Integrated Rank-Weighted Depth

Kelly Ramsay, Stephane Durocher and Alexandre Leblanc

1. Introduction

Multivariate data is ubiquitous in modern statistics. To analyze and understand these datasets, it is essential to extend univariate tools to the multivariate setting. Among many other applications, data depth extends the concept of rank to the multivariate setting. Measures of data depth provide a center-outward ordering of points in any dimension. Many definitions of data depth have been proposed, based on a wide variety of concepts such as half-spaces [26], random simplices [11], zonoid regions [21], points represented as curves [19], and many more [1, 5, 23, 30]. Many of these depth measures have complex definitions that are not intuitive or can be difficult (or impossible) to compute precisely in high dimensions [23], often requiring optimization. On the other hand, defining depth measures through integration makes them easily approximable in high dimensions, via Monte Carlo techniques.

In 2009, Cuevas and Fraiman [3] introduced an integrated depth measure called *integrated dual depth*, or ID depth for short. ID depth can be used as a measure of depth for random elements in any Banach space with a separable dual. In this paper, we focus on random elements from distributions over $\mathbb{R}^d$. We integrate with respect to the Haar measure, following traditional projection pursuit methods [3]. Lastly, we normalize the depth to the interval $[0, 1]$. This allows ID depth to be interpreted as a point's average univariate simplicial depth over all directions.

Definition 1 (Cuevas and Fraiman, 2009, [3])**.**
Let $X \sim F$ be a distribution over $\mathbb{R}^d$, $F_u(y) = \Pr(X \cdot u \leq y \cdot u)$ and S^{d-1} be the $(d-1)$-dimensional unit hypersphere. Additionally, let $D_s(y \cdot u; F_u) = F_u(y \cdot u)(1 - F_u(y \cdot u))$, the univariate simplicial depth of y in a direction u. The integrated dual depth (ID depth) of a point $y \in \mathbb{R}^d$ with respect to F is

$$D_{ID}(y; F) = \frac{4}{V_d} \int_{S^{d-1}} D_s(y \cdot u; F_u)\, du. \tag{1}$$

Cuevas and Fraiman [3] suggest the possibility of generalizing (1) to the alternate integrated depth measure based on univariate Tukey depth

$$D_T(y \cdot u; F_u) = \min(F_u(y \cdot u), 1 - F_u(y \cdot u-)).$$

We call this depth *integrated-rank weighted depth*, because the univariate depth D_T can be expressed as normalized center-outward ranks.

Definition 2.
Let $V_d = \int_{S^{d-1}} 1\, du$. The *integrated rank-weighted depth* (IRW depth), $D_{IRW}(y; F)$, of point $y \in \mathbb{R}^d$ with respect to a distribution F, is

$$\begin{aligned} D_{IRW}(y; F) &= \frac{2}{V_d} \int_{S^{d-1}} D_T(y \cdot u; F_u)\, du \\ &= \frac{1}{V_d} \int_{S^{d-1}} 1 - F_u(y \cdot u) + F_u(y \cdot u-) - |1 - F_u(y \cdot u) - F_u(y \cdot u-)|\, du. \end{aligned}$$

Both IRW and ID depth admit sample versions, where F is simply replaced with F_n, the empirical distribution function of a sample. Cuevas and Fraiman [3] mention that when F is continuous, the contours of the two

depths admit the same level sets, including the deepest point or region. They do not compare the two definitions beyond this statement. In this paper we investigate the properties of IRW depth. The aim is to identify similarities and differences between the two depth measures to ascertain whether using the integrated method with different univariate depth measures adds anything new. For example, multivariate Tukey depth is decreasing along rays, which is an important property for a depth measure [30], however, multivariate simplicial depth does not have this property [30]. The motivation is that, by using univariate Tukey depth, IRW depth may also have this property. Furthermore, we aim to extend and further study the properties of these two depth measures.

We demonstrate that many of the properties of the ID depth measure are shared with IRW depth, including invariance under similarity transformations, maximality at centre, vanishing at infinity, continuity under continuous distributions, consistency and asymptotic normality. It should be noted that many of these properties, such as asymptotic normality, cannot be shown using the proof techniques of Cuevas and Fraiman [3], making this non-trivial. We derive the asymptotic distribution of the sample IRW depth of the deepest point under half-space symmetry. We show that it converges to the integral of a folded normal process. We provide a result on the breakdown point of both the IRW and ID medians. We show that IRW and ID depth are decreasing along rays under half-space symmetric distributions. We also show that both depths are continuous under discrete and mixed distributions, except at atoms of the distribution. We further provide a weighted average representation of the two depth measures and consequentially, an algorithm for exact computation.

We conclude that the depths are similar in terms of properties and computation. Furthermore, we find that when studying further properties of the ID medians, it is useful to look at both the ID and IRW definitions. In other words, the use of D_T can allow for easier proofs of some properties. This may be useful for studying the influence function of these depth measures and their associated location estimators. In fact, it may be useful to provide some general properties that apply to all integrated versions of a class of univariate depth measures. The paper proceeds as follows:

In Section 2 we show that sample IRW and ID depths can be represented as a weighted average of a finite set of univariate ranks. This naturally leads to an algorithm for exact computation of IRW depth in any dimension, and consequentially, any function of univariate projected cumulative distribution functions, including ID depth. We discuss the approximation algorithm for high dimensions from [3]. Next, along with ID depth, IRW depth is also compared with other depth measures and can be described as a smoothed version of Tukey depth. We show that like ID depth, IRW depth satisfies most of the properties for depth measures described by Zuo and Serfling [30]. Again, like ID depth, the sample depth produced by this depth measure is shown to be universally strongly consistent and uniformly weakly consistent using the results of Nagy et al. [22] and, under mild conditions, asymptotically normal.

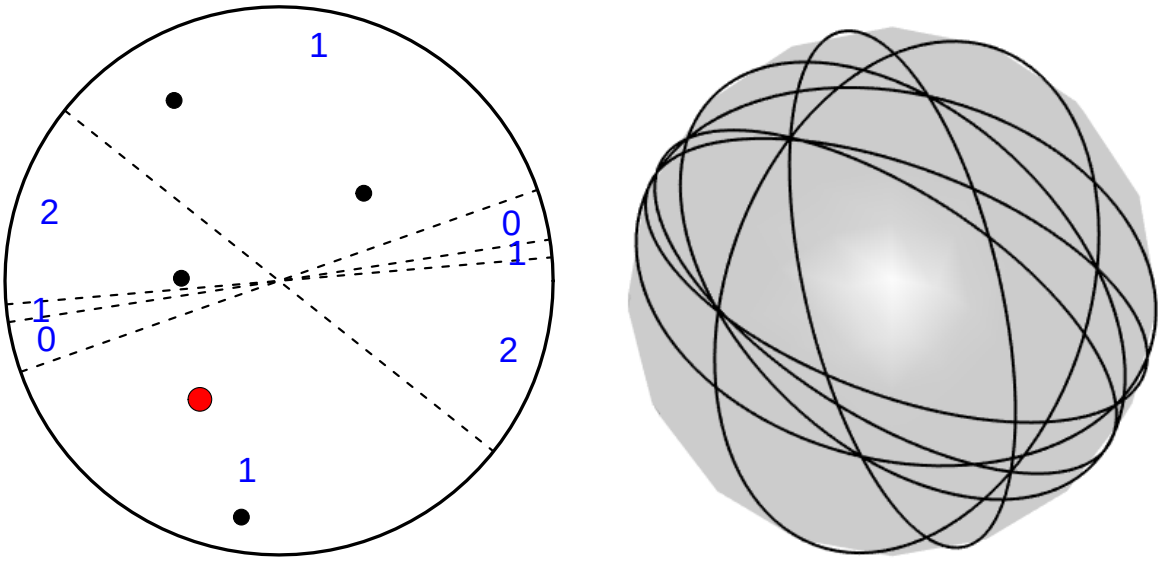
Section 3 discusses the deepest point, the maximizer of IRW depth, as a location estimator. Under continuous distributions, this maximizer is equivalent to the maximizer of ID depth [3]. We show this estimator is strongly consistent and relate it to the Tukey median. Using this relation, we obtain a lower bound on the depth of the deepest point for both the ID and IRW medians. We also bound the finite sample and asymptotic breakdown point of the IRW and ID medians below by the respective breakdown points of the Tukey median. We also mention a related in-sample location estimator that is easier to compute when d is very large.

In Section 4, we illustrate a potential use of IRW depth (and ID depth) in dd-plots, which are used to assess distribution differences between two samples in any dimension [10, 12]. We produce and analyze dd-plots for real and simulated data sets; the real data sets include the famous iris data [2, 8], as well as a very high dimensional ($d = 6033$) data set obtained from male patients with and without prostate cancer [27].

Section 5 gives a brief conclusion and some ideas for future study.

2. Integrated Rank-Weighted Depth

We start by showing that in discrete set-ups, we can represent IRW depth as a weighted average of a finite set of univariate ranks. First, note that two points have the same projection if and only if they are



(a) (b)

Figure 1: (a) Division of plane into sections where the red point's univariate depth remains the same. Its univariate rank is labeled in each section (b) Division of the unit sphere into sections where the permutations of points remain the same.

projected onto a unit vector perpendicular to the line passing through the two points. For example, consider a point y and some data set $\mathbf{X}_n$ both from $\mathbb{R}^2$. Then, y and a point $X_i \in \mathbf{X}_n$ have the same projection if and only if they are projected onto a unit vector parallel to their bisector. Let us refer to these unit vectors by their polar angle, θ_i. If there are $n^* \leq n$ unique points in $\mathbf{X}_n$, then an arbitrary point, $y \in \mathbb{R}^2$ (the query point), has n^* different bisectors; one for each point in the set. For each bisector, there are two unit vectors that are parallel to it, and so the plane is thus divided into $2n^*$ sections where a point's univariate depth remains the same. However, by symmetry of projections onto u and $-u$, we only need to consider n^* sections when determining the depth of y. This implies that y's univariate depth is constant over unit vectors in each of these sections. Figure 1a shows a partition of the plane determined by the red point's projections relative to the point set. We can rewrite IRW depth (in $\mathbb{R}^2$) as follows.

$$D_{IRW}(y; F_n) = 2 \sum_{i=1}^{n^*} w_i D_T(F_{u_i}, y \cdot u_i),$$

with

$$w_i = \frac{(\theta_{(i)} - \theta_{(i-1)})}{\pi},$$

and where $u_i \in (\theta_{(i)}, \theta_{(i-1)})$, and $\theta_{(i)}$ is the i^{th} ordered angle with $\theta_{(0)} = \theta_{(n^*)} - \pi$.

Computing the depth of a point y, with respect to some point set $\mathbf{X}_n$, is done by calculating the n^* bisectors between y and each unique point of $\mathbf{X}_n$. We can then compute the n^* unit vectors through each of these bisectors. The vectors are then converted to spherical coordinates and we calculate and sort the n^* angles θ_i that make up the subdivision of the plane. We then simply calculate the univariate depths in each section, along with normalizing the differences in sorted angles to use as weights. The time is dominated by the sorting process, thus taking $O(n \log n)$ time.

This weighted average representation and and the associated algorithm extend directly to $\mathbb{R}^d$. The expression for w_i is simply modified to represent a region on the unit hypersphere, rather than an interval of angles. Furthermore, the number of regions is $\binom{n^*}{d-1}$ if there are n^* unique points in the data set. Two

points have the same projection when the unit vector is contained in the hyperplane (containing the origin) to which the line through those two points is normal. Thus, for a query point y, we have n^* hyperplanes that divide the hypersphere into sections; see Figure 1b. We can calculate the vertices that make up this division of the unit hypersphere in $O(n^{d-1})$ time and it remains to order them as is done in $\mathbb{R}^2$, resulting in a running time of $O((d-1)n^{d-1}\log n)$. In fact, any function of the projected univariate CDFs including ID depth can be computed in this manner by simply exchanging D_T with another function of projected univariate CDFs such as ID depth.

A compelling feature of IRW depth is that computationally, in contrast to many other depth measures, it does not require solving an optimization problem. Rather, it requires computing an average. This provides a very straightforward Monte Carlo approximation algorithm for computing the depth of a point when d is very large. We can sample many, say m, unit vectors uniformly over S^{d-1}, calculate a point's univariate sample depth for each vector and take an average of the univariate depths. This estimate can be computed in $O(mnd)$ time. The law of large numbers guarantees this algorithm does indeed converge to the exact value of the depth being estimated as m tends to infinity. This is the same algorithm described by Durocher et al. [6] and by Cuevas and Fraiman [3].

2.1. Representation as a functional depth measure and resulting properties of IRW depth

A recent paper by Nagy et al. [22] derives various properties of a class of depth measures for functional data, that is, functionals that measure the centrality of a given function, say $f(t)\colon \mathbb{R}^d \to \mathbb{R}^k$, from a given set of functions. They restrict attention to continuous functions that are defined on a compact set $T \subset \mathbb{R}^d$. This class of depth measures involves integrating a k-dimensional depth, say D, over T:

$$\int_T D(f(t); P_t)\, d\lambda(t), \tag{2}$$

where P_t is a probability measure on $\mathbb{R}^k$ that depends on t and λ is a measure over $\mathbb{R}^d$. IRW depth can be represented in this manner: map the point of interest $y \in \mathbb{R}^d$ to a function which, with a slight abuse of notation, we will call $y(t) = y \cdot t$, $t \in S^{d-1}$. Now let λ be the Haar measure in $\mathbb{R}^d$, $T = S^{d-1}$, $D = D_T$ and $P_t = F_t$ in (2). We note it is also possible to represent this depth in terms of the Lebesgue measure, which is useful for technical reasons. Our functional observations in this set up come from a subset of the general class of continuous functions on T, $C(T)$ considered by Nagy et al. [22]. This subset is $\mathcal{A} = \{f(t) = a^\top t\colon a \in \mathbb{R}^d,\ t \in S^{d-1}\}$, which is the set of projection functions. (Nagy et al. actually consider observations that are vectors of functions, but we will omit that for simplicity.) The restriction to $\mathcal{A}$ makes some of the properties considered by Nagy et al. [22] not applicable in this scenario. We can say loosely that the results of Nagy et al. [22] are relevant when operations on the projection functions are equivalent to interesting operations on the points themselves. For example, scaling the projection functions is equivalent to scaling the points. We list the results from Nagy et al. [22] through these types of operational equivalencies. The implications of these properties are discussed in Section 2.2.

Theorem 1.
Let $\mathcal{P}(\mathbb{R}^d)$ be the set of probability measures on $\mathbb{R}^d$ and for a given probability measure Q, let F^Q be its distribution function. IRW depth satisfies the following properties

1. (Strong Universal Consistency) For each $\epsilon > 0$ and $Q \in \mathcal{P}(\mathbb{R}^d)$,

$$\sup_{x\in\mathbb{R}^d} \left| D_{IRW}(x; F_n^Q) - D_{IRW}(x; F^Q) \right| \xrightarrow{\text{a.s.}} 0 \text{ as } n \to \infty,$$

   where $\xrightarrow{\text{a.s.}}$ denotes almost sure convergence.

2. (Weak Uniform Consistency) For each $\epsilon > 0$

$$\sup_{Q\in\mathcal{P}(\mathbb{R}^d)} \Pr\left(\sup_{x\in\mathbb{R}^d} \left| D_{IRW}(x; F_n^Q) - D_{IRW}(x; F^Q) \right| > \epsilon \right) \to 0 \text{ as } n \to \infty.$$

3. (Weak Continuity as a Function of Q) Let $Q \in \mathcal{P}(\mathbb{R}^d)$ satisfying $Q(L) = 0$ for all hyperplanes $L \subset \mathbb{R}^d$. Then, for all $Q_\nu \to Q$ as $\nu \to \infty$

$$\sup_{x \in \mathbb{R}^d} |D_{IRW}(x; F^Q) - D_{IRW}(x; F^{Q_\nu})| \to 0 \text{ as } \nu \to \infty.$$

4. (Similarity Invariance) $D_{IRW}(y; F)$ is invariant under similarity transformations.

Properties 1, 2 and 3 follow from the bijection between $\mathcal{A}$ and $\mathbb{R}^d$. Property 4 follows from expressing IRW depth in terms of the Lebesgue measure in $\mathbb{R}^d$ the fact that a similarity transformation of the points corresponds to a measure-preserving rearrangement of the projection functions. Some properties related to measurability are satisfied via Theorem 3 and part (ii) of Theorem 12 of [22]. Other properties mentioned in Appendix A.1 of [22] either do not imply the corresponding properties in the primal setting of the point set in $\mathbb{R}^d$ or the restriction to $\mathcal{A}$ makes them not applicable.

2.2. Some Additional Properties

Like ID depth, integrated rank-weighted depth satisfies three of the four properties introduced by Liu [11] and later studied by Zuo and Serfling [30]: monotonicity with respect to centre, maximality at centre and vanishing at infinity. However, integrated rank-weighted depth is not affine invariant, it is only invariant with respect to similarity transformations, which include any combination of rotation, uniform scaling, translation and reflection. We establish these additional properties that do not follow from [22] and further investigate continuity and asymptotic properties of the depth measure. We also indicate if the result extends to ID depth. First, recall a distribution is half-space or *H-symmetric* about a point, γ when every closed half-space, H, containing γ, satisfies $\Pr(X \in H) \geq 1/2$ [31].

Theorem 2.
Integrated rank-weighted depth satisfies the following properties:

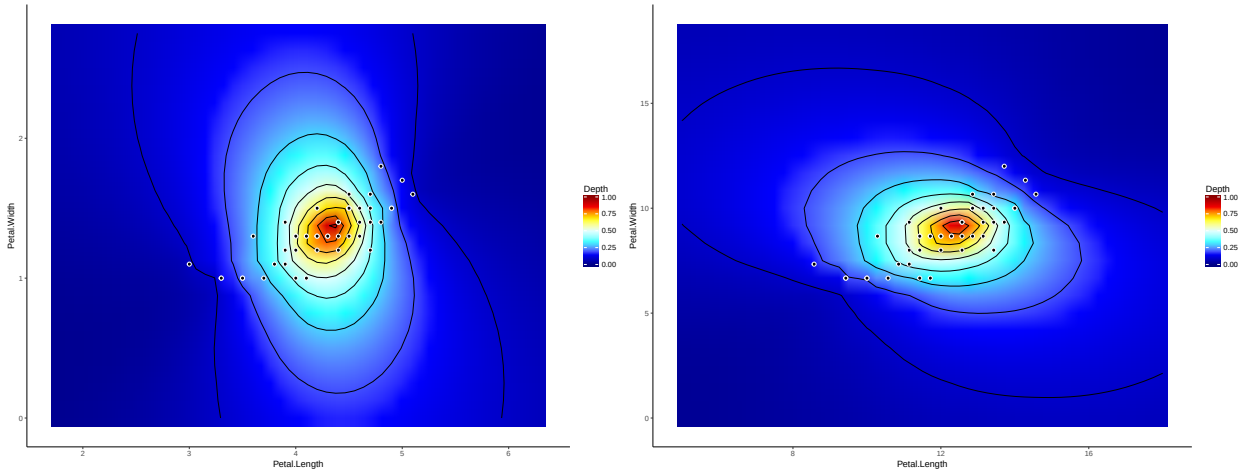
1. (Maximality at Centre) If F is H-symmetric about γ, then $D_{IRW}(y; F)$ is maximal at γ.
2. (Vanishing at Infinity) Let $c > 0$, then $\lim_{c \to \infty} D_{IRW}(cy; F) = 0$.
3. (Decreasing Along Rays) If F is H-symmetric about γ, then

$$D_{IRW}(y; F) \leq D_{IRW}(\alpha y + (1 - \alpha)\gamma; F) \leq D_{IRW}(\gamma; F) \qquad \text{for } \alpha \in [0, 1].$$

The proof of Theorem 2 can be found in Appendix A. Integrated rank-weighted depth is not invariant under all affine transformations, but it is invariant under similarity transformations. The fact that it is affected by non-uniform scaling causes axes with larger scales to have greater influence on the depth of a point than ones with smaller scales. To remove this feature one could scale the data in a robust way, such as scaling each coordinate by its median absolute deviation. This can be seen below as well as in Example 4. Alternatively, if there was a sufficient reason, one could use different scales to adjust the influence of certain variables. One could also apply a transformation to make the depths affine invariant such as described by Serfling [24], however, it is unclear what properties would be preserved and could be a subject of future study.

Example 1: Iris Data

We use the famous Iris data [2, 8]. This data set contains 150 observations, 50 from each of three species of flowers, with four different flower features. We will use the Petal Length and Petal Width measurements from the different Iris species to demonstrate the use of this depth measure throughout this paper. Figure 2a shows the depth contours for the Virginica species raw data. Figure 2b shows the depth contours for the same data, but with each coordinate scaled by its median average deviation. You can see that, in the unscaled data, a point's depth in the X-direction plays a larger role in its overall depth when compared with the scaled data. Points originally not deep in the X-direction but deep in the Y-direction become



(a) (b)

Figure 2: Example 2.2, (a) Virginica unscaled depth contours. (b) Virginica scaled depth contours.

deeper after scaling and points originally not deep in the Y-direction but deep in the X-direction become shallower after scaling. Notice also that the deepest regions are less affected by the scaling. This is due to the fact that very deep points are deep in all directions. Placing an emphasis on certain directions is then not expected to deep regions very much.

Integrated rank-weighted depth has the most general manifestation of maximality at centre; H-symmetry is the most general description of symmetry considered when analyzing depth measures [30]. On the topic of symmetry, we can also note that if the underlying population is antipodal-symmetric about some γ, i.e., $X - \gamma \overset{\mathrm{d}}{=} \gamma - X$, so is D_{IRW} [11]. We also note that decreasing along rays ensures that D_{IRW} measures depth or centrality. As we move away from the deepest point, the depth decreases. This property extends to ID depth as well.

Vanishing at infinity shows that, as a point is pulled to infinity along a ray, its depth approaches 0. It is interesting to note that a point's univariate depth is based on its rank; it does not provide any information about the spread of the points or how far the lowest depth is from the remainder of the sample. However, the average rank over all unit vectors does contain this information, as demonstrated by the maximality at centre and decreasing along rays properties. We see next that integrated rank-weighted depth has very good continuity properties with respect to the query point, even for discrete distributions. The proof can be found in Appendix A.

Proposition 1.
For a given distribution F, integrated rank-weighted and integrated dual depth are continuous for points (in $\mathbb{R}^d$) that are not atoms of F.

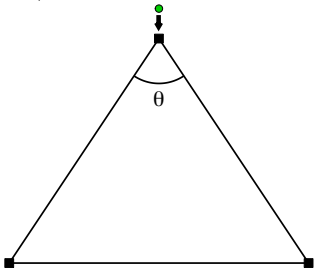
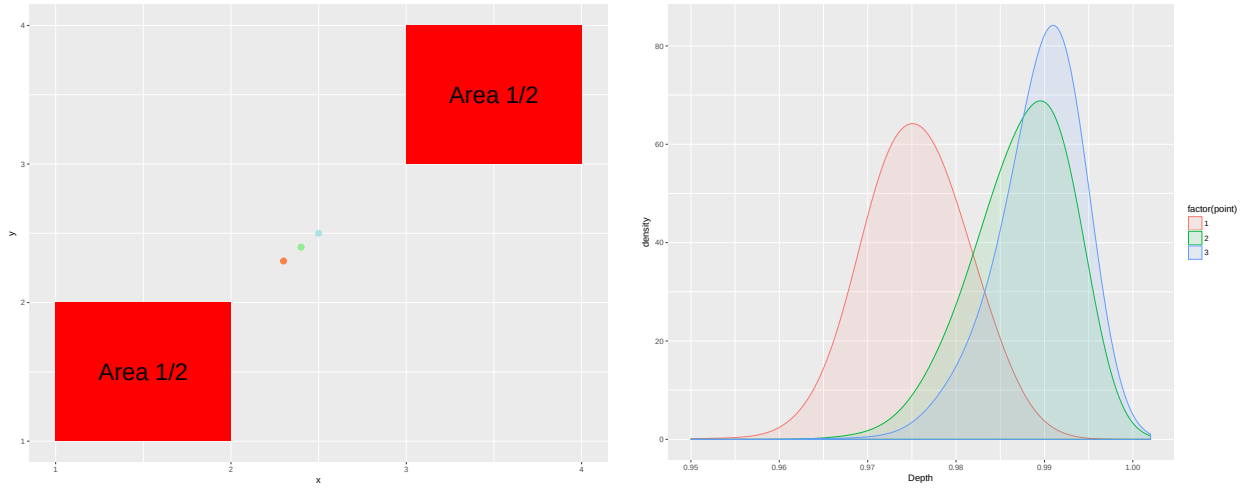


Figure 3: Example of discontinuity at the atom of F. See the text on the left.

Note that this theorem does not guarantee IRW and ID depths are discontinuous at atoms of F, just that it is a possibility. Figure 3 features three black points which we can take as a sample whose empirical distribution we will use to calculate depths. The univariate Tukey depth of any of the black points in any direction is at least 1/3. However, as the green (round) point approaches the black point along the y-axis, the proportion of angles for which it has non-zero univariate depth is approaching $\theta/\pi < 1$. Thus, the depth function is discontinuous at the



(a) (b)

Figure 4: (a) Birds eye view of density with points at which we calculate depth. (b) Asymptotic density estimates of the sample depths for the points in Figure (a).

black point. We now establish the asymptotic normality of the sample IRW depths. The proof can also be found in Appendix A.

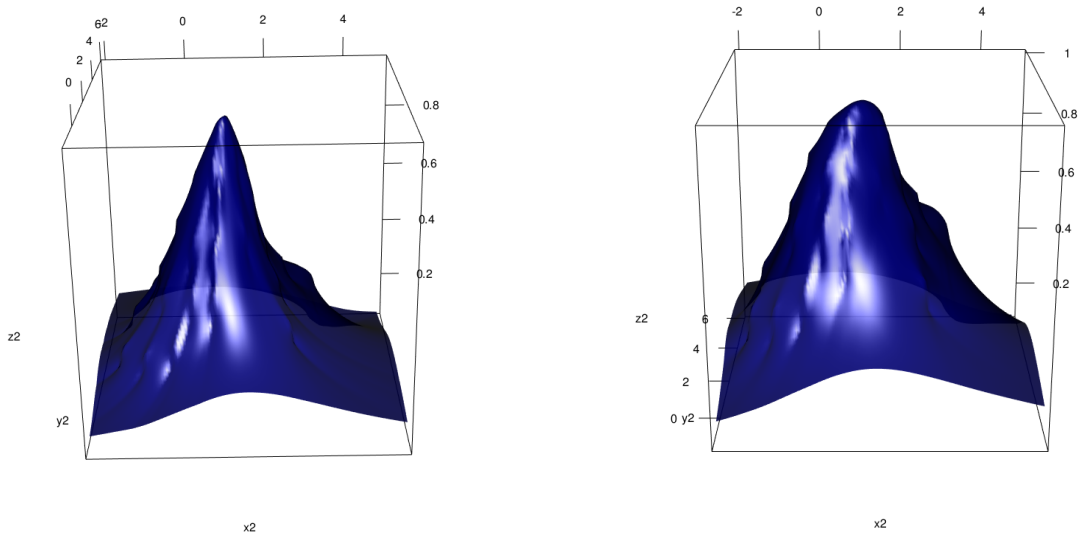
Theorem 3.
Let $A_1 = \{u \ : \ F_u(y \cdot u) < 1/2\}$, $A_2 = \{u \ : \ F_u(y \cdot u) > 1/2\}$ and $A_3 = \{u \ : \ F_u(y \cdot u) = 1/2\}$. Assume F is continuous and is such that A_3 has Haar measure 0. Then, $\sqrt{n} D_{IRW}(y; F_n)$ is asymptotically normal, that is, as $n \to \infty$:

$$\sqrt{n}(D_{IRW}(y; F_n) - D_{IRW}(y; F)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma^2(y)),$$

where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution and

$$\begin{aligned}\sigma^2(y) = \left(\frac{2}{V_d}\right)^2 \Bigg[& \int_{A_1}\int_{A_1} \Pr(X \cdot u_1 \le y \cdot u_1, X \cdot u_2 \le y \cdot u_2) - \Pr(X \cdot u_1 \le y \cdot u_1) \Pr(X \cdot u_2 \le y \cdot u_2)\ du_1 du_2 \\ & + 2 \int_{A_1}\int_{A_2} \Pr(X \cdot u_1 \le y \cdot u_1, X \cdot u_2 \ge y \cdot u_2) - \Pr(X \cdot u_1 \le y \cdot u_1) \Pr(X \cdot u_2 \ge y \cdot u_2)\ du_1 du_2 \\ & + \int_{A_2}\int_{A_2} \Pr(X \cdot u_1 \ge y \cdot u_1, X \cdot u_2 \ge y \cdot u_2) - \Pr(X \cdot u_1 \ge y \cdot u_1) \Pr(X \cdot u_2 \ge y \cdot u_2)\ du_1 du_2 \Bigg].\end{aligned}$$

In other words, D_{IRW} is asymptotically normal when there is no non-negligible proportion of unit vectors such that $F_u(y \cdot u) = 1/2$. This condition is different than the one required for asymptotically normal sample ID depths, which requires that F is continuous and y is not a point of symmetry [3]. The condition for IRW depth is more restrictive by additionally requiring that y does not lie in a 'hole' of the support of F. When A_3 is non-negligible, there is a non-degenerate region including unit vectors such that $F_u(y \cdot u) = 1/2$. Recall we estimate $F_u(y \cdot u)$ with $\min(F_{n,u}(y \cdot u), 1 - F_{n,u}(y \cdot u-))$. Over this region, $\min(F_{n,u}(y \cdot u), 1 - F_{n,u}(y \cdot u-)) < F_u(y \cdot u) = 1/2$ thus, we will always underestimate $F_u(y \cdot u)$ for $u \in A_3$. This fact creates left skewness in the asymptotic distribution that depends on the size of A_3. Figure 4a shows a density in $\mathbb{R}^2$ such that the mass is uniformly distributed between the two disjoint rectangles shown. Figure 4b shows asymptotic density estimates, using



(a) (b)

Figure 5: (a) 3D plot of the IRW depth measure with respect to a 2-dimensional point set. (b) Normalized (to (0,1]) ID depth measure with respect to the same 2-dimensional point set.

10 000 sample depths from samples of size of $n = 100$[1], for the depth of each of the coloured points in Figure 4a. Notice that the higher proportion of angles for which $F_u(y \cdot u) = 1/2$ leads to more skewness in the asymptotic distribution.

To construct confidence intervals from the sample depths, Theorem 3 could be used to justify assuming approximate normality, seeing as the skew effect appears to be minimal in almost all cases. The exception to this assumption is when y is a point of perfect symmetry. The blue point in Figure 4a is a point of symmetry, notice how skewed its asymptotic density is. If the sample depth of y is very close to 1 and one might suspect underlying symmetry about y, and a confidence interval could be computed by substituting F_n for F and simulating the asymptotic distribution via Monte Carlo techniques, rather than using the normality result.

It is not immediately clear why one would want to calculate a confidence interval on a depth value and so we provide the following example. Say we have two samples, which produce empirical distribution functions F_n and G_m, coming from populations F and G respectively. Now, consider the introduction of a new sample point, Z, for which we would like to determine whether Z comes from F or G. A common method for doing this is to calculate the depth of Z within in each sample and to assign Z to the sample in which it is deepest. Say we have that $D(Z; F_n) = 0.6$ and $D(Z; G_m) = 0.55$; in this case we would assign Z to F. However, say we also calculate the following confidence intervals for $D(Z; F_n)$ and $D(Z; G_m)$, respectively

$$(.45, .75) \qquad (.5, .6)$$

respectively. It is now less clear that we should assign Z to F, since the confidence interval for $D(Z; G_m)$ is contained in the one for $D(Z; F_n)$. Consider a situation where it is more serious to incorrectly classify a point as coming from F when it is coming from G. Here, we may want to use the following classification rule instead of the one mentioned above. If the confidence intervals overlap, assign Z to G. Otherwise, assign Z to the population in which its sample depth is the highest.

[1] The plots looked similar for a sample size of $n = 10000$.

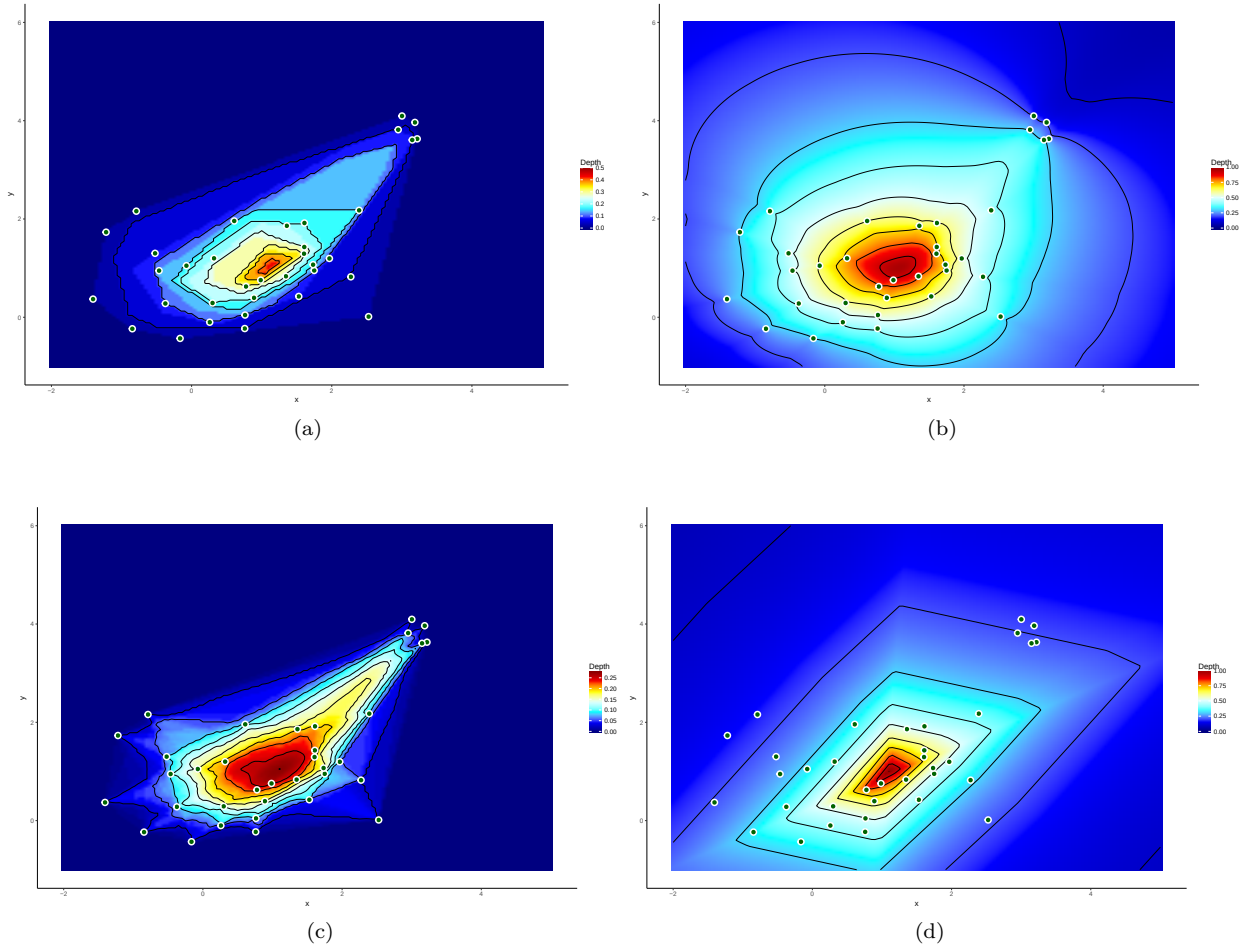


Figure 6: (a) Tukey depth contours. (b) IRW depth contours. (c) simplicial depth contours. (d) Projection depth contours using the median and median average deviation. (Same point set as Figure 5.)

Additionally, Theorem 3 provides a method for constructing hypothesis tests based on the depth values, rather than on ranks as is the convention, e.g., [14]. Suppose a quality control type scenario where we have a known distribution F and we want to test if a new sample with empirical distribution F_n comes from F. Let μ be the deepest point in F. One could use the test statistic

$$T(F_n) = \sqrt{\frac{n}{\sigma^2(\mu)}}(D_{IRW}(\mu; F_n) - D_{IRW}(\mu; F))$$

for this test, which will be approximately normal for large n. One would however, have to check that affine transformations of the data which are not similarity transformations do not overly affect the results of the test. This is briefly discussed in Section 2.2.

2.3. Comparison to other depth measures

We begin with a simple graphic comparing ID depth and IRW depth, Figure 5 shows a side-by-side comparison of the two measures for a 2-dimensional multivariate normal sample, with 5 outliers in a cluster at the top right. This data can be found in Appendix B. Note that the steeper interior of the IRW depth is quite clear. The use of D_S causes the depth measure to weight very central points higher, relative to

the deepest point, than would the use of D_T. This is due to the steeper slope associated with the Tukey measure. The steeper slope implies central regions or regions that admit depths greater than some constant c as well as associated estimators such as the average of points in such a region will differ.

Another nice property of integrated depth measures is that their contours appear much smoother than the contours of many popular depth measures. Figure 6 shows a comparison of Tukey, IRW, Liu's simplicial and projection depth [29] contours using the data from Figure 5. We do not include the contour plot of ID depth as the contours are similar in shape to those of IRW depth. Notice that IRW depth contours are very smooth compared to Tukey, projection and simplicial depth. Notice also that the deeper regions of IRW depth are not pulled towards the outlier cluster nearly as much as the deeper regions for the other depths; the outliers appear to affect the geometry of the inner contours of simplicial, projection and Tukey depth more than the inner contours of IRW depth. As a result, the deeper regions of the integrated depths appear to be less affected by outliers. It should be duly noted that Tukey, projection and Liu's simplicial depth measures are both affine invariant, which could be argued to give them an advantage over the integrated depths.

One should also note that, because they are not optimization problems, ID and IRW depths are simpler computationally than their competitors. The exact algorithm presented achieves computational time complexity better or close to its competitors, which have been much more heavily studied [7, 15–17]. The algorithms that achieve similar time complexity for competing depths are considerably more complex. In fact, it is very possible that a faster algorithm for IRW depth exists using the techniques of [15–17]. Further, when d is large, the Monte Carlo algorithm provides straightforward methods for constructing confidence intervals for the exact value of the depth value being approximated, see [6]. This is due to the fact that an integral is being approximated; this does not apply to random projection methods for depths based on optimization problems. It is difficult to assess the variability inherent in approximations based on the random projection method for depths based on optimization problems.

3. The IRW Deepest Point

It is natural to consider the deepest point as a measure of location; the deepest point refers to the point of highest centrality and is usually referred to as a median. Depth based medians are often very robust. For example, the Tukey median associated with Tukey depth (defined below and also known as half-space depth) has been shown to be robust; it has an asymptotic breakdown point of 1/3 under symmetry assumptions [4, 18, 26]. In other words it takes roughly 1/3 of the data to be corrupt before the Tukey median becomes arbitrarily bad.

Definition 3.
The *integrated rank-weighted median*, denoted $\mu(F)$, is defined as

$$\mu(F) = \underset{y\in\mathbb{R}^d}{\operatorname{argmax}}\, D_{IRW}(y;F). \tag{3}$$

If this value is not unique, then let $\mathcal{D} = \{y \in \mathbb{R}^d | D_{IRW}(y;F) = \sup_{x\in\mathbb{R}^d} D_{IRW}(x;F)\}$ and we define IRW median as

$$\mu(F) = \frac{\int_{\mathcal{D}} y\, dy}{\int_{\mathcal{D}} dy}.$$

Note that the IRW median is the same as the ID median in the case of continuous F. In this case, the sample IRW median $\mu(F_n)$ is an alternate estimator for the ID depth population median. It would be interesting to compare the efficiency and mean squared error of these estimators in different scenarios. The maximizer of IRW depth is not always unique. In such cases, we can define the IRW median as the average of the maximizing region. We suggest computing this median via convex optimization, or hill-climbing techniques, even though it is not in general convex and it has not been shown that IRW depth is decreasing along rays in all set ups. We conjecture that in fact it is always decreasing along rays. Note that when d is large, we can use the following estimator to save time in computing.

Definition 4.
The *in-sample integrated rank-weighted median* of a sample $\mathbf{X}_n$, denoted $\hat{\mu}(F_n)$, is defined as

$$\hat{\mu}(F_n) = \underset{y \in \mathbf{X}_n}{\operatorname{argmax}} D_{IRW}(y; F_n).$$

If this value is not unique, then let $\hat{\mathcal{D}} = \{y \in \mathbf{X}_n | D_{IRW}(y; F) = \sup_{x \in \mathbf{X}_n} D_{IRW}(x; F)\}$ and we define IRW median as

$$\hat{\mu}(F_n) = \frac{1}{|\hat{\mathcal{D}}|} \sum_{y \in \hat{\mathcal{D}}} y \ ,$$

where $|\hat{\mathcal{D}}|$ denotes the cardinality of $\hat{\mathcal{D}}$.

This can be paired with the Monte Carlo algorithm described in Section 2 to compute the approximate deepest point in $O(mn^2d)$ time. This algorithm can be used to produce an approximation that converges to the deepest point in the sample as $m \to \infty$.

3.1. Properties

Before discussing the properties of the estimator (3), it is useful to discuss the relationship between IRW depth and (multivariate) Tukey depth. From this point on, when we refer to Tukey depth we refer to the traditional Tukey depth normalized to the interval $[0, 1]$ and denote it by D_{TU}. It is defined as follows

$$D_{TU}(y; F) = 2 \inf_{u \in S^{d-1}} D_T(y; F).$$

The *Tukey median* of a distribution F, denoted $Tu(F)$, is the point which maximizes Tukey depth. If this point is not unique, it is the average of all such points that maximize Tukey Depth [26]. We now relate Tukey depth to IRW depth and ID depth.

Theorem 4.
For any query point $y \in \mathbb{R}^d$, integrated dual and integrated rank-weighted depths satisfy

$$D_{TU}(y; F) \leq D_{IRW}(y; F) \leq D_{ID}(y; F).$$

Corollary 1.
For any distribution F, the IRW median and consequentially, the ID median satisfy

$$\frac{2}{n} \left\lfloor \frac{n}{d+1} \right\rfloor \leq D_{IRW}(\mu(F); F) \leq D_{ID}(\mu(F); F) \leq 1.$$

Theorem 4 follows from the definitions. Corollary 1 ensures that the depth of the deepest point is at least $2/d + 1$. It follows directly from the fact [4] that

$$\sup_{y \in \mathbb{R}^d} D_{TU}(y; F) \geq \left\lfloor \frac{n}{d+1} \right\rfloor \frac{2}{n}.$$

A common measure of robustness for location estimators is the breakdown point. The breakdown point refers to the proportion of the data which must be arbitrarily corrupt in order for the estimator to become infinite. The *finite sample breakdown point*, $\epsilon^*(T, n)$ of an estimator, T, can be defined as

$$\epsilon^*(T, n) = \frac{1}{n} \max\{m \ : \ \sup_{\mathbf{Y}_m} |T(\mathbf{X}_{n-m} \cup \mathbf{Y}_m) - T(\mathbf{X}_{n-m})| = \infty\}.$$

The *asymptotic breakdown point* is simply $\lim_{n \to \infty} \epsilon^*(T, n)$. The next theorem bounds the breakdown point of the IRW median and as a consequence the ID median below. It shows both medians have a breakdown point at least as high as the Tukey median. The proof can again be found in Appendix A.

Theorem 5.
The finite sample breakdown of the IRW median satisfies

$$\epsilon^*(\mu(F_n), n) \geq \epsilon^*(Tu(F_n), n).$$

A direct consequence of Theorem 5 is the following, which is given without proof.

Corollary 2.
The finite sample breakdown of the IRW median satisfies

$$\epsilon^*(\mu(F_n), n) \geq \frac{\lceil \frac{n}{d} \rceil}{\lceil \frac{n}{d} \rceil + n}.$$

The asymptotic breakdown point of the IRW median satisfies

$$\lim_{n\to\infty} \epsilon^*(\mu(F_n), n) \geq \frac{1}{d+1}$$

for all F and

$$\lim_{n\to\infty} \epsilon^*(\mu(F), n) \geq \frac{1}{3}$$

when F is half-space symmetric.

A direct consequence of Corollary 2 is that these bounds hold for the breakdown point of the ID median. Next, we establish the consistency of the IRW medians.

Theorem 6.
Let F be a distribution on $\mathbb{R}^d$ and F_n be the resulting empirical distribution function from a sample from F. Then

$$\mu(F_n) \xrightarrow{\text{a.s.}} \mu(F) \text{ as } n \to \infty \qquad \text{and} \qquad \hat{\mu}(F_n) \xrightarrow{\text{a.s.}} \hat{\mu}(F) \text{ as } n \to \infty.$$

Theorem 1 implies both IRW medians are strongly consistent for their population counterparts. The difference between $\mu(F)$ and $\hat{\mu}(F)$ is subtle. $\hat{\mu}(F)$ is always in the support of F, whereas it is possible for $\mu(F)$ to be outside of the support of F. This can occur in scenarios where the support is a proper subset of $\mathbb{R}^d$.

Example 2: Iris Data cont.
Here we demonstrate the robustness of the IRW median. Figure 7 shows a scatter plot of the Iris data with 20% of the data in each group corrupted. The corrupted data was randomly mislabelled to one of the other two groups. The sample mean vector and IRW median before and after corruption are shown. Even with this high proportion of corrupted data, the IRW median remains close to the uncorrupted median. The group in the bottom left has mislabelled data very far from its correctly labelled ones, however, the IRW median barely moves. The green sample's median estimator is pulled half as far as the sample mean of the green sample is pulled, but both are pulled farther than in the bottom left group. The close proximity of some of the green points labelled as red points make them more difficult to interpret as outliers. It may be surprising that 'closer' mislabelled points have a greater impact on the estimator than 'far' ones. One could think that it is easier to identify 'far' observations as mislabelled rather than 'closer', more ambiguous, ones.

Note that we can write

$$\mu(F) \approx \operatorname*{argmin}_{y\in\mathbb{R}^d} \int_{S^{d-1}} \left| \frac{1}{2} - F_u(y \cdot u) \right| du.$$

This is equal in the case of continuous F. In effect, this estimator's projections have the minimum average distance $\mathbf{d}$ to the univariate median, where $\mathbf{d}(x, y) = |x - y|$. One could replace $\mathbf{d}$ with a different distance measure, such as $\mathbf{d}^*(x, y) = |x - y|^a$, $a > 0$. In fact, setting a to 2 gives the ID median.

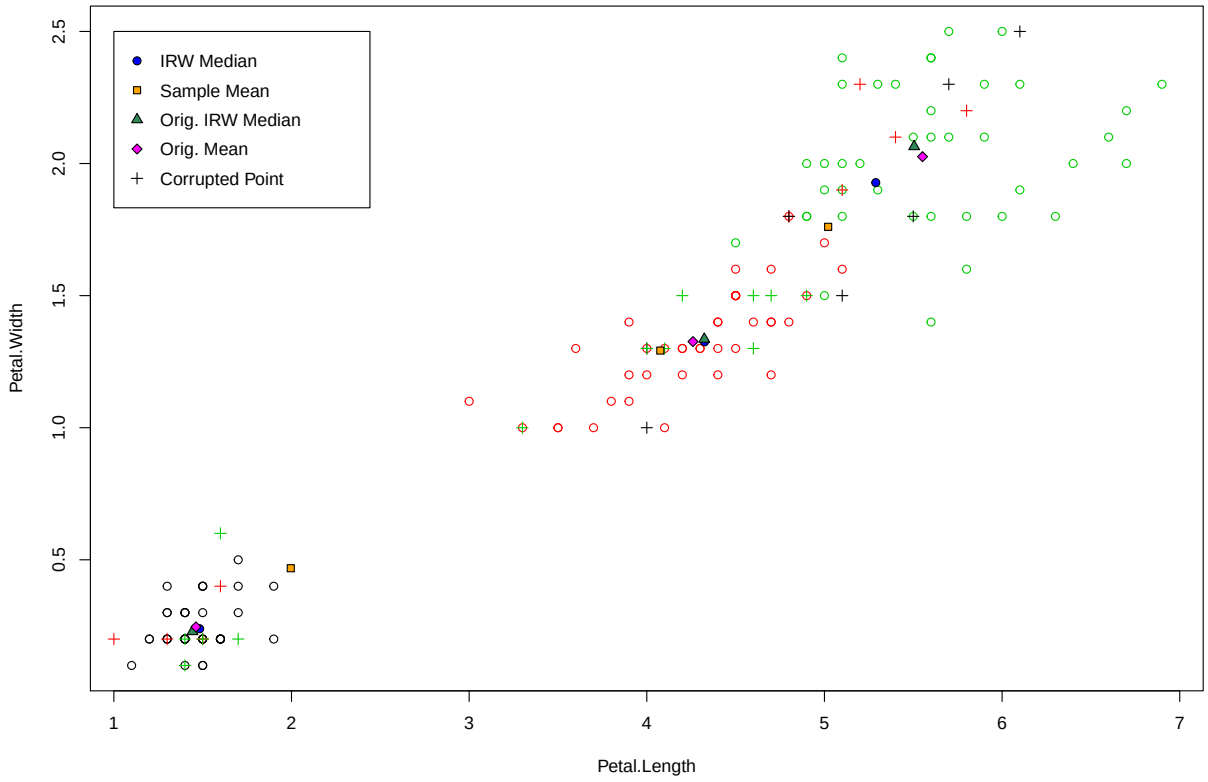


Figure 7: Example 3.1, Corrupted Iris Data and the IRW median.

4. An Application of IRW Depth

We now demonstrate and discuss a basic use of IRW depth, aside from the location estimation described above. We continue with the Iris data along with a second example featuring a very high dimensional dataset. All of the coding has been done with the R software.

Specifically, we demonstrate the use of the dd-plot, a dimension free plot used to assess how similar the parent distributions of two samples are. See the excellent description by Li and Liu in [10] or the original paper by Liu et al. [12]. To construct the plot we calculate the depth, with respect to each sample, of each point in the combined sample. Therefore for each point we have two depths which are plotted as pairs in $\mathbb{R}^2$. In other words, for each $x \in \mathbf{X}_{n_1} \cup \mathbf{X}_{n_2}$, we plot $(D_{IRW}(x; F_{n_1}), D_{IRW}(x; G_{n_2}))$, where F_{n_1} is the empirical distribution of $\mathbf{X}_{n_1}$ and G_{n_2} is the empirical distribution of $\mathbf{X}_{n_2}$. This is similar to a qq-plot except that we are plotting depths against each other rather than quantiles. These plots describe different distributional characteristics depending on how the data is transformed.

To assess location, the dd-plot should be made from the raw data or scale normalized data (using a robust measure of scale) if the scales differ widely. This is different from affine invariant depths, where the scale does not play a role in assessing location with dd-plots. Figure 8a-8c show dd-plots for samples from distributions that are the same, that differ only in location and differ only in scale respectively. In particular, in Figure 8a, the points follow a linear pattern and are somewhat evenly scattered about the $y = x$ line, which is typical when there is no distributional difference. Notice in Figure 8b that the location shift is associated with triangular shaped patterned dd-plots, with the tip of the triangle being somewhere around the bottom left or top right. To assess scale we must shift each group by its median. Note that

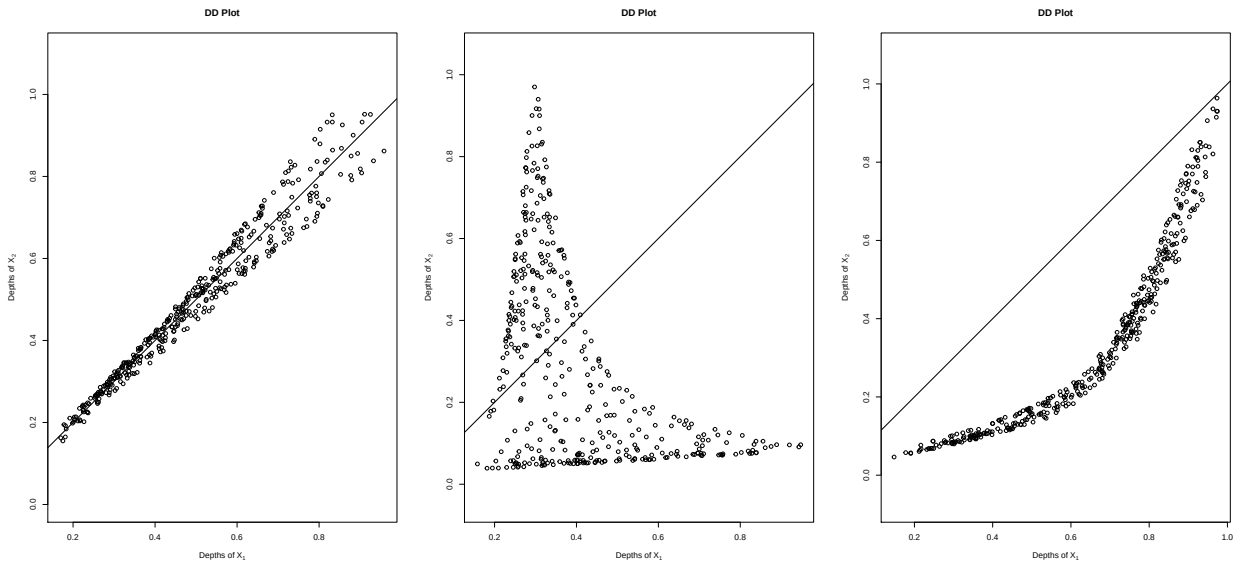
even though the depth measure is translation invariant, we will see a difference in the plots since the depth values are taken with respect to each sample, which are shifted by two different values. Scale differences are characterized by banana shaped dd-plots as in Figure 8c. Other characteristics such as kurtosis and skewness can be assessed; see [12] for more details. Li and Liu [10] also describe formal location and scale tests associated with dd-plots.

Example 3: Iris Data cont.
Figures 9a-9c shows dd-plots for the three pairs of classes with the uncontaminated, raw Iris data. The location differences between the three classes are apparent; all three exhibit that lower triangle appearance. Figure 10a-10c show the dd-plots for the same three pairs, but each class has been centred by the IRW median. Note the banana shape is most pronounced in the first and last plots, and is less distinct in the middle one. This is a reflection of a more pronounced scale difference, which is apparent in Figure 7.

Our experience is that dd-plots using IRW depth for very high dimensional, sparse data do not exhibit the same patterns as data with $n > d$. It seems that in-sample depth values from a sparse sample tend to be somewhat uniform. For example, as will be seen below, the in-sample depth values in the prostate data for the control group are in the range (0.5,0.54). This uniformity can change the way location differences appear in the plot. It is useful to do a small simulation study of dd-plots using the same n and d as the specific dataset at hand before using a dd-plot to compare samples.

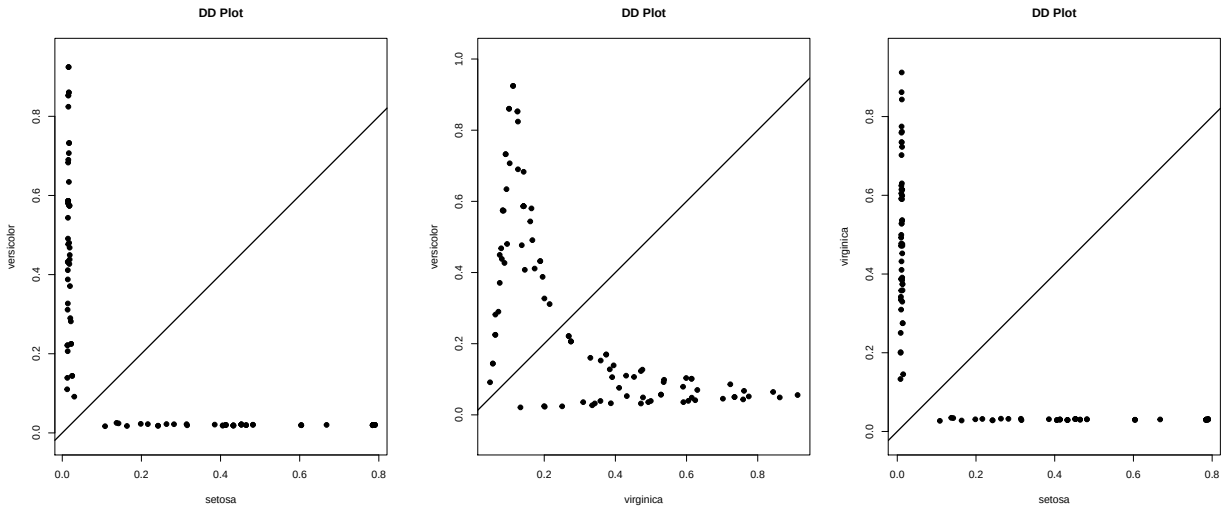
Example 4: Prostate Data
We now look at a data set which has very high dimensions compared to its sample size. We use the prostate cancer data set obtained by Welsh et al. [27], which consists of 102 observations of 6033 different gene expression values. There are 50 subjects in the control group and 52 with prostate cancer. To compute approximate depth values for the sample depth of each point, we use the Monte Carlo algorithm described in Section 2 with 100 000 uniformly generated random vectors.

Figure 11 shows the dd-plot of the raw prostate data. Notice the range of depth values in the samples is very small, which is a reflection of the large dimension and/or sparseness. In this case the lower bound on the deepest point is 1/3017, this lower bound reflects the curse of dimensionality. The points are,



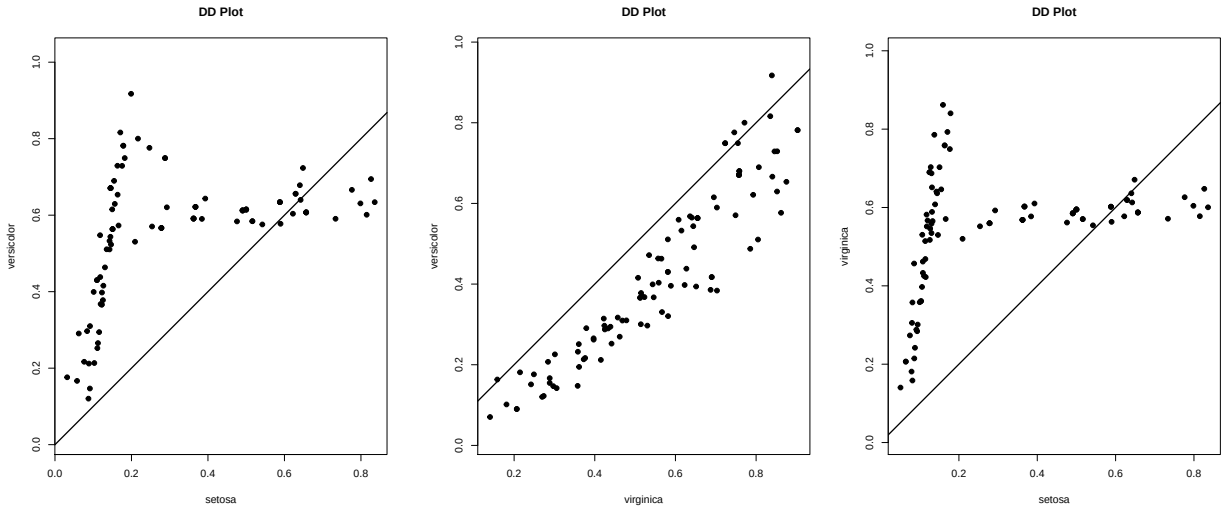
(a) (b) (c)

Figure 8: dd-plot for equal distributions, location shift and scale difference.



(a) (b) (c)

Figure 9: dd-plots for different pairs of iris species using the raw data



(a) (b) (c)

Figure 10: dd-plots for different pairs of iris species using location centred data.

however, relatively far away from the $y = x$ line, and thus we would suspect a location difference between these two groups. A dd-plot (not shown) was also made for the same data after the axes were brought to the same scale (after splitting into control and cancer). There was an even more pronounced location difference. The discrepancy between these two plots suggests that some of the features that displayed more variability in the raw data set may have less of a location difference than ones that displayed less variability.

Figure 12a and 12b show the dd-plots for the five and two most significant genes in terms of a t-test for location difference on each coordinate, respectively. Notice the much clearer triangular appearance.

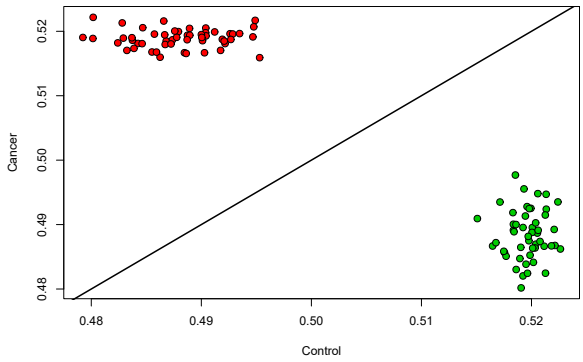
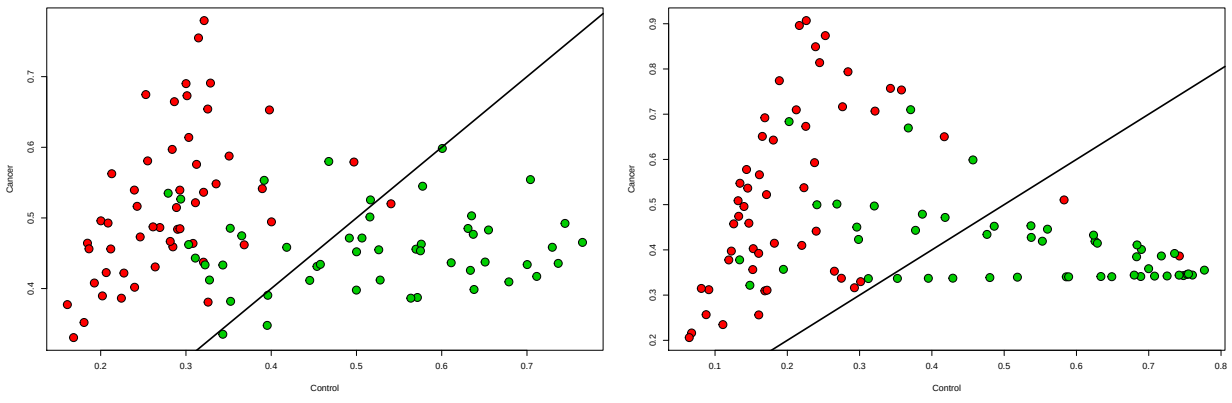


Figure 11: dd-plot of raw prostate data, red for the cancer group and green for the control, using the 6033 gene expression values.

After reducing the dimension, the plots seem to take more of the 'standard' appearance in the presence of a location shift. This leads us to consider taking many, say k, subsets of the data features or variables and producing plots for those subsets. Further, one could perform a modified version of the dd-plot test for distributional differences [10], by performing the test on each subset. The final test statistic could be the average of those k statistics, such as is done in [28]. Performing a test of this type may eliminate the sparsity issues while keeping some of the geometrical features of the data. Note simply testing each coordinate fails to account for dependencies between variables; it does not preserve geometrical features.

5. Discussion

We conclude that IRW and ID depth measures are, as suspected, very similar in terms of properties and computation. Furthermore, we find that when studying further properties of the ID medians, it is useful to look at both the ID and IRW definitions. In other words, the use of D_T can allow for easier proofs of some properties. This may be useful for studying the influence function of these depth measures and their associated location estimators. In fact, it may be useful to provide some general properties that apply to all integrated versions of a class of univariate depth measures. The integrated depths have advantages over traditional depths in terms of computation and asymptotic distributions. However, they suffer from not being affine invariant. The next approach should be to try the strong invariant coordinate system transformation procedure described by Serfling [24] to transform these integrated depths into affine invariant depths. However, one has to be careful to preserve the computational efficiency and the ability to use Monte Carlo methods or the point is moot.



(a) (b)

Figure 12: (a) dd-plot of raw prostate data using the 5 gene expression values with the largest location difference. (b) dd-plot of raw prostate data using the 2 gene expression values with the largest location difference.

If one is willing to forego invariance under non-uniform scaling, then it would be interesting to try the many depth-based statistical tools with higher dimensional data. One could, for instance, define a location estimator based on these depths, using the weighted functional technique, such as the projection weighted medians, see [29]. This would provide an estimator defined in terms of a double integral, which would allow for easy Monte Carlo approximation. Depth measures have many other uses, including but not limited to, classification, outlier trimming and scale assessment. Classification is done via simply assigning the new observation to the group in which it has the largest depth, for more information see [21]. Outlier trimming, for instance, can be done by removing observations of low depth. One could also compute a weighted sample mean, where the weights are based on the depth value. This is done with the projected medians and Stahel-Donoho estimator [29]. Scale can be assessed via a *scale curve*. A scale curve plots the volume of the convex hull of the deepest $\lfloor np \rfloor$ points, as p goes from 0 to 1. Scale curves however can be difficult to compute in high dimensions, one exception is the curve introduced by Lopez-Pintado and Romo [19]. For more ways in which depth measures can be used in analysis see [12, 13, 23].

We emphasized that integrated depths is computable in high dimensions, however the usual problems in high dimensional analysis still remain. For example, in their current form, scale curves are not easily computable in high dimensions as they involve convex hulls. On top of this, in $d > n$ scenarios the convex hull is a lower dimensional structure. Another method for measuring scale in high dimensions and sparse situations would be very useful. In fact it would be beneficial to better understand the application of depth-based analysis to sparse scenarios, such as the ones in Example 4. It would also be interesting to test the performance of the rank test introduced by Li and Liu [10] against other methods, especially ones developed for high dimensional settings.

Cuevas and Fraiman [3] introduced IDD for Banach spaces with separable dual. Another direction for future research would be to investigate how IRW depth generalizes to Banach spaces. This is a different problem in some sense because Banach spaces do not have the same geometry as $\mathbb{R}^d$. For example, the algorithm for computing IRW depth discussed in Section 2 relies on orthogonality concepts that do not hold in a general Banach space, but would hold in a Hilbert space. Further, the computing algorithm relies on the finite dimensional geometry of $\mathbb{R}^d$ in order to determine which 'patch' on the unit sphere to step to next. It is also important to verify that IRW depth still measures depth with respect to the objects being studied in such infinite dimensional spaces, or if there are better depth functionals for the problem at hand. In other words, the meaning of depth may change as one moves to studying objects in infinite dimensional space and therefore the desirable properties of the functional will change. Further, infinite dimensional Banach spaces are 'large' and in many cases it may be better to restrict attention to a non-linear subspace over which the objects are defined, and then design a suitable depth measure for this subspace, e.g., see [25].

Acknowledgement

We would like to thank Robert Serfling and our reviewers for taking time to provide some very insightful suggestions and commentary that greatly improved the quality and presentation of this paper.

Appendix A Proofs

The following notations will be used: F and G represent distributions over $\mathbb{R}^d$.

Proof. Theorem 2

Property 1*:*
Without loss of generality assume that $\gamma = 0$. Let H be a closed half-space and let h denote its boundary. Assume that h contains 0. Now, by assumption $H^c \cup h$ is also a closed half-space containing 0. By definition of half-space symmetry, we have both

$$\frac{1}{2} \leq \Pr(X \in H) \qquad \text{and} \qquad \frac{1}{2} \leq \Pr(X \in H^c \cup h).$$

Now, define u as a unit vector normal to h. Then, we have that,

$$F_u(0) = \Pr(X \cdot u \leq 0) = \min(\Pr(X \in H), \Pr(X \in H^c \cup h)) = \frac{1}{2}.$$

Since there is a bijection between S^{d-1} and half-spaces that contain 0, $F_u(0) = 1/2$ for all u. Therefore $D_{IRW}(0; F) = 1$ which is the maximum depth that can be achieved.

Property 2*:*
Let y be non-zero, also note u is always non-zero. From the elementary properties of univariate cumulative distribution functions, it is easy to see that

$$\lim_{c\to\infty} \min(F_u(c(y \cdot u)), 1 - F_u(c(y \cdot u)-)) = \min(1, 0) = 0.$$

There may be one or more u's such that $cy \cdot u = 0$, but the integrand is bounded above by 1 (and below by 0) and converges to 0 almost surely. Consequently, by the bounded convergence theorem,

$$\lim_{c\to\infty} \int_{S^{d-1}} \min(F_u(c(y \cdot u), 1 - F_u(c(y \cdot u)-)\ du = \int_{S^{d-1}} \lim_{c\to\infty} \min(F_u(c(y \cdot u), 1 - F_u(c(y \cdot u)-)\ du = 0,$$

since the range on integration is independent of c.

Property 3*:*
Without loss of generality again assume that $\gamma = 0.$, and y is on the positive x-axis; $y = (y_1, 0, \ldots, 0)$, $y_1 > 0$. Recall

$$u = (\cos\phi_1, \sin\phi_1 \cos\phi_2, \ldots, \sin\phi_{d-1}),$$

where $\phi_1, \ldots, \phi_{d-1}$ are the usual spherical coordinates. Note that $u \cdot y = y_1 \cos\phi_1$, $\phi_1 \in [0, \pi]$. Note also that $u \cdot y \geq 0$ on $\phi_1 \in [0, \pi/2]$ and $u \cdot y \leq 0$ on $\phi_1 \in [\pi/2, \pi]$. Since $F_u(0) = 1/2$, we know that $\min(F_u(y \cdot u), 1 - F_u(y \cdot u-)) = 1 - F_u(y \cdot u-)$ for $\phi_1 \in [0, \pi/2]$ and $\min(F_u(y \cdot u), 1 - F_u(y \cdot u-)) = F_u(y \cdot u)$ otherwise. Note that the same applies to $F_u(\alpha y \cdot u)$, $(\alpha \in [0, 1])$. Now for $\phi_1 \in [0, \pi/2]$, $1 - F_u(y \cdot u-) \leq 1 - F_u(\alpha y \cdot u-)$ since $y \cdot u \geq \alpha y \cdot u$. Similarly on $\phi_1 \in [\pi/2, \pi]$ $F_u(y \cdot u) \leq F_u(\alpha y \cdot u)$ since $y \cdot u \leq \alpha y \cdot u$ (recall that $y \cdot u < 0$). After projecting onto any unit vector, αy has univariate depth greater than or equal to that of y. Hence, it must have greater average depth. Therefore $D_{IRW}(y; F) < D_{IRW}(\alpha y; F)$ and the proof is complete.

□

Proof. Proposition 1

Case 1*:*
F is discrete. Let $supp(F)$ be the support of F and assume that it is finite. Let $z \in \mathbb{R}^d, \delta_1 \in \mathbb{R}$. Let y be a point not in the support of F. Choose δ_1 small enough such that the graph of vertices on the unit hypersphere where the univariate depth of y changes is the same for z. Further, assume that $0 < \|y - z\| < \delta_1$ and let N denote the number of sections into which the graph divides the unit hypersphere. We have

$$D_{IRW}(y; F) - D_{IRW}(z; F)| \leq 2 \sum_{i=1}^{N} |(w_i(y) - w_i(z))| D_T(y; F_{u_i}) \leq 2 \sum_{i=1}^{N} |(w_i(y) - w_i(z))|. \tag{4}$$

All that we need to show is that $\sum_{i=1}^{N} |w_i(y) - w_i(z)| < \epsilon/2N$.

Assume x is some vector in the support of F. We are now interested in the vertices of these sections on the unit sphere. The angles in spherical coordinates, denoted by $\phi_i, ..., \phi_{d-1}$, such that x and y have that same projection is given by $\phi_i = \prod_{j=d-1-i}^{d-1} \tan(\phi_j) \text{atan}(f_1(x, y))$ for $i > 2$ and $\phi_1 = \text{atan}(f_1(x, y))$, where f_1 is a continuous function in y. (We know this because f_1 is a product and composition of continuous trigonometric functions.) The functions atan() and tan() are also continuous, thus, for each i, there is a δ_{2i}

such that $\|y-z\| < \delta_{2i}$ implies $|w_i(y) - w_i(z)| < \epsilon/2N$. Choose $\delta = \min(\delta_1, \delta_{21}, \ldots, \delta_{2N})$. Then, if $\|y-z\| < \delta$, we have that

$$|D_{IRW}(y;F) - D_{IRW}(z;F)| \leq \ 2\sum_{i=1}^{N} |(w_i(y) - w_i(z))| D_T(y; F_{u_i})) < 2\sum_{i=1}^{N} |(w_i(y) - w_i(z))| < \epsilon.$$

Consider the case where y lies in a $(d-1)$-dimensional hyperplane (but still not in $supp(F)$) with d distinct points in the support of F, call this set of points B. In this case, if z lies outside this hyperplane determined by B, there does not exist a δ_1 small enough such that the graph of vertices on the unit hypersphere where the univariate depth of y is the same as z. Rather, if δ_1 is small enough, the univariate depth will be the same in all sections on the hypersphere, except for one. There will be one section, call it s, such that the univariate depth of z differs from the univariate depth of y by $\pm k$, $k < n/2$. This section is contained inside one of the sections in which the univariate depth of y does not change. Let N denote the number of sections into which y divides the unit hypersphere, and consequently z divides the hypersphere into $N+1$ sections. Let w_{N+1} be the size of the section s. In this case, we have

$$|D_{IRW}(y;F) - D_{IRW}(z;F)| \leq 2\sum_{i=1}^{N} |(w_i(y) - w_i(z))| + 2k \ w_{N+1}(z).$$

Note w_{N+1} is the size of the region enclosed by the d hyperplanes, each of which contains z and $d-1$ points in B. Thus, the size of this region is strictly decreasing as z approaches this plane; it is decreasing in δ_1. Thus, there is a δ_1 such that $w_{N+1}(z) < \epsilon/2N + 2k$ whenever $\|y-z\| < \delta_1$. As above, there are for each i, $\delta_{2i} > 0$ such that $|w_i(y) - w_i(z)| < \epsilon/2N + 2k$ whenever $\|y-z\| < \delta_{2i}$. Choose $\delta = \min(\delta_1, \delta_{21}, \ldots, \delta_{2N})$. Then, if $\|y-z\| < \delta$, we have that

$$|D_{IRW}(y;F) - D_{IRW}(z;F)| < 2\sum_{i=1}^{N} |(w_i(y) - w_i(z))| + 2k \ w_{N+1}(z) < (2N+2k)\frac{\epsilon}{2N+2k} = \epsilon,$$

still implying continuity. Clearly this extends to uncountable scenarios, where $\delta = \inf\{\delta_1, \delta_{21}, \ldots\}$.

Case 2:
F is absolutely continuous. Since F is continuous, $\min(F_u(y \cdot u), 1 - F_u(y \cdot u-))$ is continuous in y. Therefore $D_{IRW}(y;F)$ is continuous in y.

Case 3:
F is mixed. By the linearity property of integration and the previous two cases, the result is clear.

Clearly, this also extends to ID depth, as the weights are the same, and (4) still applies. The proof is complete. □

Proof. Theorem 5
Let $0 < \alpha, \beta < 1$. Set $m = \lfloor n\epsilon^*(T,n) \rfloor - 1$; the proportion of points for which the Tukey median is not corrupt. Without loss of generality assume that $n-m$ points lie inside the unit hypersphere and the other points may be located anywhere (we can simply scale the points down and recenter them at 0). Let F_n be the empirical distribution determined by $X \cup Y$. We know from the finite sample breakdown of the Tukey median [18], T, and the relationship between IRW depth and Tukey depth that:

$$D_{IRW}(y;F_n) = \alpha\left(2\left\lfloor\frac{m}{n}\right\rfloor\right) + (1-\alpha)2a > 2\left\lfloor\frac{m}{n}\right\rfloor,$$

where $a > 2\lfloor m/n \rfloor$. Now, it should be clear that in order for $d_u(y;F_n) > m$, the projection of y onto u must lie in the unit hypersphere. Consider an arbitrary point outside S^{d-1}, call it cu^*, where $c > 1$ and u^* is a unit vector. Note that cu^* has IRW depth that satisfies $D_{IRW}(cu^*;F_n) < \beta(2\lfloor m/n \rfloor) + (1-\beta)$, where $1-\beta$ is the proportion of unit vectors for which cu^*'s projection lies inside the unit hypersphere. $1-\beta$ is

a function of the space between the hyperplanes $u^* \cdot u = 1/c$ and $u^* \cdot u = 0$ which is approaching 0 as c approaches infinity. Clearly this is decreasing in c. We choose c large enough such that $1-\beta < (1-\alpha)a$, then $D_{IRW}(cu^*; F_n) < D_{IRW}(T, F_n)$. Therefore, the deepest point lies inside the hypersphere with center at the origin and radius c. From this, $\epsilon^*(\mu(F), n) \geq \epsilon^*(T, n)$ and $\lim_{n\to\infty} \epsilon^*(\mu(F), n) \geq \lim_{n\to\infty} \epsilon^*(T, n)$. □

Proof. Theorem 3
We use the central limit theorem. Let $A_1 = \{u \ : \ F_u(y \cdot u) < 1/2\}$, $A_2 = \{u \ : \ F_u(y \cdot u) > 1/2\}$, $A_3 = \{u \ : \ F_u(y \cdot u) = 1/2\}$, $A_{n,1} = \{u \ : \ F_{n,u}(y \cdot u) < 1/2\}$, $A_{n,2} = \{u \ : \ F_{n,u}(y \cdot u) > 1/2\}$. Since $\mathrm{Vol}(A_3) = 0$, we have

$$\begin{aligned} D_{IRW}(y; F_n) &= \frac{2}{V_d} \int_{S^{d-1}} \min(F_{n,u}(y \cdot u), 1 - F_{n,u}(y \cdot u-))\ du \\ &= \frac{2}{V_d}\Big(\int_{A_1} F_{n,u}(y \cdot u)\ du + \int_{A_2} 1 - F_{n,u}(y \cdot u-)\ du + B_{n1} + B_{n2} \Big), \end{aligned}$$

where

$$B_{n1} = \int_{A_1 \cap A_{n,2}} (1 - F_{n,u}(y \cdot u) - F_{n,u}(y \cdot u-))\ du,$$

and

$$B_{n2} = \int_{A_2 \cap A_{n,1}} (F_{n,u}(y \cdot u-) + F_{n,u}(y \cdot u) - 1)\ du.$$

Note that $|F_{n,u}(y \cdot u-) + F_{n,u}(y \cdot u) - 1| \leq 1$, which implies the following simple bounds:

$$\begin{aligned} |B_{n1}| &\leq \int_{A_1} \mathbb{1}\Big(F_{n,u}(y \cdot u) > \frac{1}{2}\Big)\ du, \\ |B_{n2}| &\leq \int_{A_2} \mathbb{1}\Big(F_{n,u}(y \cdot u) < \frac{1}{2}\Big)\ du. \end{aligned} \tag{5}$$

Now, we prove that $\sqrt{n}B_{n1}$ and $\sqrt{n}B_{n2}$ are $o_p(1)$ by showing that their expectation and variance converge to 0. Let $\delta_n = n^{-(\frac{1}{3}+\epsilon)}$, $0 < \epsilon < 1/6$, and $A_1^{\delta_n} = \{u \ : \ F_u(y \cdot u) \in (1/2 - \delta_n, 1/2)\}$. We can now write the following.

$$\begin{aligned} \mathrm{E}(|B_{n1}|) &\leq \mathrm{E}\Big(\int_{A_1} \mathbb{1}\Big(F_{n,u}(y \cdot u) > \frac{1}{2}\Big)\ du \Big), \\ &= \int_{A_1^{\delta_n}} \Pr\Big(F_{n,u}(y \cdot u) > \frac{1}{2}\Big)\ du + \int_{(A_1^{\delta_n})^c \cap A_1} \Pr\Big(F_{n,u}(y \cdot u) > \frac{1}{2}\Big)\ du. \end{aligned}$$

Note that for the second term, we can use the Bernoulli special case of Hoeffding's inequality (rf. [9]) with $t = \delta_n$. That is, if $H(n)$ is a binomially distributed random variable with n trials and success probability p, for $t > 0$

$$\Pr(H(n) \geq (p+t)n) \leq \exp\{-2t^2 n\}.$$

In our case, this inequality implies

$$\int_{(A_1^{\delta_n})^c \cap A_1} \Pr\Big(F_{n,u}(y \cdot u) > \frac{1}{2}\Big) du \leq C_1(F, y) \exp\{-2n^{\frac{1}{3}-2\epsilon}\},$$

where $C_1(F, y)$ is a finite, bounded constant that depends on F and y. Now, looking at the second term, let $\Delta_n = n^{-(\frac{1}{3}+\epsilon')}$, $0 < \epsilon' < \epsilon$. Choosing Δ_n in this way implies that, for $u \in A_1^{\delta_n}$,

$$\Delta_n < \left| \frac{\lfloor n/2 \rfloor + 1}{n} - F_u(y \cdot u) \right|.$$

Thus,

$$\int_{A_1^{\delta_n}} \Pr\left(F_{n,u}(y\cdot u) > \frac{1}{2}\right)\, du \le \int_{A_1^{\delta_n}} \sum_{|\frac{i}{n}-F_u|>\Delta_n} \binom{n}{i} (F_u(y\cdot u))^i (1-F_u(y\cdot u))^{n-i} du.$$

By [20], (see (8) on page 15), for any $k > 0$,

$$\int_{A_1^{\delta_n}} \sum_{|\frac{i}{n}-F_u|>\Delta_n} \binom{n}{i} (F_u(y\cdot u))^i (1-F_u(y\cdot u))^{n-i} du \le \int_{A_1^{\delta_n}} \frac{C_2(k)}{n^k} du \le V_d \frac{C_2(k)}{n^k}.$$

From the fact that C_2 is finite and does not depend on u, we can say the first term is $o\left(n^{-k}\right)$. Thus,

$$\mathrm{E}(|B_{n1}|) = o\left(n^{-k}\right). \tag{6}$$

Now note that from (5)

$$\left(\frac{|B_{n1}|}{\mathrm{Vol}(A_1)}\right)^2 < \frac{|B_{n1}|}{\mathrm{Vol}(A_1)} < 1,$$

which implies

$$\mathrm{E}(B_{n1}^2) < \mathrm{E}(|B_{n1}|)\mathrm{Vol}(A_1).$$

Thus,

$$\mathrm{E}(B_{n1}^2) = o\left(n^{-k}\right), \tag{7}$$

(6) and (7) together imply $\sqrt{n}B_{n1} \xrightarrow{\mathrm{p}} 0$. By a similar (but lengthier) argument, we have $\sqrt{n}B_{n2} \xrightarrow{\mathrm{p}} 0$. This leads to

$$\begin{aligned}
\sqrt{n}D_{IRW}(y;F_n) &= \frac{2\sqrt{n}}{V_d}\left[\int_{A_1} F_{n,u}(y\cdot u)\, du + \int_{A_2} 1 - F_{n,u}(y\cdot u-)\, du + o_p(n^{-\frac{1}{2}})\right] \\
&= \frac{2}{\sqrt{n}V_d}\left[\int_{A_1} \sum_{i=1}^{n} \mathbb{1}(X_i\cdot u \le y\cdot u)\, du + \int_{A_2} \sum_{i=1}^{n} \mathbb{1}(X_i\cdot u \ge y\cdot u)\, du\right] + o_p(1).
\end{aligned}$$

By switching the order of summation, we now get that

$$\begin{aligned}
\sqrt{n}D_{IRW}(y;F_n) &= \frac{2}{\sqrt{n}V_d}\sum_{i=1}^{n}\left[\int_{A_1} \mathbb{1}(X_i\cdot u \le y\cdot u)\, du + \int_{A_2} \mathbb{1}(X_i\cdot u \ge y\cdot u)\, du\right] + o_p(1), \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{2}{V_d}h(X_i) + o_p(1),
\end{aligned} \tag{8}$$

where $h(X_i)$ do not depend on n and are given by

$$h(X_i) = \frac{2}{V_d}\left[\int_{A_1} \mathbb{1}(X_i\cdot u \le y\cdot u)\, du + \int_{A_2} \mathbb{1}(X_i\cdot u \ge y\cdot u)\, du\right].$$

Now note,

$$\begin{aligned}
\mathrm{E}(h(X)) &= \frac{2}{V_d}\left[\int_{A_1} F_u(y\cdot u)\, du + \int_{A_2} 1 - F_u(y\cdot u-)\, du\right] = D_{IRW}(y;F_n)), \\
\mathrm{E}(h(X)^2) &= \left(\frac{2}{V_d}\right)^2\left[\int_{A_1}\int_{A_1} \Pr(X\cdot u_1 \le y\cdot u_1, X\cdot u_2 \le y\cdot u_2)\, du_1 du_2\right. \\
&\qquad + 2\int_{A_1}\int_{A_2} \Pr(X\cdot u_1 \le y\cdot u_1, X\cdot u_2 \ge y\cdot u_2)\, du_1 du_2 \\
&\qquad \left. + \int_{A_2}\int_{A_2} \Pr(X\cdot u_1 \ge y\cdot u_1, X\cdot u_2 \ge y\cdot u_2)\, du_1 du_2\right].
\end{aligned}$$

This implies,

$$\begin{aligned}\mathrm{Var}(h(X)) = \left(\frac{2}{V_d}\right)^2 \Bigg[& \int_{A_1}\int_{A_1} \Pr(X\cdot u_1 \le y\cdot u_1, X\cdot u_2 \le y\cdot u_2) - \Pr(X\cdot u_1 \le y\cdot u_1)\Pr(X\cdot u_2 \le y\cdot u_2)\; du_1 du_2 \\ & + 2\int_{A_1}\int_{A_2} \Pr(X\cdot u_1 \le y\cdot u_1, X\cdot u_2 \ge y\cdot u_2) - \Pr(X\cdot u_1 \le y\cdot u_1)\Pr(X\cdot u_2 \ge y\cdot u_2)\; du_1 du_2 \\ & + \int_{A_2}\int_{A_2} \Pr(X\cdot u_1 \ge y\cdot u_1, X\cdot u_2 \ge y\cdot u_2) - \Pr(X\cdot u_1 \ge y\cdot u_1)\Pr(X\cdot u_2 \ge y\cdot u_2)\; du_1 du_2 \Bigg].\end{aligned}$$

Notice (8) is written in the form $n^{-1/2}\sum_{i=1}^{n} h(X_i) + o_p(1)$, where $h(X_i)$ is independent of n (and has finite variance). By the central limit theorem and Slutsky's lemma,

$$\frac{\sqrt{n}(D_{IRW}(y;F_n) - \mathrm{E}(h(X)))}{\sqrt{\mathrm{Var}(h(X))}} \xrightarrow{\mathrm{d}} Z\ ,\ Z \sim \mathcal{N}(0,1).$$

This completes the proof. □

Appendix B Simulated Data

Data from Figure 6			
x-values	y-values	x-values	y-values
1.608	1.43	0.319	1.2
0.608	1.96	2.526	0.013
-0.464	0.948	1.346	0.834
0.763	0.045	1.604	1.293
1.75	0.952	1.529	0.424
-0.521	1.302	0.758	-0.229
1.358	1.861	2.391	2.176
-0.164	-0.43	0.985	0.756
-0.375	0.28	1.734	1.07
1.967	1.197	-0.074	1.05
-0.789	2.159	0.89	0.395
0.299	0.291	2.273	0.823
-1.22	1.731	-1.406	0.37
0.776	0.625	3.186	3.965
0.258	-0.101	3.22	3.631
1.61	1.918	2.948	3.816
-0.846	-0.234	3.148	3.608
3.002	4.096		

[1] Aloupis, G. (2006). Geometric Measures of Data Depth. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, volume 72. 1

[2] Anderson, E. (1936). The species problem in Iris. *Annals of the Missouri Botanical Garden*, 23(3):467–503. 2, 5

[3] Cuevas, A. and Fraiman, R. (2009). On depth measures and dual statistics. a methodology for dealing with general data. *Journal of Multivariate Analysis*, 100(4):753–766. 1, 2, 4, 7, 17

[4] Donoho, D. and Gasko, M. (1992). Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness. *Statistics*, 20(4):1803–1827. 10, 11

[5] Durocher, S., Fraser, R., Leblanc, A., and Skala, M. (2014). On Combinatorial Depth Measures. In *CCCG 2014 Conference Proceedings*. 1

[6] Durocher, S., Leblanc, A., and Skala, M. (2017). The Projection Median as a Weighted Average. *Journal of Computational Geometry*, 8(1):78–104. 4, 10

[7] Dyckerhoff, R. and Mozharovskyi, P. (2016). Exact computation of the halfspace depth. *Computational Statistics & Data Analysis*, 98:19–30. 10
[8] Fisher, R. A. (1936). The use of Multiple Measurements inTaxonomic Problems. *Annals of Eugenics*, 7(2):179–188. 2, 5
[9] Hoeffding, W. (1963). Probability Inequalities for Sums of Bounded Random Variables. *American Statistical Association*, 58(301):13–30. 20
[10] Li, J. and Liu, R. Y. (2004). New Nonparametric Tests of Multivariate Locations and Scales Using Data Depth. *Statistical Science*, 19(4):686–696. 2, 13, 14, 16, 17
[11] Liu, R. (1990). On a Notion of Data Depth Based on Random Simplices. *The Annals of Statistics*, 18(1):405–414. 1, 5, 6
[12] Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by liu and singh). *Ann. Statist.*, 27(3):783–858. 2, 13, 14, 17
[13] Liu, R. Y., Serfling, R. J., and Souvaine, D. L. (2008). *Data Depth: Robust Multivariate Analysis, Computational Geometry, and Applications*. DIMACS series in discrete mathematics and theoretical computer science. American Mathematical Soc. 17
[14] Liu, R. Y. and Singh, K. (1993). A Quality Index Based on Data Depth and Multivariate Rank Tests. *Journal of the American Statistical Association*, 88(421):252–260. 9
[15] Liu, X. (2017a). Approximating the projection depth median of dimensions p 3. *Communications in Statistics - Simulation and Computation*, 46(5):3756–3768. 10
[16] Liu, X. (2017b). Fast implementation of the tukey depth. *Computational Statistics*, 32(4):1395–1410.
[17] Liu, X. and Zuo, Y. (2014). Computing projection depth and its associated estimators. *Statistics and Computing*, 24(1):51–63. 10
[18] Liu, X. H., Zuo, Y., and Wang, Q. H. (2017). Finite sample breakdown point of Tukey's halfspace median. *Science China Mathematics*, 60(5):861–874. 10, 19
[19] Lopez-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734. 1, 17
[20] Lorentz, G. (1986). *Bernstein Polynomials*. AMS Chelsea Publishing Series. Chelsea Publishing Company. 21
[21] Mosler, K. and Hoberg, R. (2006). Data analysis and classification with zonoid depth. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, volume 72. 1, 17
[22] Nagy, S., Gijbels, I., Omelka, M., and Hlubinka, D. (2016). Integrated depth for functional data: statistical properties and consistency. *ESAIM: PS*, 20:95–130. 2, 4, 5
[23] Serfling, R. (2006). Depth Functions in Nonparametric Multivariate Inference. In Liu, Serfling, S., editor, *Data Depth: Robust Multivariate Analysis, Computational Geometry, and Applications*, pages 1–16. American Mathematical Society. 1, 17
[24] Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization. *Journal of Nonparametric Statistics*, 22(7):915–936. 5, 16
[25] Srivastava, A. and Klassen, E. (2016). *Functional and Shape Data Analysis*. Springer-Verlag New York. 17
[26] Tukey, J. W. (1974). Mathematics and the Picturing of Data*. In *Proceedings of the International Congress of Mathematicians*. 1, 10, 11
[27] Welsh, J. B., Sapinoso, L. M., Su, A. I., Kern, S. G., Wang-Rodriguez, J., Moskaluk, C. A., Frierson, H. F., and Hampton, G. M. (2001). Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. http://statweb.stanford.edu/ ckirby/brad/LSI/datasets-and-programs/datasets.html. 2, 14
[28] Zhang, J. and Pan, M. (2016). A high-dimension two-sample test for the mean using cluster subspaces. *Computational Statistics and Data Analysis*, 97:87–97. 16
[29] Zuo, Y. (2003). Projection-based depth functions and associated medians. *Annals of Statistics*, 31(5):1460–1490. 10, 17
[30] Zuo, Y. and Serfling, R. (2000a). General notions of statistical depth function. *Annals of Statistics*, 28(2):461–482. 1, 2, 5, 6
[31] Zuo, Y. and Serfling, R. (2000b). On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *Journal of Statistical Planning and Inference*, 84(1):55 – 79. 5