Some Intriguing Properties of Tukey's Half-Space Depth

Subhajit Dutta, Anil K. Ghosh and Probal Chaudhuri

Theoretical Statistics and Mathematics Unit, Indian Statistical Institute,

203, B. T. Road, Kolkata 700108, India.

E-mail: subhajit_r@isical.ac.in, akghosh@isical.ac.in, probal@isical.ac.in

Abstract

Tukey's half-space depth is one of the most popular depth functions available in the literature. It is conceptually very simple, satisfies several desirable properties of depth functions, and can have a natural extension for probability distributions on any Banach space, which may be finite or infinite dimensional. In this article, we derive some counter intuitive properties of half-space depth for probability distributions in finite and infinite dimensional spaces.

Keywords : Banach space, depth contours, half-space median, l_p norm, symmetric distributions.

1 Introduction

Over the last three decades, data depth has emerged as a powerful concept leading to generalization of many univariate statistical methods in multivariate set-up. A depth function measures the centrality of a point $\mathbf{x}$ w.r.t. a data set or a probability distribution, and thus helps to define order and ranks of multivariate data. There are several notions of data depth available in the literature (see, e.g., Liu *et. al.*, 1999; Zuo and Serfling, 2000; Vardi and Zhang, 2000; Mosler, 2002; Mizera and Muller, 2004; Lopez-Pintado and Romo, 2006). Tukey's half-space depth (see Tukey, 1975) is one of them, which is also one of the most popular depth functions used by many researchers in various situations. The construction of central regions based on trimming (see, e.g., Nolan, 1992), robust estimation of multivariate location (see, e.g., Donoho and Gasko, 1992), testing of multivariate statistical hypotheses (see, e.g., Chaudhuri and Sengupta, 1993) and supervised classification (see, e.g., Ghosh and Chaudhuri, 2005a, 2005b) are some examples of its wide spread application.

Like other popular depth functions, half-space depth has some nice theoretical properties. In fact, it satisfies all four desirable properties of depth functions mentioned in Zuo and Serfling (2000), i.e., affine invariance, maximality of center, monotonicity w.r.t. the deepest point and vanishing at infinity. Moreover, if the underlying population distribution F has a spherically symmetric density f, i.e., $f(\mathbf{x}) = \psi(\|\mathbf{x}\|_2)$ for some $\psi : R_+ \to R_+$, it turns out to be a decreasing function of $\|\mathbf{x}\|_2$. Here, for any $p > 0$ and $\mathbf{x} = (x_1, x_2, \ldots, x_d) \in R^d$, we define $\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p + \ldots + |x_d|^p)^{1/p}$. So, if ψ is also monotonically decreasing (i.e., f is unimodal), half-space depth becomes an increasing function of f and vice versa. Therefore, in such cases, the half-space depth contours coincide with the contours of the density function. Because of this property of half-space depth, it has been used to develop several likelihood based methods for multivariate data. For instance, classification rules based on the ordering of the half-space depth functions coincide with the optimal Bayes classifier for discriminating among elliptically symmetric unimodal populations, which differ in their centers of symmetry (see, e.g., Ghosh and Chaudhuri, 2005b). Similarly, the use of half-space depth function to order and trim multivariate data sets (see, e.g., Nolan, 1992; Donoho and Gasko, 1992) leading to the determination of central and outlying observations can have a natural justification when the density contours coincide with the half-space depth contours. Now, a natural question is whether this property of half-space depth contours holds for other l_p-symmetric distributions, i.e., when $f(\mathbf{x}) = \psi(\|\mathbf{x}\|_p)$ for some $p \neq 2$ and ψ is monotonically decreasing. In Section 2, we carry out an investigation to answer this question.

In Section 3, we consider natural extensions of half-space depth and half-space median for probability distributions in arbitrary Banach spaces using the concept of linear functionals on such spaces. Some anomalous behavior of half-space depth for probability distributions on infinite dimensional spaces and their implications are discussed in Section 4.

2 Half-space depth contours for l_p-symmetric density functions

In this section, we study the behavior of the half-space depth contours for a wide class of symmetric distributions. As it has been mentioned in the introduction, the half-space depth contours coincide with the density contours if the p.d.f. f is such that $f(\mathbf{x}) = \psi(\|\mathbf{x}\|_2)$ for some monotonically decreasing $\psi : R^+ \rightarrow R^+$, and this is an important feature of half-space depth with many useful statistical applications. Here, we will investigate the situation when $\|\cdot\|_2$ is replaced by $\|\cdot\|_p$, where p is positive and $p \neq 2$.

2.1 Depth contours for $p = \infty$

For $p = \infty$, the p.d.f. $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_d) = \psi(\max\{|x_1|, |x_2|, \ldots, |x_d|\})$ for some monotonically decreasing function ψ. Clearly, the density contours here are concentric d-dimensional hypercubes with origin at the center. Now, let us try to check whether all points on the surface of a hypercube with origin as the center have the same depth or not. First, consider the point $A = (1, 0, \ldots, 0)$ on the surface of the unit hypercube $\{\mathbf{x} : \|\mathbf{x}\|_\infty = 1\}$ (see Figure 1 for a diagram in the case d=2). It can be shown that the hyperplane $x_1 = 1$ determines the half-space depth of this point, and this depth is $P(X_1 \geq 1)$, where $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ has the p.d.f. $f(\mathbf{x})$ (see Lemma 1 in the Appendix).

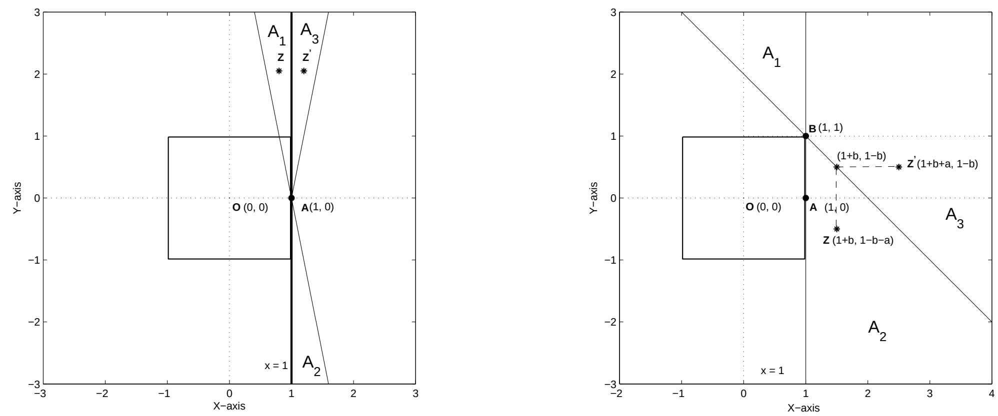

Figure 1: l_∞ contour and the line defining the half-space depth of $(1, 0)$.

Note that the line $x_1 = 1$ also passes through the point $B = (1, 1, 0, \ldots, 0)$ (see the right diagram in Figure 1 when d=2). So, A and B will have the same depth if and only if there exists no other hyperplane that passes though B in such a way that the probability of one of its half-spaces is smaller than $P(X_1 \geq 1)$. However, the hyperplane $x_1 + x_2 = 2$ passes through the point B, and we can show that $P(X_1 + X_2 \geq 2) < P(X_1 \geq 1)$ (see Lemma 2 in the Appendix). This implies that if the p.d.f. f is of the form $f(\mathbf{x}) = \psi(\|\mathbf{x}\|_\infty)$ with a monotonically decreasing ψ, the half-space depth contours cannot coincide with the corresponding density contours.

2.2 Depth contours for $1 \leq p < \infty$

Next, consider the case $1 \leq p < \infty$. Clearly, $A = (2^{1/p}c, 0, 0, \ldots, 0)$ and $B = (c, c, 0, \ldots, 0)$ are two points on the same l_p contour (see Figure 2 in the case d=2). First, let us check whether the half-space depths of these two points are equal or not. In view of Lemma 1, the depth of A is given by $P(X_1 \geq 2^{1/p}c)$ when $c > 0$. Here $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ has the p.d.f. $f(\mathbf{x}) = \psi(\|\mathbf{x}\|_p)$. We can also prove that the hyperplane $x_1 + x_2 = 2c$ determines the half-space depth of B, and this depth is $P(X_1 + X_2 \geq 2c)$ (see Lemma 3 in the Appendix).

It follows from the discussion in the preceding paragraph that the two points A and B will have the same depth only if $P(X_1 \geq 2^{1/p}c) = P(X_1 + X_2 \geq 2c)$. Note that here one can choose c arbitrarily. Therefore, the depth and the density contours can coincide only if $P(X_1 \geq 2^{1/p}c) = P(X_1 + X_2 \geq 2c)$ for all values of c, i.e., only if X_1 and $2^\alpha(X_1 + X_2)$ are identically distributed for $\alpha = (1 - p)/p$. Now, if we assume the existence of the second moments of the X_i's, the equality of the variances of X_1 and $2^\alpha(X_1 + X_2)$ and the fact that X_1 and X_2 are uncorrelated (in view of the l_p-symmetry of the density f) imply that $\alpha = -1/2$ or $p = 2$. Even if we do not assume any moment condition, the above result holds (see Lemma 4 in the Appendix). Also, it is interesting to notice that for $p < 2$, we can always choose a c such that the depth of B is more than that of A. On the other hand, for $p > 2$, it is always possible to choose a c such that A has larger depth than B.

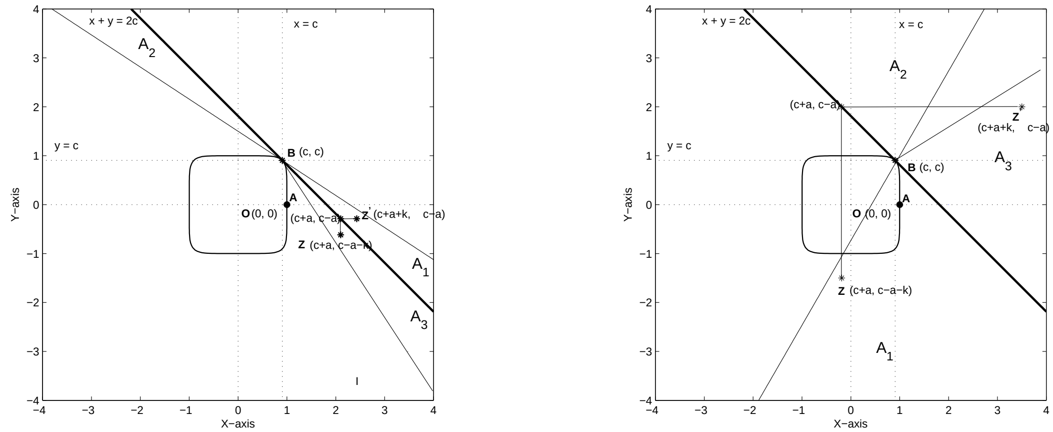

Figure 2: l_p contour and the lines defining the half-space depth of (c, c).

2.3 Depth contours for $p < 1$

Finally, we need to investigate the case $p < 1$. Note that in this case, the regions bounded by l_p contours are no longer convex sets (see Figure 3 for the case d=2). Consider three points $A = (1, 0, \ldots, 0)$, $B = (0, 1, 0, \ldots, 0)$ and $C = (\alpha, \beta, 0, \ldots, 0)$ on the same l_p contour, where $\alpha, \beta > 0$. Consider any hyperplane passing through C, which will split R^d into two half-spaces, one of which will contain the origin. Since $p < 1$, at least one of the two points A and B will lie on the half-space that does not contain the origin. Without loss of generality, we can assume that the hyperplane that determines the half-space depth of C puts B and the origin on two different half-spaces (see the bold line in Figure 3 for the case d=2). Now, we can make a parallel shift of that hyperplane away from the origin until it hits the point B (see the dotted line in Figure 3 in the case d=2). Clearly, the half-space created by this new hyperplane that has smaller probability measure, will have smaller probability than the probability of each of the two half-spaces created by the older hyperplane. Therefore, the half-space depth of B has to be smaller than that of C, and hence the depth contours cannot coincide with the density contours in this case.

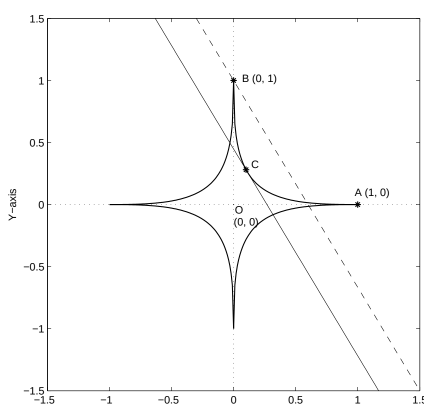

Figure 3: l_p contour for $p < 1$

Summarizing our discussion in this section, we now have the following theorem.

Theorem 1 : *Consider a probability distribution on R^d with the p.d.f. f such that $f(\mathbf{x}) = \psi(\|\mathbf{x}\|_p)$ for some monotonically decreasing function ψ. Then the half-space depth contours associated with f will coincide with the density contours if and only if $p = 2$.*

3 Half-space median and its depth

Note that for any l_p-symmetric density function $f(\mathbf{x}) = \psi(\|\mathbf{x}\|_p)$ with $0 < p \leq \infty$, the origin turns out to be the half-space median (i.e. the point with the maximum half-space depth). In fact, this is true whenever $\mathbf{X}$ and $-\mathbf{X}$ have the same distribution, or even under a slightly weaker condition that any real valued linear projection has the median zero. One should also notice that in all these cases, the half-space median coincides with the co-ordinate wise median and the depth of the half-space median, namely the origin, is 0.5. However, this holds only for a special class of multivariate distributions. For instance, for a bivariate uniform distribution on a right-angled isosceles triangle, one can easily show that the half-space depth of any point is smaller than 0.5. We can consider another interesting example of a continuous bivariate distribution, where the p.d.f. f has the support $\{(x_1, x_2) : x_1 + x_2 \geq 0,\ x_1 x_2 \leq 0\}$. In addition, if f is symmetric about the $x_1 = x_2$ line, one can easily verify that the half-space median will have depth smaller than 0.5, and the co-ordinate wise median will have zero half-space depth. We have already indicated some sufficient conditions for the depth of the half-space median to be 0.5, and in view of the two preceding examples, one would like to know some necessary and sufficient conditions for that. We now state a theorem, the proof of which is given in the Appendix.

Theorem 2 : *Suppose that $\mathbf{X}$ is a d-dimensional random vector with a probability distribution, which has its half-space median at $\boldsymbol{\mu} \in R^d$. Then, the half-space depth of $\boldsymbol{\mu}$ will be 0.5 if and only if $(\mathbf{X} - \boldsymbol{\mu})/\|\mathbf{X} - \boldsymbol{\mu}\|_2$ and $(\boldsymbol{\mu} - \mathbf{X})/\|\mathbf{X} - \boldsymbol{\mu}\|_2$ are identically distributed.*

Note that the condition that $(\mathbf{X} - \boldsymbol{\mu})/\|\mathbf{X} - \boldsymbol{\mu}\|$ and $(\boldsymbol{\mu} - \mathbf{X})/\|\mathbf{X} - \boldsymbol{\mu}\|$ are identically distributed is sufficient for the half-space median to have half-space depth 0.5 even when $\mathbf{X}$ lies in an arbitrary Banach space $\mathcal{B}$, where $\|\cdot\|$ denotes the norm in $\mathcal{B}$. If F is a probability distribution over $\mathcal{B}$, and $\mathbf{x}$ is a fixed element in $\mathcal{B}$, the half-space depth of $\mathbf{x}$ can be defined as $HD(\mathbf{x}, F) = \inf_{h \in B^*} P\{h(\mathbf{X} - \mathbf{x}) \geq 0\}$, where $h : \mathcal{B} \to R$ in a linear functional that belongs to the dual space $\mathcal{B}^*$, P stands for the probability measure on $\mathcal{B}$ corresponding to F, and $\mathbf{X}$ is a random element in $\mathcal{B}$ having the distribution F. The point $\boldsymbol{\mu} \in \mathcal{B}$ is called a half-space median if $HD(\boldsymbol{\mu}, F) = \sup_{\mathbf{x} \in \mathcal{B}} HD(\mathbf{x}, F)$. Instead of Banach spaces, if we work with a Hilbert space $\mathcal{H}$, due to Reisz representation theorem and the reflexive nature of a Hilbert space, the half-space depth of an observation $\mathbf{x} \in \mathcal{H}$ can be defined as $HD(\mathbf{x}, F) = \inf_{\mathbf{h} \in \mathcal{H}} P\{< \mathbf{h}, (\mathbf{X} - \mathbf{x}) > \ \geq 0\}$, where $< \cdot, \cdot >$ stands for the inner product defined on $\mathcal{H}$.

From the above discussion, it is clear that if we have a symmetric distribution in a Hilbert or a Banach space, the point of symmetry will have the maximum depth 0.5, and it will be the half-space median. So, in a sense, the half-space median is well defined and behaves in a nice way even in infinite dimensional spaces for symmetric probability distributions. However, in infinite dimensional spaces, even when we deal with nice symmetric distributions, the half-space depth function can have some anomalous behavior, which we will see in the next section.

4 Anomalous behavior of half-space depth in infinite dimensional spaces

We know that if we have a data cloud of n observations in a d-dimensional space, the empirical depth of an observation lying outside the convex hull formed by the data cloud is zero. For $d > n$, since the Lebesgue measure of this convex hull is zero, we have zero depth for all points in a set of probability measure one whenever we have n i.i.d. observations from an absolutely continuous distribution in R^d. In fact, for any probability measure on an infinite dimensional Banach space such that any finite dimensional hyperplane in that space has zero probability, the empirical half-space depth based on finitely many i.i.d. observations from that probability distribution will be zero almost everywhere. So, the empirical version of half-space depth does not carry any statistically useful information in such cases. Naturally, one would be curious to know what happens to the population depth function in such situations. The following theorem demonstrates that it is possible to have a nice symmetric probability distribution on the l_2 space for which the population depth function takes positive values only on a set of probability measure zero. Recall that the l_2 space of real sequences consists of infinite sequences like $(x_1, x_2, \ldots)$ such that $\sum_{i=1}^{\infty} x_i^2 < \infty$.

Theorem 3 : *Consider an infinite sequence of independent random variables* $X_1, X_2, \ldots,$ *where* $E(X_i) = 0$ *and* $E(X_i^2) = \sigma_i^2$ *for all* $i \geq 1$ *such that* $\sum_{i=1}^{\infty} \sigma_i^2 < \infty$. *Note that this implies that* $\mathbf{X} = (X_1, X_2, X_3 \ldots)$ *lies in the* l_2 *space of real sequences with probability one. Also, assume that the* X_i*'s have finite fourth moments and* $\sum_{i=1}^{\infty} E(X_i^4)/i^2\sigma_i^4 < \infty$. *For instance, all these conditions will hold if the* X_i*'s are independent Gaussian random variables. Then, for any given* $\mathbf{x} = (x_1, x_2, \ldots)$ *in that* l_2 *space, the half-space depth of* $\mathbf{x}$ *w.r.t. the distribution of* $\mathbf{X}$ *will be zero unless* $\mathbf{x}$ *lies in a subset having probability zero.*

The proof of this theorem is given in the Appendix. This theorem clearly shows that not only the empirical version, but also the population version of the half-space depth may fail to provide any meaningful statistical measure in inifinite dimensional spaces. One of the well known features of half-space depth is its characterization property. Koshevoy (2002) proved that if the half-space depth functions of two atomic measures are identical, then the measures are also identical. Under some regularity conditions, Koshevoy (2003) as well as Hassairi and Regaieg (2008) proved this characterization property of half-space depth for absolutely continuous probability distributions in finite dimensional spaces. From the above discussion, however, it is clear that in the l_2 space of real sequences, there exist several probability measures (which may even have independent Gaussian marginals) with their half-space depth functions identically equal to zero except on a subset having zero probability measure.

Appendix

Lemma 1 : *Let* $HD(\mathbf{x}, F)$ *be the half-space depth of* $\mathbf{x}$ *w.r.t. the distribution* F*, and* F *have density* f *of the form* $f(\mathbf{x}) = \psi(\|\mathbf{x}\|_p)$ *with a monotonically decreasing function* ψ *and* $0 < p \leq \infty$*. Then, for any* $\mathbf{x} = (x, 0, \ldots, 0)$ *on the co-ordinate axis, we have* $HD(\mathbf{x}, F) = P(X_1 \geq x)$ *when* $x > 0$*, and* $HD(\mathbf{x}, F) = P(X_1 \leq x)$ *when* $x \leq 0$*.*

Proof : We will prove it for $\mathbf{x}_0 = (1, 0, \ldots, 0)$. Proof for other points follows in the same way. Consider any hyperplane $\boldsymbol{\alpha}(\mathbf{x} - \mathbf{x}_0)' = 0$ other than $x_1 = 1$ that passes through $\mathbf{x}_0$ (see the left diagram in Figure 1 in the case d=2). Here $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ is a vector in R^d. Define the regions $A_1 = \{\mathbf{x} : \ x_1 < 1 \ \text{ and } \boldsymbol{\alpha}(\mathbf{x} - \mathbf{x}_0)' \geq 0\}$ and $A_2 = \{\mathbf{x} : \ x_1 \geq 1 \ \text{ and } \boldsymbol{\alpha}(\mathbf{x} - \mathbf{x}_0)' < 0\}$ (see the left diagram in Figure 1 for the case d=2). To prove the lemma, we have to show that $P(\mathbf{X} \in A_1) \geq P(\mathbf{X} \in A_2)$. Define $A_3 = \{\mathbf{x} = (x_1, x_2, \ldots, x_d) : (x_1, -x_2, -x_3, \ldots, -x_d) \in A_2\}$. Because of the symmetry of f, it is easy to check that $P(\mathbf{X} \in A_2) = P(\mathbf{X} \in A_3)$. So, it is enough to prove $P(\mathbf{X} \in A_1) \geq P(\mathbf{X} \in A_3)$. Note that for every point $\mathbf{z} = (x_1, x_2, \ldots, x_d)$ in A_1, we have a point $\mathbf{z}' = (x_1', x_2, x_3, \ldots, x_d)$ in A_3 such that $x_1' = 2x_1 - 1$. Hence, $|x_1| \leq |x_1'|$ and $\|\mathbf{z}\|_p \leq \|\mathbf{z}'\|_p$ with strict inequality being for true for all $\mathbf{z}$ not lying on the hyperplane $x_1 = 1$. This implies $f(\mathbf{z}) \geq f(\mathbf{z}')$. Since the strict inequality holds over a set of positive measure, integrating $f(\mathbf{z})$ (and $f(\mathbf{z}')$) w.r.t. $\mathbf{z}$ (and $\mathbf{z}'$), we actually get $P(\mathbf{X} \in A_1) > P(\mathbf{X} \in A_3)$. □

Lemma 2 : *Consider a p.d.f.* f *on* R^d *satisfying* $f(\mathbf{x}) = \psi(\|\mathbf{x}\|_\infty)$ *and a random vector* $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ *with that p.d.f. Then, for any* $x > 0$*, we have* $P(X_1 + X_2 \geq 2x) < P(X_1 \geq x)$*.*

Proof : Again, we will prove it for $x = 1$ only. Let us define $A_1 = \{\mathbf{x} = (x_1, x_2, \ldots, x_d) : x_1 < 1 \ \text{ and } x_1 + x_2 \geq 2\}$ and $A_2 = \{\mathbf{x} = (x_1, x_2, x_3, \ldots, x_d) : x_1 \geq 1 \ \text{ and } x_1 + x_2 < 2\}$ (these two regions are shown in the right diagram in Figure 1 for the case d=2). Also define the region $A_3 = \{\mathbf{x} = (x_1, x_2, \ldots, x_d) : \ (x_2, x_1, x_3, \ldots, x_d) \in A_1\}$. Because of the symmetry of $f(\mathbf{x})$ under permutations of the co-ordinates of $\mathbf{x}$, it is straight forward to see that $P(\mathbf{X} \in A_1) = P(\mathbf{X} \in A_3)$. Hence, it is enough to show that $P(\mathbf{X} \in A_3) < P(\mathbf{X} \in A_2)$. Now, for any $\mathbf{z} = (z_1, z_2, \ldots, z_d) \in A_2$, we have a corresponding point $\mathbf{z}' = (2 - z_2, 2 - z_1, z_3, \ldots, z_d)$ in A_3. Also note that for any $\mathbf{z} = (z_1, z_2, \ldots, z_d)$ in A_2, z_1 and z_2 are of the form $z_1 = 1 + b$ and $z_2 = 1 - b - a$ for some $a, b > 0$ (see the right diagram in Figure 1 for the case d=2). Consequently, for $\mathbf{z}^{'} = (z_1^{'}, z_2^{'}, z_3 \ldots, z_d)$, we have $z_1^{'} = 1 + b + a$ and $z_2^{'} = 1 - b$. Clearly, $\max\{|z_1|, |z_2|\} < \max\{|z_1^{'}|, |z_2^{'}|\} = 1 + a + b$, which implies $\|\mathbf{z}\|_\infty \leq \|\mathbf{z}^{'}\|_\infty$ and hence $f(\mathbf{z}) > f(\mathbf{z}^{'})$ with strict inequality on a set of positive probability measure under f. This proves that $P(\mathbf{X} \in A_2) > P(\mathbf{X} \in A_3)$. □

Lemma 3 : *Let $f(\mathbf{x}) = \psi(\|\mathbf{x}\|_p)$ for $1 \leq p < \infty$, and it is the p.d.f. of $\mathbf{X} = (X_1, X_2, \ldots, X_d)$. Consider $\mathbf{x}_0 = (c, c, 0, \ldots, 0)$ for $c > 0$. Then, its half-space depth is given by $HD(\mathbf{x}_0, F) = P(X_1 + X_2 \geq 2c)$.*

Proof : Consider the hyperplane $x_1 + x_2 = 2c$ (see Figure 2 in the case d=2). We have to show that this hyperplane determines the half-space depth of $\mathbf{x}_0$. For this, we will follow the same line of arguments as in Lemmas 1 and 2. Consider a new hyperplane $\boldsymbol{\alpha}(\mathbf{x} - \mathbf{x}_0)^{'} = 0$ passing through $\mathbf{x}_0$ (see Figure 2 for the case d=2). Define the regions $A_1 = \{\mathbf{x} = (x_1, x_2, \ldots, x_d) : \ x_1 + x_2 < 2c \ \text{and} \ \boldsymbol{\alpha}(\mathbf{x} - \mathbf{x}_0)^{'} \geq 0\}$ and $A_2 = \{\mathbf{x} = (x_1, x_2, \ldots, x_d) : \ x_1 + x_2 \geq 2c \ \text{and} \ \boldsymbol{\alpha}(\mathbf{x} - \mathbf{x}_0)^{'} < 0\}$ (see Figure 2 for the case d=2). To prove the lemma, we have to show that $P(\mathbf{X} \in A_1) \geq P(\mathbf{X} \in A_2)$. Define $A_3 = \{\mathbf{x} = (x_1, x_2, \ldots, x_d) : (x_2, x_1, x_3, \ldots, x_d) \in A_2\}$. Because of the symmetry of $f(\mathbf{x})$ under any permutation of the co-ordinates of $\mathbf{x}$, we have $P(\mathbf{X} \in A_2) = P(\mathbf{X} \in A_3)$. Therefore, it is enough to show that $P(\mathbf{X} \in A_3) \leq P(\mathbf{X} \in A_1)$.

Note that any point $\mathbf{z} \in A_1$ is of the form $\mathbf{z} = (c + a, c - a - k, x_3, \ldots, x_d)$, where $k > 0$, and a can be positive or negative (see Figure 2 in the case d=2). For any $\mathbf{z} \in A_1$, we get a corresponding point $\mathbf{z}^{'} \in A_3$ such that $\mathbf{z}^{'} = (c + a + k, c - a, x_3, \ldots, x_d)$. Now, we need to show that $\|\mathbf{z}\|_p \leq \|\mathbf{z}^{'}\|_p$, and for that, we will consider the two cases $a > 0$ and $a < 0$ separately.

When $a > 0$ (see the left diagram in Figure 2 for the case d=2), we have $0 < |c - a| < |c + a|$. Now, for $p \geq 1$ and $t, k > 0$, it is easy to check that the function $h(t) = (t + k)^p - t^p$ is non-decreasing in t. So, for $0 < t_1 < t_2$, we have $0 < h(t_1) \leq h(t_2)$. Taking $t_1 = |c - a|$ and $t_2 = |c + a|$, we get $(|c - a| + k)^p - |c - a|^p \leq (|c + a| + k)^p - |c + a|^p$. Now, using the facts that $|c + a| + k = |c + a + k|$ and $|c - a - k| \leq |c - a| + k$, one arrives at $|c - a - k|^p - |c - a|^p \leq |c + a + k|^p - |c + a|^p$. This implies that $|c - a - k|^p + |c + a|^p \leq |c + a + k|^p + |c - a|^p$, which in turn implies that $\|\mathbf{z}\|_p \leq \|\mathbf{z}^{'}\|_p$. Here also one can notice that the strict inequality holds on a set of positive probability measure under f.

For $a < 0$ (see the right diagram in Figure 2 in the case d=2), first note that $a + k > 0$ and the co-ordinates of $\mathbf{z}$ and $\mathbf{z}^{'}$ are of the form $\mathbf{z} = (c - \alpha, c - \beta, x_3, \ldots, x_d)$ and $\mathbf{z}^{'} = (c + \alpha, c + \beta, x_3, \ldots, x_d)$, respectively, where $\alpha = -a > 0$ and $\beta = a + k > 0$. Now, $|c - \alpha| < |c + \alpha|$ and $|c - \beta| < |c + \beta|$ imply that $\|\mathbf{z}\|_p < \|\mathbf{z}^{'}\|_p$. □

Lemma 4 : *Assume that we have a p.d.f. f that satisfies $f(\mathbf{x}) = \psi(\|\mathbf{x}\|_p)$ for some $p > 0$ and monotonically decreasing ψ. Let $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ be a random vector with p.d.f. f. If X_1 and $2^{(1-p)/p}(X_1 + X_2)$ are identically distributed, we must have $p = 2$.*

Proof: First note that if $f(\mathbf{x}) = \psi(\|\mathbf{x}\|_p)$, the joint p.d.f. of X_1 and X_2 is of the form $f_1(x_1, x_2) = \psi_1(|x_1|^p + |x_2|^p)$ for some $\psi_1 : R_+ \to R_+$. One can show that the p.d.f.'s of X_1 and $Y = 2^\alpha(X_1 + X_2)$, where $\alpha = (1-p)/p$, are given by $f_{X_1}(x) = \int \psi_1((|x|^p + |x_2|^p)^{1/p})dx_2$ and $f_Y(x) = 2^{-\alpha} \int \psi_1((|2^{-\alpha}x - x_2|^p + |x_2|^p)^{1/p})dx_2$, respectively.

Since both of these two p.d.f.'s are continuous functions, and X_1 and Y are identically distributed, we can equate their values at $x = 0$. Then, one gets $\int \psi_1(|x_2|)dx_2 = 2^{-\alpha} \int \psi_1(2^{1/p}|x_2|)dx_2 = 2^{-(\alpha+1/p)} \int \psi_1(|x_2|)dx_2$. Hence, we must have $\alpha = -1/p$, which implies $p = 2$. □

Proof of Theorem 2 : Note that the 'if' part is trivial in view of our discussion before the statement of the theorem. So, we shall now prove the 'only if' part.

First, we shall prove it for the bivariate case, i.e., d=2. Without loss of generality, let us assume $\boldsymbol{\mu} = \mathbf{0}$. Let Z be the angle between the positive side of the x_1-axis and the random vector $\mathbf{X}$ (measured anti-clockwise from the x_1-axis). Now, consider a straight line AB, which passes through the origin and makes an angle θ with the x_1-axis (see Figure 4). Since $\boldsymbol{\mu} = \mathbf{0}$, the two half-spaces generated by AB will have the same probability measure. Now, rotate the line AB by an angle δ to bring it in the position A_1B_1. Naturally, two half-spaces generated by A_1B_1 will also have the same probability measure 0.5. This implies that $P(\theta < Z < \theta + \delta) = P(\pi + \theta < Z < \pi + \theta + \delta)$. Since this equality holds for all θ and δ, it implies that Z and $Z + \pi$ have the same probability distribution. The result now follows from the fact that $(\mathbf{X} - \boldsymbol{\mu})/\|\mathbf{X} - \boldsymbol{\mu}\|_2 = (CosZ, \;\; SinZ)$ and $(\boldsymbol{\mu} - \mathbf{X})/\|\mathbf{X} - \boldsymbol{\mu}\|_2 = (Cos(Z + \pi), \;\; Sin(Z + \pi))$.

For $d > 2$, we need to consider $d-1$ random angles $Z_1, Z_2, \ldots, Z_{d-1}$. Note that here the direction vector $(\mathbf{X} - \boldsymbol{\mu})/\|\mathbf{X} - \boldsymbol{\mu}\|_2$ can be expressed as $(\mathbf{X} - \boldsymbol{\mu})/\|\mathbf{X} - \boldsymbol{\mu}\|_2 = (CosZ_1, \;\; SinZ_1\, CosZ_2, \;\; \ldots, \;\; SinZ_1 \ldots SinZ_{d-2}\, CosZ_{d-1}, \;\; SinZ_1 \ldots SinZ_{d-2}\, CosZ_{d-1})$. Now, consider a hyperplane H which makes angles $\theta_1, \theta_2, \ldots, \theta_{d-1}$ with the co-ordinate axes and then rotate it to H_1 such that the new angles are $\theta_1 + \delta, \theta_2, \ldots, \theta_{d-1}$. Now, the result follows from the same argument that is used in the bivariate case. □

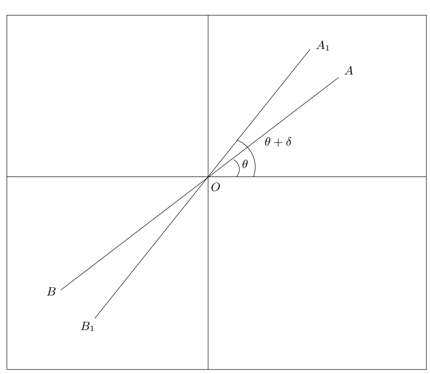

Figure 4: Two lines passing through the origin.

Lemma 5 : *For any two sequences* $\boldsymbol{\sigma}$ *and* $\mathbf{x}$ *in the* l_2 *space of real sequences, we have* $\sup_{\boldsymbol{\alpha}\in l_2}\left\{(\sum_{i=1}^{\infty}\alpha_i^2\sigma_i^2)^{-1/2}(\sum_{i=1}^{\infty}\alpha_i x_i)\right\} < \infty$ *if and only if* $\sum_{i=1}^{\infty} x_i^2/\sigma_i^2 < \infty$.

Proof: (The 'if' part). For any $\boldsymbol{\alpha} \in l_2$, $\sum_{i=1}^{\infty}\alpha_i x_i \leq (\sum_{i=1}^{\infty}\alpha_i^2\sigma_i^2)^{1/2}(\sum_{i=1}^{\infty}x_i^2/\sigma_i^2)^{1/2}$ (Cauchy Schwarz inequality) implies that $\sum_{i=1}^{\infty}\alpha_i x_i/(\sum_{i=1}^{\infty}\alpha_i^2\sigma_i^2)^{1/2} \leq \sum_{i=1}^{\infty}x_i^2/\sigma_i^2$. Now, the right side of the inequality does not depend on $\boldsymbol{\alpha}$. So, $\sum_{i=1}^{\infty}x_i^2/\sigma_i^2 < \infty$ implies the finiteness of $\sup_{\boldsymbol{\alpha}\in l_2}\left\{\sum_{i=1}^{\infty}\alpha_i x_i/(\sum_{i=1}^{\infty}\alpha_i^2\sigma_i^2)^{1/2}\right\} \leq \sum_{i=1}^{\infty}x_i^2/\sigma_i^2$.

(The 'only if' part). Next, consider the case where $\sum_{i=1}^{\infty}x_i^2/\sigma_i^2 = \infty$. Choose a sequence $\{\boldsymbol{\alpha}_n\}$ of real sequences, where $\boldsymbol{\alpha}_n = (\alpha_{n1}, \alpha_{n2}, \ldots)$ has non-zero values only at first n co-ordinates (i.e., $\alpha_{ni} = 0 \ \forall\ i > n$) and $\alpha_{ni} = x_i/\sigma_i^2$ for $i = 1, 2, \ldots, n$. Clearly, $\boldsymbol{\alpha}_n \in l_2 \ \forall\ n \geq 1$, and for each n, it is easy to check that $\sum_{i=1}^{n}\alpha_{ni}\, x_i/(\sum_{i=1}^{n}\alpha_{ni}^2\sigma_i^2)^{1/2} = (\sum_{i=1}^{n}x_i^2/\sigma_i^2)^{1/2}$. So, we get $\sup_{n\geq 1}\left\{\sum_{i=1}^{n}\alpha_{ni}\, x_i/(\sum_{i=1}^{n}\alpha_{ni}^2\sigma_i^2)^{1/2}\right\} = \infty$. This implies that $\sup_{\boldsymbol{\alpha}\in l_2}\left\{\sum_{i=1}^{\infty}\alpha_i x_i/(\sum_{i=1}^{\infty}\alpha_i^2\sigma_i^2)^{1/2}\right\}$ $= \infty$. □

Proof of Theorem 3 : Consider any $\mathbf{x}$ in the l_2 space with $\mathbf{x} \neq \mathbf{0}$. For any $\boldsymbol{\alpha}$ in the l_2 space, the random variable $Z =< \boldsymbol{\alpha}, \mathbf{X} >$ has a probability distribution with $E(Z) = 0$ and $V(Z) = \sum_{i=1}^{\infty}\alpha_i^2\sigma_i^2$. Using Chebychev's inequality, one gets $P(< \boldsymbol{\alpha}, (\mathbf{X}-\mathbf{x}) > \ \geq 0) = P(Z \ \geq \ < \boldsymbol{\alpha}, \mathbf{x} >) \leq \sum_{i=1}^{\infty}\alpha_i^2\sigma_i^2/(\sum_{i=1}^{\infty}\alpha_i x_i)^2$. So, the depth of $\mathbf{x}$ is bounded above by $\inf_{\boldsymbol{\alpha}\in l_2}\left\{\sum_{i=1}^{\infty}\alpha_i^2\sigma_i^2/(\sum_{i=1}^{\infty}\alpha_i x_i)^2\right\}$. From Lemma 5, it follows that this upper bound is zero when $\sum_{i=1}^{\infty}x_i^2/\sigma_i^2 = \infty$. Therefore, $\mathbf{x}$ will have positive depth only if $\sum_{i=1}^{\infty}x_i^2/\sigma_i^2 < \infty$.

Next, consider $Y_i = X_i^2/\sigma_i^2$ for $i \geq 1$. Then, the Y_i's are independent random variables with a common mean 1 and $\sum_{i=1}^{\infty}E(Y_i^2)/i^2 < \infty$. So, using the strong law of large numbers (see Theorem 1 in Chow and Teicher, 2005; pp. 124), we have $n^{-1}\sum_{i=1}^{n}Y_i \xrightarrow{a.s.} 1$ as $n \to \infty$. Consequently, $\sum_{i=1}^{\infty}Y_i = \sum_{i=1}^{\infty}X_i/\sigma_i^2 = \infty$ with probability one. □

References

[1] Chaudhuri, P. and Sengupta, D. (1993) Sign tests in multi-dimension : inference based on the geometry of the data cloud. *J. Amer. Statist. Assoc.*, **88**, 1363-1370.

[2] Chow, Y. S. and Teicher, H. (2005) *Probability Theory.* Springer Verlag, New York.

[3] Donoho, D. and Gasko, M. (1992) Breakdown properties of location estimates based half-space depth and projected outlyingness. *Ann. Statist.*, **20**, 1803-1827.

[4] Ghosh, A. K. and Chaudhuri, P. (2005a) On data depth and distribution free discriminant analysis using separating surfaces. *Bernoulli*, **11**, 1-27.

[5] Ghosh, A. K. and Chaudhuri, P. (2005b) On maximum depth classifiers. *Scand. J. Statist.*, **32**, 328-350.

[6] Hassairi, A. and Regaieg, O. (2008) On the Tukey depth of a continuous probability distribution. *Statist. Probab. Lett.*, **78**, 2308-2313.

[7] Koshevoy, G. A. (2002) The Tukey's depth characterizes the atomic measure. *J. Multivariate Anal.*, **83**, 360-364.

[8] Koshevoy, G. A. (2003) Lift-zonoid and multivariate depths. *Psysica, Heidelberg*, 194-202.

[9] Liu, R., Parelius, J. and Singh, K. (1999) Multivariate analysis of the data-depth : descriptive statistics and inference. *Ann. Statist.*, **27**, 783-858.

[10] Lopez-Pintado, S. and Romo, J. (2006) Depth based classification for functional data. *DIMACS Series in Mathematics and Theoretical Computer Science* (R. Liu and R. Serfling ed.), **72**, pp. 103-119.

[11] Mizera, I. and Muller, C. H. (2004) Location-scale depth. *J. Amer. Statist. Assoc.*, **99**, 949-966.

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

[13] Nolan, D. (1992) Asymptotics for multivariate trimming. *Stoc. Proc. Appl.*, **42**, 157-169.

[14] Tukey, J. (1975) Mathematics and the picturing of data. *Proc. 1975 Inter. Cong. Math., Vancouver*, pp. 523-531.

[15] Vardi, Y. and Zhang, C. H. (2000) The multivariate L_1-median and associated data depth. *Proceedings of the National Academy of Sciences, U.S.A*, **97**, 1423-1426.

[16] Zuo, Y. and Serfling, R. (2000) General notions of statistical depth function. *Ann. Statist.*, **28**, 461-482.