \ldots, X_r - \theta)\,] &= E\,[\,h(\theta + x; X_1, \ldots, X_r)\,], \\
E\,[\,h(x; \theta - X_1, \ldots, \theta - X_r)\,] &= E\,[\,h(\theta - x; X_1, \ldots, X_r)\,].
\end{aligned}$$

Let $y$ be the point in (4). Then

$$y \in (\arg\sup_{x \in \mathbf{R}^d} E\,[\,h(\theta + x; X_1, \ldots, X_r)\,]) \cap (\arg\sup_{x \in \mathbf{R}^d} E\,[\,h(\theta - x; X_1, \ldots, X_r)\,]).$$

The concavity of $h(x; x_1, \ldots, x_r)$ now shows that

$$h(\theta; X_1, \ldots, X_r) \geq \frac{1}{2} h(\theta + y; X_1, \ldots, X_r) + \frac{1}{2} h(\theta - y; X_1, \ldots, X_r).$$

Thus

$$\begin{aligned} E[h(\theta; X_1, \ldots, X_r)] &\geq \frac{1}{2} E[h(\theta + y; X_1, \ldots, X_r)] + \frac{1}{2} E[h(\theta - y; X_1, \ldots, X_r)] \\ &= \sup_{x \in \mathbf{R}^d} E[h(\theta + x; X_1, \ldots, X_r)] \\ &= \sup_{x \in \mathbf{R}^d} E[h(x; X_1, \ldots, X_r)]. \end{aligned}$$

Hence

$$D(\theta; F) = \sup_{x \in \mathbf{R}^d} D(x; F).$$

We are done. $\square$

Note that when given distributions are centrally symmetric about a point $\theta$ in $\mathbf{R}^d$, there is always a point $y \in \mathbf{R}^d$ such that

$$y \in (\arg \sup_{x \in \mathbf{R}^d} E[h(x; X_1 - \theta, \ldots, X_r - \theta)]) \cap (\arg \sup_{x \in \mathbf{R}^d} E[h(x; \theta - X_1, \ldots, \theta - X_r)]).$$

As long as $h(x; x_1, \ldots, x_r)$ is concave in its argument $x$, we also have

**Theorem 3.3.2** *If $h(x; x_1, \ldots, x_r)$ is concave in its argument $x$, then the type-A depth functions $D(x; F)$ defined in (3.5) monotonically decrease as $x$ moves outward along with the ray starting at the deepest point of $F$.*

PROOF: Let $\theta$ be the deepest point in $\mathbf{R}^d$ with respect to the underlying distribution $F$, that is,

$$D(\theta; F) = \sup_{x \in \mathbf{R}^d} D(x; F).$$

Let $x \neq \theta$ be an arbitrary point in $\mathbf{R}^d$, $\lambda \in (0,1)$ and $x_0 \equiv \theta + \lambda(x-\theta)$, then

$$D(x;F) \leq D(\theta;F).$$

The concavity of $h(x; x_1, \ldots, x_r)$ now shows that

$$h(x_0; X_1, \ldots, X_r) \;\; \geq \;\; \lambda h(x; X_1, \ldots, X_r) + (1-\lambda) h(\theta; X_1, \ldots, X_r).$$

Thus

$$\begin{aligned} E\,[\, h(x_0; X_1, \ldots, X_r)\,] &\geq \min\{E\,[\, h(x; X_1, \ldots, X_r)\,], E\,[\, h(\theta; X_1, \ldots, X_r)\,]\} \\ &= E\,[\, h(x; X_1, \ldots, X_r)\,], \end{aligned}$$

hence

$$D(x_0;F) \geq D(x;F).$$

We are done. □

**Example 3.3.1** Liu's simplicial depth function *SD(x; P)* is a typical example of *type-A depth functions.* Since

$$\begin{aligned} SD(x;P) &= P(x \in S[X_1, \ldots, X_{d+1}]) \\ &= E\,[\,\mathbf{I}\,[\,x \in S[X_1, \ldots, X_{d+1}]\,]\,] \\ &\equiv E\,[\, h(x; X_1, \ldots, X_r)\,], \end{aligned}$$

where $r = d+1$ and $h(x; x_1, \ldots, x_r) = \mathbf{I}\,[\,x \in S[x_1, \ldots, x_{d+1}]]$. It is not difficult to see that $h(x; x_1, \ldots, x_r)$ here is not a concave function in its first argument. Acturally, $SD(x;P)$ does not always satisfy the monotonicity and maximality properties for discrete distributions, as shown in Section 3.2.

**Example 3.3.2** **Majority depth** *MJD(x; P)* (Singh (1991), Liu and Singh (1993), Liu et al. (1997)). For a given random sample $X_1, \ldots, X_d$ from $P$ on $\mathbf{R}^d$, a unique hyperplane containing these points is obtained and consequently two closed halfspaces with this hyperplane as common boundary are obtained. Denote the one which carries probability mass greater than or equal to $\frac{1}{2}$ by $H^m_{X_1,\ldots,X_d}$. Then the majority depth function *MJD(x; P)* is defined as follows:

$$\mathit{MJD(x;\ P)} = P(x \in H^m_{X_1,\ldots,X_d}). \tag{3.6}$$

Clearly, the majority depth function *MJD(x; P)* is a *type-A depth function* with $r = d$ and $h(x; x_1, \ldots, x_r) \equiv \mathbf{I}\,[\, x \in H^m_{x_1,\ldots,x_r}\,]$.

Liu and Singh (1993) remarked that the majority depth function *MJD(x; P)* is affine invariant and decreases monotonically as x moves away from the center $\theta$ of any angularly symmetric distributions along any fixed ray originating from the center $\theta$.

The following result is new in two aspects: a) it generalizes the remark in Liu and Singh (1993) about the majority depth function *MJD(x; P)* to any *HS*-symmetric distributions. b) it shows that the majority depth function *MJD(x; P)* does not approach zero as $\|x\|$ approachs infinity for some distributions.

**Theorem 3.3.3** *The majority depth function MJD(x; P) defined in (3.6)*

*(1) attains maximum value at the center of any HS-symmetric distributions and decreases monotonically as x moves away from the center along any fixed ray originating from the center ;*

*(2) fails to approach zero as* $\|x\|$ *approachs infinity for some distributions.*

PROOF:

(1) a) Let $\theta$ be the center of a *HS*-symmetric distribution $F$ and $x$ be an arbitrary point in $\mathbf{R}^d$. Then by the definition of *HS*-symmetry, for any random sample $X_1, \ldots, X_d$ from $F$ we have

$$x \in H^m_{X_1,\ldots,X_d} \Longrightarrow \theta \in H^m_{X_1,\ldots,X_d}.$$

Thus

$$MJD(\theta; P) = \sup_{x \in \mathbf{R}^d} MJD(x; P).$$

b) Let $\lambda \in (0,1)$ and $x_0 \equiv \lambda\theta + (1-\lambda)x$. Then

$$\begin{aligned} MJD(x_0; P) - MJD(x; P) &= P(x_0 \in H^m_{X_1,\ldots,X_d}) - P(x \in H^m_{X_1,\ldots,X_d}) \\ &= P(x_0 \in H^m_{X_1,\ldots,X_d} \text{ and } x \overline{\in} H^m_{X_1,\ldots,X_d}) \\ &\geq 0. \end{aligned}$$

(2) We now give a *conuterexample* which shows that *MJD(x; P)* does not approach zero as $\|x\|$ approachs infinity. Let $d = 2$, $P(X = (\pm 1, 0)) = \frac{1}{3}$, $P(X = (0,1)) = \frac{1}{3}$. Then it is easy to see that

$$\lim_{\|x\| \to \infty} MJD(x; P) = \frac{2}{3}.$$

In fact, in the univariate case, one can show that

$$MJD(x; P) \to \frac{1}{2} \qquad \text{as } x \to \infty.$$

We are done. □

**Remark 3.3.1** For the type-A depth functions defined in (3.5), the corresponding sample versions are *U-statistics*.

### 3.3.2 Type-B depth functions

Let $X_1, \ldots, X_r$ be a random sample from $F$ on $R^d$. Let $h(x; X_1, \ldots, X_r)$ be a *unbounded* and nonegative function which measures the dispersion of the point cloud $\{x, X_1, \ldots, X_r\}$. Then by taking the expectation of $h(x; X_1, \ldots, X_r)$, one can measure the relative distance between $x$ and the center or the deepest point of the underlying distribution. Thus, a corresponding *bounded* depth function can be constructed as follows:

$$D(x; F) \equiv (1 + E\,[\, h(x; X_1, \ldots, X_r)\,])^{-1}\,, \tag{3.7}$$

where $h(x; X_1, \ldots, X_r)$ is a function as described above. This type of depth functions will be called *type-B depth functions*.

**Remark 3.3.2** We could have defined type-B depth functions as

$$D(x; F) \equiv E\,[\,(1 + h(x; X_1, \ldots, X_r))^{-1}\,], \tag{3.8}$$

which then fit into the type-A category. But for the sake of tractability we list them independently here as a new type of depth functions.

**Remark 3.3.3** As a measure of the dispersion of the point cloud $\{x; x_1, \ldots, x_r\}$, $h(x; x_1, \ldots, x_r)$ does not always possess affine invariance property, although in many cases it does possess *rigid-body invariance* property, that is,

$$h(Ax + b; Ax_1 + b, \ldots, Ax_r + b) = h(x; x_1, \ldots, x_r), \tag{3.9}$$

for any $d \times d$ orthogonal matrix $A$ and any vector $b \in \mathbf{R}^d$. Type-B depth functions possess the maximality at center and the monotonicity relative to the deepest point properties for suitable $h(x; x_1, \ldots, x_r)'s$.

**Theorem 3.3.4** *Suppose $\theta$ in $\mathbf{R}^d$ is the point of symmetry of distribution $F$ with respect to some notion of symmetry, as discussed in Chapter 2. The type-B depth functions $D(x; F)$ defined in (3.7) possess the maximality at center property if*

*(1) $h(x + b; x_1 + b, \ldots, x_r + b) = h(x; x_1, \ldots, x_r)$,*

*(2) $h(-x; -x_1, \ldots, -x_r) = h(x; x_1, \ldots, x_r)$,*

*(3) $h(x; x_1, \ldots, x_r)$ is convex in its argument $x$,*

*(4) there is a point $y \in \mathbf{R}^d$ such that*

$$y \in (\arg \inf_{x \in \mathbf{R}^d} E[h(x; X_1 - \theta, \ldots, X_r - \theta)]) \cap (\arg \inf_{x \in \mathbf{R}^d} E[h(x; \theta - X_1, \ldots, \theta - X_r)])$$

*where $x$, $b$ and $x_1, \ldots, x_r$ are arbitrary vectors in $\mathbf{R}^d$, $X_1, \ldots, X_r$ is a random sample from $F$.*

PROOF: Similer to that for Theorem 3.3.1. $\square$

**Theorem 3.3.5** *If $h(x; x_1, \ldots, x_r)$ is convex in its argument $x$, then the type-B depth functions $D(x; F)$ defined in (3.7) monotonically decrease as $x$ moves outward along with the ray starting at the deepest point of $F$.*

PROOF: Similer to that for Theorem 3.3.2. $\square$

**Example 3.3.3** **Simplicial volume depth** *SVD(x; F)*. Using the volume of the simplex with vertices $x, X_1, \ldots, X_d$ as a measure of the dispersion of the point cloud $\{x, X_1, \ldots, X_d\}$, one can construct a depth function as follows:

$$SVD(x; F) = (1 + E\,[\,\Delta^{\alpha}(x, X_1, \ldots, X_d)\,])^{-1}\,, \tag{3.10}$$

where $X_1, \ldots, X_d$ is a random sample from $F$, $\Delta(x, X_1, \ldots, X_d)$ is the volume of the simplex with vertices $x, X_1, \ldots, X_d$, and $\alpha \in (0, \infty)$. It is obvious that *SVD(x; F)* is a *type-B depth function.* Unfortunately, *SVD(x; F)* usually is not affine invariant since

$$\begin{aligned}\Delta^{\alpha}(Ax + b, Ax_1 + b, \ldots, Ax_d + b) &= |\det(A)\,|^{\alpha}\Delta^{\alpha}(x, x_1, \ldots, x_d)\\ &= |\det(A)\,|^{\alpha}\Delta^{\alpha}(x, x_1, \ldots, x_d),\end{aligned}$$

where $b$ is any vector in $\mathbf{R}^d$ and $\det(A)$ is the determinant of nonsingular matrix $A$ , which is not always equal to 1. This problem, however, can be fixed by a modification of (3.10) as follows:

$$SVD(x; F) \equiv \left(1 + E\,[\,\left(\frac{\Delta(x, X_1, \ldots, X_d)}{\sqrt{\det(\Sigma)}}\right)^{\alpha}\,]\right)^{-1}, \tag{3.11}$$

where $\Sigma$ is the covariance matrix of $F$. It is not difficult to verify now that *SVD(x; F)* is affine invariant. The simplicial volume depth functions also possess the maximality at center and monotonicity relative to the deepest point properties.

**Corollary 3.3.1** *The simplicial volume depth SVD(x; F) defined in (3.11) for* $\alpha \geq 1$ *possesses the monotonicity relative to the deepest point property.*

PROOF: **(i)** By Theorem 3.3.5 we need only to check the convexity of $\Delta^\alpha(x, x_1, \ldots, x_d)$ for $\alpha \in [1, \infty)$ in its argument $x$. Let x, y be two points in $\mathbf{R}^d$, $x_0 \equiv \lambda x + (1-\lambda)y$ and $\lambda \in (0,1)$. Then

$$\begin{aligned}
\Delta(x_0; x_1, \ldots, x_d) &= \left| \frac{1}{d!} \det \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_{01} & x_{11} & \cdots & x_{d1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{0d} & x_{1d} & \cdots & x_{dd} \end{vmatrix} \right| \\
&= \left| \frac{1}{d!} \det \begin{vmatrix} \lambda + (1-\lambda) & 1 & \cdots & 1 \\ \lambda x_1 + (1-\lambda) y_1 & x_{11} & \cdots & x_{d1} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda x_d + (1-\lambda) y_d & x_{1d} & \cdots & x_{dd} \end{vmatrix} \right| \\
&\leq \lambda \Delta(x; x_1, \ldots, x_d) + (1-\lambda)\Delta(y; x_1, \ldots, x_d),
\end{aligned}$$

where $x_0 = (x_{01}, \ldots, x_{0d})'$, $x = (x_1, \ldots, x_d)'$, $y = (y_1, \ldots, y_d)'$ and $x_i = (x_{i1}, \ldots, x_{id})'$ for $1 \leq i \leq d$. Now the convexity of the function $x^\alpha$ for $0 < x < \infty$ shows that

$$\Delta^\alpha(x_0; x_1, \ldots, x_d) \leq \lambda \Delta^\alpha(x; x_1, \ldots, x_d) + (1-\lambda)\Delta^\alpha(y; x_1, \ldots, x_d).$$

**(ii)** It is obvious that

$$\Delta^\alpha(x; x_1, \ldots, x_d) \to \infty \qquad \text{as } \|x\| \to \infty.$$

Thus

$$SVD(x; F) \to 0 \qquad \text{as } \|x\| \to \infty.$$

We are done. $\square$

Since $\Delta^\alpha(x, x_1, \ldots, x_d)$ is convex and rigid-body invariant, according to Theorem 3.3.4 we obtain

**Corollary 3.3.2** *The simplicial volume depth SVD(x; F) defined in (3.11) for $\alpha > 1$ possesses the maximality at center property for centrally symmetric distribution $F$ on $\mathbf{R}^d$.*

Now by the affine invariance, Corollary 3.3.1 and Corollary 3.3.2 we obtain

**Theorem 3.3.6** *The simplicial volume depth function SVD(x; F) defined in (3.11) for $\alpha > 1$ is a statistical depth function in the sense of Definition 3.2.1 for centrally symmetric distributions.*

**Remark 3.3.4** Oja (1983) introduced a location measure for centrally symmetric distributions by the use of simplicial volume as follows. The location measure is a function $\mu_\alpha$: $\mathcal{P} \to \mathbf{R}^d, 0 < \alpha < \infty$ such that

$$E\,[\,\Delta^\alpha(\mu_\alpha(P); X_1, \ldots, X_d)\,] = \inf_{\mu \in \mathbf{R}^d} E\,[\,\Delta^\alpha(\mu; X_1, \ldots, X_d)\,].$$

However, he did not develop it into a depth function, nor consider the affine invariance version (3.11).

**Example 3.3.4** $\mathbf{L_p}$ **depth** $L_pD(x;\,F)$ $(p > 0)$. The $L_p$ norm of $(X - x)$ is a measure of the distance between $x$ and $X$. By taking expectation of this distance, one can measure the distance between $x$ and the center or the deepest point of the distribution of $X$. Thus a depth measure can be constructed as follows:

$$L_pD(x; F) \equiv (1 + E\,[\,\|x - X\|_p\,])^{-1}\,, \tag{3.12}$$

where $\| \cdot \|_p$ is the usual $L_p$ norm. Clearly *$L_pD(x;\ F)$* is a *type-B depth function* with $h(x; x_1) \equiv \|x - x_1\|_p$. *$L_pD(x;\ F)$* generally does not possess the affine invariance property since

$$E\,[\, \|Ax + b \,-\, (AX + b)\|_p \,] = E\,[\, \|A(x \,-\, X\|_p \,],$$

which is not always equal to $E\,[\, \|(x \,-\, X\|_p \,]$ for any nonsingular matrix $A$. However, *$L_pD(x;\ F)$* possesses maximality at center and monotonicity relative to the deepest point properties for $p \geq 1$.

**Corollary 3.3.3** *The $L_pD(x;\ F)$ defined in (3.12) for $p \geq 1$ possesses the monotonicity relative to the deepest point property.*

PROOF: **(i)** By Theorem 3.3.5 we need only to check the convexity of $h(x; x_1) = \|x - x_1\|_p$ in argument $x$. But this follows in straightforward fashion from *Minkowski's inequality.*
**(ii)** It is obvious that

$$L_pD(x; F) \to 0 \quad \text{as } \|x\| \to \infty.$$

We are done. □

Since $h(x; x_1)$ is *location invariant* and *even*, that is, $h(x + b, x_1 + b) = h(x, x_1)$ for any vector $b \in \mathbf{R}^d$ and $h(-x, -x_1) = h(x, x_1)$, by the convexity just established and Theorem 3.3.4 we obtain

**Corollary 3.3.4** *The $L_pD(x;\ F)$ defined in (3.12) for $p \geq 1$ possesses the maximality at center property for centrally symmetric distribution $F$ on $\mathbf{R}^d$.*

Although $L_pD(x;\ F)$ ($p > 0$) generally is not affine invariant, it is easy to see that $L_2D(x;\ F)$ is rigid-body invariant. Furthermore, a modification of $L_2$ norm can lead to a afffine invariant $L_2D(x;\ F)$. For a positive definite $d \times d$ matrix $M$, define a norm $\|\cdot\|_M$ as

$$\|x\|_M \equiv \sqrt{x'Mx}, \qquad \forall x \in \mathbf{R}^d, \tag{3.13}$$

then the $L_2(x;\ F)$ defined in (3.12) can be modified as

$$L_2D(x;F) \equiv \left(1 + E\left[\ \|x - X\|_{\Sigma^{-1}}\ \right]\right)^{-1}, \tag{3.14}$$

where $\Sigma$ is the covariance matrix of $F$. Now it is not difficult to verify that the $L_2D(x;\ F)$ defined in (3.14) is affine invariant.

**Theorem 3.3.7** *The $L_2D(x;\ F)$ defined in (3.14) is a statistical depth function in the sense of Definition 3.2.1 for any distribution $F$ angularly symmetric about a unique point $\theta \in \mathbf{R}^d$.*

PROOF: Since $L_2D(x;\ F)$ defined in (3.14) is affine invariant, we need only check $(P2)$ and $(P3)$.

**(i)** We first show that $\|\cdot\|_M$ is convex for any positive definite $d \times d$ matrix $M$. Since $M$ is positive definite, there is a nonsingular matrix $S$ such that $M = S'S$. Let $x, y$ be two points in $\mathbf{R}^d$ and $\lambda \in (0,1)$. Then

$$\begin{aligned}
\|\lambda x + (1-\lambda)y\|_M^2 &= (\lambda x + (1-\lambda)y)'M(\lambda x + (1-\lambda)y) \\
&= \lambda^2 x'Mx + 2\lambda(1-\lambda)x'My + (1-\lambda)^2 y'My \\
&= \lambda^2 x'Mx + 2\lambda(1-\lambda)(Sx)'(Sy) + (1-\lambda)^2 y'My.
\end{aligned}$$

The Schwarz inequality implies that

$$
\begin{aligned}
\|\lambda x + (1-\lambda)y\|_M^2 &\leq \lambda^2 x'Mx + 2\lambda(1-\lambda)\|Sx\|\|Sy\| + (1-\lambda)^2 y'My \\
&= \lambda^2\|x\|_M^2 + 2\lambda(1-\lambda)\|x\|_M\|y\|_M + (1-\lambda)^2\|y\|_M^2 \\
&= (\lambda\|x\|_M + (1-\lambda)\|y\|_M)^2.
\end{aligned}
$$

Thus

$$
\|\lambda x + (1-\lambda)y\|_M \leq \lambda\|x\|_M + (1-\lambda)\|y\|_M.
$$

**(ii)** Now we show that there is a point $y \in \mathbf{R}^d$ satisfying the condition in (4) of Theorem 3.3.4. Equivalent, we need to show that

$$
\theta \in \arg\inf_{x\in\mathbf{R}^d} E\,[\,\|x - X\|_{\Sigma^{-1}}\,], \tag{3.15}
$$

where $\Sigma$ is the covariance matrix of $F$.

(1) We first show that

$$
E\,[\,\frac{\theta - X}{\|X-\theta\|_{\Sigma^{-1}}}\,] = 0.
$$

Since $F$ is angularly symmetric about $\theta$, then based on the results established in Chapter 2 we have

$$
P(X \in H_\theta) = P(X \in -H_\theta),
$$

for any closed halfspace $H_\theta$ with $\theta$ on the boundary. Since $\Sigma^{-1}$ is positive definite, then there is a nonsigular matrix $R$ such that $\Sigma^{-1} = R'R$. Thus

$$
P(RX \in RH_\theta) = P(RX \in -RH_\theta),
$$

for any closed halfspace $H_\theta$ with $\theta$ on the boundary. By nonsingularity and the results established in Chapter 2 we conclude that $RX$ is angularly symmetric about $R\theta$. Hence

$$\frac{R\,(X-\theta)}{\|R\,(X-\theta)\|} \stackrel{\mathrm{d}}{=} \frac{R\,(\theta-X)}{\|R\,(\theta-X)\|},$$

which is equivalent to

$$\frac{R\,(X-\theta)}{\|(X-\theta)\|_{\Sigma^{-1}}} \stackrel{\mathrm{d}}{=} \frac{R\,(\theta-X)}{\|(\theta-X)\|_{\Sigma^{-1}}}.$$

Thus

$$E\,[\,\frac{R\,(X-\theta)}{\|(X-\theta)\|_{\Sigma^{-1}}}\,] \;=\; E\,[\,\frac{R\,(\theta-X)}{\|(\theta-X)\|_{\Sigma^{-1}}}\,].$$

This implies that

$$E\,[\,\frac{\theta-X}{\|(\theta-X)\|_{\Sigma^{-1}}}\,] = 0.$$

(2) Now we are going to show that (3.15) holds true. Consider the derivative of $E\,[\,\|\mu - X\|_{\Sigma^{-1}}\,]$ with respect to $\mu \in \mathbf{R}^d$, by vector differentiation formula we have

$$\begin{aligned}
\frac{d\,(E\,[\,\|\mu-X\|_{\Sigma^{-1}}\,])}{d\mu} &= \frac{d\,(\int_{\mathbf{R}^d}\|\mu-x|_{\Sigma^{-1}}\,dF(x))}{d\mu} \\
&= \int_{\mathbf{R}^d}\frac{d\,(\|\mu-x\|_{\Sigma^{-1}})}{d\mu}\,dF(x) \\
&= \int_{\mathbf{R}^d}\frac{\Sigma^{-1}(\mu-x)}{\|\mu-x\|_{\Sigma^{-1}}}\,dF(x) \\
&= E\,[\,\frac{\Sigma^{-1}(\mu-X)}{\|\mu-X\|_{\Sigma^{-1}}}\,] \\
&= \Sigma^{-1}E\,[\,\frac{\mu-X}{\|\mu-X\|_{\Sigma^{-1}}}\,].
\end{aligned}$$

Now by convexity and (1) we conclude that

$$\theta \in \arg\inf_{x\in\mathbf{R}^d} E\,[\,\|x-X\|_{\Sigma^{-1}}\,].$$

The result then follows from Theorem 3.3.4 and Theorem 3.3.5. □

### 3.3.3 Type C depth functions

Let $O(x; F)$ be a measure of the *outlyingness* of the point $x$ in $\mathbf{R}^d$ with respect to the center or the deepest point of the distribution $F$ of $X$. Usually $O(x; F)$ is *unbounded*, but a corresponding *bounded* depth function can be constructed as follows:

$$D(x; F) \equiv (1 + O(x; F))^{-1}. \tag{3.16}$$

Such depth functions will be called *Type C depth functions.*

**Remark 3.3.5** Type B and Type C depth functions are similar in form except that in Type B depth functions the outlyingness of a point with respect to the center or the deepest point of a distribution is obtained by directly taking the expectation of some function $h(x; X_1, \ldots, X_r)$. It is convenient to have these two types separately, although one could merge them into a single type and treat them uniformly. □

For Type C depth functions, the following two results are analogues of Theorems 3.3.4 and 3.3.5 and are proved similarly. It is often convenient to write $O(x; X)$ for $O(x; F_X)$.

**Theorem 3.3.8** *Suppose $\theta$ in $\mathbf{R}^d$ is the point of symmetry of distribution $F$ with respect to some notion of symmetry, as discussed in Chapter 2. The Type C depth functions $D(x; F)$ defined in (3.16) possess the "maximality at center" property if*

(1) $O(x + b; X + b) = O(x; X + b)$,

(2) $O(-x; -X) = O(x; X)$,

(3) $O(x; X)$ *is convex in its argument* $x$,

(4) *there is a point* $y \in \mathbf{R}^d$ *such that*

$$y \in (\arg \inf_{x \in \mathbf{R}^d} O(x; X-\theta) \cap (\arg \inf_{x \in \mathbf{R}^d} O(x; \theta - X))$$

*where* $x$, $b$ *are arbitrary vectors in* $\mathbf{R}^d$. □

**Theorem 3.3.9** *If* $O(x;F)$ *is convex in the argument* $x$, *then the Type C depth functions* $D(x;F)$ *defined in* (3.16) *decrease monotonically as* $x$ *moves outward along any ray starting at a deepest point of* $F$. □

**Example 3.3.5** **Projection depth** *PD*$(x;F)$. A depth measure based on the outlyingness of a point with respect to the center or the deepest point of an underlying distribution can be defined as follows. Define the outlyingness of a point $x$ to be the worst case outlyingness of $x$ with respect to the one-dimensional median in any one-dimensional projection, that is,

$$O(x;F) \equiv \sup_{\|u\|=1} \frac{|\, u'x - \operatorname{Med}(u'X) \,|}{\operatorname{Mad}(u'X)}, \tag{3.17}$$

where Med denotes the univariate median as defined in Chapter 2, Mad denotes the univariate median absolute deviation defined for $Y \in \mathbf{R}$ as

$$\operatorname{Mad}(Y) = \operatorname{Med}(|Y - \operatorname{Med}(Y)|,$$

and $\|\cdot\|$ is the Euclidean norm. Then a corresponding depth measure, which will be called *projection depth*, is defined as

$$PD(x;F) \equiv (1 + O(x;F))^{-1}, \tag{3.18}$$

where $F$ is the distribution of $X$. Clearly, this depth function is a Type C depth function.

□

**Remark 3.3.6** For a given one-dimensional dataset $X = \{X_1, \ldots, X_n\}$, $O_n(x) \equiv (x - \mathrm{Med}_{1\le i\le n}\{X_i\}) / \mathrm{Mad}_{1\le i\le n}\{X_i\}$ has long been used as a robust outlyingness measure of a point $x \in \mathbf{R}$ with respect to the center (the median) of the dataset. See Mosteller and Tukey (1977), pp. 205-208. Here

$$\begin{aligned}
\mathrm{Med}_{1\le i\le n}\{X_i\} &= \frac{1}{2}\left(X_{(\lfloor \frac{n+1}{2} \rfloor)} + X_{(\lfloor \frac{n+2}{2} \rfloor)}\right), \\
\mathrm{Mad}_{1\le i\le n}\{X_i\} &= \mathrm{Med}_{1\le i\le n}\{|X_i - \mathrm{Med}_{1\le j\le n}\{X_j\}|\},
\end{aligned}$$

and $X_{(1)} < \ldots < X_{(n)}$ are the ordered $X_1, \ldots, X_n$. Donoho and Gasko (1992) generalized the one-dimensional outlyingness measure to arbitrary dimension $d$. The sample version of the projection depth function $PD(x; F)$ is given by

$$PD_n(x) = (1 + O_n(x))^{-1}, \tag{3.19}$$

for a random sample $X = \{X_1, \ldots, X_n\}$ from F in $\mathbf{R}^d$. Liu (1992) considered the sample version of projection depth, but did not provide any treatment of it. □

**Theorem 3.3.10** *The projection depth function* $PD(x; F)$ *defined in* (3.18) *is a statistical depth function in the sense of Definition 3.2.1.*

PROOF:

(1) Affine Invariance. For any nonsingular $d \times d$ matrix $A$, any vector $b \in \mathbf{R}^d$, and

$O(x;X)$ defined in (3.17), we have

$$
\begin{aligned}
O(Ax+b\,;AX+b) &= \sup_{\|u\|=1} \frac{|\, u'(Ax+b) - \text{ Med}(u'(AX+b))\,|}{\text{Mad}(u'(AX+b))} \\
&= \sup_{\|u\|=1} \frac{|\,(u'A)x - \text{ Med}((u'A)X)\,|}{\text{Mad}((u'A)X)} \\
&= \sup_{\|u\|=1} \frac{|\, v'x - \text{ Med}(v'X)\,|}{\text{Mad}(v'X)},
\end{aligned}
$$

where $v = A'u/\|A'u\|$. Since $A$ is nonsingular, then

$$
O(Ax+b\,;AX+b) = \sup_{\|u\|=1} \frac{|\, u'x - \text{ Med}(u'X)\,|}{\text{Mad}(u'X)},
$$

that is,

$$
PD(Ax+b; F_{AX+b}) = PD(x; F_X).
$$

(2) Maximality at Center.

(i) Suppose that $F$ is not *HS*-symmetric about any point. Then we simply define the center of the distribution to be the point $\theta$ where $D(x;F)$ attains maximum value, and thus

$$
PD(\theta; F) = \sup_{x\in\mathbf{R}^d} PD(x;F).
$$

(ii) Suppose that $F$ is *HS*-symmetric about a unique point $\theta \in \mathbf{R}^d$. Then by Theorem 2.2.2 established in Chapter 2, we have

$$
\text{Med}(u'X) = u'\theta,
$$

for any unit vector $u \in \mathbf{R}^d$. Thus

$$
PD(\theta; F) = \sup_{x\in\mathbf{R}^d} PD(x;F).
$$

(3) Monotonicity Relative to Deepest Point.

We show that $O(x;X)$ is convex in its first argument. Let $\theta$ and $x$ be two arbitrary points in $\mathbf{R}^d$, $0 < \alpha < 1$, and put $x_0 \equiv (1-\alpha)\theta + \alpha x$. Then we have

$$\begin{aligned}
|\,u'x_0 - \mathrm{Med}(u'X)\,| &= |\,U'((1-\alpha)\theta + \alpha x) - \mathrm{Med}(u'X)\,| \\
&= |\,(1-\alpha)(u'\theta - \mathrm{Med}(u'X)) + \alpha(u'x - \mathrm{Med}(u'X))\,| \\
&\le (1-\alpha)\,|\,(u'\theta - \mathrm{Med}(u'X))\,| + \alpha\,|\,(u'x - \mathrm{Med}(u'X))\,|,
\end{aligned}$$

and hence

$$\begin{aligned}
O(x_0\,;X) &= \sup_{\|u\|=1} \frac{|\; u'x_0 - \mathrm{Med}(u'X)\;|}{\mathrm{Mad}(u'X)} \\
&\le \sup_{\|u\|=1} \frac{(1-\alpha)\;|\;(u'\theta - \mathrm{Med}(u'X))\;| + \alpha\;|\;(u'x - \mathrm{Med}(u'X))\;|}{\mathrm{Mad}(u'X)} \\
&\le (1-\alpha)O(\theta;F) + \alpha O(x;F).
\end{aligned}$$

Now by Theorem 3.3.9 we conclude that *PD(x; F)* decreases monotonically as $x$ moves outward along any fixed ray originating from the deepest point of the distribution F.

(4) Vanishing at Infinity.

It is obvious that

$$PD(x;\,F) \to 0, \qquad \text{as } \|x\| \to \infty.$$

This completes the proof. □

**Example 3.3.6 Mahalanobis depth** $MHD(x;F)$. Mahalanobis (1936) introduced a distance between two points $x$ and $y$ in $\mathbf{R}^d$ with respect to a positive definite $d \times d$

matrix $M$ as:

$$d_M^2(x,y) = (x-y)'M(x-y).$$

Based on this *Mahalanobis distance*, one can define *Mahalanobis depth* as follows:

$$MHD(x;F) = \Big(1 + (x-\mu(F))'\Sigma^{-1}(x-\mu(F))\Big)^{-1}, \tag{3.20}$$

where $F$ is the distribution of $X$ on $\mathbf{R}^d$, $\mu(F)$ is a location measure, and $\Sigma$ is the covariance matrix of $F$. Obviously, the Mahalanobis depth function $MHD(x;F)$ is a Type C depth function, taking

$$O(x;F) \equiv (x-\mu(F))'\Sigma^{-1}(x-\mu(F)) \tag{3.21}$$

in (3.16). $O(x;F)$ and $O(x;X)$ are used interchangeably. □

**Theorem 3.3.11** *The Mahalanobis depth function $MHD(x;F)$ defined in* (3.20) *is a statistical depth function in the sense of Definition 3.2.1 for any symmetric distribution $F$ if $\mu(F)$ in* (3.20) *is affine equivariant and agrees with the point of symmetry of $F$.*

PROOF: Assume, w.l.o.g., that the probability mass of $F$ is not concentrated on any subspace of $R^d$ with dimension less than $d$.

(1) Affine Invariance. For any nonsingular $d \times d$ matrix $A$, any vector $b \in \mathbf{R}^d$ and $O(x;X)$ defined in (3.21), by affine equivariance of $\mu(\cdot)$ we have

$$\begin{aligned}
O(Ax+b\,;AX+b) &= (Ax+b-\mu(AX+b))'\Sigma_{AX+b}^{-1}(Ax+b-\mu(AX+b)) \\
&= (x-\mu(X))'A'(A')^{-1}\Sigma^{-1}A^{-1}A(x-\mu(X)) \\
&= (x-\mu(X))'\Sigma^{-1}(x-\mu(X)) \\
&= O(x;X).
\end{aligned}$$

Thus $MHD(x;F)$ is affine invariant.

(2) Maximality at Center. This follows directly from the fact that $\mu(F)$ agrees with the point of symmetry of $F$.

(3) Monotonicity Relative to Deepest Point.

By Theorem 3.3.9 we need only check the convexity of $O(x;X)$ in its first argument $x$. Let $x$, $y$ be two points in $\mathbf{R}^d$, $\alpha \in (0,1)$, $M \equiv \Sigma^{-1}$, $\beta \equiv \alpha(1-\alpha)$ and $x_0 \equiv \alpha x + (1-\alpha)y$. Then

$$
\begin{aligned}
O(x_0;X) &= d_M^2(x_0, \mu(X)) \\
&= d_M^2(\alpha(x-\mu(X)) + (1-\alpha)(y-\mu(X)), 0) \\
&= \alpha^2 d_M^2(x, \mu(X)) + (1-\alpha)^2 d_M^2(y, \mu(X)) + 2\beta(x-\mu(X))'\Sigma^{-1}(y-\mu(X)) \\
&= \alpha d_M^2(x, \mu(X)) + (1-\alpha) d_M^2(y, \mu(X)) - \beta(x-y)'\Sigma^{-1}(x-y) \\
&\leq \alpha d_M^2(x, \mu(X)) + (1-\alpha) d_M^2(y, \mu(X)).
\end{aligned}
$$

Thus $O(x;X)$ defined in (3.21) is convex.

(4) Vanishing at Infinity. It is obvious that

$$MHD(x;F) \to 0, \qquad \text{as } \|x\| \to \infty.$$

This completes the proof. □

**Remark 3.3.7** Liu (1992, 1993) introduced the *Mahalanobis depth* as

$$M_h D(x;F) = \left(1 + (x-\mu)'\Sigma^{-1}(x-\mu)\right)^{-1},$$

where $\mu$ and $\Sigma$ are the mean and covariance matrix of $F$. Unfortunately, $M_hD(\cdot\,; F)$ is not "robust" because the population mean has been used in the construction of $M_hD(\cdot\,; F)$ as the relative center of $F$. Also $M_hD(x; F)$ can fail to achieve maximum value at the point of symmetry of angularly symmetric distributions, although, as remarked by Liu (1992), it does satisfy $(P2)$ and $(P3)$ for centrally symmetric distributions. □

### 3.3.4 Type Ddepth functions

Based on the "tailedness" of a point with respect to a given distribution, one can measure the relative depth of the point with respect to the center or the deepest point of the distribution. Let $\mathcal{C}$ be a class of closed subsets of $\mathbf{R}^d$ and $P$ be a probability measure on $\mathbf{R}^d$. Define a depth measure as follows:

$$D(x; P, \mathcal{C}) \equiv \inf_{C} \{ \ P(C) \mid x \in C \in \mathcal{C} \ \}, \tag{3.22}$$

which will be called *Type D depth function*.

A sample version of the Type D depth function, denoted by $D_n(x; P_n, \mathcal{C})$ and $D_n(x)$ for short, will be defined as

$$D_n(x; P_n, \mathcal{C}) \equiv \inf_{C} \{ \ P_n(C) \mid x \in C \in \mathcal{C} \ \}, \tag{3.23}$$

for a random sample $X_1, \ldots, X_n$ from $P$, where $P_n$ is the empirical probability measure corresponding to $P$, defined as $P_n(C) \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}(X_i \in C)$.

We will confine our attention to the class of closed and convex Borel sets $\mathcal{C}$ satisfying the following conditions:

**(C1)** if $C \in \mathcal{C}$, then $\overline{C^c} \in \mathcal{C}$,

**(C2)** if $C \in \mathcal{C}, x \in C^o$, then there exists a $C_1 \in \mathcal{C}$ such that $x \in \partial C_1, C_1 \subset C^o$,

**(C3)** if $C \in \mathcal{C}$, then $AC + b \in \mathcal{C}$ for any orthogonal $d \times d$ matrix $A$ and vector $b \in \mathbf{R}^d$,

where $C^c, C^o$ and $\overline{C}$ denotes the complement, the interior and the closure of $C$ rescep-tively.

**Example 3.3.7** Clearly, the class of all closed halfspaces $\mathcal{H}$ on $\mathbf{R}^d$ is one of examples which satisfy $(C1), (C2)$ and $(C3)$. The halfspace depth function $HD(x; P)$ defined in (3.2) is a typical example of Type D depth functions.

**Theorem 3.3.12** *Let $\mathcal{C}$ be a class of closed and convex Borel sets satisfying $(C1), (C2)$ and $(C3)$ and $P$ be a probability measure on $\mathbf{R}^d$. Then*

*(1) $D(x; P, \mathcal{C})$ is upper-semicontinuous,*

*(2) $D^\alpha \equiv \{x \in \mathbf{R}^d \mid D(x; P, \mathcal{C}) \geq \alpha\}$ are compact, convex and nested (i.e., $D_{\alpha_1} \subset D_{\alpha_2}$ if $\alpha_1 > \alpha_2$) for $\alpha \in [0, 1]$.*

PROOF:

(1) We first show that

$$\{x \in \mathbf{R}^d \mid D(x; P, \mathcal{C}) \geq \alpha\} = \cap\{C \mid P(C) > 1 - \alpha, C \in \mathcal{C}\} \tag{3.24}$$

a) if $x \in \{x \in \mathbf{R}^d \mid D(x; P, \mathcal{C}) \geq \alpha\}$ and there exists a $C \in \mathcal{C}$ such that

$$P(C) > 1 - \alpha, \qquad x \not\in C.$$

Then

$$x \in C^c, \qquad P(C^c) < \alpha.$$

By $(C1)$ and $(C2)$, there is a $C_1 \in \mathcal{C}$ such that

$$x \in \partial C_1, \qquad C_1 \subset C^c,$$

thus

$$P(C_1) < \alpha,$$

hence $D(x; P, \mathcal{C}) < \alpha$, which is a contradiction to the assumption that $x \in \{x \in \mathbf{R}^d \mid D(x; P, \mathcal{C}) \geq \alpha\}$. Thus

$$\{x \in \mathbf{R}^d \mid D(x; P, \mathcal{C}) \geq \alpha\} \quad \subset \quad \cap\{C \mid P(C) > 1 - \alpha, C \in \mathcal{C}\}.$$

b) if $x \in \cap\{C \mid P(C) > 1 - \alpha, C \in \mathcal{C}\}$, and there is a $C \in \mathcal{C}$ such that

$$x \in C, \qquad P(C) < \alpha.$$

Then by $(C3)$, there exists a $C_1 \in \mathcal{C}$ such that

$$x \in C_1^o, \qquad P(C_1) < \alpha,$$

thus

$$x \not\in \overline{C_1^c}, \qquad P\left(\overline{C_1^c}\right) > 1 - \alpha,$$

which contradicts the assumption that $x \in \cap\{C \mid P(C) > 1 - \alpha, C \in \mathcal{C}\}$. Thus

$$\{x \in \mathbf{R}^d \mid D(x; P, \mathcal{C}) \geq \alpha\} \quad \supset \quad \cap\{C \mid P(C) > 1 - \alpha, C \in \mathcal{C}\}.$$

Now by a) and b) we conclude that $D^\alpha$ is closed and thus $D(x; P, \mathcal{C})$ is upper-semicontinuous.

(2) The nestedness of $D^\alpha$ is trival. The boundedness of $D^\alpha$ follows from the fact that

$$D(x; P, \mathcal{C}) \to 0 \quad \text{as } \|x\| \to \infty. \tag{3.25}$$

The compactness of $D^\alpha$ now follows from the boundedness and the closeness of $D^\alpha$. The convexity follows from (3.24) and the fact that the intersection of convex sets is convex. We are done. □

**Remark 3.3.8** a) The above theorem still holds true if $(C2)$ is replaced by

**(C2')** $P(\partial C) = 0, \quad \forall\, C \in \mathcal{C}.$

b) $D^\alpha$ is called $\alpha$th *depth contour*. Discussion on depth contours will be given later.

**Definition 3.3.1** *Let $\mathcal{C}$ be a class of closed convex sets satisfying $(C1), (C2)$ and $(C3)$ and $P$ be a probability measure on $\mathbf{R}^d$. Denote $C_y$ the member of $\mathcal{C}$ with $y$ on the boundary. $P$ is called $\alpha$-$\mathcal{C}$-symmetric about a point $\theta \in \mathbf{R}^d$ if*

$$P(C_\theta) \geq \alpha, \quad \forall\, C_\theta \in \mathcal{C}$$

*and there is no point $x \in \mathbf{R}^d$ and $\beta > \alpha$ such that*

$$P(C_x) \geq \beta, \quad \forall\, C_x \in \mathcal{C}.$$

**Remark 3.3.9** a) It is easy to see that *HS*-symmetry is one of examples of $\alpha$-$\mathcal{C}$-*symmetry* with $\alpha = 1/2$ and $\mathcal{C} = \mathcal{H}$.

b) Clearly, not every distribution is *HS*-symmetric. However, there always exists an $\alpha$

for any distribution $F$ such that $F$ is *$\alpha$-$\mathcal{C}$-symmetric* about some point $\theta \in \mathbf{R}^d$, although $\theta$ may not be unique.

**Theorem 3.3.13** *Type D depth functions defined in (3.22) are statistical depth functions in the sense of Definition 3.2.1.*

PROOF:

(1). Affine Invariance. For any nonsingular $d \times d$ matrix $A$, any vector $b \in \mathbf{R}^d$ we have

$$X \in C \Longleftrightarrow AX + b \in AC + b.$$

Thus

$$\begin{aligned} D(x; P_X, \mathcal{C}) &= \inf_C \{P(X \in C) \mid x \in C \in \mathcal{C}\} \\ &= \inf_C \{P(AX + b \in AC + b) \mid Ax + b \in AC + b \in A\mathcal{C} + b\} \\ &= D(Ax + b; P_{AX+b}, A\mathcal{C} + b), \end{aligned}$$

that is, $D(x; P, \mathcal{C})$ is affine invariant.

(2). Maximality at Center. Based on Remark 3.3.9, there always exists an $\alpha$ such that $P$ is $\alpha$-$\mathcal{C}$-symmetric about a point $\theta \in \mathbf{R}^d$. Now by (3.22) and Definition 3.3.1 we have that

$$D(\theta; P, \mathcal{C}) = \sup_{x \in \mathbf{R}^d} D(x; P, \mathcal{C}) = \alpha,$$

that is, $D(x; P, \mathcal{C})$ attains maxmium value at the center (the point of symmetry of $\alpha$-$\mathcal{C}$-symmetry) of any distribution.

(3). Monotonicity Relative to the Deepest Point.

**(i)** Let $\theta$ be the deepest point of $P$, that is,

$$D(\theta; P, \mathcal{C}) = \sup_{x \in \mathbf{R}^d} D(x; P, \mathcal{C}).$$

Let $x \neq \theta$ be a point in $\mathbf{R}^d$, $\lambda \in (0,1)$, $x_0 \equiv \lambda\theta + (1-\lambda)x$ and $\beta \equiv D(x; P, \mathcal{C})$. Then by (2) of Theorem 3.3.12 we have

$$x \in D_\beta, \qquad \theta \in D_\beta.$$

The convexity of $D_\beta$ now shows that

$$x_0 \in D_\beta,$$

that is,

$$D(x_0; P, \mathcal{C}) \geq D(x; P, \mathcal{C}).$$

Thus $D(x; P, \mathcal{C})$ monotonically decreases as $x$ moves outward along with any fixed ray originating from the deepest point of any distribution.

**(ii)** It is obvious that

$$D(x; P, \mathcal{C}) \to 0 \quad \text{ as } \|x\| \to \infty.$$

We are done. □

When $\mathcal{C} = \mathcal{H}$, we obtain following result, which is a generalization of Theorem 3.2.1.

**Corollary 3.3.5** *The halfspace depth function HD(x; P) defined in (3.2) is a statistical depth function in the sense of Definition 3.2.1 for any distributions which are* $\alpha$*-*$\mathcal{H}$*-symmetric on* $\mathbf{R}^d$*.*

**Remark 3.3.10** It is not difficult to show from the proof of the above two Theorems that the sample depth funciton $D_n(x; P_n, \mathcal{C})$ of the Type D depth function $D(x; P, \mathcal{C})$ possess the following properties:

(i) upper-semicontinuity;

(ii) affine invariance;

(iii) monotonicity relative to the deepest point;

(iv) uniformly decreasing to zero as $\|x| \to \infty$.

The above list of four types of depth functions is by no means exhaustive, other general structures for depth functions may exist.

Liu et al. (1997) gives seven examples of depth functions. Among them, six have been defined in both population case and sample case. All these six depth functions have been included as special examples of our four types of depth functions except the *likelihood depth function LD(x; F)*. The likelihood depth function *LD(x; F)*, however, satisfies neither the affine invariance property $(P1)$ nor the maximality at center property $(P2)$ or the monotonicity relative to the deepest point property $(P3)$.

Koshevoy and Mosler (1997) introduced a new depth concept, called *zonoid data depth*, based on their *zonoid trimming*. The zonoid data depth function has nice properties, but may fail to satisfy maximality at center property for angularly symmetric or *HS*-symmetric distributions, because zonoid data depth function attains maximum value always at expectation point $E(X)$ for any random variable $X$ on $R^d$. Zonoid data depth is not a "robust" concept of data depth, since a single corrupted data point will move "the center point of zonoid data depth" to infinity.

## 3.4 The Behavior of Sample Depth Functions

As "reasonable" estimators of population depth functions for each fixed $x$, the sample versions of depth functions are desired to be consistent. Donoho and Gasko (1992) proved that the sample version of the halfspace depth function, denoted by $HD_n(x)$, uniformly converges to the halfspace depth function $HD(x;P)$ with probability 1, that is,

$$\sup_{x\in\mathbf{R}^d} \mid HD_n(x) - HD(x;\ P) \mid \to 0 \qquad \text{a.s.} \qquad as\ n \to \infty.$$

Liu (1990) established the almost sure uniform convergence of the sample simplicial depth function $SD_n(x;P)$ to the simplicial depth function *SD(x; P)* for any absolutely continuous distribution $F$ on $\mathbf{R}^d$ with bounded density $f$. Liu and Singh (1993) remarked that the uniform consistency of the sample versions of the majority depth function *MJD(x; P)* and the Mahalanobis depth function $M_hD(x;P)$ holds under proper conditions.

We are going to investigate the almost sure uniform convergence property of the sample versions of the *median depth function MD(x; F)* and the *Type Ddepth functions*.

### 3.4.1 Uniform Consistency of the Sample Median Depth Functions

The following results shall be used in establishing the almost sure uniform convergence of the sample median depth functions.

**Lemma 3.4.1** *Let* $X_1, \ldots, X_n$ *be a random sample from* $X \in \mathbf{R}^d$ *and* $\mathrm{Med}(u'X)$,

$\mathrm{Mad}(u'X)$, $\mathrm{Med}_{1\le i\le n}\{u'X_i\}$ *and* $\mathrm{Mad}_{1\le i\le n}\{u'X_i\}$ *be defined as in Section 3. Then*

$$(1) \qquad \sup_{\|u\|=1} \mathrm{Mad}(u'X) \;\le\; 2 \sup_{\|u\|=1} \mathrm{Med}(|u'X|) < +\infty,$$

$$(2) \quad \sup_{\|u\|=1} \mathrm{Mad}_{1\le i\le n}\{u'X_i\} \;\le\; 2 \sup_{\|u\|=1} \mathrm{Med}_{1\le i\le n}\{|u'X_i|\} < +\infty \;\text{ a.s. } \;\text{as } n \ge N,$$

*for some N.*

PROOF:

(1) This follows immediately from the triangle and the Schwarz inequalities.

(2) Let $M$ be a number such that

$$P(\|X\| > M) < \frac{1}{4}.$$

The definition of sample median implies that,

$$P(\,\mathrm{Med}_{1\le i\le n}\{\|X_i\|\} > M) \le \begin{pmatrix} n \\ \lfloor\frac{n+1}{2}\rfloor \end{pmatrix} (P(\|X_1\|) > M)^{\lfloor\frac{n+1}{2}\rfloor}.$$

Applying Sterling's formula, we obtain

$$\begin{aligned} P(\,\mathrm{Med}_{1\le i\le n}\{\|X_i\|\} > M) &\le \frac{2^n n^n e^{-n}\sqrt{2\pi n}e^{-r(n)}}{n^n e^{-n}(\pi n)e^{-2r(\frac{n}{2})}} (P(\|X_1\|) > M)^{\frac{n}{2}} \\ &= \frac{2e^{2r(\frac{n}{2})-r(n)}}{\sqrt{2\pi n}} \left(\sqrt{4P(\|X_1\|) > M)}\right)^n \\ &\le 2\left(\sqrt{4P(\|X_1\|) > M)}\right)^n, \end{aligned}$$

where $1-\frac{1}{12n+1} < 12n\,r(n) < 1$. Thus

$$\sum_{n=1}^{\infty} P(\,\mathrm{Med}_{1\le i\le n}\{\|X_i\|\} > M) \le 2\sum_{n=1}^{\infty}\left(\sqrt{4P(\|X_1\|) > M)}\right)^n < \infty.$$

Now the Borel-Cantelli lemma reveals that there is a $N$ such that

$$\mathrm{Med}_{1\le i\le n}\{\|X_i\|\} \le M \quad \text{a.s.} \quad \text{for } n \ge N.$$

The Schwarz inequality yields

$$\sup_{\|u\|=1} \mathrm{Med}_{1\le i\le n}\{|u'X_i|\} \le \mathrm{Med}_{1\le i\le n}\{\|X_i\|\} \le M \quad \text{a.s.} \quad \text{for } n \ge N.$$

Hence

$$\begin{aligned}
\sup_{\|u\|=1} \mathrm{Mad}_{1\le i\le n}\{u'X_i\} &\le \sup_{\|u\|=1} \mathrm{Med}_{1\le i\le n}(|u'X_i| + \mathrm{Med}_{1\le i\le n}\{u'X_i\}) \\
&\le 2\sup_{\|u\|=1} \mathrm{Med}_{1\le i\le n}\{|u'X_i|\} \\
&\le 2M < \infty \quad \text{a.s.} \quad \text{for } n \ge N.
\end{aligned}$$

we are done. □

**Lemma 3.4.2** *Let $X_1, \ldots, X_n$ be a sample from $F$ and MD(x; F) and $MD_n(x;\hat{F_n})$ be deifned as in Section 3. Then*

(1) *MD(x; F)* $\longrightarrow 0$ *as* $\|x\| \to \infty$,

(2) $MD_n(x;\hat{F_n}) \xrightarrow{a.s.} 0$ *as* $\|x\| \to \infty$ for sufficiently large n.

PROOF:

(1) This follows directly from the monotonicity property of $MD(x;F)$. Refer to the proof of Theorem **??**.

(2) For any $x \in \mathbf{R}^d$, let $v = \frac{x}{\|x\|}$, then

$$\sup_{\|u\|=1} \frac{|\, u'x - \mathrm{Med}_{1\le i\le n}\{u'X_i\}\,|}{\mathrm{Mad}_{1\le i\le n}\{u'X_i\}} \ge \sup_{\|u\|=1} \frac{|\, |u'x| - \mathrm{Med}_{1\le i\le n}\{|u'X_i|\}\,|}{\mathrm{Mad}_{1\le i\le n}\{u'X_i\}}$$

$$
\begin{aligned}
&\geq \frac{|\,|v'x| - \ \mathrm{Med}_{1\leq i\leq n}\{|v'X_i|\}\,|}{\mathrm{Mad}_{1\leq i\leq n}\{v'X_i\}} \\
&= \frac{|\,\|x\| - \ \mathrm{Med}_{1\leq i\leq n}\{|v'X_i|\}\,|}{\mathrm{Mad}_{1\leq i\leq n}\{v'X_i\}}.
\end{aligned}
$$

Now by Lemma 3.4.1, for sufficiently large $n$ and $\|x\|$, we have

$$
\begin{aligned}
\sup_{\|u\|=1} \frac{|\,u'x - \ \mathrm{Med}_{1\leq i\leq n}\{u'X_i\}\,|}{\mathrm{Mad}_{1\leq i\leq n}\{u'X_i\}} \quad &\geq \quad \frac{\|x\| - \sup_{\|u\|=1}\ \mathrm{Med}_{1\leq i\leq n}\{|u'X_i|\}}{\sup_{\|u\|=1}\mathrm{Mad}_{1\leq i\leq n}\{v'X_i\}} \\
&\xrightarrow{a.s.} \quad \infty \quad \text{as } \|x\| \to \infty.
\end{aligned}
$$

Hence

$$
MD_n(x;\hat{F}_n) \xrightarrow{a.s.} 0 \quad \text{as } \|x\| \to \infty,
$$

for sufficiently large $n$.

We are done. $\square$

**Lemma 3.4.3** *Let $X_1,\ldots,X_n$ be a sample from $F$ and $\mathrm{Med}(u'X)$, $\mathrm{Mad}(u'X)$, $\mathrm{Med}_{1\leq i\leq n}\{u'X_i\}$ and $\mathrm{Mad}_{1\leq i\leq n}\{u'X_i\}$ be deifned as in Section 3.*

(1) *If $\mathrm{Med}(u'X)$ is unique (i.e. $\forall\epsilon > 0$, $P(u'X \leq \mathrm{Med}(u'X) - \epsilon) < \frac{1}{2}$) for any unit vector $u \in \mathbf{R}^d$, then*

$$
\mathrm{Med}_{1\leq i\leq n}\{u'X_i\} \xrightarrow{a.s.} \mathrm{Med}(u'X) \qquad \text{as } n \to \infty.
$$

(2) *If in addition to the assumption of (1), $\mathrm{Mad}(u'X)$ is unique (i.e. $\forall\epsilon > 0$, $P(|u'X - \mathrm{Med}(u'X)| \leq \mathrm{Mad}(u'X) - \epsilon) < \frac{1}{2}$ and $P(|u'X - \mathrm{Med}(u'X)| \leq \mathrm{Mad}(u'X) + \epsilon) > \frac{1}{2}$) for any unit vector $u \in \mathbf{R}^d$, then*

$$
\mathrm{Mad}_{1\leq i\leq n}\{u'X_i\} \xrightarrow{a.s.} \mathrm{Mad}(u'X) \qquad \text{as } n \to \infty.
$$

(3) *If for any* $\epsilon > 0$, $\delta_\epsilon \equiv \sup_{\|u\|=1} P(u'X \leq \mathrm{Med}(u'X) - \epsilon) < \frac{1}{2}$, *then*

$$\sup_{\|u\|=1} \mid \mathrm{Med}_{1\leq i\leq n}\{u'X_i\} - \mathrm{Med}(u'X) \mid \xrightarrow{a.s.} 0 \qquad \textit{as } n \to \infty.$$

(4) *If for any* $\epsilon > 0$, *in addition to the assumption of (3)*, $\delta_\epsilon \equiv \sup_{\|u\|=1} P(|u'X - \mathrm{Med}(u'X)| \leq \mathrm{Mad}(u'X) - \epsilon) < \frac{1}{2}$, *and* $\delta'_\epsilon \equiv \inf_{\|u\|=1} P(|u'X - \mathrm{Med}(u'X)| \leq \mathrm{Mad}(u'X) + \epsilon) > \frac{1}{2}$, *then*

$$\sup_{\|u\|=1} \mid \mathrm{Mad}_{1\leq i\leq n}\{u'X_i\} - \mathrm{Mad}(u'X) \mid \xrightarrow{a.s.} 0 \qquad \textit{as } n \to \infty.$$

PROOF:

(1) This follows directly from Theorem 2.3.1 of Serfling (1980).

(2) Denote "Med", "Mad", "$\mathrm{Med}_{1\leq i\leq n}$" and "$\mathrm{Mad}_{1\leq i\leq n}$" by "$l$", "$s$", "$l_n$" and "$s_n$", respectively. By the triangle inequality,

$$\begin{aligned} s_n\{u'X_i\} &= l_n\{\mid u'X_i - l_n\{u'X_j\} \mid\} \\ &\leq l_n\{\mid u'X_i - l\,(u'X) \mid\} + \mid l\,(u'X) - l_n\{u'X_i\} \mid . \end{aligned}$$

Likewise,

$$l_n\{\mid u'X_i - l\,(u'X) \mid\} \;\leq\; l_n\{\mid u'X_i - l_n\{u'X_j\} \mid\} + \mid l\,(u'X) - l_n\{u'X_i\} \mid .$$

Thus

$$\mid s_n\{u'X_i\} - l_n\{\mid u'X_i - l\,(u'X) \mid\} \mid \;\leq\; \mid l\,(u'X) - l_n\{u'X_i\} \mid .$$

Hence, by the result established in (1), we have

$$\begin{aligned}\lim_{n\to\infty} \mathrm{Mad}_{1\le i\le n}\{u'X_i\} &\overset{a.s.}{=} \lim_{n\to\infty} \mathrm{Med}_{1\le i\le n}\{\mid u'X_i - \mathrm{Med}(u'X) \mid\} \\ &\overset{a.s.}{=} \mathrm{Mad}(u'X).\end{aligned}$$

(3) For any $\epsilon > 0$, let $I_u \equiv P(|l_n\{u'X_i\} - l(u'X)| > \epsilon)$, then

$$I_u = P(l_n\{u'X_i\} > l(u'X) + \epsilon) + P(l_n\{u'X_i\} < l(u'X) - \epsilon).$$

Now applying a similar approach of Theorem 2.3.2 of Serfling (1980), we have

$$\begin{aligned}P(l_n\{u'X_i\}>l(u'X)+\epsilon) &\le P\left(\sum_{i=1}^{n}\mathbf{I}(u'X_i > l(u'X) + \epsilon) \ge \frac{n}{2}\right) \\ &= P\left(\sum_{i=1}^{n}(V_i - EV_i) \ge n\left(\frac{1}{2} - P(u'X > l(u'X) + \epsilon)\right)\right) \\ &= P\left(\sum_{i=1}^{n}(V_i - EV_i) \ge n\left(P(u'X \le l(u'X) + \epsilon) - \frac{1}{2}\right)\right),\end{aligned}$$

where $V_i = \mathbf{I}(u'X > l(u'X) + \epsilon)$, and

$$\begin{aligned}P(l_n\{u'X_i\}<l(u'X)-\epsilon) &\le P\left(\sum_{i=1}^{n}\mathbf{I}(u'X_i < l(u'X) - \epsilon) \ge \frac{n}{2}\right) \\ &= P\left(\sum_{i=1}^{n}(W_i - EW_i) \ge n\left(\frac{1}{2} - P(u'X < l(u'X) - \epsilon)\right)\right),\end{aligned}$$

where $W_i = \mathbf{I}(u'X < l(u'X) - \epsilon)$. Let $\delta'_\epsilon \equiv \inf_{\|u\|=1} P(u'X \le l(u'X) + \epsilon)$ and $\delta^u_\epsilon \equiv \min\left\{P(u'X \le l(u'X) + \epsilon) - \frac{1}{2},\ \frac{1}{2} - P(u'X < l(u'X) - \epsilon)\right\}$. Now utilizing the result of Hoeffding (1963) (see Lemma 2.3.2 of Serfling (1980)), we obtain

$$P(|l_n\{u'X\} - l(u'X)| > \epsilon) \le 2e^{-2n(\delta^u_\epsilon)^2}$$

Write

$$\begin{aligned}
\delta_\epsilon &= \sup_{\|u\|=1}(1 - P(u'X > l(u'X) - \epsilon)) \\
&= 1 - \inf_{\|u\|=1} P(u'X > l(u'X) - \epsilon) \\
&= 1 - \inf_{\|u\|=1} P(-u'X > l(-u'X) - \epsilon) \\
&= 1 - \inf_{\|u\|=1} P(u'X < l(u'X) + \epsilon),
\end{aligned}$$

then

$$\begin{aligned}
\delta'_\epsilon &\geq \inf_{\|u\|=1} P(u'X < l(u'X) + \epsilon) \\
&= 1 - \delta_\epsilon > \frac{1}{2}.
\end{aligned}$$

Hence

$$\begin{aligned}
\delta^u_\epsilon &\geq \min\left\{\delta'_\epsilon - \frac{1}{2}, \frac{1}{2} - \delta_\epsilon\right\} \\
&\equiv \delta > 0.
\end{aligned}$$

Since for any $\epsilon > 0$, there always exists a unit vector $u \in \mathbf{R}^d$ such that

$$P\left(\sup_{\|v\|=1} |l_n\{v'X_i\} - l(v'X)| > \epsilon\right) \leq P(|l_n\{u'X\} - l(u'X)| > \epsilon).$$

We have

$$P\left(\sup_{\|v\|=1} |l_n\{v'X_i\} - l(v'X)| > \epsilon\right) \leq 2e^{-2n\delta^2}.$$

The Borel-Cantelli lemma now implies that

$$\sup_{\|u\|=1} \mid \mathrm{Med}_{1\leq i\leq n}\{u'X_i\} - \mathrm{Med}(u'X) \mid \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.$$

(4) By the proof of (2), we obtain

$$| \, s_n\{u'X_i\} - l_n\{| \, u'X_i - l\,(u'X)\,|\} \, | \;\leq\; | \, l\,(u'X) - l_n\{u'X_i\} \, | \, .$$

Thus, by (3)

$$\begin{aligned}\sup_{\|u\|=1} | \, s_n\{u'X_i\} - l_n\{| \, u'X_i - l\,(u'X)\,|\} \, | \;&\leq\; \sup_{\|u\|=1} | \, l\,(u'X) - l_n\{u'X_i\} \, | \\ &\xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.\end{aligned}$$

Utilizing the same argument of (3), we also have

$$\sup_{\|u\|=1} | \, l_n\{|u'X_i - l(u'X)|\} - s(u'X) \, | \;\xrightarrow{a.s.}\; 0 \quad \text{as } n \to \infty.$$

By the triangle inequility, we have

$$\begin{aligned}\sup_{\|u\|=1} | \, s_n\{u'X_i\} - s(u'X) \, | \;&\leq\; \sup_{\|u\|=1} |s_n\{u'X_i\} - l_n\{|u'X_i - l(u'X)|\}| \\ &\quad + \sup_{\|u\|=1} |l_n\{|u'X_i - l(u'X)|\} - s(u'X)|.\end{aligned}$$

Hence, we have

$$\sup_{\|u\|=1} | \, \mathrm{Mad}_{1\leq i\leq n}\{u'X_i\} - \mathrm{Mad}(u'X) \, | \;\stackrel{a.s.}{\to}\; 0 \quad \text{as } n \to \infty.$$

We are done. □

**Theorem 3.4.1** *Let $X_1, \ldots, X_n$ be a sample from $F$ and MD(x; F) and $MD_n(x; \hat{F}_n)$ be deifned as in Section 3. Suppose in addition to the assumptions of (4) of Lemma 3.4.3,* $\inf_{\|u\|=1} \mathrm{Mad}(u'X) > 0$. *Then*

$$\sup_{x\in\mathbf{R}^d} \left| MD(x; F) - MD_n(x; \hat{F}_n) \right| \xrightarrow{a.s.} 0 \quad \text{as } \|n\| \to \infty.$$

PROOF: By Lemma 3.4.2 we need only to show that

$$\sup_{\|x\|\le M}\left|MD(x;F)-MD_n(x;\hat{F}_n)\right| \xrightarrow{a.s.} 0 \quad \text{as } n\to\infty,$$

for a given $M$.

We have

$$\begin{aligned}
&\left| \sup_{\|u\|=1}\left|\frac{|u'x-l_n\{u'X_i\}|}{s_n\{u'X_i\}}\right| - \sup_{\|u\|=1}\left|\frac{|u'x-l(u'X)|}{s(u'X)}\right| \right| \\
&\le \sup_{\|u\|=1}\left|\frac{|u'x-l_n\{u'X_i\}|}{s_n\{u'X_i\}} - \frac{|u'x-l(u'X)|}{s(u'X)}\right| \\
&\le \sup_{\|u\|=1}\left|\frac{u'x-l_n\{u'X_i\}}{s_n\{u'X_i\}} - \frac{u'x-l(u'X)}{s(u'X)}\right| \\
&\le \sup_{\|u\|=1}\left|\frac{u'x}{s_n\{u'X_i\}} - \frac{u'x}{s(u'X)}\right| + \sup_{\|u\|=1}\left|\frac{l(u'X)}{s(u'X)} - \frac{l_n\{u'X_i\}}{s_n\{u'X_n\}}\right|.
\end{aligned}$$

Denote the first and the second terms in the right hand side of the last inequality by **I** and **II** respectively, then when $\|x\|\le M$

$$\begin{aligned}
\mathbf{I} &= \sup_{\|u\|=1}\frac{|u'x\,(s(u'X)-s_n\{u'X_i\})|}{s_n\{u'X_i\}s(u'X)} \\
&\le \frac{M\Delta_s^n}{\inf_{\|u\|=1}s_n\{u'X_i\}\inf_{\|u\|=1}s(u'X)},
\end{aligned}$$

where $\Delta_s^n=\sup_{\|u\|=1}|s(u'X)-s_n\{u'X_i\}|$, and

$$\begin{aligned}
\mathbf{II} &= \sup_{\|u\|=1}\frac{|(l(u'X)-l_n\{u'X_i\})s_n\{u'X_i\}+l_n\{u'X_i\}(s_n\{u'X_i\}-s(u'X))|}{s_n\{u'X_i\}s(u'X)} \\
&\le \frac{\Delta_l^n\sup_{\|u\|=1}s_n\{u'X_i\}+\sup_{\|u\|=1}|l_n\{u'X_i\}|\,\Delta_s^n}{\inf_{\|u\|=1}s_n\{u'X_i\}\inf_{\|u\|=1}s(u'X)},
\end{aligned}$$

where $\Delta_l^n=\sup_{\|u\|=1}|l(u'X)-l_n\{u'X_i\}|$. Since $\inf_{\|u\|=1}s(u'X)>0$, by Lemma 3.4.1 and Lemma 3.4.3, for any $\epsilon>0$ there is a $K$ such that

$$\sup_{\|x\|\le M}\left| \sup_{\|u\|=1}\left|\frac{|u'x-l_n\{u'X_i\}|}{s_n\{u'X_i\}}\right| - \sup_{\|u\|=1}\left|\frac{|u'x-l(u'X)|}{s(u'X)}\right| \right| \le \mathbf{I}+\mathbf{II}$$

$$\leq \quad K\epsilon,$$

for sufficiently large $n$. The definition of *MD(x; F)* and $MD_n(x; \hat{F_n})$ now implies that

$$
\begin{aligned}
&\sup_{\|x\|\leq M} \left| MD(x;F) - MD_n(x;\hat{F_n}) \right| \\
&\quad \leq \quad \sup_{\|x\|\leq M} \left| \sup_{\|u\|=1} \left| \frac{|u'x - l_n\{u'X_i\}|}{s_n\{u'X_i\}} \right| - \sup_{\|u\|=1} \left| \frac{|u'x - l(u'X)|}{s(u'X)} \right| \right| \\
&\quad \xrightarrow{a.s.} \quad 0 \qquad \text{as } n \to \infty,
\end{aligned}
$$

for any given $M$.

We are done. $\square$

### 3.4.2 The Behavior of the Sample Type D Depth Functions

The following result, due to Vapnik and Chervonenkis (1971) and Steele (1978), will be used in the proof of the almost sure uniform convergence of the sample Type Ddepth functions.

**Lemma 3.4.4** *Let $\mathcal{S}$ be any class of measurable subsets of $\mathbf{R}^d$, $P$ be a probability measure on $\mathbf{R}^d$ and $X_1, \ldots, X_n$ be a random sample from $P$. Then*

$$\sup_{A\in\mathcal{S}} |P_n(A) - P(A)| \xrightarrow{a.s.} 0 \qquad \text{as } n \to \infty$$

*if and only if*

$$\frac{1}{n} E[\log \Delta^{\mathcal{S}}(X_1, \ldots, X_n)] \longrightarrow 0 \qquad \text{as } n \to \infty,$$

*where $P_n$ is the empirical measure of $P$ defined by $P_n(A) \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}(X_i \in A)$ and $\Delta^{\mathcal{S}}(W) \equiv card(\{W \cap A \mid A \in \mathcal{S}\})$ for any finite subset $W$ of $\mathbf{R}^d$.*

**Remark 3.4.1** Vapnik and Chervonenkis (1971) acturally only proved that $P_n(A)$ converges uniformly to $P_n(A)$ in probability. Steele (1978) proved that the uniform convergence of $P_n(A)$ to $P(A)$ holds true with probability 1.

**Theorem 3.4.2** *Let $\mathcal{C}$ be a class of closed and convex Borel sets on $\mathbf{R}^d$, $P$ be a probability measure on $\mathbf{R}^d$ and $D(x; P, \mathcal{C})$ and $D_n(x; P_n, \mathcal{C})$ be defined as in (3.22) and (3.23). Then*

$$\sup_{x \in \mathbf{R}^d} |D_n(x; P_n, \mathcal{C}) - D(x; P, \mathcal{C})| \xrightarrow{a.s.} 0 \quad \textit{as } n \to \infty.$$

PROOF: The class of convex Borel sets on $\mathbf{R}^d$ form a Vapnik-Chervonenkis class (see Vapnik and Chervonenkis (1971) and Steele (1978) ). Now by the Lemma 3.4.4, it is not difficult to see that

$$\begin{aligned} \sup_{x \in \mathbf{R}^d} |D_n(x; P_n, \mathcal{C}) - D(x; P, \mathcal{C})| &= \sup_{x \in \mathbf{R}^d} \left| \inf_{C_x \in \mathcal{C}} P_n(C_x) - \inf_{C_x \in \mathcal{C}} P(C_x) \right| \\ &\leq \sup_{C \in \mathcal{C}} |P_n(C) - P(C)| \\ &\xrightarrow{a.s.} 0 \quad \text{as } n \to \infty, \end{aligned}$$

where $C_x$ is the set $C$ with $x$ on its boundary.

We are done. □

We $\mathcal{C} = \mathcal{H}$, we obtain the result of Donoho and Gasko (1992), the almost sure uniform convergence of the sample halfspace depth function $HD_n(x; P_n)$ to the population halfspace depth function $HD(x; P)$.

**Corollary 3.4.1** *The sample halfspace depth function $HD_n(x; P_n)$ converges uniformly* *to the population halfspace depth function* $HD(x;P)$ *with probability* $1$*, that is*

$$\sup_{x\in\mathbf{R}^d}|HD_n(x;P_n)-HD(x;P)|\xrightarrow{a.s.}0 \qquad as \;\; n\to\infty.$$

## Chapter 4

## STATISTICAL DEPTH CONTOURS

### 4.1 Introduction

Statistical depth functions discussed in Chapter 3 can be immediately utilized to construct "depth contours" or "multivariate quantile contours". Depth contours can provide a good geometric view of the structure of an underlying multivariate distribution and can reveal the shape of multivariate datasets. They are analogous to univariate "$\alpha$th quantile trimmed regions", the intervals $[q_{1-\alpha}, q_\alpha]$, for $\frac{1}{2} \le \alpha < 1$, where $q_p = \inf\{x : F(x) \ge p\}$ for $0 < p < 1$ and a distribution function $F$ on $\mathbf{R}$. Depth contours permit one to generalize the univariate *L-statistics* and *R-statistics* in the multivariate setting. (See Serfling (1980) for discussion of classical *L-statistics* and *R-statistics.*) They also are useful in multivariate robust statistical procedures in connection with generalized multivariate medians and data ordering.

The properties of halfspace depth contours have been studied by many authors, including Eddy (1985), Nolan (1992), Donoho and Gasko (1992) and Masse and Theodorescu (1994). Under some assumptions on depth functions and sample depth functions, He and Wang (1997) provided a convergence result on depth contours of some depth functions for elliptical distributions.

In this chapter, a general definition of depth contour is introduced and properties of depth contours of various depth functions are then explored. Finally, some convergence results are established for depth contours of general depth functions.

### 4.2 Statistical Depth Contours

We mentioned in Section 3.3.4 the depth contour concept for Type D depth functions. Now we give a general definition of depth contours.

**Definition 4.2.1** Let $D(x)$ be any depth function for a given distribution and $D_n(x)$ a sample depth function corresponding to $D(x)$. Define

$$D^\alpha \equiv \{x \in \mathbf{R}^d \mid D(x) \geq \alpha\} \quad \text{and} \quad D_n^\alpha \equiv \{x \in \mathbf{R}^d \mid D_n(x) \geq \alpha\}, \tag{4.1}$$

where $n$ is the sample size and $\alpha > 0$. Then $D^\alpha$ and $D_n^\alpha$ are called the $\alpha$th *depth contour* and *sample* $\alpha$th *depth contour*, respectively. If $D(x)$ is a statistical depth function, then $D^\alpha$ and $D_n^\alpha$ are called the $\alpha$th *statistical depth contour* and *sample* $\alpha$th *statistical depth contour*, respectively.

**Remarks 4.2.1** (1) Strictly speaking, it is more suitable to call $\partial D^\alpha$ and $\partial D_n^\alpha$ the $\alpha$th depth contour and sample $\alpha$th depth contour and $D^\alpha$ and $D_n^\alpha$ the $\alpha$th depth trimmed region and sample $\alpha$th depth trimmed region. When there is no confusion, we will still use the above terminology.

(2) It is not difficult to see that $D^\alpha$ and $D_n^\alpha$ are the analogues of the univariate $\alpha$th quantile trimmed region and the univariate sample $\alpha$th quantile trimmed region in the

multivariate setting. So, $D^\alpha$ and $D_n^\alpha$ are also called the *multivariate* $\alpha$th *quantile contour* and the *multivariate sample* $\alpha$th *quantile contour*.

(3) By the nonnegativity of depth functions, $D^\alpha$ and $D_n^\alpha$ will be given by the whole space $\mathbf{R}^d$ if $\alpha = 0$. Thus we will assume $\alpha > 0$ unless stated otherwise. $\square$

**Definition 4.2.2** A set $E$ in a topological space $X$ is said to be *connected* if $E$ is not the union of two nonempty sets $A$ and $B$ such that

$$\overline{A} \cap B = \phi = A \cap \overline{B},$$

where $\overline{S}$ is the closure of a set $S$.

**Theorem 4.2.1** (i) *The depth contours and the sample depth contours are nested:* $D^{\alpha_1} \subset D^{\alpha_2}$ *and* $D_n^{\alpha_1} \subset D_n^{\alpha_2}$ *if* $\alpha_1 \geq \alpha_2$;

(ii) *the depth contours and the sample depth contours are affine equivariant if the depth function and the sample depth function are affine invariant*;

(iii) *statistical depth contours are connected.*

PROOF: (i) This follows directly from the definitions of $D^\alpha$ and $D_n^\alpha$.

(ii) This follows immediately from the affine invariance property of the depth functions.

(iii) The "monotonicity relative to deepest point" property of statistical depth functions implies that there are no "holes" in $D^\alpha$, and thus it is connected. The proof is complete.

$\square$

The following corollary is a generalization of Proposition 2.5 in Masse and Theodorescu (1994).

**Corollary 4.2.1** *The halfspace depth contours are affine equivariant.*

**Remark 4.2.1** Connectedness is clearly a desired property for depth contours. As we have already seen in Remark 3.2.2, however, the simplicial depth function $SD(x; P)$ may fail to satisfy the "monotonicity relative to deepest point" property for discrete distributions. Therefore the depth contours induced by the simplicial depth function $SD(x; P)$ are not in general connected. □

In Theorem 3.3.4 we have shown that the depth contours of Type D depth functions are compact. This result holds true for the most of the other depth functions discussed in Section 3.3.

**Theorem 4.2.2** *The depth contours are compact for the simplicial depth, the simplicial volume depth* $SVD^{\alpha}(x; F)$ $(\alpha \geq 1)$ *, the* $L_p$ *depth* $(p \geq 1)$*, the projection depth, and the Mahalanobis depth defined in Section* 3.3.

PROOF: a) The continuity of the simplicial volume depth, the $L_p$ depth $(p \geq 1)$, the projection depth, and the Mahalanobis depth implies the closedness of the depth contours of these depth functions. The property that these depth functions monotonically decrease to 0 as $\|x\| \to \infty$ implies the boundedness of their depth contours. The compactness then follows immediately.

b) Theorem 1 of Liu (1990) implies the boundedness of the depth contour of the simplicial depth function. For absolutely continuous distributions on $\mathbf{R}^d$, the closedness of the simplicial depth contours follows from Theorem 2 of Liu (1990). To prove the closedness of the simplicial depth contours for discrete distributions, let $x_n$ be a sequence in $\mathbf{R}^d$ which converges to $x$, assume that $x_n \in D^\alpha$ and $D(x) < \alpha$. Then there is at least one simplex $S[y_1, \ldots, y_{d+1}]$ such that $x_n \in S[y_1, \ldots, y_{d+1}]$ for sufficiently large $n$ and $x \notin S[y_1, \ldots, y_{d+1}]$. But this event occurs with probability 0, since $x_n \to x$ as $n \to \infty$. Thus the closedness of the depth contour holds true for the discrete case. The compactness now follows. The proof is complete. $\square$

**Remark 4.2.2** Compactness is another desired property for depth contours. Applying an argument similar to that used in b) of Theorem 4.2.2, one can prove that the depth contours of the majority depth function $MJD(x; P)$ are closed. But, as shown in Theorem 3.3.3, the majority depth function $MJD(x; P)$ fails to decrease monotonically to 0 as as $\|x\| \to \infty$ . Hence the depth contours of the majority depth are not compact. $\square$

Now let us examine some properties of sample depth contours. We select, as an example, the sample Type D depth contours and discuss their properties. For the Type D depth function $D(x; P, \boldsymbol{\mathcal{C}})$, the *sample* $\alpha$th *depth contour* is

$$D_n^\alpha \equiv \{x \in \mathbf{R}^d \mid D_n(x; \boldsymbol{\mathcal{C}}) \geq \alpha\}. \tag{4.2}$$

We have

**Theorem 4.2.3** *Let $\boldsymbol{\mathcal{C}}$ be a class of closed and connected Borel sets satisfying* (C1), (C2)*and* (C3) *in Section 3.3.4, and let $P$ be a probability measure on $\mathbf{R}^d$. Then the sample depth contours of Type D depth functions are connected and compact.*

PROOF: Following the proof of Theorem 3.3.12, we can establish

$$D_n^\alpha = \{x \in \mathbf{R}^d \mid D_n(x; \boldsymbol{\mathcal{C}}) \geq \alpha\} = \cap\{C \mid P_n(C) > 1 - \alpha, C \in \boldsymbol{\mathcal{C}}\} \tag{4.3}$$

The closedness and connectedness of $D_n^\alpha$ follow. It is not difficult to see that

$$D_n(x; \boldsymbol{\mathcal{C}}) \to 0 \quad \text{ as } \|x\| \to \infty.$$

The compactness of $D_n^\alpha$ now follows immediately. □

When $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{H}}$, we obtain the following result, which slightly extends Lemma 2.2 in Donoho and Gasko (1992).

**Corollary 4.2.2** *The sample depth contours of the halfspace depth function $HD(x; P)$ are connected, convex, compact and nested.*

**Remark 4.2.3** The depth contours $D^\alpha$ and the sample depth contours $D_n^\alpha$ of the halfspace depth function $HD(x; P)$ are called the "$\alpha$-trimmed regions" and the "empirical $\alpha$-trimmed regions" in Nolan (1992). □

Depth contours possess exactly the same shape as that of constant density contours for special distributions such as elliptical distributions. We first give a definition of elliptical distribution (see Muirhead (1982), p. 34).

**Definition 4.2.3** A real vector $X \in \mathbf{R}^d$ is said to have an *elliptical distribution* with parameters $\mu$ and $\Sigma$, denoted by $X \sim E(\mu, \Sigma)$, if its density function is of the form

$$f(x) = c\,|\Sigma|^{-\frac{1}{2}}\, g\left((x-\mu)'\Sigma^{-1}(x-\mu)\right)$$

for some function g, where $\Sigma > 0$ is positive definite.

Liu and Singh (1993) proved for elliptically distributed random vectors that the boundaries of the depth contours of strictly decreasing and affine invariant depth functions are the surfaces of ellipsoids.

**Lemma 4.2.1** (Liu and Singh 1993) *Suppose that $X \in \mathbf{R}^d$ is elliptically distributed, $X \sim E(\mu, \Sigma)$, and that $D(x)$ is affine invariant and strictly decreasing on any ray originating from the center $\mu$. Then the contours $\{x : D(x) = c\}$ are of the form $(x-\mu)'\Sigma^{-1}(x-\mu) = d_c$ for some $d_c$ in $\mathbf{R}$.*

The following result is a generalization of Lemma 4.2.1.

**Lemma 4.2.2** *Suppose that $X \in \mathbf{R}^d$ is elliptically distributed, $X \sim E(\mu, \Sigma)$, and that $D(x)$ is affine invariant and attains maximum value at $\mu$. Then*

(i) *$D^\alpha$ is of the form $D^\alpha = \{x \in \mathbf{R}^d \mid (x-\mu)'\Sigma^{-1}(x-\mu) \le r_\alpha^2\}$ for some $r_\alpha$, and $D(x) = f((x-\mu)'\Sigma^{-1}(x-\mu))$ for some nonincreasing function $f$.*

(ii) *$D(x)$ is strictly decreasing on any ray originating from the center if and only if $\{x \in \mathbf{R}^d | D(x) = \alpha\} = \{x \in \mathbf{R}^d \mid (x-\mu)'\Sigma^{-1}(x-\mu) = r_\alpha^2\}$.*

PROOF: (i) Utilizing an argument similar to that for Lemma 3.1 of Liu and Singh (1993), one can obtain the first part of (i). Since the points on the boundary of $D^\alpha$ possess the same depth, it follows that

$$D(x) = f((x-\mu)'\Sigma^{-1}(x-\mu)).$$

The monotonicity of $f$ follows from the fact that $D(\lambda\mu + (1-\lambda)x) \ge D(x)$, since $(\lambda\mu + (1-\lambda)x) \in D^{\alpha_0}$, where $\alpha_0 = D(x)$.

(ii) Sufficiency follows directly from Lemma 4.2.1. We need to show that $D(x)$ is strictly decreasing if

$$\{x \in \mathbf{R}^d | D(x) = \alpha\} = \{x \in \mathbf{R}^d \mid (x-\mu)'\Sigma^{-1}(x-\mu) = r_\alpha^2\}.$$

Let $y \neq \mu$ be a point in $\mathbf{R}^d$, and $y_0 = \lambda\mu + (1-\lambda)y$ for some $\lambda \in (0,1)$. Then $y \in \partial D^{\alpha_y}$ and $y_0$ is in the interior of $D^{\alpha_y}$ for some $\alpha_y$ such that

$$(y-\mu)'\Sigma^{-1}(y-\mu) = r_{\alpha_y}^2.$$

Hence $D(y_0) > D(y)$. The proof is complete. □

**Remark 4.2.4** The maximality at $\mu$ condition on $D(x)$ in Lemma 4.2.2 could be replaced by a convexity condition on $D^\alpha$, which is also sufficient to prove the necessity part of (ii) in the Lemma. □

**Theorem 4.2.4** *Suppose $X \in \mathbf{R}^d$ is elliptically distributed, $X \sim E(\mu, \Sigma)$. Then the depth contours of the simplicial depth, the majority depth, the simplicial volume depth $SVD^\alpha$ $(\alpha \geq 1)$, the $L^p$ depth $(p \geq 1)$, the projection depth, the Mahalanobis depth, and the Type D depth defined in Section 3.3 are ellipsoids. The boundaries of these depth contours are the surfaces of the ellipsoids.*

PROOF: By Lemma 4.2.2 and the affine invariance of these depth functions, we need only check the strictly decreasing property of these depth functions under the elliptical distribution assumption.

a) An argument similar to that of Theorem 3 of Liu (1990) gives the strict monotonicity property of the simplicial depth function.

b) The strict decreasing property of the majority depth function follows from the fact that

$$
\begin{aligned}
P(\lambda\mu + (1-\lambda)x \in H^m_{X_1,\ldots,X_d}) - P(x \notin H^m_{X_1,\ldots,X_d}) &= \\
P(\lambda\mu + (1-\lambda)x \in H^m_{X_1,\ldots,X_d},\ x \notin H^m_{X_1,\ldots,X_d}) > 0,
\end{aligned}
$$

for any $\lambda \in (0,1)$ and $x \neq \mu$ in $\mathbf{R}^d$.

c) For the simplicial volume depth function $SVD^\alpha(x;F)$ ($\alpha \geq 1$), following the proof of Corollary 3.3.1, we have

$$
\Delta(x_0; x_1, \ldots, x_d) \leq \lambda\Delta(\mu; x_1, \ldots, x_d) + (1-\lambda)\Delta(x; x_1, \ldots, x_d),
$$

for any $\lambda \in (0,1)$, $x_0 = \lambda\mu + (1-\lambda)x$, $x \neq \mu$ and $x, x_1, \ldots, x_d$ in $\mathbf{R}^d$. And

$$
P(\Delta(x_0; X_1, \ldots, X_d) < \lambda\Delta(\mu; X_1, \ldots, X_d) + (1-\lambda)\Delta(x; X_1, \ldots, X_d)) > 0,
$$

for a random sample $X_1, \ldots, X_d$ from $X$. The convexity of the function $x^\alpha$ for $0 < x < \infty$ and the maximality of $SVD^\alpha(x;F)$ at $\mu$ now imply that

$$
\Delta^\alpha(x_0; x_1, \ldots, x_d) \leq \Delta^\alpha(x; x_1, \ldots, x_d),
$$

and

$$
P\left(\Delta^\alpha(x_0; X_1, \ldots, X_d) < \Delta^\alpha(x; X_1, \ldots, X_d)\right) > 0.
$$

Hence, $SVD^\alpha(x;F) < SVD^\alpha(x_0;F)$. The strict decreasing property of the simplicial volume depth function thus follows.

d) For the $L^p$ depth function $L^p(x; F)$ $(p \geq 1)$, Minkowski's inequality implies that

$$\|\lambda\mu + (1-\lambda)x - x_1\| \leq \lambda\|\mu - x_1\| + (1-\lambda)\|x - x_1\|$$

and

$$P(\|\lambda\mu + (1-\lambda)x - X\| < \lambda\|\mu - X\| + (1-\lambda)\|x - X\|) > 0,$$

for any $\lambda \in (0, 1)$, $x \neq \mu$ and $x, x_1$ in $\mathbf{R}^d$. Hence

$$L^p(\lambda\mu + (1-\lambda)x; F) > L^p(x; F).$$

The strict monotonicity property of the $L^p$ depth function thus follows.

e) The strict decreasing property of the projection depth function follows immediately from the fact that $\mathrm{Med}(u'X) = u'\mu$ for any unit vector $u$ in $\mathbf{R}^d$.

f) Following the proof of (3) of Theorem 3.3.11, we have

$$O(x_0; X) < \lambda O(\mu; X) + (1-\lambda)O(x; X),$$

for any $\lambda \in (0, 1)$, $x \neq \mu$ in $\mathbf{R}^d$ and $x_0 = \lambda\mu + (1-\lambda)x$. Hence $MHD(x_0; F) > MHD$(x; F). The strict decreasing property of the Mahalanobis depth function thus follows.

g) Let $\lambda \in (0, 1)$, $x \neq \mu$ in $\mathbf{R}^d$ and $x_0 = \lambda\mu + (1-\lambda)x$. To consider the depth of point $x_0$ and $x$, we need only take the infimum of $P(C)$ over all $C \in \mathcal{C}$ which do not contain the center $\mu$. Now for any such $C_{x_0}$, by the condition (C2) in Section 3.3.4 there exists a $C_x$ such that $C_x \subset C_{x_0}$ and $P(C_x) < P(C_{x_0})$. Hence, $D(x; P, \boldsymbol{\mathcal{C}}) > D(x_0; P, \boldsymbol{\mathcal{C}})$. The strict monotonicity property of Type D depth functions thus follows. This completes the proof. $\square$

**Theorem 4.2.5** *Suppose that $X \in \mathbf{R}^d$ is elliptically distributed, $X \sim E(\mu, \Sigma)$, and that $D(x)$ is affine invariant. Then $D(x)$ strictly decreases as $x$ moves away from the center $\mu$ along any ray if and only if*

$$D(x) = f\left((x-\mu)'\Sigma^{-1}(x-\mu)\right),$$

*for some strictly decreasing continuous function f.*

PROOF: The necessity is trival. We need only prove the sufficiency.

a) By Lemma 4.2.2, there is a function $f$ such that

$$D(x) = f\left((x-\mu)'\Sigma^{-1}(x-\mu)\right).$$

b) To show that $f$ is strictly decreasing, let $q_2 > q_1 > 0$. Then there is an $x \in \mathbf{R}^d$, an $\alpha \in (0,1)$ and an $x_0 = \alpha\mu + (1-\alpha)x$ such that

$$q_1 = (x_0-\mu)'\Sigma^{-1}(x_0-\mu) \qquad \text{and} \qquad q_2 = (x-\mu)'\Sigma^{-1}(x-\mu).$$

Now $f(q_1) = D(x_0) > D(x) = f(q_2)$, proving that f is strictly decreasing.

c) To show that $f$ is continuous, we note, by Lemma 4.2.2, that $D(x)$ is upper and lower semicontiuous. Since $(x-\mu)'\Sigma^{-1}(x-\mu)$ is also continuous, the continuity of $f$ follows from Lemma 4.2.3. This completes the proof. □

**Lemma 4.2.3** *Let X, Y and be topological spaces, g: $X \rightarrow Y$ be continuous, and $h = f \circ g$: $X \rightarrow Y$ be continuous. Then f: $Y \rightarrow Y$ is continuous.*

PROOF: If $V$ is open in $Y$ and $f(x_0) \in V$ for some point $x_0 \in Y$, then there exist open sets $W$ and $U$ in $X$ such that $y_0 \in W \cap U$, $g(y_0) = x_0$ and $g(W) \cup h(U) \subset V$. It follows that $f(g(U \cup W)) = h(U \cup W) \subset V$ and $x_0 \in g(U \cap W)$, proving the continuity of f. □

The following lemma, which has some alternative assumptions, can also be used to prove the continuity of $f$ in Theorem 4.2.5.

**Lemma 4.2.4** *Suppose that* $g$*:* $\mathbf{R}^d \to \mathbf{R}$ *is continuous.*

**(i)** *Let* $f$*:* $\mathbf{R} \to \mathbf{R}$ *be nonincreasing.*

(1) *If* $f \circ g$ *is upper semicontinuous, then* $f$ *is left continuous;*

(2) *if* $f \circ g$ *is lower semicontinuous, then* $f$ *is right continuous.*

**(ii)** *Let* $f$*:* $\mathbf{R} \to \mathbf{R}$ *be nondecreasing.*

(1) *If* $f \circ g$ *is upper semicontinuous, then* $f$ *is right continuous;*

(2) *if* $f \circ g$ *is lower semicontinuous, then* $f$ *is left continuous.*

PROOF: We only prove (1) of (i); the proof for the other statements is similar.

Let $x_n \uparrow x$. The continuity of g implies that there is a sequence $y_n$ and a point $y$ in $\mathbf{R}^d$ such that $y_n \to y$ and $g(y_n) = x_n$, $g(y) = x$. The continuity of $g$, the nonincreasing property of $f$, and the upper semicontinuous property of $f \circ g$ imply that

$$f(x-) \leq f(x_n) = f(g(y_n)) < f(g(y)) + \epsilon = f(x) + \epsilon \leq f(x-) + \epsilon.$$

Hence $f(x-) = f(x)$. □

## 4.3 The Behavior of Sample Depth Contours in General

We first establish, in a very general setting, an *almost sure result* about sample depth contours.

**Theorem 4.3.1** *Let $D(x)$ be any nonnegative depth function and $D_n(x)$ a corresponding sample depth function. Let $D^\alpha(x)$ and $D_n^\alpha(x)$ be defined as in Definition 4.2.1. Assume that*

(C1) $D(x) \to 0$, *as* $\|x\| \to \infty$,

(C2) $\sup_{x \in S} |D_n(x) - D(x)| \xrightarrow{a.s.} 0$, *for any bounded set $S \subset \mathbf{R}^d$.*

*Then, for any $\epsilon > 0$, $\delta < \epsilon$, $\alpha \geq 0$, and $\alpha_n \to \alpha$,*

$$D^{\alpha+\epsilon} \subset D_n^{\alpha_n+\delta} \subset D_n^{\alpha_n} \subset D_n^{\alpha_n-\delta} \subset D^{\alpha-\epsilon} \quad a.s.$$

*for sufficiently large $n$, and uniformly if $\alpha_n \to \alpha$ uniformly as $n \to \infty$.*

PROOF: (I) Clearly, $D_n^{\alpha_n+\delta} \subset D_n^{\alpha_n} \subset D_n^{\alpha_n-\delta}$.

(II) To show that $D_n^{\alpha_n-\delta} \subset D^{\alpha-\epsilon}$, assume that $\alpha - \epsilon > 0$ (the inclusion relation holds trivally when $\alpha - \epsilon \leq 0$, since then $D^{\alpha-\epsilon} = \mathbf{R}^d$). Since $\alpha_n \to \alpha$, there is an $N_1$ such that when $n \geq N_1$

$$|\alpha_n - \alpha| < \frac{\epsilon - \delta}{2}.$$

Since by (C1), $D(\lambda x)$ uniformly approaches 0 as $\lambda$ increases to $\infty$, there exists a bounded region $S \subset \mathbf{R}^d$ such that $D^{\alpha-\epsilon} \subset D^{\alpha-2\epsilon} = S$. By (C2), there is an $N_2(\geq N_1)$ such that when $n \geq N_2$

$$(*) \qquad \sup_{x \in S} |D_n(x) - D(x)| \leq \frac{\epsilon - \delta}{2} \quad \text{a.s..}$$

Let $x \in D_n^{\alpha_n-\delta} \cap (D^{\alpha-2\epsilon} - D^{\alpha-\epsilon})$. Then when $n \geq N_2$

$$D_n(x) - D(x) > \alpha_n - \delta - (\alpha - \epsilon) \geq \alpha - \frac{\epsilon - \delta}{2} - \delta - (\alpha - \epsilon) \geq \frac{\epsilon - \delta}{2},$$

contradicting $(*)$. Thus either $D_n^{\alpha_n-\delta} \subset D^{\alpha-\epsilon}$ or $D_n^{\alpha_n-\delta} \cap D^{\alpha-2\epsilon} = \Phi$. But the latter is impossible, for if it is true and we let $x \in D_n^{\alpha_n-\delta}$, then when $n \geq N_2$

$$D_n(x) - D(x) > \alpha_n - \delta - (\alpha - 2\epsilon) \geq \alpha - \frac{\epsilon - \delta}{2} - \delta - (\alpha - 2\epsilon) \geq \frac{3(\epsilon - \delta)}{2},$$

for any $n$, violating (C2). Hence $D_n^{\alpha_n-\delta} \subset D^{\alpha-\epsilon}$.

(III) Applying a similar argument as for (II), one can show that

$$D^{\alpha+\epsilon} \subset D_n^{\alpha_n+\delta}.$$

The proof is complete. □

**Corollary 4.3.1** *Suppose, in addition to the assumptions of Theorem 4.3.1, that* $P(\{x \in \mathbf{R}^d | D(x) = \alpha\}) = 0$. *Then*

$$D_n^{\alpha_n} \xrightarrow{a.s.} D^\alpha \quad as \;\; n \to \infty.$$

*The convergence is unform in $\alpha$ if $\alpha_n \to \alpha$ uniformly as $n \to \infty$.*

PROOF: It is easy to see that

$$\{x \in \mathbf{R}^d | D(x) > \alpha\} = \bigcup_{\epsilon \in Q^+} D^{\alpha+\epsilon} \subset \bigcap_{\epsilon \in Q^+} D^{\alpha-\epsilon} = \{x \in \mathbf{R}^d | D(x) \geq \alpha\},$$

where $Q^+$ is the set of positive rational numbers. By Theorem 4.3.1, we can show that

$$\bigcup_{\epsilon \in Q^+} D^{\alpha+\epsilon} \subset \liminf_{n\to\infty} D_n^{\alpha_n} \subset \limsup_{n\to\infty} D_n^{\alpha_n} \subset \bigcap_{\epsilon \in Q^+} D^{\alpha-\epsilon} \quad \text{a.s.}$$

$P(\{x \in \mathbf{R}^d | D(x) = \alpha\}) = 0$ then implies that

$$\lim_{n\to\infty} D_n^{\alpha_n} \overset{a.s.}{=} D^\alpha.$$

The proof is complete. $\square$

Applying the general result about depth contours stated in Theorem 4.3.1 to elliptical distributions and affine invariant depth functions, we obtain

**Corollary 4.3.2** *Suppose $D(x)$ is nonnegative and affine invariant. Suppose also that*

(0) $X \sim E(\mu, \Sigma)$,

(1) $D(x) \to 0$ *as* $\|x\| \to \infty$,

(2) $\sup_{x \in S} |D_n(x) - D(x)| \xrightarrow{a.s.} 0$ *as* $n \to \infty$, *for any bounded set* $S \subset \mathbf{R}^d$,

(3) $D_n^\alpha$ *is convex and closed.*

*Then*

(4) *the identity* $\{x \in \mathbf{R}^d | D(x) = \alpha\} = \{x \in \mathbf{R}^d | e(x) = r_\alpha^2\}$, *for some* $r_\alpha \in \mathbf{R}$ *and* $e(x) = (x-\mu)'\Sigma^{-1}(x-\mu)$, *holds if and only if for any* $\alpha \in (0,1)$ *and* $\epsilon > 0$, *there exists* $\delta > 0$ *such that for sufficiently large* $n$

$$(*) \qquad D^{h(q(\alpha-\epsilon))} \subset D_n^{\beta_n(\alpha)+\delta} \subset D_n^{\beta_n(\alpha)-\delta} \subset D^{h(q(\alpha+\epsilon))} \quad a.s.\ ,$$

*and uniformly in* $\alpha \in [0, \alpha_0]$ *for* $\alpha_0 < 1$, *where* $h(x)$, $q(\alpha)$ *and* $\beta_n$ *are defined by* $P(\{x \in \mathbf{R}^d \mid e(x) \le q(\alpha)\}) = \alpha$, $P_n(\{x \in \mathbf{R}^d \mid D_n(x) \ge \beta_n(\alpha)\}) = \lfloor \alpha n \rfloor$, *and* $D(x) = h(e(x))$.

PROOF: (I) *Sufficiency.* Convexity of $D_n^\alpha$ and (2) imply the convexity of $D^\alpha$. Remark 4.2.4 and Lemma 4.2.2 now show that $D(x)$ is strictly decreasing as $x$ moves away from the $\mu$ along any fixed ray. Theorem 4.2.5 then shows that $h(x)$ is strictly decreasing and continuous. By Lemma 3 of He and Wang (1997), $\lim_{n\to\infty} \beta_n(\alpha) = h(q(\alpha))$ uniformly in $\alpha$. The continuity and monotonicity of $q(\alpha)$ and $h(x)$ imply that $h(q(\alpha+\epsilon)) =$

$h(q(\alpha)) - f_1(\epsilon)$ and $h(q(\alpha - \epsilon)) = h(q(\alpha)) + f_2(\epsilon)$ for some continuous and positive functions $f_1(\epsilon)$ and $f_2(\epsilon)$. Sufficiency now follows from Theorem 4.3.1.

(II) *Necessity.* By Remark 4.2.4 and Lemma 4.2.2, we need only show that $h(x)$ is strictly decreasing as $x$ moves away from $\mu$ along any fixed ray. The nonincreasing property of $h(x)$ follows from $(*)$. Assume that $e(x) = q(\alpha + \epsilon)$, $e(x_0) = q(\alpha - \epsilon)$, $D(y) = \beta_n(\alpha)$ and

$$(**) \qquad D^{h(q(\alpha - \frac{\epsilon}{2}))} \subset D_n^{\beta_n(\alpha)+\delta} \subset D_n^{\beta_n(\alpha)-\delta} \subset^{h(q(\alpha + \frac{\epsilon}{2}))},$$

for sufficiently large $n$. By (2) and $(**)$, we have $h(q(\alpha - \epsilon) = D(x_0) \geq D_n(x_0) - \frac{\delta}{2} \geq D_n(y) + \delta - \frac{\delta}{2} \geq D_n(x) + 2\delta - \frac{\delta}{2} \geq D(x) - \frac{\delta}{2} + 2\delta - \frac{\delta}{2} = D(x) + \delta = h(q(\alpha + \epsilon) + \delta$, for sufficiently large $n$. By the continuity and monotonicity of $q(\alpha)$, we conclude that $h(x)$ is strictly decreasing. Necessity thus follows. $\square$

**Remarks 4.3.1**

a) He and Wang (1997) proved the following. Assume, for nonnegative and affine invariant $D(x)$, that

(0) $X \sim E(\mu, \Sigma)$,

(1) $D(x) \to 0$ as $\|x\| \to \infty$,

(2) $\lim_{n\to\infty} \sup_{x \in S} |D_n(x) - D(x)| = 0$ a.s. for any compact set $S \subset \mathbf{R}^d$,

(3) $D_n^{\alpha}$ is convex and closed,

(4) $\{x \in \mathbf{R}^d | D(x) = \alpha\} = \{x \in \mathbf{R}^d | e(x) = r_\alpha^2\}$ for some $r_\alpha \in \mathbf{R}$ and $e(x) = (x - \mu)'\Sigma^{-1}(x - \mu)$, and

(5) $D_n(x)$ attains maximum value at $\mu$.

Then

(6) $D(x)$ is a strictly monotone function of $e(x)$ (which implies that for any $c > 0$, $P(\{x : D(x) = c\} = 0)$ holds

if and only if for any $\alpha \in (0, 1)$ and $\epsilon > 0$, there exists $\delta > 0$ such that as $n \to \infty$,

$$(***) \qquad D^{h(q(\alpha-\epsilon))} \subset D_n^{\beta_n+\delta} \subset D_n^{\beta_n-\delta} \subset D^{h(q(\alpha+\epsilon))} \quad a.s. .$$

and uniformly in $\alpha \in [0, \alpha_0]$, $\alpha_0 < 1$, where $h(x)$, $q(\alpha)$ and $\beta_n$ are defined as $P(\{x \in \mathbf{R}^d \mid e(x) \leq q(\alpha)\}) = \alpha$, $P_n(\{x \in \mathbf{R}^d \mid D_n(x) \geq \beta_n(\alpha)\}) = \lfloor \alpha n \rfloor$ and $D(x) = h(e(x))$.

Although conditions (1) and (5) are not explicitly stated in He and Wang (1997), they implicitly used them in their proof, so we have listed them here.

b) Corollary 4.3.2 is an improvement and extension of the main result of He and Wang (1997), since it establishes $(***)$ only under the assumptions (0)–(4). Also, condition (2) of Corollary 4.3.2 is weaker than condition (2) of He and Wang (1997).

c) On the other hand, condition (6) of He and Wang (1997) seems to be redundant, since convexity of $D_n^\alpha$ and condition (2) imply the convexity of $D^\alpha$, which, combined with condition (4), Remark 4.2.4 and Lemma 4.2.2, implies condition (6). $\square$

We now turn to the almost sure convergence of sample depth contours of some specific depth functions. We select, as examples, the simplicial depth, the projection depth, and the general Type D depth, and investigate their contour convergence.

**Theorem 4.3.2** *Suppose $X \in \mathbf{R}^d$ is elliptically distributed, $X \sim E(\mu, \Sigma)$. Then for* *the simplicial depth, the projection depth, and the general Type D depth, we have*

$$\lim_{n\to\infty} D_n^{\alpha_n} \overset{a.s.}{=} D^{\alpha},$$

*for any sequence $\alpha_n$ with $\alpha_n \to \alpha$ as $n \to \infty$ and random sample $X_1, \ldots, X_n$ from $X$, where $D^{\alpha}$ is an ellipsoid of the same shape as that of the constant density contours of the parent distribution. Futher, the convergence is uniform in $\alpha$ if $\alpha_n \to \alpha$ uniformly as $n \to \infty$.*

PROOF: By results in Chapter 3, the depth functions $SD(x; P)$, $PD(x; F)$ and $D(x; P, \boldsymbol{\mathcal{C}})$ satisfy (C1) of Theorem 4.3.1, and the almost sure uniform convergence of sample depth functions to population depth functions also holds for the sample projection depth function $PD_n(x)$ and the sample Type D depth function $D_n(x; \boldsymbol{\mathcal{C}})$. The almost sure uniform convergence of the sample simplicial depth function also holds (see, e.g., Corollary 6.8 of Arcones and Gine (1993)). Now, by Theorem 4.2.4, the condition in Corollary 4.3.1 is also satisfied for all three of these depth functions, and the $D^{\alpha}$'s are ellipsoids. Thus the proof is complete. □

**Remark 4.3.1** It is not difficult to see that the contours in Theorem 4.3.2 satisfy

$$\lim_{n\to\infty} \rho(D_n^{\alpha_n}, D^{\alpha}) \overset{a.s.}{=} 0,$$

and uniformly in $\alpha$ if $\alpha_n \to \alpha$ uniformly as $n \to \infty$, where $\rho$ represents the *Hausdorff distance* between two sets, that is,

$$\rho(A, B) = \inf\{\epsilon \mid \epsilon > 0,\ A \subset B^{\epsilon}, B \subset A^{\epsilon}\},$$

where $A^\epsilon = \{x \,|d(x,A) < \epsilon\}$ and $d(x,A) = \inf\{d(x,y) \,|y \in A\}$. $\square$

When $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{H}}$ in the Type D depth functions, we obtain the following almost sure convergence result for sample halfspace depth contours, which is a slight extension of Donoho and Gasko (1992), Lemma 2.5.

**Corollary 4.3.3** *Suppose $X \in \mathbf{R}^d$ is elliptically distributed, $X \sim E(\mu, \Sigma)$. Then the sample depth contours $D^{\frac{\lfloor \alpha n \rfloor}{n}}$ of the halfspace depth converge almost surely and uniformly in $\alpha \in [0, \frac{1}{2}]$, as $n \to \infty$, to $D^\alpha$, an ellipsoid of the same shape as that of the parent distribution.*

For the multivariate normal distribution, a special case of elliptical distribution, we have

**Corollary 4.3.4** *Suppose $X \in \mathbf{R}^d$ is normally distributed, $X \sim N(\mu, \Sigma)$. Then for the halfspace depth,*

$$D^{\frac{\lfloor \alpha n \rfloor}{n}} \xrightarrow{a.s.} D^\alpha = \left\{ x \in \mathbf{R}^d \mid (x-\mu)'\Sigma^{-1}(x-\mu) \le \left(\Phi^{-1}(1-\alpha)\right)^2 \right\}, \tag{1}$$

*and uniformly in $\alpha \in [0, \frac{1}{2}]$, where $\Phi^{-1}(p)$ denotes the* pth *quantile of the standard normal distribution. Also*

$$P\left(D^{\frac{\lfloor \alpha n \rfloor}{n}}\right) \xrightarrow{a.s.} 1-\beta, \tag{2}$$

*and uniformly in $\alpha \in [0, \frac{1}{2}]$, where $\beta$ is determined by $(\Phi^{-1}(1-\alpha))^2 = \chi^2_d(\beta)$, and $\chi^2_d(p)$ denotes the pth quantile of the chi-square distribution with $d$ degree of freedom.*

PROOF: (1) Suppose $Y \in \mathbf{R}^d$ is normally distributed and $Y \sim N(0, I)$. Then, by affine invariance of the halfspace depth, it is not difficult to see that the depth contour $D^\alpha$ under $Y$ is a sphere with radius $\Phi^{-1}(1-\alpha)$. Let $X = \Sigma^{\frac{1}{2}} Y + \mu$. Then $X \sim N(\mu, \Sigma)$ and the affine invariance implies that

$$D^\alpha = \{x \in \mathbf{R}^d \mid (x-\mu)'\Sigma^{-1}(x-\mu) \leq \left(\Phi^{-1}(1-\alpha)\right)^2\},$$

which, combined with Corollary 4.3.3, gives (1).

(2) Since

$$\begin{aligned} P\left(\liminf_{n\to\infty} D^{\frac{\lfloor \alpha n \rfloor}{n}}\right) &\leq \liminf_{n\to\infty} P\left(D^{\frac{\lfloor \alpha n \rfloor}{n}}\right) \\ &\leq \limsup_{n\to\infty} P\left(D^{\frac{\lfloor \alpha n \rfloor}{n}}\right) \leq P\left(\limsup_{n\to\infty} D^{\frac{\lfloor \alpha n \rfloor}{n}}\right), \end{aligned}$$

it follows that

$$\begin{aligned} \lim_{n\to\infty} P\left(D^{\frac{\lfloor \alpha n \rfloor}{n}}\right) &\overset{a.s.}{=} P\left(\lim_{n\to\infty} D^{\frac{\lfloor \alpha n \rfloor}{n}}\right) \\ &\overset{a.s.}{=} P\left((X-\mu)'\Sigma^{-1}(X-\mu) \leq \left(\Phi^{-1}(1-\alpha)\right)^2\right). \end{aligned}$$

Now since $X \sim N(\mu, \Sigma)$, we have $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2_d$, and thus (2) follows. $\square$

Applying Corollary 4.3.4 for $X \sim N(0, I)$, and Corollary 4.3.1, we obtain Theorem 1 of Yeh and Singh (1997).

**Corollary 4.3.5** *Suppose $F$ is absolutely continuous in $\mathbf{R}^d$ and $E\|X\|^2 < \infty$, then*

$$\partial C \subseteq \liminf_{n\to\infty}(W^*_{n,1-\alpha}) \subseteq \limsup_{n\to\infty}(W^*_{n,1-\alpha}) \subseteq C,$$

*for $C = \{x \in \mathbf{R}^d \mid \|x\| \leq \gamma_{1-\alpha}\}$ and some $\gamma_{1-\alpha}$ such that $P(N(0, I) > \gamma_{1-\alpha}) = \alpha$. Here $W^*_{n,1-\alpha}$ is a $100(1-\alpha)\%$ bootstrap confidence region obtained by first deleting $100\alpha\%$*

*exterior bootstrap points $Z_n^*$ based on halfspace depth and then forming the convex hull of the remaining points,*

$$Z_n^* = n^{\frac{1}{2}} S_{\hat{\theta}_n}^{-\frac{1}{2}} (\hat{\theta}_n^* - \hat{\theta}_n),$$

$\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ *is a consistent estimator of a d-dimensional parameter of interest,* $S_{\hat{\theta}_n}$ *is a consistent estimator of the dispersion matrix of* $\hat{\theta}_n$*, and the asterisk is used to denote the statistic computed under the bootstrap sample.*

## 4.4 Summary

In this chapter, a general definition of depth contour has been introduced and properties of depth contours of various depth functions explored. Convergence results have been established for sample depth contours in a very general setting and developed for some specific cases. Some results obtained here improve and generalize recent results in the literature.

## Chapter 5

## DEPTH-BASED LOCATION MEASURES

### 5.1 Introduction

It has long been known that in one dimension the median, regarded as the center of a given distribution, is one of the most favorable robust location measures for the distribution. One would naturally suppose the same to be true in higher dimensions, making the higher dimensional median a natural location measure for multivariate distributions. Questions we consider here are: (1) What are desirable properties that multivariate location measures should satisfy? (2) How to generalize the concept of the univariate median into higher dimensional settings? (3) Do the generalized medians satisfy desirable properties for multivariate location measures?

In Section 2 of this chapter, desirable properties for multivariate nonparametric location measures are examined. A new property for multivariate nonparametric location measures, called "center locating" condition, is introduced. Relationships among conditions for location measures are explored. It turns out that the "center locating" condition is a favorable requirement for multivariate nonparametric location measures. The "stochastic order preserving" condition required in Bickel and Lehmann (1975) and Oja (1983) for location measures does not guarantee that the measures are able to identify the locations of underlying distributions.

In Section 3, statistical depth functions, which are inherently bound up with the notion of center of a multivariate distribution, are applied to introduce corresponding notions of multivariate median.

Section 4 is devoted to the performance analysis of some multivariate nonparametric location measures with respect to the "center locating" condition. Various multivariate medians derived from depth functions, as the candidates for multivariate nonparametric location measures, are studied. It is found, for example, that for halfspace symmetric distributions the Liu simplicial median fails to satisfy the "center locating" condition and, consequently, should be used with caution as a multivariate location measure. On the other hand, the halfspace median, the $L_1$ median, the $\tilde{L}_2$ median and the projection median are found to be good choices as multivariate nonparametric location measures, for halfspace symmetric distributions.

## 5.2 Properties for Multivariate Nonparametric Location Measures

### 5.2.1 Conditions of Bickel and Lehmann (1975) and Oja (1983)

Bickel and Lehmann (1975) introduced a definition of location measure on $\mathbf{R}$, and Oja (1983) extended the definition to higher dimensions.

Let $\boldsymbol{\mathcal{P}}$ be a class of probability distributions in $\mathbf{R}^d$ and $\psi\colon \boldsymbol{\mathcal{P}} \to \mathbf{R}^d$ be a functional on $\boldsymbol{\mathcal{P}}$. It is convenient to write $\psi(X)$ for $\psi(F)$ when $F \in \boldsymbol{\mathcal{P}}$ is the distribution function of the random vector $X \in \mathbf{R}^d$. For a functional $\psi : \boldsymbol{\mathcal{P}} \to \mathbf{R}^d$ to be a multivariate nonparametric location measure, Bickel and Lehmann (1975) and Oja (1983) require the following two conditions:

**(C1)** $\psi$ is affine equivariant, that is, $\psi(AX+b) = A\psi(X)+b$, for any $d \times d$ nonsingular matrix $A$ and vector $b \in \mathbf{R}^d$,

**(C2)** $\forall\, P, Q \in \boldsymbol{\mathcal{P}}$, if $P \overset{st}{\leq} Q$, then $\psi(P) \leq \psi(Q)$,

where $P \overset{st}{\leq} Q$ denotes that $P$ is *stochastically smaller* than $Q$, i.e., that $\int f\, dP \leq \int f\, dQ$ for any real bounded coordinatewise increasing $f$ on $\mathbf{R}^d$, and $x = (x_1, \ldots, x_d)' \leq y = (y_1, \ldots, y_d)'$ denotes that $x_i \leq y_i$ for $i = 1, 2, \ldots, d$.

**Remarks 5.2.1** (1) Condition (C1) is a typical requirement on location measures. It requires that the location measure $\psi$ should not depend on the underlying coordinate system, nor on the scales of the underlying measurements. Note that (C1) really means that $\psi(F_{AX+b}) = A\psi(F_X) + b$ for any $d \times d$ nonsingular matrix $A$ and vector $b \in \mathbf{R}^d$, where $F_X$ denotes the distribution of a random vector $X \in \mathbf{R}^d$. On the other hand, condition (C2) requires that the location measure $\psi$ should take on larger values for random vectors which typically are "larger". (C2) is often called a "stochastic order preserving" condition.

(2) It is worth noting that satisfaction of (C1) and (C2) does not guarantee that the location measure $\psi$ indentfies the location of a given distribution. For example, under the definitions of Bickel and Lehmann (1975) and Oja (1983), the mean functional $E[X]$ on $\boldsymbol{\mathcal{P}}$ is always a location measure; see Theorem 5.2.1. The mean functional, however, may not always be a good location measure for some distributions like angularly symmetric

distributions as shown in Examples 5.2.1, and especially may be poor for asymmetric distributions. Also, it has long been recognized that the mean functional is not robust. $\square$

According to the definitions of Bickel and Lehmann (1975) and Oja (1983), we have

**Theorem 5.2.1** *The mean functional on* $\boldsymbol{\mathcal{P}}$ *satisfies* $(C1)$ *and* $(C2)$ *and thus is always a location measure in the sense of Bickel and Lehmann* (1975) *and Oja* (1983).

PROOF: Assume $\psi(F_X) = E[X]$ for any random vector $X$ with its distribution $F_X$ in $\boldsymbol{\mathcal{P}}$. Then clearly $\psi$ satisfies (C1). Now suppose $F, G \in \boldsymbol{\mathcal{P}}$ and $F \overset{st}{\leq} G$. Then by a result of Strassen (1965), there are two $\mathbf{R}^d$-valued random vectors $X$ and $Y$ on the same probability space, with respective distributions $F$ and $G$, such that $X \overset{a.s.}{\leq} Y$. Thus $\psi(F) = E[X] \overset{a.s.}{\leq} E[Y] = \psi(G)$, that is, (C2) holds. The proof is complete. $\square$

### 5.2.2 A further condition of interest

As a location measure, $\psi$ should be able to identify the location of an underlying distribution when it is defined as the center of a distribution which is symmetric under a given notion of symmetry. Specifically, we introduce the "center locating" codition

**(C3)** $\psi(P) = \theta$, for any $P \in \boldsymbol{\mathcal{P}}$ which is symmetric about a unique point $\theta \in \mathbf{R}^d$ under a given notion of symmetry.

### 5.2.3 Relationships among the three conditions

We devote this section to the discussion of the relationships among the conditions (C1), (C2) and (C3). In general, the three conditions are independent.

**Example 5.2.1** (C1) and (C2) but not (C3).

Define $\psi(F) = [E(X)]$ for any random vector $X$ with distribution $F \in \boldsymbol{\mathcal{P}}$. Let $F_0 \in \boldsymbol{\mathcal{P}}$ be angularly symmetric about a unique point $\theta \in \mathbf{R}^d$, $X_0$ be a random vector in $\mathbf{R}^d$ with distribution $F_0$, and $E(X_0) \neq \theta$. Then it is not difficult to see that $\psi$ satisfies (C1). By Lemma 5.2.1 $\psi$ also satisfies (C2). But according to the assumption that $E[X_0] \neq \theta$, $\psi$ does not satisfy (C3) for $F_0$. $\square$

**Example 5.2.2** (C1) and (C3) but not (C2).

Define $\psi$ to be the Tukey/Donoho halfspace median (Section 5.3). Consider the set $\boldsymbol{\mathcal{P}}$ of halfspace symmetric distributions in $\mathbf{R}^d$. Then by Theorem 3.2.1, we see that $\psi(F)$ satisfies (C1) and (C3) for any $F \in \boldsymbol{\mathcal{P}}$. However, $\psi(F)$ may not satisfy (C2) for $F \in \boldsymbol{\mathcal{P}}$ as shown in the following example. Let

$$x_1 = (2,2),\ x_1' = (-4,-4),\ x_2 = (2,3),\ x_2' = (-2,-5),$$

$$y_1 = (6,3),\ y_1' = (-2,-1),\ y_2 = (3,6),\ y_2' = (-2,-1),$$

$$\theta_1 = (1,1),\ \theta_2 = (0,0),$$

$$P(X = x_1) = P(X = x_1') = P(Y = y_1) = P(Y = y_1') = \frac{1}{5},$$

$$P(X = x_2) = P(X = x_2') = P(Y = y_2) = P(Y = y_2') = \frac{3}{10},$$

$$X(\omega_i) = x_i,\ X(\omega_i') = x_i',\ Y(\omega_i) = y_i,\ Y(\omega_i') = y_i', \quad \text{for } i = 1, 2.$$

Then it is not difficult to show that $X$ and $Y$ are angularly symmetric about the unique points $\theta_1$ and $\theta_2$ respectively, and that $X \overset{a.s.}{\leq} Y$. However, $\theta_1 \not\leq \theta_2$, and thus (C3) is not satisfied. □

**Example 5.2.3** (C2) and (C3) but not (C1).

Define $\psi(F) = \min_{t \in \mathbf{R}^d} E[\|X - t\|_2]$. Consider the set $\boldsymbol{\mathcal{P}}$ of centrally symmetric distributions. Then it is easy to see that $\psi(F)$ does not satisfy (C1) for some $F \in \boldsymbol{\mathcal{P}}$ since $\psi$ is only rigid-body equivariant. However, a result of Strassen (1965) and the following result shows that $\psi$ does satisfy (C2) and (C3). □

**Theorem 5.2.2** *Suppose that $F \in \boldsymbol{\mathcal{P}}$ is angularly symmetric about a unique point $\theta \in \mathbf{R}^d$. Then $\psi(F) = \min_{t \in \mathbf{R}^d} E[\|X - t\|_2]$ agrees with $\theta$.*

PROOF: Consider $E[\|X - t\|_2]$ as a function of $t \in \mathbf{R}^d$. By vector differentiation

$$\begin{aligned}
\frac{d(E[\|X - t\|_2])}{dt} &= \frac{d(\int (\|x - t\|_2)\, dF_X(x))}{dt} \\
&= \int \frac{d(\|x - t\|_2)}{dt}\, dF_X(x) \\
&= \int \frac{x - t}{\|x - t\|_2}\, dF_X(x).
\end{aligned}$$

Thus

$$(*) \qquad \frac{d(E[\|X - t\|_2])}{dt} = E\left[\frac{X - t}{\|X - t\|_2}\right]$$

Since $X$ is angularly symmetric about $\theta$, that is

$$\frac{X - \theta}{\|X - \theta\|_2} \overset{d}{=} \frac{\theta - X}{\|\theta - X\|_2},$$

we have

$$E\left[\frac{X-\theta}{\|X-\theta\|_2}\right] = E\left[\frac{\theta-X}{\|X-\theta\|_2}\right] = 0$$

Now the convexity of $\|\cdot\|_p$ for $p \geq 1$ (follows directly from Minkowski's inequality) and $(*)$ imply that $\psi(F) = \theta$. □

The examples above show that in general conditions (C1), (C2) and (C3) are mutually independent. For centrally symmetric distributions, however, there do exist some dependent relationships among conditions (C1), (C2) and (C3). The first lemma below follows in a straightforward fashion from a result of Strassen (1965).

**Lemma 5.2.1** *Suppose a functional* $\psi : \boldsymbol{\mathcal{P}} \to \mathbf{R}^d$ *satisfies* $\psi(F_X) = E[X]$ *for any* $F \in \boldsymbol{\mathcal{P}}$. *Then* (C2) *is satisfied.*

**Lemma 5.2.2** *Suppose* $T : \boldsymbol{\mathcal{P}} \to \mathbf{R}^d$ *is an odd and translation equivariant functional, that is,* $T(F_{X+b}) = T(F_X) + b$ *and* $T(F_{-X}) = -T(F_X)$ *for any vector* $b \in \mathbf{R}^d$, *and suppose that* $F_X$ *is centrally symmetric about a point* $\theta$ *in* $\mathbf{R}^d$. *Then* $T(F_X) = \theta$ $(= E[X])$.

PROOF: Since $X - \theta \stackrel{d}{=} \theta - X$,

$$T(F_{X-\theta}) = T(F_{\theta-X}) \quad \text{and} \quad E[X] = \theta.$$

By the translation equivariance of $T$ we have

$$T(F_X) - \theta = T(F_{-X}) + \theta.$$

Then $T(F_{-X}) = -T(F_X)$ implies that

$$T(F_X) = \theta = E[X].$$

The proof is complete. □

**Theorem 5.2.3** *Suppose that $\boldsymbol{\mathcal{P}}$ is a set of centrally symmetric multivariate distributions on $\mathbf{R}^d$ and $\psi$ is a functional on $\boldsymbol{\mathcal{P}}$ satisfying* (C1). *Then $\psi$ satisfies* (C2) *and* (C3).

PROOF: This follows immediately from Lemma 5.2.2 and Lemma 5.2.1. □

**Remark 5.2.1** Theorem 5.2.3 implies that condition (C2) in the definitions of location measure of Bickel and Lehmann (1975) and Oja (1983) is redundant for centrally symmetric multivariate distributions. On the other hand, under a weaker notion of symmetry, satisfaction of (C2) does not guarantee that (C3) is also satisfied, as shown in Example 5.2.1. This fact again reflects the relevance of condition (C3). □

## 5.3 Depth-Based Multivariate Nonparametric Location Measures

Statistical depth functions introduced in Chapter 3 are inherently bound up with notions of center of multivariate distributions. They could be applied immediately to introduce notions of multivariate median. Indeed, multivariate medians induced by these depth functions are good candidates for location measures of multivariate distributions.

**Definition 5.3.1** Suppose that $D(\cdot,\cdot) : \mathbf{R}^d \times \boldsymbol{\mathcal{P}} \to \mathbf{R}^1$ is a depth functional. Then the point which maximizes $D(\cdot, F)$ on $\mathbf{R}^d$ for a given $F \in \boldsymbol{\mathcal{P}}$ is called the *multivariate median* of $F$ (induced by $D(\cdot,\cdot)$).

**Remark 5.3.1** The halfspace median, the simplicial median, the Oja median and the $L_1$ median are among the foremost existing multivariate medians in the literature; see Donoho and Gasko (1992), Liu (1990), Oja (1983) and Small(1990). These medians may be derived respectively from the halfspace depth function (Tukey 1974, Donoho and Gasko 1992), the simplicial depth function (Liu 1990), the simplicial volume depth function (with $\alpha = 1$) and the $L_p$ depth function (with $p = 2$). The latter two depth functions were defined in Chapter 3. ("Spatial median" is another term for the $L_1$ median in the literature.) $\square$

**Remark 5.3.2** If there is a set $S$ of multiple points which maximize $D(\cdot, F)$ on $\mathbf{R}^d$ for some $F \in \boldsymbol{\mathcal{P}}$, then any sensible rule may be employed to obtain a multivariate median. For example, $\int_S x\, dx / \int_S dx$ may be defined as a multivariate median of $F$. Most medians induced by depth functionals in the sense of Definition 5.3.1 are not unique, the $L_1$ median being an exception. The uniqueness of the $L_1$ median has been studied by Kemperman (1987) and Milasevic and Ducharme (1987). It turns out that the $L_1$ median is unique if the probability mass is not carried by any straight line in $R^d$. It is not difficult to show that the same result holds for the median induced by the $\tilde{L}_2$ depth defined in Chapter 3. $\square$

**Definition 5.3.2** The medians induced by the simplicial volume depth, the $L_p$ depth, the projection depth, and the Type D depth will be called the *simplicial volume median*, the *$L_p$ median*, the *projection median*, and the *Type D median*, respectively.

**Theorem 5.3.1** *Multivariate medians induced by statistical depth functions satisfy condition* (C1)*. For centrally symmetric multivariate distributions, these medians also satisfy conditions* (C2) *and* (C3)*.*

PROOF: (1) Since for any $F \in \boldsymbol{\mathcal{P}}$, all the corresponding depth functions $D(\cdot, F)$ which induce these medians are affine invariant, thus

$$y \in \arg\max_{x \in \mathbf{R}^d} D(x, F_X) \Longleftrightarrow Ay + b \in \arg\max_{x \in \mathbf{R}^d} D(x, F_{AX+b}),$$

for any $d \times d$ matrix $A$ and vector $b \in \mathbf{R}^d$. Hence all the medians induced by statistical depth functions are affine equivariant.

(2) By Theorem 5.2.3, all the medians satisfy conditions (C2) and (C3). $\square$

## 5.4 Performance of multivariate nonparametric location measures

Employing Theorems 3.3.10 and 3.3.13 we immediately have

**Theorem 5.4.1** *For halfspace symmetric multivariate distributions, the projection median and the Type D median satisfy the "center locating" condition* (C3)*.*

Taking $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{H}}$ in the definition of Type D median, the above theorem yields the following result.

**Corollary 5.4.1** *For halfspace symmetric multivariate distributions, the halfspace median satisfies the "center locating" condition* (C3)*.*

Now we focus our attention on the simplicial median and the $\tilde{L}_2$ median. We will also examine the "center locating" condition for the $L_1$ median (the spatial median), which is only rigid-body equivariant, but is a popular multivariate median in the literature. We start with the case of a class of *angularly symmetric* distributions.

**Theorem 5.4.2** *For distributions absolutely continuous and angularly symmetric about a unique point in* $\mathbf{R}^d$*, the simplicial median satisfies the "center locating" condition* (C3).

PROOF: (modeled after the proof of Theorem 3 of Liu (1990)). Suppose $X \in \mathbf{R}^d$ is continuous and angularly symmetric about a unique point in $\theta \in \mathbf{R}^d$. For any $x \in \mathbf{R}^d$, $x \neq \theta$, we have that

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

where $D(\cdot)$ denotes the simplicial depth function with respect to $X$, $X_1 \ldots, X_{d+1}$ is a sample from $X$, $S[X_1, \ldots, X_{d+1}]$ is the simplex consisting of vertices $X_1 \ldots, X_{d+1}$. Let $\vec{\theta x}$ be the vector in $\mathbf{R}^d$ starting at $\theta$ ending at $x$, then

$$\begin{aligned}
& D(\theta) - D(x) \\
&= P(\vec{\theta x} \text{ leaves but not enters } S[X_1, \ldots, X_{d+1}]) \\
&\quad - P(\vec{\theta x} \text{ enters but not leaves } S[X_1, \ldots, X_{d+1}]) \\
&= P(\vec{\theta x} \text{ leaves } S[X_1, \ldots, X_{d+1}]) - P(\vec{\theta x} \text{ enters } S[X_1, \ldots, X_{d+1}]) \\
&= (d+1) \int_{\vec{\theta x} \cap HP(x_1, \ldots, x_d) \neq \Phi} P(X_{d+1} \in H_\theta(x_1, \ldots, x_d)\, dF(x_1) \ldots dF(x_d) \\
&\quad - (d+1) \int_{\vec{\theta x} \cap HP(x_1, \ldots, x_d) \neq \Phi} P(X_{d+1} \in H_x(x_1, \ldots, x_d)\, dF(x_1) \ldots dF(x_d)
\end{aligned}$$

where $HP(x_1,\ldots,x_d)$ is the hyperplane determined by $x_1,\ldots,x_d$ and $H_y(x_1,\ldots,x_d)$ is the halfspace containing $y$ and with $x_1,\ldots,x_d$ on the boundary. Angular symmetry about $\theta$ of $X$ now implies that

$$H_\theta(x_1,\ldots,x_d) \geq \frac{1}{2} \geq H_x(x_1,\ldots,x_d).$$

Thus $D(\theta) \geq D(x)$, completing the proof. $\square$

**Theorem 5.4.3** *For distributions discrete and angularly symmetric about a unique point in* $\mathbf{R}^d$*, the simplicial median satisfies the "center locating" condition* (C3)*.*

PROOF: Assume that $X \in \mathbf{R}^d$ is discrete and angularly symmetric about a unique point $\theta \in \mathbf{R}^d$. We need to show that for any $x \in \mathbf{R}^d$

$$(*) \qquad P(x \in S[X_1,\ldots,X_{d+1}]) \leq P(\theta \in S[X_1,\ldots,X_{d+1}]).$$

For simplicity of description, we only consider $d = 2$ here. However, the following proof is readily generalized to the case $d > 2$.

To prove $(*)$, it is clear that we need only consider the simplices which contain $x$ but not $\theta$, or contain $\theta$ but not $x$. The definition of the simplicial median and Theorem 2.3.3 yield that there are only two cases which need be taken into account.

Case 1: The vertices of the triangles which contain $x$ lie on two lines intersecting at $\theta$.

Assume $AB$ and $CD$ intersect at $\theta$, $x_{p_i}$ lies on $\theta B$, $x_{p_i'}$ lies on $A\theta$, $x_{q_i}$ lies on $\theta D$, and $x_{q_i'}$ lies on $C\theta$, $i \in \{1,2,\ldots\}$. It is often found convenient to use the probability mass $p_i$ on

$x_{p_i}$ and the point $x_{p_i}$ interchangeably. Assume that

$$\sum p_i = p, \ \ \sum p_i' = p', \ \ \sum q_j = q, \ \ \sum q_j' = q'.$$

Then by Theorem 2.3.3

$$p = p', \ \ q = q'.$$

Consider an arbitrary point $p_i$, and assume that $p_i x$ intersects $\theta D$ at point $y_{p_i}$ and that

$$P(X \in (\theta, \ y_{p_i})) = q_{i1}, \ \ P(X \in [y_{p_i}, \ D)) = q_{i2}.$$

Then it is not difficult to show that

$$P(x \in \Delta(X_1, X_2, X_3) \notin \theta)/3! = \sum_i p_i q_{i1} q_{i2} + \sum_j q_j p_{j1} p_{j2},$$

where $p_{j1}$ and $p_{j2}$ are defined similarly to $q_{i1}$ and $q_{i2}$. On the other hand, it is also not difficult to see that

$$P(\theta \in \Delta(X_1, X_2, X_3) \notin x)/3! \geq \sum_i p_i q_{i1} q' + \sum_j q_j p_{j1} p'.$$

Thus

$$P(x \in \Delta(X_1, X_2, X_3)) \leq P(\theta \in \Delta(X_1, X_2, X_3)).$$

Case 2: The vertices of the triangles which contain $x$ lie on three lines intersecting at $\theta$.

Assume $AB$, $CD$ and $EF$ intersect at $\theta$, $x_{p_i}$ lies on $\theta B$, $x_{p_i'}$ lies on $A\theta$, $x_{q_i}$ lies on $\theta D$, $x_{q_i'}$ lies on $C\theta$, $x_{r_i}$ lies on $\theta F$, and $x_{r_i'}$ lies on $E\theta$, $i \in \{1, 2, \ldots\}$. Assume that

$$\sum p_i = p \ \ \sum p_i' = p', \ \ \sum q_j = q,$$

$$\sum q'_j = q', \quad \sum r_k = r, \quad \sum r'_k = r'.$$

By Theorem 2.3.3 we have that

$$p = p', \quad q = q', \quad r = r'.$$

Consider an arbitrary point $p_i$, connect $p_i$ and $x$, intersecting $\theta D$ at $y_{p_i}$, $\theta F$ at $z_{p_i}$. Assume that

$$P(X \in (\theta, y_{p_i})) = q^1_{p_i}, \quad P(X \in [y_{p_i}, D) = q^2_{p_i},$$

$$P(X \in (\theta, z_{p_i})) = r^1_{p_i}, \quad P(X \in [z_{p_i}, F)) = r^2_{p_i}.$$

Then it is not difficult to show that

$$\begin{aligned} & P(x \in \Delta(X_1, X_2, X_3) \notin \theta)/3! \\ \leq \;& \sum p_i q^1_{p_i} r^2_{p_i} + \sum p_i q^2_{p_i} r^1_{p_i} + \sum p'_i q^2_{p'_i} r^1_{p'_i} + \sum p'_i r^2_{p'_i} q^1_{p'_i} \end{aligned}$$

where $r^1_{p'_i}$, $r^2_{p'_i}$, $q^1_{p'_i}$ and $q^2_{p'_i}$ are defined similarly to $q^k_{p_i}$ and $r^k_{p_i}$ for $k = 1, 2$. It is also not very difficult to prove that

$$\begin{aligned} & P(\theta \in \Delta(X_1, X_2, X_3) \notin x))/3! \\ \geq \;& \sum p'_i \max_i \{q^1_{p_i}\} r' + \sum p_i q' r^1_{p_i} + \sum p_i \sum p'_i \sum r'_k + \sum q_j \sum q'_j \sum r'_k. \end{aligned}$$

(Note that it is very important that we have avoided the repeated use of the triangles containing $\theta$ which have been employed in the proof of Case 1.) Thus

$$\begin{aligned} & P(\theta \in \Delta(X_1, X_2, X_3) \notin x)/3! \\ \geq \;& \sum p_i \max_i \{q^1_{p_i}\} r' + \sum p_i q' r^1_{p_i} + (p^2 + q^2) r \end{aligned}$$

$$\geq \sum p_i q^1_{p_i} r^2_{p_i} + \sum p_i q^2_{p_i} r^1_{p_i} + 2pqr$$

$$\geq \sum p_i q^1_{p_i} r^2_{p_i} + \sum p_i q^2_{p_i} r^1_{p_i} + \sum p'_i q^2_{p'_i} r^1_{p'_i} + \sum p'_i q^1_{p'_i} r^2_{p'_i},$$

and thus

$$P(\theta \in \Delta(X_1, X_2, X_3)) \geq P(x \in \Delta(X_1, X_2, X_3)).$$

Now Case 1 and Case 2 combined imply that

$$\theta = \arg\max_{x \in \mathbf{R}^d} P(x \in S[X_1, \ldots, X_{d+1}]).$$

The proof is complete. □

For the $L_1$ and $\tilde{L}_2$ medians, by Theorems 3.3.7 and 5.2.2, we have

**Theorem 5.4.4** *For distributions angularly symmetric about a unique point in* $\mathbf{R}^d$*, the* $L_1$ *and* $\tilde{L}_2$ *medians satisfy the "center locating" condition* (C3)*.*

Next we examine the "center locating" condition (C3) for the simplicial median and the $L_2$ and $\tilde{L}_2$ medians, in the case of *halfspace symmetric* distributions. By Theorems 5.4.2 and 5.4.3 we have seen that the simplicial median is able to locate the center of angularly symmetric distributions. However, for halfspace symmetric distributions, Remark 3.2.2 in Chapter 3 reveals that

**Theorem 5.4.5** *For distributions halfspace symmetric about a unique point in* $\mathbf{R}^d$*, the simplicial median does not satisfy the "center locating" condition* (C3) *in general.*

For the $L_1$ and $\tilde{L}_2$ medians, we have

**Theorem 5.4.6** *For distributions halfspace symmetric about a unique point in* $\mathbf{R}^d$*, the* $L_1$ *and* $\tilde{L}_2$ *medians satisfy the "center locating" condition* (C3).

PROOF: We will prove the above result only for the $L_1$ median, the proof for the $\tilde{L}_2$ median being similar. For descriptive simplicity, we consider only the case $d = 2$. However, the proof is immediately generalizable to the case $d > 2$. By Theorems 2.3.4 and 5.4.4, we need only consider the discrete case with nonzero probability mass on the center.

Assume $X$ is halfspace symmetric about a unique point $\theta$ in $\mathbf{R}^d$ with $P(X = \theta) = p_\theta > 0$. By Theorem 4.17 of Kemperman (1987) and rigid-body equivariance of the $L_1$ median, we may assume, w.l.o.g., that $\theta$ is the origin and all probability mass is carried by the unit circle.

We need to show that

$$E[\|X - \theta\|] \le E[\|X - x|] \quad \forall x \in \mathbf{R}^d.$$

Assume, w.l.o.g., that $x = (0, y)$, that $x_1, \ldots, x_n, \ldots$ fall on the left open semi-circle with $x_n \neq (-1, 0)$ for $n = 1, 2, \ldots$, and that $P(X = x_i) = p_i > 0$, $d_i = \|x - x_i\|$ and $d_i \le d_j$ for $i \le j$. Assume that $x_\alpha = (-1, 0)$, $x_\beta = (0, 1)$ and $x_\gamma = (0, -1)$, with $P(X = x_\alpha) = p_\alpha$, $P(X = x_\beta) = p_\beta$ and $P(X = x_\gamma) = p_\gamma$.

It is found convenient sometimes to denote a point and the mass at that point by the same notation. Denote the point $-x_i$ by $x_i'$ for $i = 1, 2, \ldots$. Starting at $x_\gamma$, travel clockwise along the circle until hitting the first point, say $x_i$, denote by $\mathrm{arc}(x_\beta, x_i']$ the open-closed arc which is the symmetric part about the origin of the arc just covered, and assume $P(X \in \text{arc}(x_\beta, x_i']) = p_i'$. Starting from $x_i$ now, travel clockwise along the circle until hitting the next point, say $x_j$, assume that $P(X \in \text{arc}(x_i', x_j']) = p_j'$. Continue in this fashion and stop before hitting $x_\alpha$. Starting at $x_\beta$ now, travel counterclockwise along the circle performing similar steps as before, that is, labeling the probability mass on the open-closed arcs which are the symmetric parts about the origin of the arcs covered, and stopping before hitting $x_\alpha$. Denote by $p_\alpha'$ the probability mass on the open-open arc on the right semi-circle, which contains $(1,0)$ and is the symmetric part about the origin of the open-open arc on the left semi-circle which contains $x_\alpha$ and is never covered above.

Now it is not difficult to show that

$$
\begin{aligned}
& E[\|X - x\|] - E[\|X - \theta\|] \\
\geq\ & \left(\sqrt{1+y^2} - 1\right)(p_\beta + p_\gamma) + y(p_\alpha + p_\theta - p_\alpha') + \sum_{i=1}[(d_i - 1)p_i + (d_i' - 1)p_i'] \\
=\ & \left(\sqrt{1+y^2} - 1\right)(p_\beta + p_\gamma) \\
& +(d_1 - 1)\left(p_\alpha + p_\theta - p_\alpha' + \sum_{i=1} p_i - \sum_{i=1} p_i'\right) + (d_1 + d_1' - 2)p_1' \\
& +(y + 1 - d_1)(p_\alpha + p_\theta - p_\alpha') + \sum_{i=2}[(d_i - d_1)p_i + (d_i' + d_1 - 2)p_i')] \\
=\ & \left(\sqrt{1+y^2} - 1\right)(p_\beta + p_\gamma) \\
& +(d_1 - 1)\left(p_\alpha + p_\theta - p_\alpha' + \sum_{i=1} p_i - \sum_{i=1} p_i'\right) + (d_1 + d_1' - 2)p_1'
\end{aligned}
$$

$$
\begin{aligned}
& (d_2 - d_1)\left(p_\alpha + p_\theta - p'_\alpha + \sum_{i=2} p_i - \sum_{i=2} p'_i\right) + (d_2 + d'_2 - 2)p'_2 \\
& +(y + 1 - d_2)(p_\alpha + p_\theta - p'_\alpha) + \sum_{i=3}[(d_i - d_2)p_i + (d'_i + d_2 - 2)p'_i)] \\
= \quad & \left(\sqrt{1+y^2} - 1\right)(p_\beta + p_\gamma) \\
& +(d_1 - 1)\left(p_\alpha + p_\theta - p'_\alpha + \sum_{i=1} p_i - \sum_{i=1} p'_i\right) + (d_1 + d'_1 - 2)p'_1 \\
& (d_2 - d_1)\left(p_\alpha + p_\theta - p'_\alpha + \sum_{i=2} p_i - \sum_{i=2} p'_i\right) + (d_2 + d'_2 - 2)p'_2 \\
& \vdots \\
& +(d_n - d_{n-1})\left(p_\alpha + p_\theta - p'_\alpha + \sum_{i=n} p_i - \sum_{i=n} p'_i\right) + (d_n + d'_n - 2)p'_n \\
& +(y + 1 - d_n)(p_\alpha + p_\theta - p'_\alpha) + \sum_{i=n+1}[(d_i - d_n)p_i + (d'_i + d_n - 2)p'_i)] \\
= \quad & \left(\sqrt{1+y^2} - 1\right)(p_\beta + p_\gamma) \\
& +(d_1 - 1)\left(p_\alpha + p_\theta - p'_\alpha + \sum_{i=1} p_i - \sum_{i=1} p'_i\right) + (d_1 + d'_1 - 2)p'_1 \\
& (d_2 - d_1)\left(p_\alpha + p_\theta - p'_\alpha + \sum_{i=2} p_i - \sum_{i=2} p'_i\right) + (d_2 + d'_2 - 2)p'_2 \\
& \vdots \\
& +(d_n - d_{n-1})\left(p_\alpha + p_\theta - p'_\alpha + \sum_{i=n} p_i - \sum_{i=n} p'_i\right) + (d_n + d'_n - 2)p'_n \\
& \vdots
\end{aligned}
$$

By Theorem 2.3.2, we immediately have that

$$\begin{aligned} p_\alpha + p_\theta &\geq p'_\alpha, \\ p_\alpha + p_\theta + \sum_{i=m}^{n} p_i &\geq p'_\alpha + \sum_{i=m}^{n} p'_i \quad \forall\, m \leq n \ \text{ and } \ m \geq 1. \end{aligned}$$

In conjunction with the fact that $d_i + d'_i \geq 2 \quad \forall\, i \geq 1$, we conclude that

$$E[\|X - x\|] - E[\|X - \theta\|] \geq 0 \qquad \forall\, x \in \mathbf{R}^d.$$

The proof is complete. □

We conclude this section with the following remark.

**Remark 5.4.1** The projection median, the Type D median (including the halfspace median), and the $\tilde{L}_2$ median are good choices of multivariate nonparametric location measures for halfspace symmetric multivariate distributions. The simplicial median, on the other hand, is a good candidate for multivariate nonparametric location measure only for angularly symmetric distributions. It may fail to identify the location (the point of symmetry) of halfspace symmetric distributions. The sample simplicial median thus may be inconsistent. In sum, the simplicial median should be used with great caution as a multivariate nonparametric location measure in practice. □

## 5.5 Summary

In this chapter, a desirable condition for multivariate nonparametric location measures, called the "center locating" condition, has been introduced. Interrelationships among this condition and conditions of Bickel and Lehmann (1975) and Oja (1983) have been explored. Depth-based multivariate medians, as multivariate nonparametric location measures, have been introduced and their performance studied. It turns out that for halfspace symmetric distributions the Liu simplicial median may fail to satisfy the "center locating" condition and, consequently, should be used with great caution as a multivariate nonparametric location measure in general. On the other hand, the halfspace median, the $L_1$ median, the $\hat{L}_2$ median and the projection median are good choices of multivariate nonparametric location measures for halfspace symmetric multivariate distributions.

# Chapter 6

# DEPTH-BASED SCATTER MEASURES

## 6.1 Introduction

Bickel and Lehmann (1976) introduced a notion of "more dispersed" to compare one symmetric distribution with another in $\mathbf{R}$. In Bickel and Lehmann (1979), they extended to the case of arbitrary (i.e. not necessarily symmetric) distributions and introduced a notion of "more spread out". Eaton (1982) and Oja (1983) generalized the above concepts to higher dimensions. Comparisons of dispersions/spreads of distributions not only have theoretical interest but also have practical significance in applications such as medical trials, quality control, and system reliability.

In this chapter, statistical depth functionals are employed to introduce a notion of "more scattered" for comparison of one multivariate distribution with another in $\mathbf{R}^d$. Relationships among the notions of Bickel and Lehmann (1976), Eaton (1982), Oja (1983) and ours are explored. It turns out that our notion is more general than those of Oja (1983) and Eaton (1982) under some typical conditions, and is a generalization of that of Bickel and Lehmann (1976) in $\mathbf{R}$. The properties related to our depth-based notion of "more scattered" are studied thoroughly. Finally, depth-based "scatter measures" are defined and some examples of scatter measures are presented and studied.

## 6.2 Notions of "More Scattered"

In this section, various existing notions of "more scattered" are given, and a new notion of "more scattered" based on statistical depth functionals is introduced. Here we use the term "more scattered" to represent various terms (such as "more dispersed", "more spread out", "more concentrated") used by different authors.

(D1) Let $P$ be the distribution of a random variable $X$ in $\mathbf{R}$, symmetric about $\mu$, and $Q$ be the distribution of a random variable $Y$ in $\mathbf{R}$, symmetric about $\nu$. Then $X$ is said to be *more scattered* about $\mu$ than $Y$ about $\nu$, in the sense of Bickel and Lehmann (1976), if

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

where a random variable $Z_1$ with distribution function $G_1(x)$ is said to be *stochastically smaller* than another random variable $Z_2$ with distribution function $G_2(x)$, denoted by $Z_1 \overset{st}{\leq} Z_2$, if $G_1(x) \geq G_2(x)$ for every $x$ in $\mathbf{R}$.

The above concept of "more scattered" is essentially equivalent to the "peakedness" ordering introduced by Birnbaum (1948).

(D2) Bickel and Lehmann (1979) extended the "more dispersed" concept to arbitrary (i.e. not necessarily symmetric) distributions, and defined a distribution $F$ to be *more scattered* than a distribution $G$ if

$$F^{-1}(v) - F^{-1}(u) \geq G^{-1}(v) - G^{-1}(u), \quad \forall\, 0 < u < v < 1,$$

where $F^{-1}(\alpha) = \sup\{x\colon F(x) \leq \alpha\}$.

The essence of this definition was introduced by Brown and Tukey (1946). Note that the above two notions of Bickel and Lehmann do not coincide for symmetric distributions.

(D3) Eaton (1982) extended the "more scattered" concept of Bickel and Lehmann (1976) for $\mathbf{R}$ to arbitrary dimension: $P$ is said to be *more scattered* about $\mu$ than $Q$ about $\nu$ if

$$P(X \in C + \mu) \leq Q(Y \in C + \nu)$$

for any convex set $C$ in $\mathbf{R}^d$ with $C = -C$.

(D4) Oja (1983) generalized the notion of "more scattered" of Bickel and Lehmann (1979) for $\mathbf{R}$ to arbitrary dimension as follows: $Q$ is said to be *more scattered* than $P$ if $P$ is the distribution of a random vector $X$ and $Q$ the distribution of a random vector $\phi(X)$, for some function $\phi$ such that

$$\Delta(\phi(x_1), \phi(x_2), \ldots, \phi(x_{d+1})) \geq \Delta(x_1, x_2, \ldots, x_{d+1})$$

for any $x_1, x_2, \ldots, x_{d+1}$ in $\mathbf{R}^d$, where $\Delta$ is the volume function of $d$-dimensional simplices.

We now introduce a notion of "more scattered" in $\mathbf{R}^d$ based on the statistical depth functionals discussed in Chapter 3. Denote by $\boldsymbol{\mathcal{F}}$ a class of multivariate distributions on $\mathbf{R}^d$.

**Definition 6.2.1** 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 $\boldsymbol{\mathcal{F}}$ if

$$\Delta(\{x \in \mathbf{R}^d \,|\, D(x; P) \geq \alpha\}) \geq \Delta(\{y \in \mathbf{R}^d \,|\, D(y; Q) \geq \alpha\})$$

for any $\alpha > 0$, where $\Delta$ is a volume function for statistical depth contours.

**Remark 6.2.1** For the sake of descriptive simplicity, in the latter discussion of this chapter the statistical depth functional in the above definition will be the *halfspace* depth functional. □

## 6.3 Relationships Among Various Notions of "More Scattered"

Now we examine the relationships among the "more scattered" notions of Bickel and Lehmann (1976), Eaton(1982), Oja(1983), and ours given in Section 6.2. We will use the notation "$P \overset{sc}{\geq} Q$" to denote "$P$ is more scattered than $Q$" in a given sense.

First we present some characterizations of the "more scattered" notion of Bickel and Lehmann (1976). Some of these results will be used in the latter part of this section.

**Theorem 6.3.1** *Let $P$ be the probability measure corresponding to the distribution $F$ of a random variable $X$ in $\mathbf{R}$ symmetric about $\mu$ and $Q$ be the probability measure corresponding to the distribution $G$ of a random variable $Y$ in $\mathbf{R}$ symmetric about $\nu$. Then the following statements are equivalent:*

(1) *$X$ is more scattered about $\mu$ than $Y$ about $\nu$ in the sense of* (D1);

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

(3) $t\,(F_{X-\mu}(t) - G_{Y-\nu}(t)) \leq 0, \quad \forall\, t;$

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

*where* $F^{-1}(p) = \inf\{x : F(x) \geq p\}$.

PROOF: (I) (1) $\Leftrightarrow$ (2). This follows in a straightforward fashion from the concept of "stochastically smaller" for comparison of random variables.

(II) (1) $\Leftrightarrow$ (3). we have established that (1) holds iff (2) holds. By the symmetry of $X-\mu$ and $Y-\nu$, (2) holds iff $F_{X-\mu}(t) \le G_{Y-\nu}(t)$ for $t>0$ and $F_{X-\mu}(t) \ge G_{Y-\nu}(t)$ for $t<0$, that is, iff (3), proving (1) $\Leftrightarrow$ (3).

(III) (1) $\Leftrightarrow$ (4). By the symmetry of $X-\mu$ and $Y-\nu$, it is not difficult to see that (4) $\Leftrightarrow$ (3). Hence (1) $\Leftrightarrow$ (4). $\square$

**Remark 6.3.1** Clearly, by the above theorem, for symmetric distributions (D1) is equivalent to (D3) in $\mathbf{R}$. Thus definition (D3) of Eaton (1982) generalizes definition (D1) of Bickel and Lehmann (1976). $\square$

**Theorem 6.3.2** *For symmetric distributions in* $\mathbf{R}$*, Definition 6.2.1 is equivalent to* (D1) *or* (D3).

PROOF: (i) It is often convenient to use the distribution of a random variable and the corresponding probability measure interchangeably. Suppose $P$ is symmetric about $\mu$, $Q$ is symmetric about $\nu$, and $P$ is "more scattered" than $Q$ in the sense of Definition 6.2.1. Then

$$\begin{aligned}
\Delta(\{x \in \mathbf{R} \mid D(x;P) \ge \alpha\}) &= \Delta([\mu - p_\alpha, \mu + p_\alpha]) \\
&\ge \Delta([\nu - q_\alpha, \nu + q_\alpha]) \\
&= \Delta(\{x \in \mathbf{R} \mid D(x;Q) \ge \alpha\}),
\end{aligned}$$

for any $\alpha > 0$, where $P([p_\alpha, +\infty)) = \alpha = Q([q_\alpha, +\infty))$. Hence, we have

$$p_\alpha \ge q_\alpha, \ \ \forall\, \alpha > 0.$$

Therefore,

$$P([\mu - a, \mu + a]) \le Q([\nu - a, \nu + a]), \ \ \forall\, a > 0.$$

Thus, $P$ is "more scattered" than $Q$ in the sense of (D3). By Theorem 6.3.1, $P$ is also "more scattered" than $Q$ in the sense of (D1).

(ii) Suppose that $P$ is "more scattered" than $Q$ in the sense of (D1) or (D3). By Theorem 6.3.1,

$$P([\mu - a, \mu + a]) \le Q([\nu - a, \nu + a]), \ \ \forall\, a > 0.$$

Thus if $P([\mu - p_\alpha, \mu + p_\alpha]) = Q([\nu - q_\alpha, \nu + q_\alpha]) = 1 - 2\alpha$, for $\alpha > 0$, then

$$p_\alpha \ge q_\alpha, \ \ \forall\, \alpha > 0.$$

Hence

$$\begin{aligned} \Delta(\{x \in \mathbf{R} \mid D(x;P) \ge \alpha\}) &= \Delta([\mu - p_\alpha, \mu + p_\alpha]) \\ &\ge \Delta([\nu - q_\alpha, \nu + q_\alpha]) \\ &= \Delta(\{x \in \mathbf{R} \mid D(x;Q) \ge \alpha\}), \end{aligned}$$

for any $\alpha > 0$. Therefore, $P$ is "more scattered" than $Q$ in the sense of Definition 6.2.1. This completes the proof. □

**Remark 6.3.2** Theorem 6.3.2 implies that Definition 6.2.1 is indeed a generalization of Bickel and Lehmann (1976). □

**Remark 6.3.3** In the following, let $F$ and $G$ be strictly increasing in $\mathbf{R}$.

(1) Definition (D4) of Oja (1983) is equivalent to definition (D2) of Bickel and Lehmann

(1979); see Theorem 1 of Bickel and Lehmann (1979) for the proof.

(2) Definition (D4) of Oja (1983) implies definition (D1) of Bickel and Lehmann (1976), but the converse does not hold since the "more scattered" notions of Bickel and Lehmann (1976) and (1979) are inconsistent. □

**Theorem 6.3.3** *Suppose that $P$ and $Q$ are centrally symmetric about the origin in $\mathbf{R}^d$. Then $Q \overset{sc}{\leq} P$ in the sense of* D3 *implies $Q \overset{sc}{\leq} P$ in the sense of Definition 6.2.1.*

PROOF: We are going to show that for any $\alpha > 0$

$$(*) \qquad D^\alpha(Q) = \{x \in \mathbf{R}^d \mid D(x;Q) \geq \alpha\} \subset D^\alpha(P) = \{x \in \mathbf{R}^d \mid D(x;P) \geq \alpha\}.$$

For any $x \in \mathbf{R}^d$, suppose that $D(x;Q) = \alpha$ and $D(x;P) < \alpha$. Then, by symmetry and the definition of halfspace depth, there exists a closed halfspace $H_x$ with $x$ on the boundary and its reflection $H_{-x}$ about the origin such that $P(H_x) < \alpha$ and $P(H_{-x}) < \alpha$. Hence by (D3)

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

which implies that $Q(H_x) = Q(H_{-x}) < \alpha$. This contradicts to the assumption that $D(x;Q) = \alpha$. Therefore, $D(x;P) \geq \alpha$. Now we have that for any $x$ belongs to $D^\alpha(Q)$, $x$ also belongs to $D^\alpha(P)$. Thus $(*)$ holds, proving that $Q \overset{sc}{\leq} P$ in the sense of Definition 6.2.1. □

**Theorem 6.3.4** *Suppose that $P \overset{sc}{\leq} Q$ in the sense of* (D4), *and $\phi$ in the* (D4) *is an affine transformation. Then $P \overset{sc}{\leq} Q$ in the sense of Definition 6.2.1.*

PROOF: (1) $\mathbf{R}^1$ case. Let $D^\alpha(P) = \{y \in \mathbf{R}^1 \mid D(y; P) \geq \alpha\}$. Then, by Theorem 3.3.12, we have that $D^\alpha(P) = [y_1, y_2]$ for some $y_1$ and $y_2$. By the affine invariance property of statistical depth functionals,

$$\begin{aligned} D^\alpha(Q) &= \{x \in \mathbf{R}^1 \mid D(x; Q) \geq \alpha\} \\ &= \{x \in \mathbf{R}^1 \mid D(x; Q) \geq \alpha, x = \phi(y), y \in D^\alpha(P)\} \\ &= [\phi(y_1), \phi(y_2)]. \end{aligned}$$

Thus $\Delta(D^\alpha(Q)) \geq \Delta(D^\alpha(P))$ by (D4), that is, $P \overset{sc}{\leq} Q$ in the sense of Definition 6.2.1.

(2) $\mathbf{R}^d$ ($d > 1$) case. By Theorem 3.3.12, we have that $D^\alpha(P) = \{y \in \mathbf{R}^d \mid D(y; P) \geq \alpha\}$ is a convex compact set in $\mathbf{R}^d$. Thus $D^\alpha(P)$ can be approximated by $d$-dimensional simplices, that is,

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

for some $d$-dimensional simplices $S_i$'s. The affine invariance of statistical depth functionals implies that

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

Now since $\Delta(\phi(x_1), \ldots, \phi(x_{d+1}) \geq \Delta(x_1, \ldots, x_{d+1})$ for any $x_1, \ldots, x_{d+1} \in \mathbf{R}^d$, thus

$$\Delta(D^\alpha(Q)) \geq \Delta(D^\alpha(P)),$$

proving that $P \overset{sc}{\leq} Q$ in the sense of Definition 6.2.1. The proof is complete. □

**Remark 6.3.4** Theorem 6.3.3 and Theorem 6.3.4 imply that the "more scattered" Definition 6.2.1 is weaker (more general) than those of Eaton (1982) and Oja (1983), under typical conditions. □

## 6.4 Properties of Depth-Based Notion of "More Scattered"

In this section, we investigate properties related to the notion of "more scattered" in the sense of Definition 6.2.1. For the sake of simplicity, the halfspace depth functional will again be used in subsequent discussion.

**Theorem 6.4.1** *Suppose that $P$ and $Q$ are two distributions in $\boldsymbol{\mathcal{F}}$. Then $P \overset{sc}{\geq} Q$ in the sense of Definition 6.2.1 if $P(H) \geq Q(H)$ for any closed halfspace $H$ in $\mathbf{R}^d$ with $P(H) \leq \frac{1}{2}$.*

PROOF: We show that

$$(*) \qquad D^\alpha(Q) = \{x \in \mathbf{R}^d \mid D(x;Q) \geq \alpha\} \subset D^\alpha(P) = \{x \in \mathbf{R}^d \mid D(x;P) \geq \alpha\},$$

for any $\alpha > 0$. For any $x \in \mathbf{R}^d$, suppose that $D(x;Q) = \beta \geq \alpha$, and $D(x;P) < \alpha$. Then there exists a closed halfspace $H_x$ with $x$ on its boundary such that

$$\alpha > P(H_x) \geq D(x;P),$$

by the definition of halfspace depth function. Hence $\beta > Q(H_x)$, contradicting the assumption that $D(x;Q) = \beta$. Thus, $D(x;P) \geq \beta$, that is, $x \in D^\alpha(Q)$ implies that $x \in D^\alpha(P)$. Therefore $(*)$ holds, completing the proof. □

**Remark 6.4.1** If $P$ and $Q$ are symmetric about the origin in $\mathbf{R}$, then the condition in the above Theorem is also necessary. $\square$

**Corollary 6.4.1** *Let $X$ be a random vector in $\mathbf{R}^d$. Then*

(1) $X \overset{sc}{\geq} aX$ *in the sense of Definition 6.2.1 for any $a \in (0, 1]$.*

(2) $X \overset{sc}{\geq} b$ *in the sense of Definition 6.2.1 for any vector $b \in \mathbf{R}^d$.*

PROOF: (1) For any closed halfspace $H$, we have that

$$P(aX \in H) = P(X \in \frac{1}{a}H) \leq P(X \in H).$$

Theorem 6.4.1 now gives the desired result.

(2) For any closed halfspace $H$ which does not contain $b$, we have that

$$P(X \in H) \geq P(b \in H) = 0,$$

Thus by the proof of Theorem 6.4.1, we conclude that $X \overset{sc}{\geq} b$ in the sense of Definition 6.2.1. $\square$

**Remark 6.4.2** Results in Corollary 6.4.1 are generalizations of corresponding results in Bickel and Lehmann (1976) and (1979). $\square$

Applying Theorem 6.4.1, we obtain a result of Bickel and Lehmann (1976).

**Corollary 6.4.2** *Suppose that $P$ and $Q$ are symmetric about the origin in $\mathbf{R}$, with densities $f$ and $g$ respectively. Then $P \overset{sc}{\geq} Q$ in the sense of Definition 6.2.1 if $f(x)/g(x)$ is increasing for any $x > 0$.*

PROOF: Since $\forall x > 0$, $f(x)/g(x)$ is increasing, by symmetry of $P$ and $Q$, we have

$$P(H) = P([x, \infty)) \geq Q([x, \infty)) = Q(H),$$

for any closed halfspace $H = [x, \infty)$ for some $x$. Similar result holds for any closed halfspace $H = (-\infty, y]$ with some $y$ in $\mathbf{R}$. The desired result now follows immediately from Theorem 6.4.1. $\square$

**Theorem 6.4.2** *Suppose that $F_1 \overset{sc}{\geq} F_2$ in the sense of Definition 6.2.1, and $D^\alpha(F_1) \supset D^\alpha(F_2)$ for any $\alpha > 0$. Then $F_\theta \overset{sc}{\geq} F_2$ in the sense of Definition 6.2.1 for any $\theta \in (0, 1)$ if*

$$F_\theta(x) = (1 - \theta)F_1(x) + \theta F_2(x)$$

*for any $x \in \mathbf{R}^d$ and $\theta \in (0, 1)$.*

PROOF: Let $P_1, P_2$, and $P_\theta$ be the probability measures corresponding to $F_1, F_2$, and $F_\theta$, respectively. We show that

$$(*) \quad D^\alpha(F_\theta) = \{x \in \mathbf{R}^d \mid D(x; F_\theta) \geq \alpha\} \supset D^\alpha(F_2) = \{x \in \mathbf{R}^d \mid D(x; F_2) \geq \alpha\},$$

for any $\alpha > 0$. For any $x \in D^\alpha(F_2)$, assume that $D(x; F_2) = \beta \geq \alpha$. Then $D(x; F_\theta) \geq \alpha$, that is, $x \in D^\alpha(F_\theta)$. Since, if not, suppose that $D(x; F_\theta) < \alpha$, then there exists a closed halfspace $H_x$ with $x$ on its boundary such that $\alpha > P_\theta(H_x)$. Since $D^\alpha(F_1) \supset D^\alpha(F_2)$, thus $P_1(H_x) \geq \alpha$. Now since

$$P_\theta(H_x) \quad = \quad \int_{H_x} d\, F_\theta(x)$$

$$
\begin{aligned}
&= \int_{H_x} d\left((1-\theta)F_1(x) + \theta F_2(x)\right) \\
&= (1-\theta)P_1(H_x) + \theta P_2(H_x).
\end{aligned}
$$

Thus

$$\theta P_2(H_x) < \alpha - (1-\theta)P_1(H_x) < \theta\alpha,$$

thus $P_2(H_x) < \alpha$, contradicting the assumption that $D(x; F_2) = \beta \geq \alpha$. Hence $x \in D^\alpha(F_2)$ implies $x \in D^\alpha(F_\theta)$. Thus $(*)$ holds. The proof is complete. □

Theorem 6.4.2 immediately yields an important result as follows. A standard multivariate normal distribution $N_d(\mathbf{0}, \mathbf{I})$ contaminated with another multivariate normal distribution $N_d(\mathbf{0}, \sigma^2\mathbf{I})$ with $\sigma > 1$ is more scattered than the uncontaminated standard multivariate normal distribution $N_d(\mathbf{0}, \mathbf{I})$ in the sense of Definition 6.2.1.

In $\mathbf{R}$, Theorem 6.3.2 and the above theorem give the following, which was established in Bickel and Lehmann (1976).

**Corollary 6.4.3** *If $F$ and $G$ are symmetric about zero, and $G$ is more scattered than $F$, and if*

$$H_\theta(x) = \theta G(x) + (1-\theta)F(x),$$

*then $H_\theta$ is more scattered than $F$ for any $0 < \theta < 1$.*

**Theorem 6.4.3** *Assume that*

$1^o$ *$X_1$ and $X_2$ are independent with distributions $F_i$ $(i = 1, 2)$,*

*$Y_1$ and $Y_2$ are independent with distributions $G_i$ $(i = 1, 2)$,*

$2^o$ *$F_i$ and $G_i$ are symmetric about the origin in $\mathbf{R}$,*

$3^o$ *$F_1$ and $G_2$ have unimodal densities (the corresponding densities $f_1(x)$ and $g_2(x)$ are not increasing for $x > 0$),*

$4^o$ *$Y_i$ is more scattered than $X_i$ for $i = 1, 2$.*

*Then $Y_1 + Y_2$ is more scattered than $X_1 + X_2$ in the sense of Definition 6.2.1.*

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

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

The unimodality of $F_1$ implies that $f_1(x+c) - f_1(x-c) \le 0$ for any $x > 0$, and $4^o$ and Theorem 6.3.1 imply that $F_2(x) \ge G_2(x)$ for any $x > 0$. The symmetry of $F_2$ and $G_2$ about the origin implies that $F_2(0) = G_2(0) = \frac{1}{2}$. Thus

$$\begin{aligned} P(|X_1 + X_2| < c) &= 2\int_0^\infty (F_1(x+c) - F_1(x-c))\, d\, F_2(x) \\ &\ge 2\int_0^\infty (F_1(x+c) - F_1(x-c))\, d\, G_2(x) \\ &= 2\,((F_1(x+c) - F_1(x-c))G_2(x)|_0^\infty) \\ &\quad -2\left(\int_c^\infty G_2(x-c)d, F_1(x) - \int_{-c}^\infty G_2(x+c)d, F_1(x)\right) \\ &= 2\left((\frac{1}{2} - F_1(c)) - \int_0^\infty (G_2(x-c) - G_2(x+c))d\, F_1(x)\right) \\ &\quad +2\left(\int_0^c (G_2(x-c) + G_2(-x+c))d\, F_1(x)\right) \end{aligned}$$

$$\begin{aligned}
&= 2\left((\frac{1}{2} - F_1(c)) - \int_0^\infty (G_2(x-c) - G_2(x+c))d\,F_1(x)\right) \\
&\quad +2\left((F_1(c) - \frac{1}{2})\right) \\
&= 2\int_0^\infty (G_2(x+c) - G_2(x-c))d\,F_1(x).
\end{aligned}$$

Repeating the argument used above, we have that

$$\begin{aligned}
P(|X_1 + X_2| < c) &= 2\int_0^\infty (G_2(x+c) - G_2(x-c))d\,F_1(x) \\
&\geq 2\int_0^\infty (G_2(x+c) - G_2(x-c))d\,G_1(x) \\
&= P(|Y_1 + Y_2| < c).
\end{aligned}$$

The result follows immediately from Theorem 6.3.1 and Theorem 6.3.2. The proof is complete. □

**Remarks 6.4.1** (1) Theorem 6.4.3 is a generalization of Theorem 1 of Bickel and Lehmann (1976), where they required the independence of $X_1$ and $Y_2$. We have used here a similar approach to the proof of Theorem 1 of Bickel and Lehmann (1976). Under the independence assumption of $X_1$ and $Y_2$, however, the proof of the above result is much simpler.

(2) Theorem 6.4.3 is also a generalization of the main result of Birnbaum (1948), where the continuity of $Y_1$ and $X_2$ was required. □

**Theorem 6.4.4** *Let $X$ be a random vector with density $f(x)$ such that*

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

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

*Let* $Y$ *be a random vector independently distributed and symmetric about the origin in* $\mathbf{R}^d$. *Then* $X + Y \overset{sc}{\geq} X + aY$ *in the sense of Definition 6.2.1 for any* $a \in [0, 1]$.

PROOF: Let G be the distribution of $Y$. We are going to show that

$$(*) \quad D^\alpha(X+Y) = \{x \in \mathbf{R}^d \mid D(x; X+Y) \geq \alpha\}$$

$$\supset D^\alpha(X+aY) = \{x \in \mathbf{R}^d \mid D(x; X+aY) \geq \alpha\},$$

Suppose that $x \in D^\alpha(X+aY)$ and $D(x; X+aY) = \beta \geq \alpha$. We claim that $x \in D^\alpha(X+Y)$, that is, $D(x; X+Y) \geq \alpha$. Since, if not, assume that $D(x; X+Y) < \alpha$. Then there exists a closed halfspace $H_x$ such that $\alpha > P(X+Y \in H_x)$. On the other hand, it is not difficult to see that

$$\begin{aligned} P(X+Y \in H_x) &= \int_{\mathbf{R}^d} P(X \in H_x - t) d\,G(t) \\ &= \int_{\mathbf{R}^d/2} (P(X \in H_x + t) + P(X \in -H_x + t)) d\,G(t), \end{aligned}$$

where $\mathbf{R}^d/2$ is the closed halfspace with the origin on its boundary which is hyperparallel to the boundary of $H_x$, and $\mathbf{R}^d/2$ contains $H_x$. By Theorem 1 of Anderson (1955) (also see Mudholker (1966) and Anderson (1996)), we have that

$$\begin{aligned} P(X+Y \in H_x) &= \int_{\mathbf{R}^d/2} (P(X \in H_x + t) + P(X \in -H_x + t)) d\,G(t) \\ &\geq \int_{\mathbf{R}^d/2} (P(X \in H_x + at) + P(X \in -H_x + at)) d\,G(t) \\ &= \int_{R^d} P(X \in H_x - at) d\,G(t) \\ &= P(X + aY \in H_x). \end{aligned}$$

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 $x \in D^\alpha(X + Y)$, that is, $(*)$ holds. This completes the proof. □

**Corollary 6.4.4** *Let $X$ and $Y$ be the random vectors satisfying the conditions in Theorem 6.4.4. Then $X + Y \overset{sc}{\geq} bX + aY$ in the sense of Definition 6.2.1 for any $0 \leq a \leq b \leq 1$ $(b \neq 0)$.*

PROOF: By Theorem 6.4.4, $X + Y$ is more scattered than $X + \frac{a}{b}Y$. Applying Corollary 6.4.1, we have that $X + \frac{a}{b}Y$ is more scattered than $bX + aY$. Since $\overset{sc}{\geq}$ is a transitive relation, thus $X + Y \overset{sc}{\geq} bX + aY$ in the sense of Definition 6.2.1 for any $0 \leq a \leq b \leq 1$ $(b \neq 0)$. □

**Corollary 6.4.5** *Let $X \sim N_d(u, \Sigma_1)$ and $Y \sim N_d(v, \Sigma_2)$ and $\Sigma_2 - \Sigma_1$ is positive semidefinite. Then $Y \overset{sc}{\geq} X$ in the sense of Definition 6.2.1.*

PROOF: Let $Z \sim N_d(0, \Sigma_2 - \Sigma_1)$ and be independent of $X$. Then $Y - v \overset{d}{=} (X - u) + Z$. Employing Theorem 6.4.4, we obtain that $Y - v$ is more scattered than $X - u$. Affine invariance of statistical depth functionals together with translation invariance of the volume function $\Delta$ now gives the desired result. □

Corollary 6.4.5 immediately yields the following result.

**Corollary 6.4.6** *Let*

$1^o$ *$\{X_i\}$ $(i = 1, 2, \ldots, n)$ be independent random vectors with $X_i \sim N_d(u_i, \Phi_i)$,*

$2^o$ *$\{Y_i\}$ $(i = 1, 2, \ldots, n)$ be independent random vectors with $Y_i \sim N_d(v_i, \Psi_i)$.*

*If* $\sum_{i=1}^{n}(B_i\Psi_iB_i' - A_i\Phi_iA_i') \geq 0$ *(positive semidefinite), then* $\sum_i^n B_iY_i \overset{sc}{\geq} \sum_i^n A_iX_i$ *in the sense of Definition 6.2.1.*

The following result is an extension of Corollary 6.4.5.

**Theorem 6.4.5** *Suppose that* $X$ *and* $Y$ *are elliptically distributed with* $X \sim E_d(h; u, \Sigma_1)$ *and* $Y \sim E_d(h; v, \Sigma_2)$*. Then* $Y \overset{sc}{\geq} X$ *in the sense of Definition 6.2.1 if and only if* $\Sigma_2 - \Sigma_1 \geq 0$.

PROOF: Following the proof of Lemma 4.2.2 in Chapter 4, we can establish

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

and

$$\begin{aligned} D^{\alpha}(Y) &= \{y \in \mathbf{R}^d \,|\, D(y; Y) \geq \alpha\} \\ &= \{y \in \mathbf{R}^d \,|\, (y-u)'\Sigma_2^{-1}(y-u) \geq r_\alpha^2\}. \end{aligned}$$

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 $\Sigma_1^{-1} - \Sigma_2^{-1} \geq 0$. Hence $Y \overset{sc}{\geq} X$ in the sense of Definition 6.2.1 if and only if $\Sigma_2 - \Sigma_1 \geq 0$. □

**Remark 6.4.3** By Theorem 6.4.5, the conditions $\Sigma_2 - \Sigma_1$ in Corollary 6.4.5 and $\sum_{i=1}^{n}(B_i\Psi_iB_i' - A_i\Phi_iA_i') \geq 0$ in Corollary 6.4.6 are also necessary for the corresponding statements. □

A random vector $X$ is said to have a *spherically symmetric* distribution if $X \stackrel{d}{=} TX$ for any orthogonal matrices $T$ (see Muirhead (1982), p. 32).

**Theorem 6.4.6** *Let $X_i$ and $Y_i$ $(i = 1, 2)$ be independent with spherically symmetric distributions $F_i$ and $G_i$. Suppose that*

*$1^o$ $Y_i \stackrel{sc}{\geq} X_i$ in the sense of Definition 6.2.1 for $i = 1, 2$,*

*$2^o$ $F_1$ and $G_2$ have unimodal densities.*

*Then $Y_1 + Y_2 \stackrel{sc}{\geq} X_1 + X_2$ in the sense of Definition 6.2.1.*

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, $D(x; Y_1 + Y_2) \geq \alpha$. Since, if not, assume that $D(x; Y_1 + Y_2) < \alpha$, then 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 6.4.4, and utilizing spherical symmetry, we have

$$
\begin{aligned}
I &= \int_{\mathbf{R}^d} P(Y_2 \in H_x - t) d\, G_1(t) \\
&= \int_{\mathbf{R}^d/2} (P(Y_2 \in t + H_x) + P(Y_2 \in t - H_x)) d\, G_1(t) \\
&= d \int_{\mathbf{R}^d_+} (P(Y_2 \in t + H_x) + P(Y_2 \in t - H_x)) d\, G_1(t) \\
&= d \int_0^\infty \cdots \left[ \int_0^\infty P(Y_2; t, H_x) d \left( \frac{\partial^{d-1} G_1(t)}{\partial t_2 \cdots \partial t_d} \right) \right] d\, t_2 \cdots d\, t_d,
\end{aligned}
$$

where $\mathbf{R}^d_+$ is the first quadrant of $\mathbf{R}^d$, $t = (t_1, \ldots, t_d)'$, and $P(Y_2; t, H_x) = P(Y_2 \in t + H_x) + P(Y_2 \in t - H_x)$. By Theorem 1 of Anderson (1955), $P(Y_2 \in t + H_x) + P(Y_2 \in t - H_x)$

is a increasing function of $t_1$. Since $Y_1$ is more scattered than $X_1$, then by Theorem 6.3.1 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 $t_1 \geq 0$ ($t_i \geq 0, i = 2, \ldots, d$). Integral by parts now yields

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

Now utilizing a similar argument as above, we obtain

$$\begin{aligned} P(Y_1 + Y_2 \in H_x) &= P(X_1 + Y_2 \in H_x) \\ &\geq P(X_1 + X_2 \in H_x). \end{aligned}$$

Hence $\alpha > P(X_1 + X_2 \in H_x)$, contradicting to the assumption that $D(x; X_1 + X_2) \geq \alpha$. Therefore $x \in D^\alpha(X_1 + X_2)$ implies $x \in D^\alpha(Y_1 + Y_2)$, proving that $Y_1 + Y_2 \overset{sc}{\geq} X_1 + X_2$ in the sense of Definition 6.2.1. The proof is complete. $\square$

Theorem 6.4.6 and Corollary 6.4.6 are established under the assumptions that the underlying distributions are spherically distributed or multivariate normally distributed. We conclude this section with an open question: Under what weaker assumption(s) on distributions can we establish an analogue of Theorem 6.4.3 in the multivariate setting?

## 6.5 Depth-Based Multivariate Nonparametric Scatter Measures

In this section, we introduce a definition of depth-based multivariate nonparametric scatter measure and provide some examples.

Denote by $\mathcal{F}$ a class of distributions on $\mathbf{R}^d$. Denote by $\mathbf{R}^+$ the nonnegative real numbers in $\mathbf{R}$.

**Definition 6.5.1** A functional $\varphi : \mathcal{F} \to \mathbf{R}^+$ is said to be a multivariate nonparametric measure of scatter in $\mathbf{R}^d$ if it satisfies

(i) $\varphi(P) \geq \varphi(Q)$, for any $P$ and $Q$ in $\mathcal{F}$ such that $P \overset{sc}{\geq} Q$ in the sense of Definition 6.2.1;

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

**Remarks 6.5.1** (1) Bickel and Lehmann (1976) and (1979) introduced dispersion measures and spread measures in $\mathbf{R}$ respectively. Their definitions are similar to Definition 6.5.1 except the notion of "more scattered" above is in the sense of (D1) and (D2) respectively. Oja (1983) using his "more scattered" notion extended the definition of Bickel and Lehmann (1979) to higher dimensions.

(2) By Theorem 6.3.2, for symmetric distributions the above definition is equivalent to that of Bickel and Lehmann (1976) in $\mathbf{R}$.

(3) According to Theorem 6.3.4, if $\phi$ is an affine transformation in (D4), then a scatter measure in Oja's sense is also a scatter measure in the sense of Definition 6.5.1. $\square$

**Example 6.5.1** Define

$$\varphi^\alpha(F) = \Delta(D^\alpha(F)),$$

for any $\alpha > 0$ and $F \in \boldsymbol{\mathcal{F}}$, where $\Delta$ is a volume function and $D^\alpha(F) = \{x \in \mathbf{R}^d \mid D(x;F) \geq \alpha\}$ is a depth contour for some statistical depth functional $D(\cdot,\cdot)$. Then as shown in Theorem 6.5.1 below, $\varphi^\alpha$ is a multivariate nonparametric measure of scatter. Note that in $\mathbf{R}$, when $\alpha = \frac{1}{4}$, $\varphi^\alpha(F)$ gives the interquartile range of $F$, which was suggested in Bickel and Lehmann (1979) as a scatter measure. $\square$

**Theorem 6.5.1** *$\varphi^\alpha$ defined in Example 6.5.1 is a multivariate nonparametric scatter measure in the sense of Definition 6.5.1 for any fixed $\alpha > 0$.*

PROOF: By Definition 6.2.1, it is easy to see that for any fixed $\alpha$ $\varphi^\alpha$ satisfies condition (i) in Definition 6.5.1. On the other hand, Theorem 2.20 of Rudin (1987) or Theorem 12.2 of Billingsley (1986) implies that $\varphi^\alpha$ satisfies condition (ii) in Definition 6.5.1. Hence $\varphi^\alpha$ is a multivariate nonparametric scatter measure in the sense of Definition 6.5.1, for any fixed $\alpha > 0$. $\square$

## 6.6 Summary

In this chapter, statistical depth functions have been utilized to introduce a notion of "more scattered" for comparison of one multivariate distribution with another in $\mathbf{R}^d$. Relationships among this new notion and the notions of Bickel and Lehmann (1976), Eaton (1982) and Oja (1983) have been explored. It turns out that this notion is a generalization of that of Bickel and Lehmann (1976) in $\mathbf{R}$, and is more general than those of Eaton (1982) and Oja (1983) in $\mathbf{R}^d$ under some typical conditions. The properties related to this depth-based new notion have been studied, some depth-based multivariate nonparametric scatter measures have been defined, and some examples have been presented and studied.

## 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] Arcones, M. A., Chen, Z., and Giné, E. (1994). Estimators related to $U$-Processes with applications to multivariate medians: asymptotic normality. *Ann. Statist.* **22** 1460-1477.

[4] Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1993). *A First Course in Order Statistics.* John Wiley & Sons, New York.

[5] Barnett, V. (1976). The ordering of multivariate data (with comments). *J. R. Statist. Soc. A* **139** 318-354.

[6] Bickel, P. J., and Lehmann, E. L. (1975). Descriptive statistics for nonparametric models. (I. Introduction, II. Location). *Ann. Statist.* **3** 1039-1069.

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

[8] Bickel, P. J., and Lehmann, E. L. (1979). Descriptive statistics for nonparametric models. (IV. Spread). In J. Jureckova, Ed., *Contributions to Statistics* (pp. 33-40). Academia, Prague.

[9] Billingsley, P. (1986). *Probability and Measure.* (Second Edition). John Wiley & Sons, New York.

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

[11] Chen, Z. (1995). Robustness of the half-space median. *J. Statist. Plann. Inference.* **46** 175-181.

[12] David, H. A. (1981). *Order Statistics.* John Wiley & Sons, New York.

[13] Donoho, D. L., and Gasko, M. (1992). Breakdown properties of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269-1292.

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

[15] Eddy, W. F. (1985). Ordering of multivariate data. In L. Billard, Ed., *Computer Science and Statistics: The Interface* (pp. 25-30). North-Holland, Amsterdam.

[16] Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric Multiariate and Related Distributions.* Chpman and Hall, London.

[17] He, X., and Wang, G. (1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.* **25** 495-504.

[18] Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks.* John Wiley & Sons, New York.

[19] Hoeffding, W. (1963). Probability inequalities for sum of bounded random variables. *J. Amer. Statis. Assoc.* **36** 369-408.

[20] Kemperman, J. H. B. (1987). The median of a finite measure on a Banach space. In Y. Dodge, Ed., *Statistical Data Analysis Based on the $L_1$-Norm and Related Methods* (pp. 217-230). Elsevier Science, North-Holland, Amsterdam.

[21] Koshevoy, G., and Mosler, K. (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.* **25** 1998-2017.

[22] Liu, R. Y. (1988). On a notion of simplicial depth. *Proc. Natl. Acad. Sci. USA.* **85** 1732-1734.

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

[24] Liu, R. Y. (1992). Data depth and multivariate rank tests. In Y. Dodge, Ed., *$L_1$-Statistics Analysis and Related Methods* (pp. 279-294). Elsevier Science, North-Holland, Amsterdam.

[25] Liu, R. Y., Parelius, J. M., and Singh, K. (1997). Multivariate Analysis by data depth: descriptive statistics, graphics and inference. Preprint.

[26] 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.

[27] Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proc. of the National Academy India.* **12** 49-55.

[28] Massé, J. C., and Theodorescu, R. (1994). Halfplane trimming for bivariate distributions. *J. Multivariate Anal.* **48** 188-202.

[29] Milasevic, P., and Ducharme, G. R. (1987). Uniqueness of the spatial median. *Ann. Statist.* **15** 1332-1333.

[30] 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.

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

[32] Nolan, D. (1992). Asymptotics for multivariate trimming. *Stochastic Process. Appl.* **42** 157-169.

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

[34] Randles, R. H. (1989). A distribution-free multivariate sign test based on interdirections. *J. Amer. Statist. Assoc.* **84** 1045-1050.

[35] Reiss, R. D. (1989). *Approximate Distributions of Order Statistics With Applications to Nonparametric Statistics.* Springer, New York.

[36] Rousseeuw, P. J., and Hubert, M. (1996). Regression depth. Preprint.

[37] Rousseeuw, P. J., and Struyf, A. (1997). Computing location depth and regression depth in higher dimensions. Preprint.

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

[39] Ruts, I., and Rousseeuw, P. J. (1996). Computing depth contours of bivariate point clouds. *Computational Statistics and Data Analysis.* **23** 153-168.

[40] Serfling, R. (1980). *Approximation theorems of mathematical statistics.* John Wiley & Sons, New York.

[41] Singh, K. (1991). A notion of majority depth. Tech. Report, Rutgers University, Dept. of Statistics.

[42] Small, C. G. (1990). A survey of multidimensional medians. *Internat. Stat. Rev.* **58** 263-277.

[43] Steele, J. M. (1978). Empirical discrepancies and subadditive processes. *Ann. Probab.* **6** 118-127.

[44] Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statis.* **36** 423-439.

[45] Tukey, J. W. (1974). Mathematics and picturing data. *Proc. Int. Congr. Mathematics, Vancouver.* **2** 523-531.

[46] Tukey, J. W. (1977). *Exploratory Data Analysis.* Addison-Wesley, Reading, Mass.

[47] Vapnik, V. N., and Červonenkis, A. YA. (1971). On the uniform conergence of relative frequencies of eents to their probabilities. *Theory Probab. Appl.* **16** 264-280.

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