From Depth to Local Depth : A Focus on Centrality

Davy Paindaveine* and Germain Van Bever

Université libre de Bruxelles, Brussels, Belgium

Abstract

Aiming at analysing multimodal or non-convexly supported distributions through data depth, we introduce a local extension of depth. Our construction is obtained by conditioning the distribution to appropriate depth-based neighborhoods, and has the advantages, among others, to maintain affine-invariance and to apply to all depths in a generic way. Most importantly, unlike their competitors, that (for extreme localization) rather measure probability mass, the resulting *local depths* focus on centrality and remain of a genuine depth nature at any locality level. We derive their main properties, establish consistency of their sample versions, and study their behavior under extreme localization. We present two applications of the proposed local depth (for classification and for symmetry testing), and we extend our construction to the regression depth context. Throughout, we illustrate the results on some, artificial and real, univariate and multivariate data sets.

Keywords and phrases: Classification, multimodality, non-convex support, regression depth, statistical depth functions, symmetry testing.

*Davy Paindaveine is Professor of Statistics, Université libre de Bruxelles, ECARES and Département de Mathématique, Avenue F. D. Roosevelt, 50, CP 114/04, B-1050 Bruxelles, Belgium (E-mail: dpaindav@ulb.ac.be). He is also member of ECORE, the association between CORE and ECARES. Germain Van Bever is FNRS PhD candidate, Université libre de Bruxelles, ECARES and Département de Mathématique, Campus de la Plaine, Boulevard du Triomphe, CP 210, B-1050 Bruxelles, Belgium (E-mail: gvbever@ulb.ac.be).

1 INTRODUCTION

Data depth was originally introduced as a way to generalize the concept of median to the multivariate setup but is also known to be a powerful data analytic tool able to reveal diverse features of the underlying distribution. Indeed, not only does depth provide a robust multivariate location functional (through the deepest point), it also yields information about spread, shape, and symmetry (through depth regions, Serfling (2004)), and even characterizes the underlying distribution under very mild conditions (see Kong and Zuo, 2010 and the references therein). Celebrated instances of such depths include Tukey's halfspace depth (Tukey, 1975), Liu's simplicial depth (Liu, 1990), the projection depth (Zuo, 2003), or the Mahalanobis depth (see, e.g., Zuo and Serfling, 2000a). Depth methods allow to address several inference problems, including, e.g., testing for location and scale differences based on the DD-plot (first introduced as a graphical display for data exploration; Liu et al. (1999), Li and Liu (2004)), diagnostics of non-normality (Liu et al., 1999), and outlier detection (Chen et al., 2009), etc. More recently, depth was used extensively in a classification context (see, among many others, Ghosh and Chaudhuri, 2005; Li et al., 2012; Dutta and Ghosh, 2012; Paindaveine and Van Bever, 2012).

Depth deals with centrality. Its first purpose is to provide a *center-outward ordering* from the deepest point(s) towards less deep, exterior points. Classical depth functions indeed associate with any center of symmetry (should it exist) a maximal depth value. Together with the fact that depth decreases along any halfline originating from any deepest point, this leads to nested star-shaped (in most cases, convex) depth regions, whatever the underlying distribution may be (depth/quantile regions that may be non-convex are defined in Wei (2008)). That is the reason why it is often reported that depth is suitable for unimodal convexly-supported distributions only; see, e.g., Zuo and Serfling (2000a); Lok and Lee (2011); Izem et al. (2008); Hlubinka

et al. (2010). Distributions that are multimodal or have a non-convex support, however, are met in many fields of applications, among which, obviously, those involving mixture models or clustering problems (see, e.g., McLachlan and Basford (1988) or McLachlan and Peel (2000)). This motivates extending the concept of depth to make it flexible enough to deal with such distributions.

A few such extensions are available in the literature, under the name of *local depths*. In particular, Agostinelli and Romanazzi (2011) introduced local versions of the halfspace and simplicial depths. For halfspace depth, locality is achieved by replacing halfspaces with finite-width slabs, while, for simplicial depth, it is obtained by restricting to simplices with a volume smaller than some fixed threshold. These local depths—after adequate standardization—provide a continuum between (global) depths and the density of the underlying distribution. Density and depth, however, are antinomic in spirit : for instance, the symmetry center of a centrally symmetric bimodal distribution always assumes maximal depth while the density may very well be zero there; also, uniform distributions have non-trivial depth contours but do not show proper equidensity contours.

Similarly, other proposals for local depth—or, more generally, other extensions of depth aiming at distributions with possibly non-convex supports—converge, as locality becomes extreme, to either a density measure (Hlubinka et al. (2010)) or a constant value (Chen et al. (2009)), hence lose their nature of a centrality measure. The purpose of this paper is to introduce a new concept of local depth that, at any locality level, remains of a genuine depth nature and provides a measure of *local* centrality. Our construction will actually allow to turn, in a common generic way, *any* (global) depth into a corresponding local depth. This is another advantage over the competing local depths, that focus on a specific depth (Hlubinka et al. (2010); Chen et al. (2009)) or require a specific definition for each global depth considered

(Agostinelli and Romanazzi (2011)). The proposed local depth is defined as global depth conditional on some neighborhood of the point of interest. To make this local concept purely based on depth, we use the neighborhoods that were recently introduced (for classification purposes) in Paindaveine and Van Bever (2012). As we will show, the resulting local depths allow for interesting inferential applications.

The outline of the paper is as follows. In Section 2, we illustrate our local depth concept on two real data sets, that highlight the need for this extension from global to local centrality. In Section 3, we first review the basics of depth (Section 3.1). We then describe the depth-based neighborhoods from Paindaveine and Van Bever (2012) and show how they allow to define local depth (Section 3.2). We also establish consistency of the corresponding sample local depth (Section 3.3). Section 4 is dedicated to the limit behavior of the proposed local depth as locality becomes extreme. Section 5 illustrates the results of the previous sections on several univariate and multivariate examples. Section 6 presents two inferential applications of the proposed local depth concept. In Section 7, we show that our construction extends to regression depth. Computational aspects are discussed in Section 8. Finally, the Appendix collects technical proofs.

2 MOTIVATING EXAMPLES

As mentioned above, we introduce a concept of local depth that can cope with multimodal and/or non-convexly supported distributions. Here we illustrate this on the basis of two real data sets, that are freely available in the well-known R package *MASS* (the first one provides a univariate bimodal example, whereas the second one involves a bivariate distribution with a non-convex support).

2.1 Geyser data

The *Geyser* data set is related to eruption data from the *Old Faithful geyser* in the Yellowstone National Park, Wyoming, USA (see Härdle, 1991). It contains $n = 299$ measurements of two variables : "duration" (duration, in minutes, of the eruption) and "waiting" (waiting time, still in minutes, between two eruptions). As we want to start with a univariate data set, we focus here on the bimodal variable "waiting".

Besides a histogram of the waiting times, Figure 1 reports the halfspace and simplicial depths of 100 equispaced values in the range of interest, together with the proposed local halfspace and simplicial depths at locality levels $\beta = .7$ (intermediate localization) and .3 (more extreme localization) ; in the present univariate setup, we simply define the local depth of a waiting time x, at locality level β, as the (global) depth of x with respect to the $\lceil n\beta \rceil$ observed waiting times that are closest to x. For the sake of comparison, we also report the local halfspace and simplicial depths from Agostinelli and Romanazzi (2011), at locality levels $\tau = 23$ and 7 (for proper comparison, these τ-values, as in Agostinelli and Romanazzi (2011), were selected as the .7- and .3-quantiles of the $\binom{n}{2}$ distances between observed waiting times ; in order to avoid these local depth functions collapsing to zero as $\tau \to 0$, they were scaled so that, for each depth and τ, the deepest waiting time receives depth 1/2).

With the exception of the Agostinelli and Romanazzi (2011) ($\tau = 23$)-local halfspace depth, all local depths clearly show the obvious multimodality that is missed by global depth. For more extreme localization, all local depths reveal both local modes about 55 and 80. Unlike ours, that attribute comparable depth values to both local modes, the Agostinelli and Romanazzi (2011) local depths, that, at such locality levels, are not local centrality measures but rather density measures, clearly reflect the heterogeneous probability masses around the two local modes.

For $\beta = .3$, the proposed local depths show a third local center (about $x_0 = 65$

minutes), which is in line with the fact that, at this locality level, the distribution is nearly symmetric about x_0, so that it should receive a large (local) centrality measure. If needed, discriminating between the two "true" local modes and this "artificial" mode about x_0 may e.g. be based on the corresponding depth-based neighborhoods involved (See Section 3.2 below), that are much wider at x_0 than at both "true" modes. Detecting modes, however, is not one of the primary applications of the proposed local depth concept ; see Section 6 for such applications.

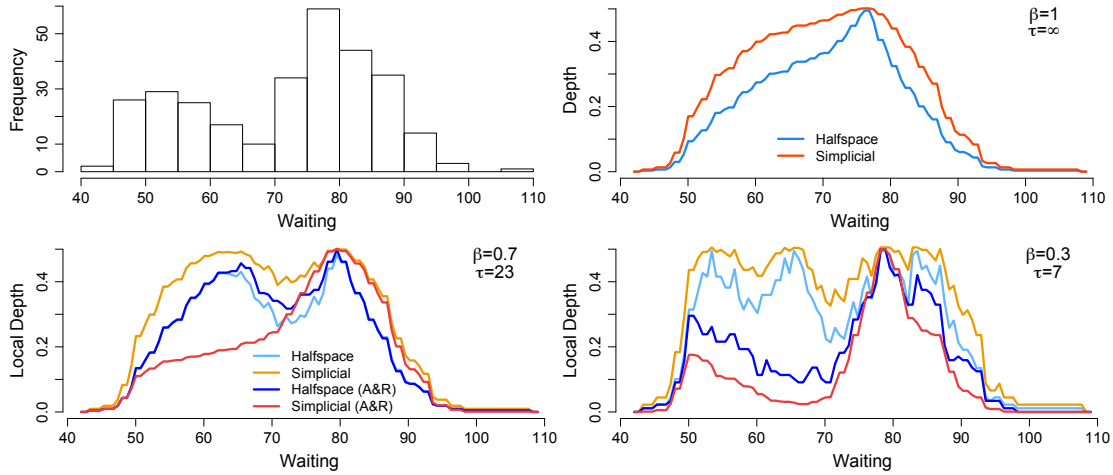

Figure 1: (Upper left:) Histogram of the variable "Waiting" from the *Geyser* data set. (Upper right): Plots of halfspace (blue) and simplicial (orange-red) depths over 100 equispaced points. (Lower:) the proposed local halfspace (light blue) and simplicial (orange) depths at locality levels $\beta \in \{.7, .3\}$, along with their halfspace (dark blue) and simplicial (red) counterparts from Agostinelli and Romanazzi (2011).

2.2 Boston data

The *Boston* data set was first introduced in Harrison and Rubinfeld (1978). It contains 506 observations related to housing and was first used to estimate the "need for clean air" in the Boston area. The data set originally contains 14 different variables. For the sake of illustration, we restrict here to two variables, namely "NOX" (annual average of nitrogen oxide concentration, in parts per ten million) and "DIS" (the

weighted mean of distances to five Boston employment centers, in miles). The left panel of Figure 2 shows a scatter plot of the resulting 506 bivariate data points.

This scatter plot shows that the data set has a non-convex support ; this entails that there may be points whose respective depth values do not reflect properly what one would naturally consider the relative centrality of these points in the data set. To illustrate this, we consider four particular locations, marked in orange, blue, red, and green in the scatter plot. Both for halfspace and simplicial depths, the green location is considered more central than the blue one, which is somehow paradoxical since the green location is much closer to the boundary of the support. Similarly, the red location—that actually is the halfspace deepest one—is about twice as (halfspace or simplicial) deep as the blue location, while visual inspection suggests that the latter is more central than the former (or at least is of comparable centrality).

Parallel to the univariate case, the β-local depth of a point $\mathbf{x} \in \mathbb{R}^2$ is still obtained as the global depth of $\mathbf{x}$ with respect to the data points sitting in a neighborhood of $\mathbf{x}$ containing a proportion β of the observations (the exact definition of this neighborhood, that is actually of a depth-based nature, will be provided in Section 3.2). The right panel of Figure 2 shows the plots of the proposed local (halfspace and simplicial) depths for the four locations above, as a function of the locality level β. As β moves away from one (that still corresponds to going from global depth to more and more local depth), the paradoxes above vanish : both the green location (that is close to the boundary of the support) and the red location show decreasing local depths that eventually fall below the local depth of the blue one. Note that, except for small β-values (to which little attention should be paid, as local depth is then evaluated on the basis of few observations in each neighborhood), the orange location has uniformly low local depth, which is expected since it is close to the boundary of the convex hull of the data (would this point be outside the convex hull, its local

depth would be zero for any β).

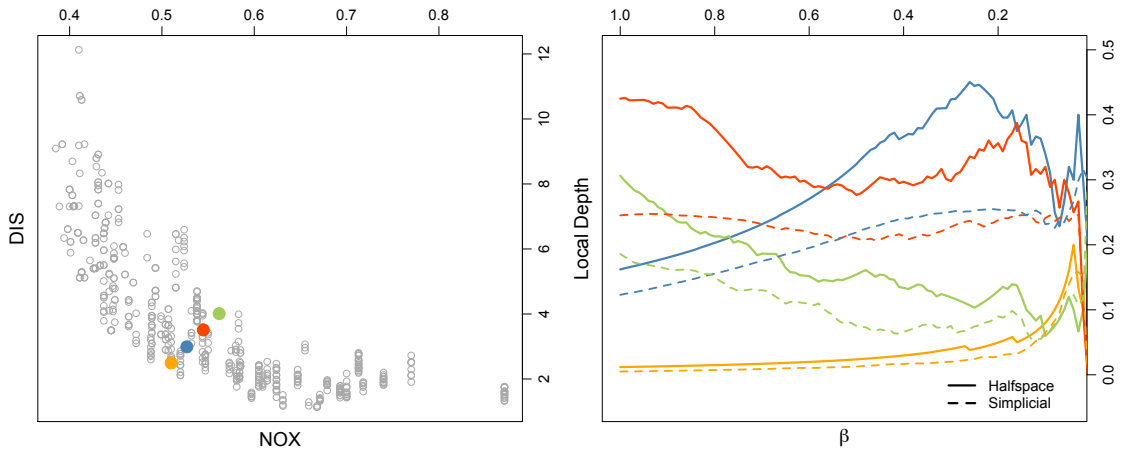

Figure 2: (Left:) Scatterplot of the NOX and DIS variables from the Boston data set, with four particular locations. (Right:) Plots, as a function of the locality level β, of the proposed local halfspace (solid curves) and simplicial (dashed curves) depths of these locations.

A comparison with the local halfspace and simplicial depths from Agostinelli and Romanazzi (2011) is provided in the Supplementary Materials ; in particular, it is shown there that the lack of affine-invariance of their local halfspace depth makes the results very sensitive to unit changes.

3 FROM GLOBAL TO LOCAL DEPTH

In this section, we first review the concept of depth (Section 3.1). We then explain how it can be used to construct neighborhoods of any $\mathbf{x} \in \mathbb{R}^d$, and propose a local version of any depth (Section 3.2). Finally, we define the sample local depths that were already put at work in Section 2, and establish their consistency (Section 3.3).

3.1 Depth functions

A depth function $D(\cdot\,, P)$ associates with any $\mathbf{x} \in \mathbb{R}^d$ a measure $D(\mathbf{x}, P)(\geq 0)$ of its *centrality* with respect to the probability measure P over $\mathbb{R}^d$ (the more *central* $\mathbf{x}$ is, the *deeper* it is). The two most celebrated depths are the following.

Definition 3.1 (Tukey, 1975) *Denoting by S^{d-1} the set of unit vectors in $\mathbb{R}^d$, the* halfspace depth *of $\mathbf{x}$ with respect to P is the "minimal" probability of all halfspaces containing $\mathbf{x}$, i.e., $D_H(\mathbf{x}, P) = \inf_{\mathbf{u}\in S^{d-1}} P\left[\mathbf{u}'(\mathbf{X} - \mathbf{x}) \geq 0\right]$, where $\mathbf{X} \sim P$.*

Definition 3.2 (Liu, 1990) *Letting $S(\mathbf{x}_1, \ldots, \mathbf{x}_{d+1})$ be the convex hull of $\mathbf{x}_1, \ldots, \mathbf{x}_{d+1}$, the* simplicial depth *of $\mathbf{x}$ with respect to P is $D_S(\mathbf{x}, P) = P\left[\mathbf{x} \in S(\mathbf{X}_1, \ldots, \mathbf{X}_{d+1})\right]$, where $\mathbf{X}, \ldots, \mathbf{X}_{d+1}$ are i.i.d. from P.*

There are numerous other depths, including the (standardized) spatial depth (Chaudhuri, 1996; Serfling, 2010), the projection depth (Zuo, 2003), the simplicial volume and Mahalanobis depths (Zuo and Serfling, 2000a), the zonoid depth (Koshevoy and Mosler (1997)), etc. The halfspace depth and—under absolute continuity—the simplicial depth are *statistical depth functions* in the following sense.

Definition 3.3 (Zuo and Serfling, 2000a) *A bounded mapping $D(\cdot\,, P)$ from $\mathbb{R}^d$ to $\mathbb{R}^+$ is a* statistical depth function *if it satisfies the four following properties :*

(P1) affine-invariance: *for any $d \times d$ invertible matrix $\mathbf{A}$, any d-vector $\mathbf{b}$, and any distribution P on $\mathbb{R}^d$, $D(\mathbf{A}\mathbf{x} + \mathbf{b}, P^{\mathbf{A},\mathbf{b}}) = D(\mathbf{x}, P)$, where $P^{\mathbf{A},\mathbf{b}}$ stands for the distribution of $\mathbf{A}\mathbf{X} + \mathbf{b}$ when $\mathbf{X}$ has distribution P;*

(P2) maximality at center: *if $\boldsymbol{\theta}$ is a center of (central, angular or halfspace) symmetry of P, then it holds that $D(\boldsymbol{\theta}, P) = \sup_{\mathbf{x}\in\mathbb{R}^d} D(\mathbf{x}, P)$;*

(P3) monotonicity relative to deepest point: *for any P having deepest point $\boldsymbol{\theta}$, $D(\mathbf{x}, P)$ $\leq D((1-\lambda)\boldsymbol{\theta} + \lambda\mathbf{x})$ for any $\mathbf{x}$ in $\mathbb{R}^d$ and any $\lambda \in [0, 1]$;*

(P4) vanishing at infinity: *for any* P, $D(\mathbf{x}, P) \to 0$ *as* $||\mathbf{x}|| \to \infty$.

For any depth function, the *depth regions* $R_\alpha(P) = \{\mathbf{x} \in \mathbb{R}^d \,|\, D(\mathbf{x}, P) \geq \alpha\}$ (of *order* $\alpha > 0$) are of paramount importance as they reveal very diverse characteristics from P : location, dispersion, dependence structure, etc. (see, e.g., Liu et al., 1999). Clearly, these regions are nested, and inner regions contain points with larger depth. When defining local depth below, it will be more appropriate to index the family $\{R_\alpha(P)\}$ by means of probability contents : for any $\beta \in [0, 1]$, we define

$$R^\beta(P) = \bigcap_{\alpha \in A(\beta)} R_\alpha(P), \qquad \text{with } A(\beta) = \big\{\alpha \geq 0 : P\left[R_\alpha(P)\right] \geq \beta\big\}, \tag{3.1}$$

the smallest depth region with P-probability larger than or equal to β ; we use subscripts and superscripts to denote depth regions associated with some fixed order (α) and some fixed probability content (β), respectively.

3.2 Depth-based neighborhoods and local depth

Unlike all local depth concepts available in the literature, the proposed local depth will involve *neighborhoods* of any location $\mathbf{x} \in \mathbb{R}^d$. The depth regions $R_\alpha(P)$ or $R^\beta(P)$ provide neighborhoods of the deepest point(s) only, hence cannot be used for that purpose. However, in view of (P2)-(P3) in Definition 3.3, neighborhoods of any $\mathbf{x} \in \mathbb{R}^d$ can be obtained (as in Paindaveine and Van Bever (2012)) by replacing $P = P^{\mathbf{X}}$ with its symmetrized version $P_{\mathbf{x}} = \frac{1}{2}P^{\mathbf{X}} + \frac{1}{2}P^{2\mathbf{x}-\mathbf{X}}$. In line with most depths, the resulting (depth-based) neighborhoods $R_\alpha(P_{\mathbf{x}})$ or $R^\beta(P_{\mathbf{x}})$ are of a nonparametric nature. The parameter α (resp., β) plays the role of the locality parameter, smaller neighborhoods corresponding to larger values of α (resp., to smaller values of β).

Definition 3.4 *The order-α (resp., probability-β) depth-based neighborhood of* $\mathbf{x}$ *with respect to the distribution* P *is* $R_{\mathbf{x},\alpha}(P) = R_\alpha(P_{\mathbf{x}})$ *(resp.,* $R^\beta_{\mathbf{x}}(P) = R^\beta(P_{\mathbf{x}})$*).*

Alternatively, one could (centro-)symmetrize $P = P^{\mathbf{X}}$ about $\mathbf{x}$ by mapping it to $g_1(P^{\mathbf{X}}) = \frac{1}{2}P^{\mathbf{x}-\mathbf{X}} + \frac{1}{2}P^{\mathbf{x}+\mathbf{X}}$ or to $g_2(P^{\mathbf{X}}) = \frac{1}{4}P^{\mathbf{X}} + \frac{1}{4}P^{\mathrm{Rot}_{\mathbf{x}}^{\pi/2}(\mathbf{X})} + \frac{1}{4}P^{\mathrm{Rot}_{\mathbf{x}}^{\pi}(\mathbf{X})} + \frac{1}{4}P^{\mathrm{Rot}_{\mathbf{x}}^{3\pi/2}(\mathbf{X})}$, where $\mathrm{Rot}_{\mathbf{x}}^{\omega}$ stands for the rotation about $\mathbf{x}$ with angle ω (in radians). However, g_1 provides depth-based neighborhoods that fail to be spherically symmetric about $\mathbf{x}$ in case $P^{\mathbf{X}}$ itself is spherically symmetric about $\mathbf{x}(\neq \mathbf{0})$, and g_2 leads to neighborhoods that are not affine-equivariant and require more computational efforts in the sample case. This motivates using the symmetrization $P^{\mathbf{X}} \mapsto \frac{1}{2}P^{\mathbf{X}} + \frac{1}{2}P^{2\mathbf{x}-\mathbf{X}}$.

Now, conditioning on the depth-based neighborhoods from Definition 3.4 provides a local version of any depth D. More precisely, we adopt the following definition.

Definition 3.5 *Let $D(\cdot\,, P)$ be a depth function. The corresponding local depth function at locality level $\beta(\in (0,1])$—or simply, β-*local depth function*—is*

$$LD^{\beta}(\cdot\,, P) \,:\, \mathbb{R}^d \to \mathbb{R}^+ : \mathbf{x} \mapsto LD^{\beta}(\mathbf{x}, P) = D(\mathbf{x}, P_{\mathbf{x}}^{\beta}),$$

where $P_{\mathbf{x}}^{\beta}[\,\cdot\,] = P[\,\cdot\,|R_{\mathbf{x}}^{\beta}(P)]$ is the conditional distribution of P, conditional on $R_{\mathbf{x}}^{\beta}(P)$.

As announced, we favor the β-parametrization over the α-parametrization when defining local depth. The reason is twofold. First, the maximal depth order $\alpha_*(P) = \max_{\mathbf{x}\in\mathbb{R}^d} D(\mathbf{x}, P)$, hence also the range of relevant α-values, depends on P. Second, and more importantly, the neighborhood $R_{\mathbf{x},\alpha}(P)$ may have P-probability zero for α close to $\alpha_*(P)$ (an example is obtained for $\mathbf{x} = \mathbf{0} \in \mathbb{R}^d$ and P being the distribution of $\mathbf{X}$ conditional on $[\|\mathbf{X}\| > 1]$, where $\mathbf{X}$ is standard d-variate normal), in which case $LD_{\alpha}(\mathbf{x}, P) = D(\mathbf{x}, P[\,\cdot\,|R_{\mathbf{x},\alpha}(P)])$ would not be properly defined. In contrast, β-local depth is always well-defined, and the range of β-values does not depend on P, nor on D : β always goes from 0 (extreme localization) to 1 (no localization).

Unlike its competitors, this construction of local depth applies in a generic way to any depth D, and it ensures affine-invariance at any locality level β (which follows from Property (P1)). For $\beta = 1$, the local depth reduces to its global an-

tecedent D, which shows that our concept extends usual (global) depth. The properties of $LD^\beta(\cdot\,, P)$ for extreme locality (i.e., as $\beta \to 0$) will be considered in Section 4.

3.3 Sample local depth and consistency

We now turn to the sample case. To do so, consider d-variate mutually independent observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with common distribution P, and denote by $P^{(n)}$ the corresponding empirical distribution. Classically, sample (global) depths are obtained by substituting $P^{(n)}$ for P in $D(\cdot\,, P)$, which leads, e.g., to the sample halfspace depth $D_H(\mathbf{x}, P^{(n)}) = \frac{1}{n} \inf_{\mathbf{u}\in S^{d-1}} \#\big\{i = 1, \ldots, n : \mathbf{u}'(\mathbf{X}_i - \mathbf{x}) \geq 0\big\}$, and the sample simplicial depth $D_S(\mathbf{x}, P^{(n)}) = \binom{n}{d+1}^{-1} \sum_{1\leq i_1<i_2<\ldots<i_{d+1}\leq n} \mathbb{I}\big[\mathbf{x} \in S(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_{d+1}})\big]$, where $\mathbb{I}[B]$ stands for the indicator function of B. Sample depth regions are defined accordingly : $R_\alpha(P^{(n)})$ is defined as the collection of $\mathbf{x}$'s with $D(\mathbf{x}, P^{(n)})$ larger than or equal to α, and $R^\beta(P^{(n)})$ as the intersection of all $R_\alpha(P^{(n)})$ with $P^{(n)}$-probability larger than or equal to β. We refer to He and Wang (1997) and Zuo and Serfling (2000c) for results on sample depth regions.

As in the population case, our sample local depth concept will require considering, for any $\mathbf{x} \in \mathbb{R}^d$, the symmetrized distribution $P^{(n)}_{\mathbf{x}}$, that is the empirical distribution associated with $\mathbf{X}_1, \ldots, \mathbf{X}_n, 2\mathbf{x} - \mathbf{X}_1, \ldots, 2\mathbf{x} - \mathbf{X}_n$. We adopt the following definition.

Definition 3.6 *Let $D(\cdot\,, P)$ be a depth function. The corresponding sample local depth function at locality level $\beta(\in (0, 1])$ is $LD^\beta(\cdot\,, P^{(n)}) : \mathbb{R}^d \to \mathbb{R}^+ : \mathbf{x} \mapsto LD^\beta(\mathbf{x}, P^{(n)}) = D(\mathbf{x}, P^{\beta,(n)}_{\mathbf{x}})$, where $P^{\beta,(n)}_{\mathbf{x}}$ denotes the empirical measure associated with those data points among $\mathbf{X}_i$, $i = 1, \ldots, n$ that sit in $R^\beta_{\mathbf{x}}(P^{(n)})(= R^\beta(P^{(n)}_{\mathbf{x}}))$.*

By definition, $R^\beta_{\mathbf{x}}(P^{(n)})$ is the smallest sample depth region that contains at least a proportion β of the $2n$ random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n, 2\mathbf{x} - \mathbf{X}_1, \ldots, 2\mathbf{x} - \mathbf{X}_n$, or equivalently (symmetrization indeed implies that these depth regions are centro-

symmetric about $\mathbf{x}$), a proportion β of the n original data points $\mathbf{X}_i$. Note that, for $k \in \{1, 2, \ldots, n-1\}$, ties may imply that $R_{\mathbf{x}}^{k/n}(P^{(n)})$ contains more than k of the $\mathbf{X}_i$'s.

Some applications of local depth may require selecting one or a few β-values. This choice crucially depends on the application at hand, so that no universal β-selection strategy exists. Therefore it is desirable, in every specific application, to define appropriate such strategies, at least whenever the results strongly depend on β. This will be illustrated in Section 6.

Theorem 3.1 below provides consistency of sample local depth under absolute continuity assumptions. Of course, we need assuming consistency for the original global depth $D(\cdot\,, P)$: for any absolutely continuous P and any $\mathbf{x} \in \mathbb{R}^d$, $|D(\mathbf{x}, P^{(n)}) - D(\mathbf{x}, P)| \stackrel{a.s.}{\to} 0$ as $n \to \infty$. Actually, we will need the following reinforcement.

(Q1) *weak continuity*: for any absolutely continuous P, any sequence of probability measures (P_n) that converges weakly to P as $n \to \infty$, and any $\mathbf{x} \in \mathbb{R}^d$, we have that $|D(\mathbf{x}, P_n) - D(\mathbf{x}, P)| \to 0$ as $n \to \infty$.

This reinforcement is needed to cope with the complex dependence of the sample local depth $LD^{\beta}(\mathbf{x}, P^{(n)}) = D(\mathbf{x}, P_{\mathbf{x}}^{\beta,(n)})$ on $P^{(n)}$. Note indeed that the dependence of $P_{\mathbf{x}}^{\beta,(n)}[\,\cdot\,] = P^{(n)}[\,\cdot\,|R^{\beta}(P_{\mathbf{x}}^{(n)})]$ on empirical measures is twofold.

Theorem 3.1 (Consistency) *Fix $\mathbf{x} \in \mathbb{R}^d$ and let $D(\cdot\,, P)$ satisfy Property (P2), (P3), and (Q1). Then, for any absolutely continuous P and any sequence $\beta_n \to \beta$, we have that $LD^{\beta_n}(\mathbf{x}, P^{(n)}) \stackrel{a.s.}{\to} LD^{\beta}(\mathbf{x}, P)$ as $n \to \infty$.*

Property (Q1) actually holds for many depths. In particular, Proposition 1 of Mizera and Volauf (2002) and Theorem 2.2 (ii) of Zuo (2003) establish (Q1) for the halfspace and projection depths, respectively. For simplicial depth, Dümbgen (1992) proved the stronger property $\sup_{\mathbf{x}\in\mathbb{R}^d} |D_S(\mathbf{x}, P_n) - D_S(\mathbf{x}, P)| \to 0$ as $P_n \to P$ weakly.

4 EXTREME LOCALIZATION

As locality becomes extreme, all available extensions of depth converge to either a density measure or to a constant value, hence lose their nature of a centrality measure. We now show that the proposed local depths improve on this.

4.1 Assumptions and extreme local regions

For the sake of convenience, we list here the assumptions—all on the original depth D—we will need in this section.

(Q1$^+$) *uniform weak continuity*: for any two sequences (P_n) and (P'_n) of absolutely continuous distributions for which $|P_n[B] - P'_n[B]| \to 0$ as $n \to \infty$ for any Borel set B, $|D(\mathbf{x}, P_n) - D(\mathbf{x}, P'_n)| \to 0$ as $n \to \infty$ for any $\mathbf{x} \in \mathbb{R}^d$;

(Q2) *unique maximization at the symmetry center*: if P is absolutely continuous (with density f, say) and is centrally symmetric about $\boldsymbol{\theta}$ in the closure $\mathrm{Supp}(f)$ of $\mathrm{Supp}_+(f) = \{\mathbf{x} \in \mathbb{R}^d \,|\, f(\mathbf{x}) > 0\}$, then $D(\boldsymbol{\theta}, P) > D(\mathbf{x}, P)$ for all $\mathbf{x}$;

(Q3) *P-independent depth at the symmetry center*: if P is absolutely continuous and centrally symmetric about $\boldsymbol{\theta}$, then $c_D = D(\boldsymbol{\theta}, P)$ (that, under (P2), is equal to $\max_{\mathbf{x}\in\mathbb{R}^d} D(\mathbf{x}, P)$) is independent of P (which justifies the notation c_D).

Recalling that $LD^\beta(\mathbf{x}, P) = D(\mathbf{x}, P^\beta_\mathbf{x})$, where $P^\beta_\mathbf{x}$ is obtained from P by conditioning it on $R^\beta_\mathbf{x}(P)$, it seems natural to expect that, under Property (Q1),

$$\lim_{\beta\to 0} LD^\beta(\mathbf{x}, P) = D(\mathbf{x}, P^0_\mathbf{x}), \tag{4.1}$$

where $P^0_\mathbf{x}$ denotes the possible weak limit of $P^\beta_\mathbf{x}$. Unfortunately, the situation is not so simple, as we show below. We start with a result on $R^0_\mathbf{x}(P) := \bigcap_{\beta>0} R^\beta_\mathbf{x}(P)$ of $P^0_\mathbf{x}$.

Lemma 4.1 *Let $D(\cdot, P)$ satisfy (P2), (P3), (Q1), and (Q2). Fix an absolutely continuous P (with density f, say). Then, (i) for any $\mathbf{x} \in \mathrm{Supp}(f)$, for all $\varepsilon > 0$,*

there exists $\beta > 0$ *such that* $R^{\beta}_{\mathbf{x}}(P) \subset B_{\mathbf{x}}(\varepsilon) := \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon\}$*, so that* $R^{0}_{\mathbf{x}}(P) = \{\mathbf{x}\}$*; (ii) if one further assumes that (Q2) also holds for symmetry centers* $\boldsymbol{\theta} \notin \mathrm{Supp}(f)$*, then, for any* $\mathbf{x} \notin \mathrm{Supp}(f)$*,* $\mathbf{x}$ *belongs to the interior of* $R^{0}_{\mathbf{x}}(P)$*.*

This result leads to treating separately the cases $\mathbf{x} \in \mathrm{Supp}(f)$ and $\mathbf{x} \notin \mathrm{Supp}(f)$.

4.2 Extreme behavior in the support of the distribution

For $\mathbf{x} \in \mathrm{Supp}(f)$, (4.1) cannot hold because, as we show in the Supplementary Materials, $P^{0}_{\mathbf{x}}$ does not exist. This explains why we have to reinforce (Q1) into (Q1$^+$), under which the following result holds (see the Appendix for the proof).

Theorem 4.1 *Let* $D(\cdot, P)$ *satisfy (P2), (P3), (Q1*$^+$*), (Q2), and (Q3). Fix an absolutely continuous* P *(with density* f*, say). Let* $\mathbf{x} \in \mathrm{Supp}_+(f)$ *be a continuity point of* f*. Then* $LD^{\beta}(\mathbf{x}, P) \to c_D$ *as* $\beta \to 0$*, where* c_D *is the constant in (Q3).*

Therefore, unlike most of its competitors, our local depth concept is not of a density nature under extreme localization ; irrespective of the density at $\mathbf{x} \in \mathrm{Supp}_+(f)$, the limiting local depth at $\mathbf{x}$ takes the constant (maximal) value c_D, supporting the intuition that, for extreme locality, points inside the support get arbitrarily central.

More precise results can be derived for univariate halfspace and simplicial depths (see the Supplementary Materials for a proof).

Theorem 4.2 *Fix* $x \in \mathrm{Supp}_+(f)$*. Then, (i) provided that* f *admits a continuous derivative* f' *in a neighborhood of* x*, we have that, as* $\beta \to 0$*,* $LD^{\beta}_{H}(x, P) = \frac{1}{2} - \frac{|f'(x)|}{8f^2(x)}\beta + o(\beta)$*; (ii) provided that* f *admits a continuous second derivative* f'' *in a neighborhood of* x*, we have that, as* $\beta \to 0$*,* $LD^{\beta}_{S}(x, P) = \frac{1}{2} - \frac{(f'(x))^2}{16f^4(x)}\beta^2 + o(\beta^2)$*.*

This shows that, for small β-values, local depth is not characterized by $f(x)$, but rather by $|f'(x)|/f^2(x)$, that measures *local asymmetry* at x. This further indicates

that our local depth provides a centrality measure for x, and not a density measure at x. From Theorem 4.2, LD_S^β is seen to converge to $1/2 (= c_{D_S} = c_{D_H}$ for $d = 1)$ faster than LD_H^β does. As a consequence, one may expect having to consider larger values for simplicial depth than for halfspace depth to find out about the above local asymmetry features ; this may actually be seen in Figure 3 below.

We point out that, in contrast with Theorem 4.1, a point $\mathbf{x}$ at the boundary of the support may assume, as $\beta \to 0$, any limiting local depth value between the minimal possible value 0 and the maximal possible value c_D. This is shown on a bivariate example in the Supplementary Materials.

4.3 Extreme behavior outside the support of the distribution

For $\mathbf{x} \notin \text{Supp}(f)$, the weak limit $P_{\mathbf{x}}^0 = \lim_{\beta \to 0} P_{\mathbf{x}}^\beta$, when it does exist, coincides with the probability measure obtained by conditioning P on $R_{\mathbf{x}}^0$ (which, according to Lemma 4.1(ii), is a neighborhood of $\mathbf{x}$). Since the interior of $R_{\mathbf{x}}^0$ has zero P-probability, the support of $P_{\mathbf{x}}^0$ is contained in the boundary $\partial R_{\mathbf{x}}^0$ of $R_{\mathbf{x}}^0$, so that $P_{\mathbf{x}}^0$ may *not* be absolutely continuous.

Quite fortunately, for the halfspace and simplicial depths, Property (Q1) extends to P's that are not absolutely continuous ; see Remark 2.5 in Zuo (2003). For these depths, we may therefore conclude that $\lim_{\beta \to 0} LD^\beta(\mathbf{x}, P) = D(\mathbf{x}, P_{\mathbf{x}}^0)$ as in (4.1). For most $\mathbf{x} \notin \text{Supp}(f)$, the support of $P_{\mathbf{x}}^0$ will be contained in an open halfspace having $\mathbf{x}$ on its boundary hyperplane, in which case $\lim_{\beta \to 0} LD^\beta(\mathbf{x}, P) = D(\mathbf{x}, P_{\mathbf{x}}^0) = 0$ for both halfspace and simplicial depths. It is only in some very specific points $\mathbf{x} \notin \text{Supp}(f)$, that typically are symmetry centers of the corresponding limiting region $R_{\mathbf{x}}^0$, that $\lim_{\beta \to 0} LD^\beta(\mathbf{x}, P) = D(\mathbf{x}, P_{\mathbf{x}}^0)$ will be non-zero. Quite interestingly, the resulting value needs not be the maximal value c_D, but is obtained from $P_{\mathbf{x}}^0$ in a natural way.

We illustrate this in the univariate case $d = 1$, where the limiting region R_x^0 is

always an interval of the form $[x - h_x^0, x + h_x^0]$. From the general discussion above, we know that the support of the limiting distribution P_x^0 is included in $\partial R_x^0 = \{x - h_x^0, x + h_x^0\}$. Denoting by p_x^- and p_x^+ the respective probabilities P_x^0 assigns to $x - h_x^0$ and $x + h_x^0$, we obtain that

$$\lim_{\beta \to 0} LD_H^\beta(x, P) = D_H(x, P_x^0) = \min(p_x^-, p_x^+), \tag{4.2}$$

$$\lim_{\beta \to 0} LD_S^\beta(x, P) = D_S(x, P_x^0) = 2p_x^- p_x^+. \tag{4.3}$$

The probabilities (p_x^-, p_x^+) can be computed from the identities

$$p_x^- + p_x^+ = 1 \quad \text{and} \quad \frac{p_x^+}{p_x^-} = \lim_{\varepsilon \gtrdot 0} \frac{P[X \in (x + h_x^0, x + h_x^0 + \varepsilon)]}{P[X \in (x - h_x^0 - \varepsilon, x - h_x^0)]} \quad (\in [0, \infty]).$$

An explicit example is provided in Section 5.

5 Examples

We restrict our attention to local halfspace and simplicial depths, that, in the univariate case, admit the following explicit expressions (resulting from the well-known formulae $D_H(x, P) = \min(F(x), 1 - F(x))$ and $D_S(x, P) = 2F(x)(1 - F(x))$, where $F(x) = P[(-\infty, x]]$ is the cumulative distribution function associated with P).

Proposition 5.1 *Let $x_\beta := x - \inf\{h > 0 : F(x + h) - F(x - h) \geq \beta\}$, where F denotes the cumulative distribution function associated with the absolutely continuous distribution P. Then the local halfspace and simplicial depths of x with respect to P are given by $LD_H^\beta(x, P) = \frac{1}{\beta} \min\left[F(x) - F(x_\beta), F(2x - x_\beta) - F(x)\right]$ and $LD_S^\beta(x, P) = \frac{2}{\beta^2}\left(F(x) - F(x_\beta)\right)\left(F(2x - x_\beta) - F(x)\right)$, respectively.*

We first considered univariate Gaussian and uniform mixtures, obtained with $X \sim \frac{1}{2}\mathcal{N}(\mu_a = -2, 2) + \frac{1}{2}\mathcal{N}(\mu_b = 2, 1)$ and $X \sim \frac{1}{2}\text{Unif}(-5, -1) + \frac{1}{2}\text{Unif}(1, 3)$, respectively. Figures 3 and 4 report the plots of the corresponding β-local halfspace and simplicial depth functions for several β-values, along with the plot of the density f of X.

For the Gaussian mixture, global depth functions are unimodal, while local depth functions allow for local maxima. In line with the univariate example from Section 2, small β-values give raise to three local maxima : two located about the modes μ_a and μ_b, and a third one (also for simplicial depth, although it is less visible than for halfspace depth) at $\mu \in (\mu_a, \mu_b)$, say. The large local centrality measure μ gets for small β results from the approximate symmetry (about μ) of X in $R^\beta_\mu(P)$; however, the large volume of $R^\beta_\mu(P)$, compared to $R^\beta_{\mu_a}(P)$ and $R^\beta_{\mu_b}(P)$, allows to discriminate between both types of local maxima. Finally, the plot associated with $\beta = 0.01$ illustrates Theorem 4.1.

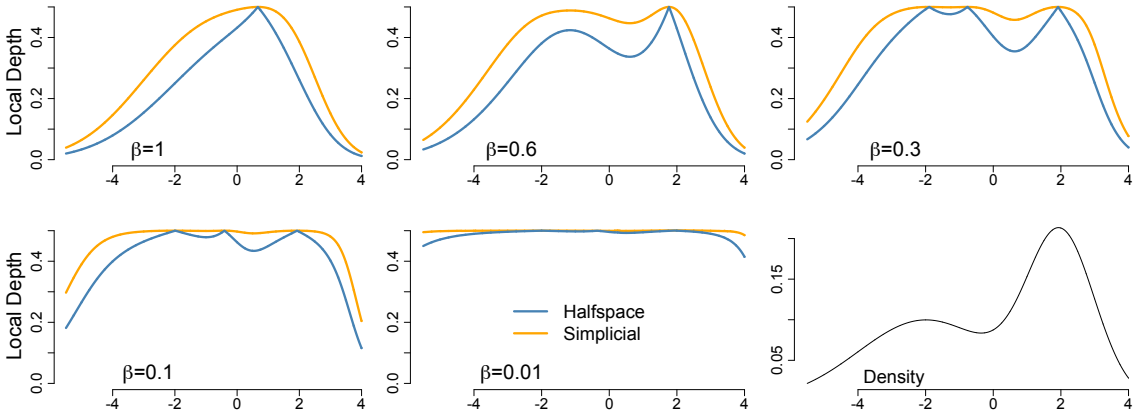

Figure 3: Plots of several β-local halfspace (blue) and simplicial (orange) depth functions for a mixture of Gaussian distributions ($X \sim \frac{1}{2}\mathcal{N}(-2,2)+\frac{1}{2}\mathcal{N}(2,1)$), along with a plot of the corresponding density.

Regarding the uniform mixture, the same comments can be repeated, and we therefore focus on the specifics of this example. It holds $\lim_{\beta\to 0} LD^\beta_i(x,P) = 1/2$ $(i = H, S)$ for all $x \in \mathrm{Supp}_+(f)$ (Theorem 4.1) and $\lim_{\beta\to 0} LD^\beta_i(x,P) = 0$ $(i = H, S)$ for $x \in \mathrm{Supp}(f) \setminus \mathrm{Supp}_+(f) = \{-5,-1,1,3\}$. For points $x \in \mathbb{R} \setminus \mathrm{Supp}(f) = (-\infty,-5)\cup(-1,1)\cup(3,\infty)$, it is easy to check that $(p^-_x, p^+_x) = (0,1)$ if $x \in (-\infty,-5)\cup(0,1)$, $(p^-_x, p^+_x) = (1,0)$ if $x \in (-1,0)\cup(3,\infty)$ and $(p^-_x, p^+_x) = (1/3, 2/3)$ if $x = 0$, which according to (4.2)-(4.3), results into $\lim_{\beta\to 0} LD^\beta_i(x,P) = 0$ $(i = H, S)$ for

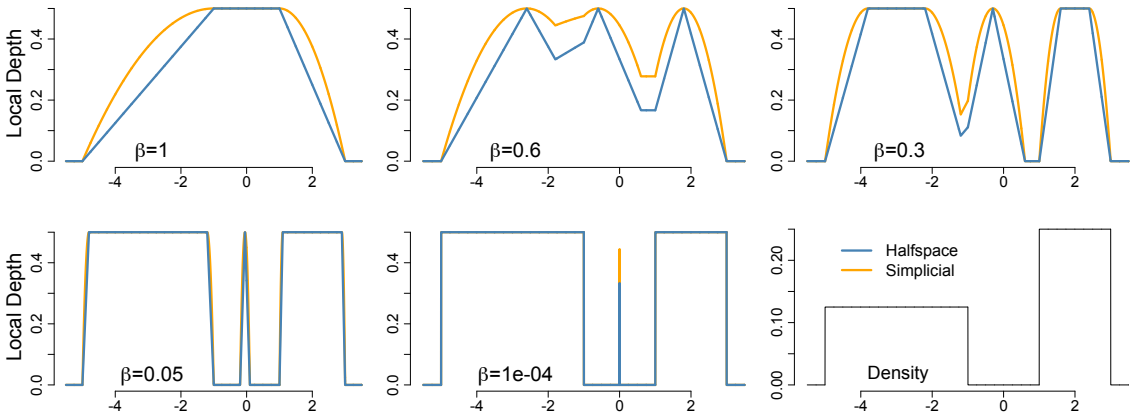

Figure 4: Plots of several β-local halfspace (blue) and simplicial (orange) depth functions for a mixture of uniform distributions ($X \sim \frac{1}{2}\text{Unif}(-5,-1)+\frac{1}{2}\text{Unif}(1,3)$), along with a plot of the corresponding density.

all non-zero such values of x, and into $\lim_{\beta\to 0} LD_H^\beta(x,P) = D_H(x,P_x^0) = 1/3$ and $\lim_{\beta\to 0} LD_S^\beta(x,P) = D_S(x,P_x^0) = 4/9$ for $x = 0$. This thoroughly explains the plot corresponding to $\beta = 10^{-4}$ in Figure 4.

Turning to the multivariate case, the two following (simulated) examples involve (i) a bimodal distribution—for which we generated $n = 1,000$ independent observations $\mathbf{X}_i = \sqrt{0.3}\, h(\mathbf{Z}_i)\mathbf{Z}_i + T_i\boldsymbol{\mu}_a + (1-T_i)\boldsymbol{\mu}_b$, where $\boldsymbol{\mu}_a = \binom{0}{0}$, $\boldsymbol{\mu}_b = \binom{2}{0}$, the $\mathbf{Z}_i$'s are i.i.d. standard bivariate normal, $h(\mathbf{z})$ is the indicator that the Euclidean norm of $\mathbf{z}$ is smaller than 0.6, and the T_i's are i.i.d $\text{Bin}(0,1/2)$, independent from the $\mathbf{Z}_i$'s—and (ii) a non-convexly supported distribution, based on $n = 500$ independent observations $\binom{X_i}{Y_i}$, where $X_i \sim \text{Unif}(-1,1)$ and $Y_i|[X_i = x] \sim \text{Unif}(1.5(1-x^2), 2(1-x^2))$. Figures 5 and 6 show heatplots of the local halfspace depth functions at several locality levels β, along with observations in the upper left panels.

In Figure 5, one can see that, as β moves away from one, the multimodal nature of the distribution is revealed (a task in which global halfspace depth clearly fails). At any β, a third local maximum is present around $\boldsymbol{\mu} = (\boldsymbol{\mu}_a + \boldsymbol{\mu}_b)/2$, resulting from the symmetry of the distribution about $\boldsymbol{\mu}$ at any locality level β (i.e., $P_{\boldsymbol{\mu}}^\beta$ is centrally

symmetric about $\boldsymbol{\mu}$ for any β). Discriminating between the true modes around $\boldsymbol{\mu}_a$, $\boldsymbol{\mu}_b$, and this "artificial" mode in $\boldsymbol{\mu}$ may again be based on a comparison of the volumes of the neighborhoods $R^{\beta}_{\mathbf{x}}(P^{(n)})$, for $\mathbf{x} = \boldsymbol{\mu}_a, \boldsymbol{\mu}_b, \boldsymbol{\mu}$. Incidentally, in some applications (including, in particular, classification ; see Section 6.1), such artificial modes, due to the zero (or small) probability mass there, will have no (or low) impact in practice.

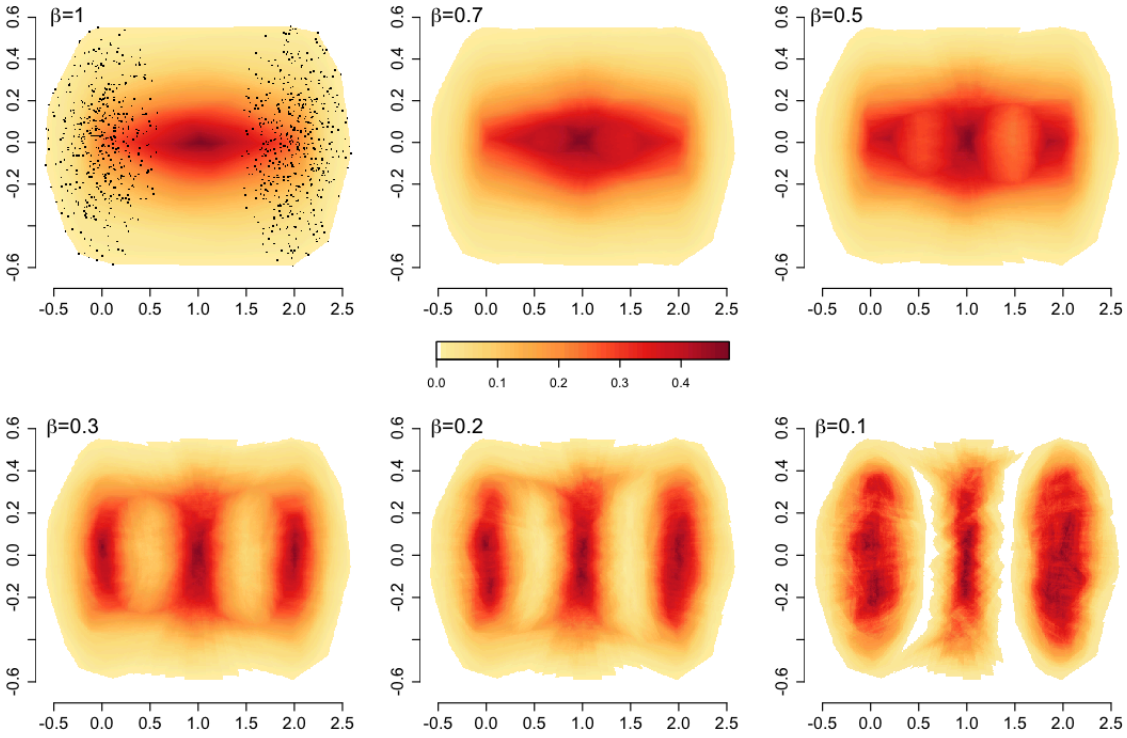

Figure 5: Heatplots of local halfspace depth functions at locality levels $\beta = 1$ (global halfspace depth), 0.7, 0.5, 0.3, 0.2, and 0.1, for $n = 1,000$ independent observations from the bivariate mixture distribution described in Section 5.

Parallel to the Boston example in Section 2, Figure 6 illustrates that global depth cannot deal with non-convexly supported distributions, since in particular the global deepest point is very close the boundary on the support. As β decreases, local depth much better reflects centrality in the present setup. Small β-values illustrate Theorem 4.1, since local depth is then almost constant in the support. We point out that this would hold irrespective of the (non-vanishing) density over the same support.

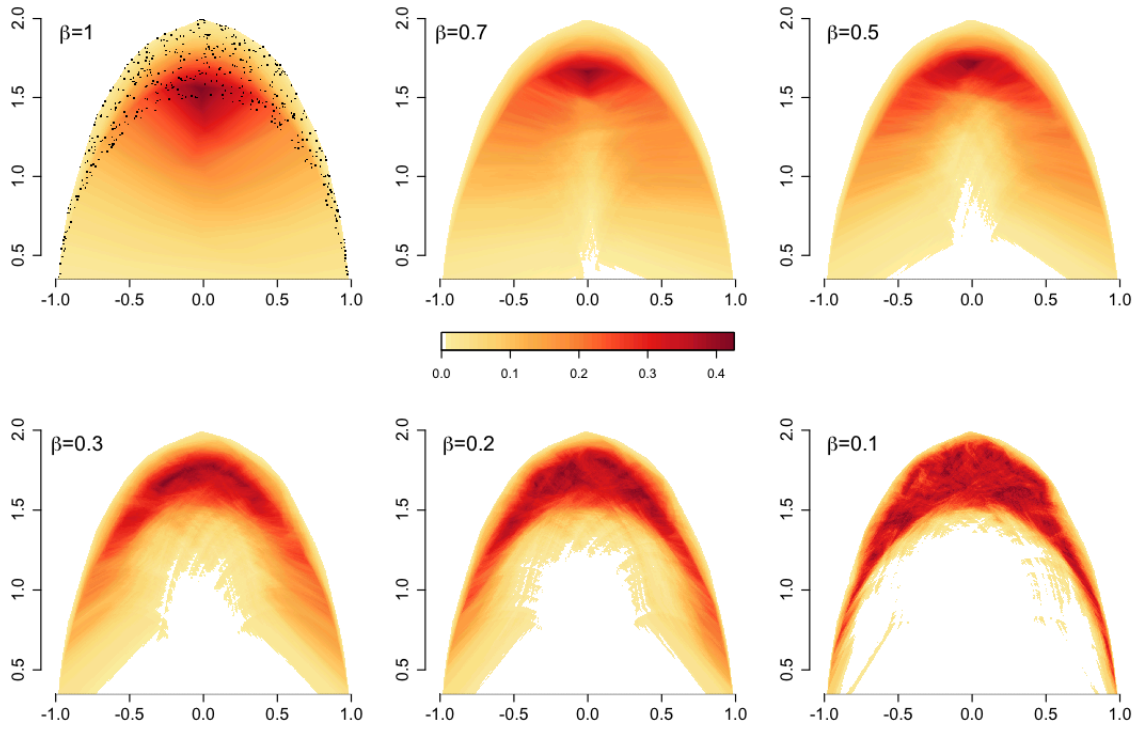

Figure 6: Heatplots of local halfspace depth functions at locality levels $\beta = 1$ (global halfspace depth), 0.7, 0.5, 0.3, 0.2, and 0.1, for $n = 500$ independent observations from the distribution with a non-convex ("moon-shaped") support described in Section 5.

6 INFERENTIAL APPLICATIONS

In this section, we describe two applications of the proposed local depth concept. The first one is related to classification, while the second one deals with symmetry testing.

6.1 Max-depth classification

Consider the classical problem in which a random d-vector is to be classified as arising from any of two probability measures P_0 or P_1, on the basis of the value $\mathbf{x}$ it assumes. This is to be achieved on the basis of a "training sample", made of two mutually independent random samples $(\mathbf{X}_{01}, \ldots, \mathbf{X}_{0n_0})$ and $(\mathbf{X}_{11}, \ldots, \mathbf{X}_{1n_1})$ from P_0 and P_1,

respectively. Depth-based classifiers typically match $\mathbf{x}$ to the population with respect to which it is most central : denoting by $P_j^{(n)}$, $j = 0, 1$, the empirical distribution associated with $(\mathbf{X}_{j1}, \ldots, \mathbf{X}_{jn_j})$, $\mathbf{x}$ is classified into Population 0 (resp., Population 1) if $D(\mathbf{x}, P_0^{(n)}) > D(\mathbf{x}, P_1^{(n)})$ (resp., $D(\mathbf{x}, P_0^{(n)}) < D(\mathbf{x}, P_1^{(n)})$), while ties are decided at random.

This *max-depth* approach was first proposed in Liu et al. (1999), and was then investigated in Ghosh and Chaudhuri (2005). In the same vein, Li et al. (2012) recently proposed the "Depth vs Depth" (DD) classifiers that improve on the max-depth ones by constructing appropriate polynomial separating curves in the DD-plot, that is, in the scatter plot of $(D(\mathbf{X}_i, P_0^{(n)}), D(\mathbf{X}_i, P_1^{(n)}))$, $i = 1, \ldots, n$ (the original max-depth classifiers simply use the main bisector as a separating curve).

As we showed in Section 2, global depth may fail to properly measure centrality for non-convexly supported distributions. Consequently, max-depth classifiers may perform poorly when P_0 and/or P_1 have a non-convex support (which is confirmed in our simulations below). Since the proposed local depths can deal with such non-convexity, one may think of defining max-*local*-depth classifiers obtained by substituting, in max-depth classifiers, β-local depth (for some β) for (global) depth. In practice, β may be chosen through cross-validation, that is, by minimizing in $\beta \in (0, 1]$, the resulting empirical misclassification rate evaluated on the training sample.

We conducted the following simulation exercise both to show that max-depth classifiers may indeed behave poorly under non-convexly supported distributions and to investigate the performances of the proposed max-local-depth classifiers. Three bivariate distributional setups were investigated :

Setup 1 (multinormality): P_j, $j = 0, 1$, is bivariate normal with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$, with $\boldsymbol{\mu}_0 = \binom{0}{0}$, $\boldsymbol{\mu}_1 = \binom{2}{2}$, $\boldsymbol{\Sigma}_0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, and $\boldsymbol{\Sigma}_1 = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$;

Setup 2 (moon- and ball-supported distributions): P_0 is the distribution of $\binom{X}{Y}$,

where $X \sim \text{Unif}(-1,1)$ and $Y|[X=x] \sim \text{Unif}(1.5(1-x^2), 2(1-x^2))$, whereas P_1 is the uniform distribution on the ball with center $\binom{0}{1.3}$ and radius 0.7;

Setup 3 (ring- and rectangle-supported distributions): P_0 is the distribution of $R\mathbf{U}$, where $R \sim \text{Unif}(1,2)$ and $\mathbf{U} = \binom{\cos\Theta}{\sin\Theta}$, with $\Theta \sim \text{Unif}(0, 2\pi)$, are independent, while P_1 is the uniform distribution on the rectangle $(-1.5, 1.5) \times (-25, 25)$.

Exactly as in Li et al. (2012), we generated, for each setup, 100 training samples of size $n_0 = n_1 = 200$, and recorded, on corresponding test samples of size $n_{test} = 1,000$ (500 observations from each population), the misclassification frequencies of the following classifiers (all depth-based classifiers below are based on halfspace depth) : (i) the Linear and Quadratic Discriminant Analysis classifiers (LDA/QDA); (ii) the standard kNN classifier, where k is chosen through cross-validation (kNN); (iii) the max-depth classifier from Ghosh and Chaudhuri (2005) (max-D); (iv) its (exact) DD-refinements from Li et al. (2012) based on linear and quadratic separating curves (DD1 and DD2); (v) our cross-validated max-local-depth classifier (max-LD ($\beta = \beta_{\text{CV}}$)); (vi) various max-local-depth classifiers based on a fixed β (max-LD).

Figure 7 shows boxplots of the resulting misclassification frequencies, and reports, in each setup, the median of the 100 β-values selected through cross-validation. Our cross-validated max-local-depth classifier shows similar performances as its depth-based competitors under ellipticity (Setup 1), but clearly outperforms these under non-convex populations (Setups 2 and 3), with the only exception of the classifier DD2 in Setup 3 with whom it competes equally. The β-values selected through cross-validation nicely reflect the non-convexity of the underlying setup, hence the need to restrict to observations that are close to the point to be classified (β small) or the allowance to base classification on all observations (β close to 1). This is seen in the three setups, where the medians of the 100 selected β-values are, respectively, .9 (convex setup), .125 and .3 (non-convex setups).

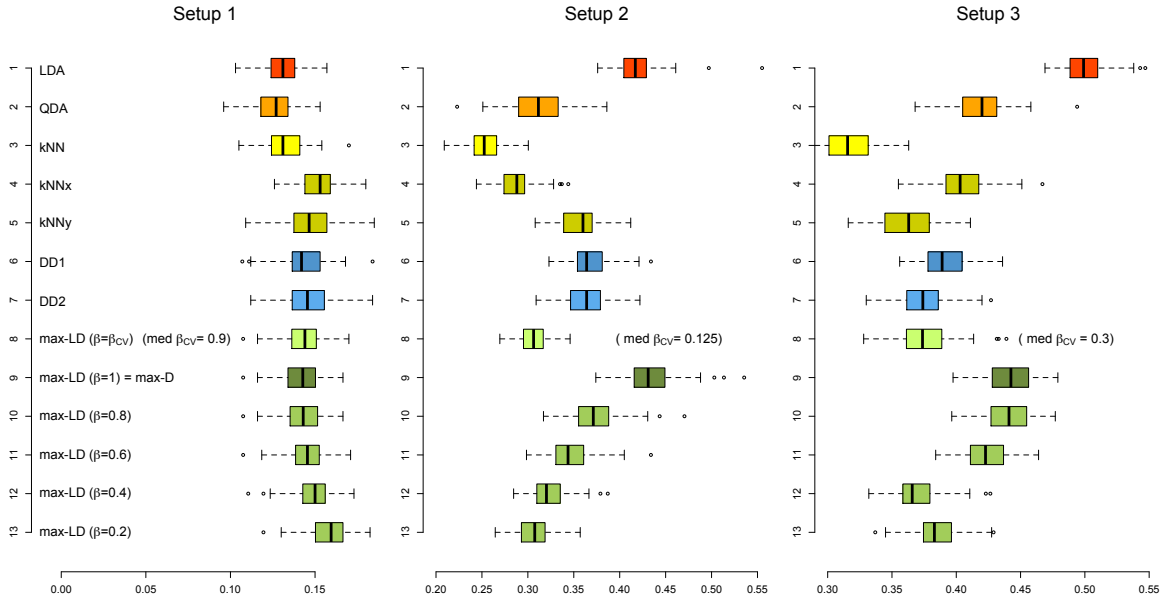

Figure 7: Boxplots of missclassification frequencies from 100 replications, in Setups 1 to 3 described in Section 6.1, with training sample sizes $n_0 = n_1 = 200$ and test sample size $n_{\text{test}} = 1,000$ (500 observations from each population), of the LDA/QDA classifiers, the standard (cross-validated) kNN classifier (kNN), the exact linear (DD1) and quadratic (DD2) DD-classifiers, the proposed cross-validated max-local-depth classifiers (max-LD ($\beta = \beta_{\text{CV}}$)), as well as some max-local-depth classifiers with fixed β, for $\beta = 1$ (max-depth classifier) and $\beta = 0.8, 0.6, 0.4, 0.2$. Results are also provided, under (kNNx) (resp., (kNNy)), for kNN classifiers applied to samples obtained by multiplying by 10 the first (resp., second) coordinate of each observation.

Comparison with classical benchmarks is also of interest. As expected, our cross-validated max-local-depth classifier dominates LDA/QDA classifiers under non-convexity. On the contrary, the (universally consistent) kNN classifiers seem to dominate the proposed classifiers, hence also our depth-based competitors from Li et al. (2012) (which may seem unexpected in view of the Monte Carlo comparisons conducted there). Unlike depth-based classifiers, however, kNN classifiers fail to be affine-invariant, hence may show significantly poorer performances under unit changes. This

is illustrated in our simulations where it is seen that, in all setups, misclassification rates of kNN classifiers suffer from multiplying one of both coordinates by a factor 10.

6.2 Testing for central symmetry

There are many graphical methods based on depth—or on the companion concept of multivariate quantiles—to assess departures from angular symmetry, central symmetry, or other types of multivariate symmetry ; see Liu et al. (1999) and Serfling (2004). There are, however, few genuine tests of symmetry based on depth. To the best of our knowledge, the only such tests, available in any dimension d, are

- the test from Rousseeuw and Struyf (2002), that is a test for angular symmetry about a specified center $\mathbf{x}_0$ rejecting the null for large values of $T^{(n)}_{\mathbf{x}_0}$, with $T^{(n)}_{\mathbf{x}} = \frac{1}{2} - D_H(\mathbf{x}, P^{(n)})$. Quite remarkably, $T^{(n)}_{\mathbf{x}_0}$ is distribution-free under the null, which allows to approximate arbitrary well the exact fixed-n critical values through simulations;

- the test from Dutta et al. (2011), that may be seen as the companion test for the null of angular symmetry about an unspecified center, as it rejects this null for large values of $T^{(n)} = T^{(n)}_{\hat{\boldsymbol{\theta}}}$, where $\hat{\boldsymbol{\theta}}$ denotes the halfspace deepest point of $P^{(n)}$ (or, if unicity fails, the barycenter of the collection of deepest points). Critical values are obtained from bootstrap-type samples (as in Dutta et al. (2011), we will use the term "bootstrap", although the corresponding tests are rather of a permutation nature).

The motivation for both tests comes from the following characterization result : for an absolutely continuous P, $D_H(\mathbf{x}_0, P) \leq 1/2$, and equality holds iff P is angularly symmetric about $\mathbf{x}_0$; see Zuo (1998), Zuo and Serfling (2000b), Rousseeuw and Struyf (2004), and Dutta et al. (2011).

Since the null of central symmetry is at least as relevant for applications as the null of angular symmetry, it is unfortunate that there is no depth-based tests of central symmetry available in any dimension d. As we now show, the proposed local depth

concept allows to define (universally consistent) tests of central symmetry. This relies on the following result, that characterizes central symmetry through local depth (see the Appendix for the proof).

Theorem 6.1 *Let P be an absolutely continuous distribution over $\mathbb{R}^d$. Then P is centrally symmetric about $\mathbf{x}_0 (\in \mathbb{R}^d)$ if and only if $LD_H^\beta(\mathbf{x}_0, P) = 1/2$ for all $\beta \in (0, 1]$.*

Testing central symmetry about $\mathbf{x}_0$ may then be based on the Cramèr-Von Mises (CM) or Kolmogorov-Smirnov (KS) statistics

$$CM^{(n)}_{\mathbf{x}_0;\beta_n} = \int_{\beta_n}^1 \left(LD_H^\beta(\mathbf{x}_0, P^{(n)}) - 1/2\right)^2 d\beta \tag{6.1}$$

$$KS^{(n)}_{\mathbf{x}_0;\beta_n} = \sup_{\beta\in[\beta_n,1]} \left|LD_H^\beta(\mathbf{x}_0, P^{(n)}) - 1/2\right|, \tag{6.2}$$

where the sequence (β_n) is such that $\beta_n \to 0$ and $n\beta_n \to \infty$ (such a sequence typically allows to achieve universal consistency while discarding, at any given sample size n, the levels at which local depth can only be poorly estimated, due to the small numbers of observation in each neighborood). Critical values are obtained as in Dutta et al. (2011). More precisely, one first generates "bootstrap" samples of the form $\mathbf{X}^*_{(m)} = (\mathbf{x}_0 + s_{(m)1}(\mathbf{X}_1 - \mathbf{x}_0), \ldots, \mathbf{x}_0 + s_{(m)n}(\mathbf{X}_n - \mathbf{x}_0))$, $m = 1, \ldots, M$, where $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ denotes the original sample and the $s_{(m)i}$'s are mutually independent variables taking values ± 1 with equal probability $1/2$. The α-level critical value for $CM^{(n)}_{\mathbf{x}_0;\beta_n}$ is then simply the order-α quantile in the series $CM^{(n)}_{\mathbf{x}_0;\beta_n}(\mathbf{X}^*_{(m)})$, $m = 1, \ldots, M$ (discreteness may require randomization to achieve null size α). Critical values for $KS^{(n)}_{\mathbf{x}_0;\beta_n}$ are computed in the exact same way.

We conducted a simulation study in order to investigate the finite-sample behavior of these tests. For any of the following setups and any corresponding value of a, we generated $1,000$ independent random samples $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ of size $n = 400$ from the same distribution as the generic random vector $\mathbf{X}$:

Setup 1: $\mathbf{X} = R\binom{\cos\Theta}{\sin\Theta}$, where $\Theta \sim \text{Unif}(0, 2\pi)$ and $R|[\Theta = \theta] \sim \text{Unif}(0, \theta^a)$, for $a = 0$ (central symmetry) and $a = .125, .250, .375, .500$ (angular symmetry);

Setup 2: $\mathbf{X} = R\binom{\cos\Theta}{\sin\Theta}$, where $R \sim \text{Unif}(0, 1)$ and $(\Theta/2\pi)^{1/(1+a)} \sim \text{Unif}(0, 1)$, for $a = 0$ (central symmetry) and $a = .15, .30, .45, .60$ (no angular symmetry);

Setup 3: $\mathbf{X} = R\binom{\cos\Theta}{\sin\Theta} + \binom{a}{a}$, where $R \sim \text{Unif}(0, 1)$ and $\Theta \sim \text{Unif}(0, 2\pi)$, for $a = 0$ (central symmetry) and $a = .125, .250, .375, .500$ (no angular symmetry).

Figure 8 plots the resulting rejection frequencies (at nominal level 5%) of the angular symmetry test based on $T^{(n)}_{\mathbf{x}_0}$ and of the central symmetry tests based on $CM^{(n)}_{\mathbf{x}_0;\beta_n}$ and $KS^{(n)}_{\mathbf{x}_0;\beta_n}$, for $\beta_n = .15, .16, \ldots, .30$; exact critical values were used for $T^{(n)}_{\mathbf{x}_0}$ (see Rousseeuw and Struyf (2002)), while critical values for $CM^{(n)}_{\mathbf{x}_0;\beta_n}$, and $KS^{(n)}_{\mathbf{x}_0;\beta_n}$ were obtained as described above from $M = 1,000$ bootstrap samples.

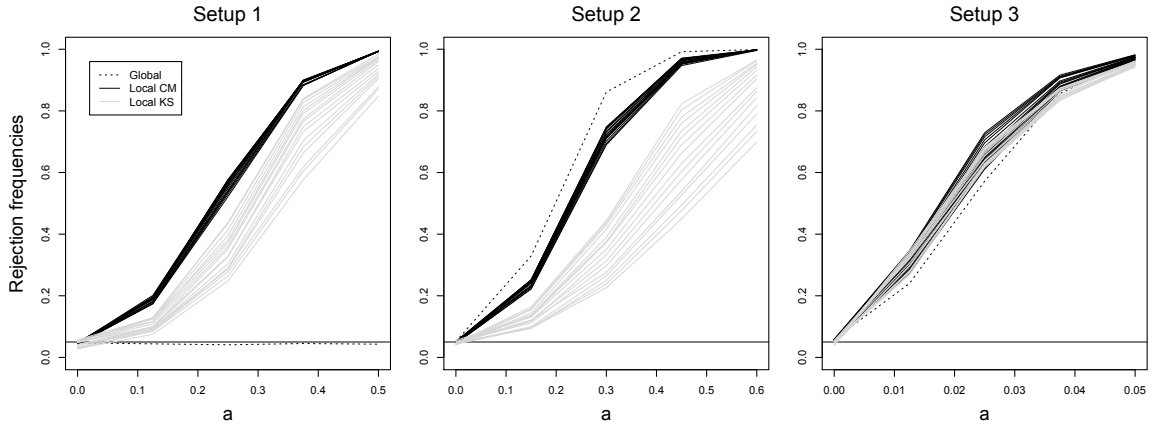

Figure 8: Rejection frequencies, in each of the three setups described in Section 6.2, of the angular symmetry test from Rousseeuw and Struyf (2002), and of the proposed Cramèr-Von Mises (CM) and Kolmogorov-Smirnov (KS) central symmetry tests, for $\beta_n = .15, .16, \ldots, .30$; results are based on 1,000 replications and the sample size is $n = 400$.

The results show that the bootstrap procedure indeed leads to central symmetry tests that have the correct size under the null. As expected, these tests succeed in

detecting central asymmetry in all setups, while the angular symmetry test, of course, shows no power in Setup 1 (which confirms that it is inappropriate as a test for central symmetry). The angular symmetry test seems to dominate the central symmetry ones in Setup 2, and the opposite holds in Setup 3. Most importantly, the proposed Cramér Von Mises local-depth-based tests, that dominate their Kolmogorov-Smirnov counterparts, show empirical powers that barely depend on β_n ; consequently, in contrast with classification in Section 6.1, it is not needed here to design a β-selection procedure (one just needs using a β_n-value that is small, but large enough to make it so that the actual sample size ($n\beta_n$) used in the most extreme local depth involved (level β_n) does not fall below 50, say).

Of course, tests for central symmetry about an unspecified center may be obtained, as in Dutta et al. (2011), by rejecting the null for large values of $CM^{(n)}_{\hat{\boldsymbol{\theta}};\beta_n}$ and $KS^{(n)}_{\hat{\boldsymbol{\theta}};\beta_n}$.

7 LOCAL REGRESSION DEPTH

Our construction extends to the regression depth context; see Rousseeuw and Hubert (1999). Let us recall that regression depth measures how well a regression hyperplane $y = \boldsymbol{\theta}'\binom{1}{\mathbf{x}}$ in $\mathbb{R}^p$—equivalently, the corresponding parameter value $\boldsymbol{\theta}$—fits the observations $\binom{\mathbf{X}_i}{Y_i}$, $i = 1, \ldots, n$, taking values in $\mathbb{R}^{p-1} \times \mathbb{R}$. Letting $\mathbf{0} = (0, \ldots, 0)' \in \mathbb{R}^p$, regression depth can be defined as

$$RD(\boldsymbol{\theta}, P^{(n)}) = D_H(\mathbf{0}, P^{(n)}_{\text{Regr};\boldsymbol{\theta}}),$$

where $P^{(n)}_{\text{Regr};\boldsymbol{\theta}}$ denotes the empirical distribution of the collection of random p-vectors $\left(Y_i - \boldsymbol{\theta}'\binom{1}{\mathbf{X}_i}\right)\binom{1}{\mathbf{X}_i}$, $i = 1, \ldots, n$. Replacing global halfspace depth with its local version proposed in this paper readily provides a local regression depth concept.

Definition 7.1 *The* local regression depth *of* $\boldsymbol{\theta}$ *with respect to* $P^{(n)}$*, at locality level* β $(\in (0, 1])$*—or simply,* β-local regression depth*—is* $LRD^{\beta}(\boldsymbol{\theta}, P^{(n)}) = LD^{\beta}_H(\mathbf{0}, P^{(n)}_{\text{Regr};\boldsymbol{\theta}})$.

To illustrate this local concept, we generated $n = 500$ independent regression observations $\binom{X_i}{Y_i}$ from a balanced mixture of simple linear regression models, according to $Y = \theta_1 + \theta_2 X + \varepsilon$, where $X \sim \text{Unif}(0, 5)$, $\varepsilon \sim \mathcal{N}(0, .1)$, and $\boldsymbol{\theta} = \binom{\theta_1}{\theta_2}$ uniformly distributed over $\{\boldsymbol{\theta}_a = \binom{.75}{0}, \boldsymbol{\theta}_b = \binom{-.25}{1}\}$, are mutually independent. Figure 9 shows the heatplots of the β-local regression depth for $\beta = 1$ (classical regression depth), 0.8, 0.6, 0.4 and 0.2, along with a scatter plot of the data.

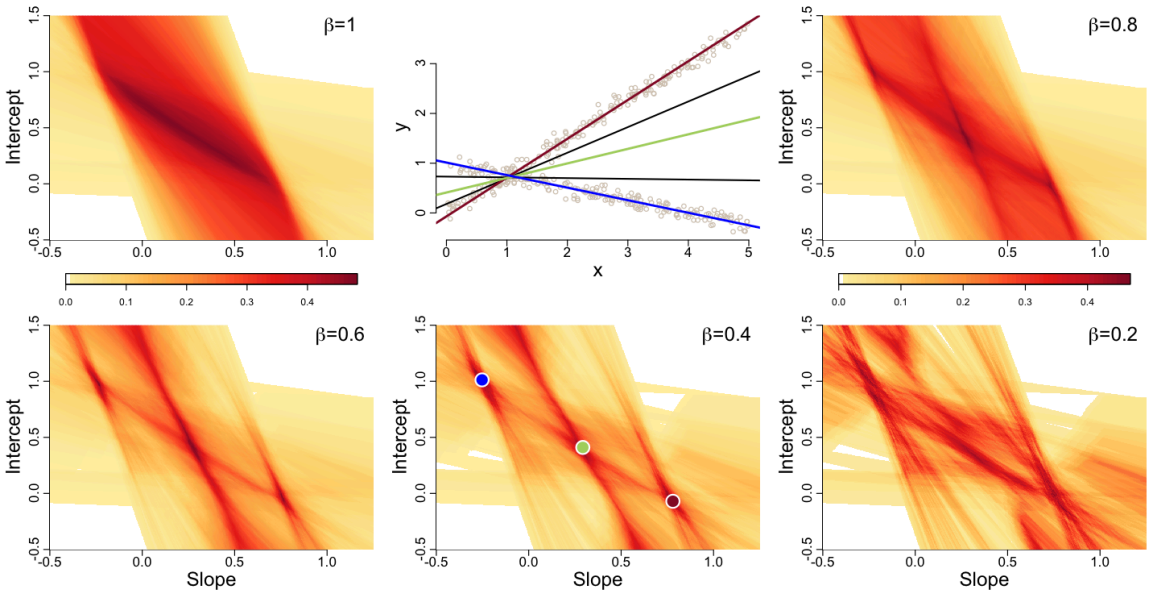

Figure 9: (Upper center:) Scatter plot of the 500 data points generated from the mixture of linear regression models described in Section 7. Maxima of global regression depth (black lines) and local maxima of $\beta = 0.4$-local regression depth (brown, green, and blue lines) are pictured. (Others:) Heatplots of local regression depth functions at locality levels $\beta = 1$ (global regression depth), 0.8, 0.6, 0.4, and 0.2. Local maxima are highlighted in the plot for $\beta = 0.4$.

All maximizers of global regression depth lie approximately on a segment in the slope-intercept space, which corresponds to a collection of regression lines passing through a fixed point $\binom{x}{y}$; we plotted in the observation space the regression lines associated with the maximizers with smallest and largest slopes (in solid lines). Clearly,

this shows that, as in the location case, global regression depth misses the mixture or "bimodal" structure of the model. In contrast, β-local regression depths clearly show local maxima about $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$, and, parallel to the location examples from the previous sections, also in a third intermediate parameter value $\boldsymbol{\theta}$, between $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$, that corresponds to a symmetry center. The regression lines associated with $\boldsymbol{\theta}_a$, $\boldsymbol{\theta}_b$, and $\boldsymbol{\theta}$ are plotted in the observation space ; the corresponding parameter values are reported in the heatplot for $\beta = 0.4$.

8 COMPUTATIONAL ASPECTS

In the location case, the evaluation of $LD^{\beta}(\mathbf{x}, P^{(n)})$ at a fixed point $\mathbf{x} \in \mathbb{R}^d$ with respect to the empirical distribution $P^{(n)}$ associated with observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ proceeds along the following few simple steps :

1. Evaluate $D(\mathbf{X}_i, P_{\mathbf{x}}^{(n)})$, $i = 1, \ldots, n$, where $P_{\mathbf{x}}^{(n)}$ is the empirical distribution associated with the symmetrized observations $\mathbf{X}_1, \ldots, \mathbf{X}_n, 2\mathbf{x}-\mathbf{X}_1, \ldots, 2\mathbf{x}-\mathbf{X}_n$;

2. Rank the (original) observations according to $D(\mathbf{X}_{(1)}, P_{\mathbf{x}}^{(n)}) \geq D(\mathbf{X}_{(2)}, P_{\mathbf{x}}^{(n)}) \geq \ldots \geq D(\mathbf{X}_{(n)}, P_{\mathbf{x}}^{(n)})$ (this ranking is not unique in case of ties, but this will not affect the final value of local depth);

3. Determine $n_\beta(P_{\mathbf{x}}^{(n)}) = \max\left\{\ell = \lceil n\beta \rceil, \ldots, n : D(\mathbf{X}_{(\ell)}, P_{\mathbf{x}}^{(n)}) = D(\mathbf{X}_{(\lceil n\beta \rceil)}, P_{\mathbf{x}}^{(n)})\right\}$;

4. Compute $LD^{\beta}(\mathbf{x}, P^{(n)}) = D(\mathbf{x}, P_{\mathbf{x}}^{\beta,(n)})$, where $P_{\mathbf{x}}^{\beta,(n)}$ is the empirical measure associated with $\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(n_\beta(P_{\mathbf{x}}^{(n)}))}$.

The computation of local regression depth is obtained by substituting above, $\mathbf{0}(\in \mathbb{R}^d)$ for $\mathbf{x}$ and $\left(Y_i - \boldsymbol{\theta}'\binom{1}{\mathbf{X}_i}\right)\binom{1}{\mathbf{X}_i}$ for $\mathbf{X}_i$, $i = 1, \ldots, n$, and by restricting to halfspace depth $D = D_H$.

The procedure in Steps 1-4 makes clear that the proposed sample local depths can be computed from global depth routines only (all illustrations in this paper were

simply obtained from the R package *depth*). This is another advantage over the competing local depths, that do require developing specific routines or packages ; see, e.g., the R package *localdepth*, from Agostinelli and Romanazzi (2011).

Note that the evaluation of $LD^{\beta}(\mathbf{x}, P^{(n)})$ may be time consuming since it requires computing $n+1$ depth values (n depth values, in a sample of $2n$ data points, in Step 1, and one depth value, in a sample of $n_{\beta}(P_{\mathbf{x}}^{(n)})(\leq n)$ data points, in Step 4). Quite fortunately, there has been much progress in the computation of depth in the recent years ; see in particular Hallin et al. (2010) for halfspace depth, and Liu and Zuo (2011a,b) and Liu et al. (2011) for projection depth.

Of course, computing "the whole local depth field" $\{LD^{\beta}(\mathbf{x}, P^{(n)}) : \mathbf{x} \in \mathbb{R}^d\}$ — in practice, computing local depth on a fine grid in a compact set — may still be very demanding. Generating the heat plots in Figures 5, 6, and 9 relied on a trivial method, where evaluation of $LD^{\beta}(\mathbf{x}, P^{(n)})$ started from scratch at any newly considered $\mathbf{x}$, which, indeed, may be slow for moderate to large sample sizes n. However, the value of $LD^{\beta}(\mathbf{x}+\mathbf{\Delta}, P^{(n)})$, with $\mathbf{\Delta}$ small, might be computed from the previous evaluation of $LD^{\beta}(\mathbf{x}, P^{(n)})$, by exploiting the fact that the distributions $P_{\mathbf{x}}^{(n)}$ and $P_{\mathbf{x}+\mathbf{\Delta}}^{(n)}$, hence also the empirical measures $P_{\mathbf{x}}^{\beta,(n)}$ and $P_{\mathbf{x}+\mathbf{\Delta}}^{\beta,(n)}$ (leading to the corresponding local depth values in Step 4 above), are close to each other. How to turn this into a practical algorithm allowing to compute efficiently the local depth field clearly remains a non-trivial question, that is beyond the scope of this methodological paper.

Now, most importantly, practical applications of local depth typically do not require evaluating the whole local depth field, but rather requires computing local depth at one or a reasonably small number of locations $\mathbf{x}$ only. This is the case for both applications considered in Section 6 : classification indeed requires evaluating local depth only at points to be classified (and at data points if β is selected through cross-validation), whereas symmetry testing only involves the local depth of the null

symmetry center. Incidentally, we stress that, for symmetry testing, (i) the discrete nature of halfspace depth implies that (6.1)-(6.2) can be obtained from a finite number of β-values only; (ii) the bootstrap procedure there can be implemented in practice, since the M bootstrap samples, by symmetry, lead to the same results in Steps 1-3, that therefore need to be performed only once (only Step 4, in which a single depth value is computed, needs to be performed for each bootstrap sample).

Finally, we point out that computing local depth of a fixed point for ℓ distinct β-values typically requires much less time than computing ℓ times local depth for one fixed β-value. One can indeed take advantage of the fact that Step 1 above is common to the various computations of β-local depths (there is some analogy with quantile regression, where the information used to compute a fixed regression quantile may be exploited when computing regression quantiles at other quantile levels).

Acknowledgments

Davy Paindaveine's research is supported by an A.R.C. contract from the Communauté Française de Belgique and by the IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy). Germain Van Bever's research is supported through a Mandat d'Aspirant FNRS (Fonds National pour la Recherche Scientifique), Communauté Française de Belgique. The authors are grateful to three anonymous referees, an Associate Editor, and the Editor Xuming He for their careful reading and insightful comments that led to substantial improvements of the manuscript. They also wish to thank Claudio Agostinelli for stimulating discussions.

A Appendix

This appendix collects proofs of technical results. We start with the proof of Theorem 3.1, which requires the following preliminary result. Throughout this section, $R^{\beta}_{\mathbf{x}}$

will denote $R^{\beta}_{\mathbf{x}}(P)$, when no ambiguity is possible.

Lemma A.1 *Let $D(\cdot, P)$ be a depth function satisfying Property (Q1). Then, for any $\mathbf{x} \in \mathbb{R}^d$, any Borel set $B \subset \mathbb{R}^d$, and any absolutely continuous distribution P, the mapping $\beta \mapsto P[R^{\beta}_{\mathbf{x}} \cap B]$ is continuous over $(0, 1]$.*

Proof of Lemma A.1. Note first that Property (Q1) implies that, for any absolutely continuous P, $\mathbf{x} \mapsto D(\mathbf{x}, P)$ is a continuous function : indeed, if $\mathbf{X}$ is a random d-vector with distribution $P = P^{\mathbf{X}}$, then Property (P1) entails that, for any sequence $\mathbf{x}_n$ converging to $\mathbf{x}$, $|D(\mathbf{x}_n, P) - D(\mathbf{x}, P)| = |D(\mathbf{x}, P^{\mathbf{X}+(\mathbf{x}-\mathbf{x}_n)}) - D(\mathbf{x}, P)| \to 0$ as $n \to \infty$, since $P^{\mathbf{X}+(\mathbf{x}-\mathbf{x}_n)}$ converges weakly to P. Together with the fact that P is absolutely continuous, this implies that $P[R^{\beta}_{\mathbf{x}}(P)] = \beta$ for any $\beta \in (0, 1]$.

Now, fix $\beta_0 \in (0, 1]$ and a Borel set B. Consider a decreasing sequence (β_n) converging to β_0. The numbers $\gamma_n = P[R^{\beta_n}_{\mathbf{x}} \cap B]$ form a monotone decreasing sequence that is lower bounded by $\gamma_0 = P[R^{\beta_0}_{\mathbf{x}} \cap B]$. Hence they admit a limit $\lim_{n\to\infty} \gamma_n \geq \gamma_0$. Letting $\bar{\gamma}_n = P[R^{\beta_n}_{\mathbf{x}} \cap B^c]$, with $B^c = \mathbb{R}^d \setminus B$, we similarly obtain that $\lim_{n\to\infty} \bar{\gamma}_n \geq \bar{\gamma}_0 = P[R^{\beta_0}_{\mathbf{x}} \cap B^c]$. If $\lim_{n\to\infty} \gamma_n > \gamma_0$, then we have $\lim_{n\to\infty} \beta_n = \lim_{n\to\infty}(\gamma_n + \bar{\gamma}_n) > \gamma_0 + \bar{\gamma}_0 = \beta_0$, a contradiction. Hence, we must have that $\lim_{n\to\infty} \gamma_n = \gamma_0$, i.e., that $\beta \mapsto P[R^{\beta}_{\mathbf{x}} \cap B]$ is right continuous at β_0. The result then follows since left continuity can be established along the same lines. □

Proof of Theorem 3.1. In view of (Q1), it is sufficient, in order to show that

$$\left|LD^{\beta_n}(\mathbf{x}, P^{(n)}) - LD^{\beta}(\mathbf{x}, P)\right| = \left|D(\mathbf{x}, P^{\beta_n,(n)}_{\mathbf{x}}) - D(\mathbf{x}, P^{\beta}_{\mathbf{x}})\right| \overset{a.s.}{\to} 0 \quad \text{as } n \to \infty,$$

to prove that $P^{\beta_n,(n)}_{\mathbf{x}}[B] \overset{a.s.}{\to} P^{\beta}_{\mathbf{x}}[B]$ for any Borel set B. Fix then such a B and $\varepsilon > 0$. Lemma A.1 implies that there exist $\delta, \eta > 0$ such that

$$[P[R^{\beta-\delta}_{\mathbf{x}} \cap B] - \eta, P[R^{\beta+\delta}_{\mathbf{x}} \cap B] + \eta] \subset [P[R^{\beta}_{\mathbf{x}} \cap B] - \beta\varepsilon, P[R^{\beta}_{\mathbf{x}} \cap B] + \beta\varepsilon]. \quad \text{(A.1)}$$

Now, Theorem 3 in Zuo and Serfling (2000c) implies that there exists n_0 such that $R_{\mathbf{x}}^{\beta-\delta} \subset R_{\mathbf{x}}^{\beta_n,(n)} \subset R_{\mathbf{x}}^{\beta+\delta}$ a.s. for all $n \geq n_0$ (throughout the proof, $R_{\mathbf{x}}^{\beta_n,(n)}$ stands for $R^{\beta_n}(P_{\mathbf{x}}^{(n)})$), which of course yields that, a.s. for all $n \geq n_0$,

$$P^{(n)}[R_{\mathbf{x}}^{\beta-\delta} \cap B] \leq P^{(n)}[R_{\mathbf{x}}^{\beta_n,(n)} \cap B] \leq P^{(n)}[R_{\mathbf{x}}^{\beta+\delta} \cap B], \tag{A.2}$$

The SLLN entails that $P^{(n)}[R_{\mathbf{x}}^{\beta\pm\delta} \cap B] \overset{a.s.}{\to} P[R_{\mathbf{x}}^{\beta\pm\delta} \cap B]$ as $n \to \infty$; consequently, there exists n_1 such that, a.s. for all $n \geq n_1$,

$$\big[P^{(n)}[R_{\mathbf{x}}^{\beta-\delta} \cap B], P^{(n)}[R_{\mathbf{x}}^{\beta+\delta} \cap B]\big] \subset \big[P[R_{\mathbf{x}}^{\beta-\delta} \cap B] - \eta, P[R_{\mathbf{x}}^{\beta+\delta} \cap B] + \eta\big]. \tag{A.3}$$

Combining (A.1)-(A.3), we proved that, a.s. for all $n \geq \max(n_0, n_1)$,

$$P[R_{\mathbf{x}}^{\beta} \cap B] - \beta\varepsilon \leq P^{(n)}[R_{\mathbf{x}}^{\beta_n,(n)} \cap B] \leq P[R_{\mathbf{x}}^{\beta} \cap B] + \beta\varepsilon,$$

or equivalently, $P_{\mathbf{x}}^{\beta}[B] - \varepsilon \leq \frac{1}{\beta} P^{(n)}[R_{\mathbf{x}}^{\beta_n,(n)} \cap B] \leq P_{\mathbf{x}}^{\beta}[B] + \varepsilon$. In other words, we have proved that, as $n \to \infty$,

$$\frac{1}{\beta}\, P^{(n)}[R_{\mathbf{x}}^{\beta_n,(n)} \cap B] \overset{a.s.}{\to} P_{\mathbf{x}}^{\beta}[B]. \tag{A.4}$$

Taking $B = \mathbb{R}^d$ in (A.4) yields $P^{(n)}[R_{\mathbf{x}}^{\beta_n,(n)}] \overset{a.s.}{\to} \beta$, which, jointly with (A.4), establishes that $P_{\mathbf{x}}^{\beta_n,(n)}[B] = P^{(n)}[B|R_{\mathbf{x}}^{\beta_n,(n)}] \overset{a.s.}{\to} P_{\mathbf{x}}^{\beta}[B]$, as was to be proved. □

Proof of Lemma 4.1. (i) Fix $\mathbf{x} \in \mathrm{Supp}(f)$ and $\varepsilon > 0$. By Lemma A.1 in Paindaveine and Van Bever (2012) (whose proof, under the properties (Q1)-(Q2) introduced in the present paper, trivially extends to the case where the symmetry center $\boldsymbol{\theta}$ belongs to $\mathrm{Supp}(f) \setminus \mathrm{Supp}_+(f)$), there exist $\delta > 0$ and $\alpha < \alpha_{\mathbf{x}}^* := \max_{\mathbf{y}\in\mathbb{R}^d} D(\mathbf{y}, P_{\mathbf{x}})$ such that $B_{\mathbf{x}}(\delta) \subset R_{\mathbf{x},\alpha} \subset B_{\mathbf{x}}(\varepsilon)$. Since $\mathbf{x} \in \mathrm{Supp}(f)$, we then have that $\beta_0 := P[R_{\mathbf{x},\alpha}] \geq P[B_{\mathbf{x}}(\delta)] > 0$. From the definition of $R_{\mathbf{x}}^{\beta_0}$, it follows that $R_{\mathbf{x}}^{\beta_0} \subset R_{\mathbf{x},\alpha} \subset B_{\mathbf{x}}(\varepsilon)$.

(ii) Fix $\mathbf{x} \notin \mathrm{Supp}(f)$ and let $\varepsilon > 0$ be such that $P[B_{\mathbf{x}}(\varepsilon)] = 0$. If one assumes that (Q2) also holds for $\boldsymbol{\theta} \notin \mathrm{Supp}(f)$, then it is easy to check that the proof of Lemma

A.1(i) in Paindaveine and Van Bever (2012) further extends to the case where the symmetry center does not belong to $\text{Supp}(f)$. Therefore there still exist $\delta > 0$ and $\alpha < \alpha^*_{\mathbf{x}}$ such that $B_{\mathbf{x}}(\delta) \subset R_{\mathbf{x},\alpha} \subset B_{\mathbf{x}}(\varepsilon)$. The definition of $R^\beta_{\mathbf{x}}$ implies that $R_{\mathbf{x},\alpha} \subset R^\beta_{\mathbf{x}}$ for any $\beta > 0$. It follows that $\mathbf{x} \in B_{\mathbf{x}}(\delta) \subset R_{\mathbf{x},\alpha} \subset R^0_{\mathbf{x}} = \cap_{\beta>0} R^\beta_{\mathbf{x}}$, hence that $\mathbf{x}$ is an interior point of $R^0_{\mathbf{x}}$. □

Lemma A.2 *Under the assumptions of Theorem 4.1, $\beta/\text{Vol}(R^\beta_{\mathbf{x}}) \to f(\mathbf{x})$ as $\beta \to 0$.*

Proof of Lemma A.2. Fix $\varepsilon > 0$ and let $r = r(\varepsilon)$ be such that $f(\mathbf{x}) - \varepsilon \leq f(\mathbf{y}) \leq f(\mathbf{x}) + \varepsilon$ for any $\mathbf{y} \in B_{\mathbf{x}}(r) = \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z} - \mathbf{x}\| < r\}$. Lemma A.1 in Paindaveine and Van Bever (2012) ensures that there exists $\beta_0 > 0$ such that $R^{\beta_0}_{\mathbf{x}} \subset B_{\mathbf{x}}(r)$. Therefore, for any $\beta \in (0, \beta_0)$, one has $(f(\mathbf{x}) - \varepsilon)\text{Vol}(R^\beta_{\mathbf{x}}) \leq \beta = \int_{R^\beta_{\mathbf{x}}} f(\mathbf{y}) d\mathbf{y} \leq (f(\mathbf{x}) + \varepsilon)\text{Vol}(R^\beta_{\mathbf{x}})$, or equivalently $f(\mathbf{x}) - \varepsilon \leq \beta/\text{Vol}(R^\beta_{\mathbf{x}}) \leq f(\mathbf{x}) + \varepsilon$. The result follows. □

Proof of Theorem 4.1. Fix $\mathbf{x} \in \mathbb{R}^d$ such that f is positive and continuous at $\mathbf{x}$. For any β, let $B \mapsto P^{\text{sym},\beta}_{\mathbf{x}}[B] = P_{\mathbf{x}}[B|R^\beta_{\mathbf{x}}]$ be the symmetrized (about $\mathbf{x}$) version of P, conditional to $R^\beta_{\mathbf{x}}$—recall that we let $P_{\mathbf{x}} = \frac{1}{2}P^{\mathbf{x}} + \frac{1}{2}P^{2\mathbf{x}-\mathbf{x}}$. We have

$$|LD^\beta(\mathbf{x}, P) - c_D| = |D(\mathbf{x}, P^\beta_{\mathbf{x}}) - c_D| = |D(\mathbf{x}, P^\beta_{\mathbf{x}}) - D(\mathbf{x}, P^{\text{sym},\beta}_{\mathbf{x}})|,$$

where we used the fact that $D(\cdot, P)$ satisfies (Q3). In view of (Q1$^+$), it is therefore sufficient to prove that, for any Borel set B, $P^\beta_{\mathbf{x}}[B] - P^{\text{sym},\beta}_{\mathbf{x}}[B] \to 0$ as $\beta \to 0$. To do so, fix such a B and, denoting by $f^{\text{sym}}_{\mathbf{x}}$ the density of $P^{\text{sym}}_{\mathbf{x}}$, write

$$P^\beta_{\mathbf{x}}[B] - P^{\text{sym},\beta}_{\mathbf{x}}[B] = \frac{1}{\beta}\int_{R^\beta_{\mathbf{x}} \cap B} (f(\mathbf{y}) - f_{\mathbf{x}}(\mathbf{y}))\, d\mathbf{y} = \frac{1}{2\beta}\int_{R^\beta_{\mathbf{x}} \cap B} (f(\mathbf{y}) - f(2\mathbf{x} - \mathbf{y}))\, d\mathbf{y}.$$

If $\mathbf{x}$ lies in the interior of B, Lemma 4.1(i) shows that there exists $\beta_0 > 0$ such that, for all $\beta \leq \beta_0$, we have $R^\beta_{\mathbf{x}} \cap B = R^\beta_{\mathbf{x}}$. Clearly, this implies that for all $\beta \leq \beta_0$, the integral above, hence also $P^\beta_{\mathbf{x}}[B] - P^{\text{sym},\beta}_{\mathbf{x}}[B]$, is equal to zero. If $\mathbf{x}$ does not belong to the closure of B, then the same lemma implies that $R^\beta_{\mathbf{x}} \cap B$ is empty for β small

enough, which leads to the same conclusion. It remains to consider the case where $\mathbf{x}$ belongs to the boundary of B. For such an $\mathbf{x}$, we may write

$$P^{\beta}_{\mathbf{x}}[B] - P^{\mathrm{sym},\beta}_{\mathbf{x}}[B] \;=\; \frac{\mathrm{Vol}(R^{\beta}_{\mathbf{x}} \cap B)}{2\beta}\,(\mathcal{I}_{\beta} - \mathcal{I}^{\mathrm{refl}}_{\beta}),\ \text{where}$$

$$\mathcal{I}_{\beta} = \frac{1}{\mathrm{Vol}(R^{\beta}_{\mathbf{x}} \cap B)} \int_{R^{\beta}_{\mathbf{x}} \cap B} f(\mathbf{y})\, d\mathbf{y} \ \text{ and } \ \mathcal{I}^{\mathrm{refl}}_{\beta} = \frac{1}{\mathrm{Vol}(R^{\beta}_{\mathbf{x}} \cap B^{\mathrm{refl}})} \int_{R^{\beta}_{\mathbf{x}} \cap B^{\mathrm{refl}}} f(\mathbf{y})\, d\mathbf{y},$$

and where $B^{\mathrm{refl}} = 2\mathbf{x} - B$ denotes the reflection of B about $\mathbf{x}$. The same reasoning as in the proof of Lemma A.2 allows to show that both $\mathcal{I}_{\beta}$ and $\mathcal{I}^{\mathrm{refl}}_{\beta}$ converge to $f(\mathbf{x})$ as $\beta \to 0$. The result then follows from the fact that $\mathrm{Vol}(R^{\beta}_{\mathbf{x}} \cap B)/\beta \leq \mathrm{Vol}(R^{\beta}_{\mathbf{x}})/\beta$ remains bounded as $\beta \to 0$ (Lemma A.2). □

Proof of Theorem 6.1. (Necessity:) For any β, the region $R^{\beta}_{\mathbf{x}_0}(P)$ is centrally symmetric about $\mathbf{x}_0$: $R^{\beta}_{\mathbf{x}_0}(P) = 2\mathbf{x}_0 - R^{\beta}_{\mathbf{x}_0}(P)$. Hence the central symmetry of P about $\mathbf{x}_0$ implies that $P^{\beta}_{\mathbf{x}_0}[\,\cdot\,] = P[\,\cdot\,|R^{\beta}_{\mathbf{x}_0}(P)]$ is also centrally symmetric about $\mathbf{x}_0$. This implies that $LD^{\beta}_{H}(\mathbf{x}_0, P) = D_H(\mathbf{x}_0, P^{\beta}_{\mathbf{x}_0}) = 1/2$ for any β.

(Sufficiency:) For any β, $LD^{\beta}_{H}(\mathbf{x}_0, P) = D_H(\mathbf{x}_0, P^{\beta}_{\mathbf{x}_0}) = 1/2$ implies that $P^{\beta}_{\mathbf{x}_0}$ is angularly symmetric about $\mathbf{x}_0$. In other words, for any β, $P^{\beta}_{\mathbf{x}_0}[C] = P^{\beta}_{\mathbf{x}_0}[2\mathbf{x}_0 - C]$, for any C in the collection $\mathcal{C}_{\mathbf{x}_0}$ of cones originating from $\mathbf{x}_0$. This of course rewrites $P[C \cap R^{\beta}_{\mathbf{x}_0}(P)] = P[(2\mathbf{x}_0 - C) \cap R^{\beta}_{\mathbf{x}_0}(P)]$, $\forall C \in \mathcal{C}_{\mathbf{x}_0}$, $\forall \beta \in (0, 1]$. Since the regions $R^{\beta}_{\mathbf{x}_0}(P)$ are symmetric with respect to $\mathbf{x}_0$, this implies that

$$P[C \cap (R^{\beta_2}_{\mathbf{x}_0}(P) \setminus R^{\beta_1}_{\mathbf{x}_0}(P))]$$
$$= P[2\mathbf{x}_0 - (C \cap (R^{\beta_2}_{\mathbf{x}_0}(P) \setminus R^{\beta_1}_{\mathbf{x}_0}(P)))] \quad \forall C \in \mathcal{C}_{\mathbf{x}_0},\ \forall \beta_1 < \beta_2 \in (0, 1].$$

This proves the result since the sigma-algebra generated by the subsets $C \cap (R^{\beta_2}_{\mathbf{x}_0} \setminus R^{\beta_1}_{\mathbf{x}_0})$, $C \in \mathcal{C}_{\mathbf{x}_0}$, $0 < \beta_1 < \beta_2 \leq 1$, coincides with the Borel sigma-algebra on $\mathbb{R}^d$. □

Supplementary Materials

The Supplementary Materials report (i) a comparison with the local halfspace and simplicial depths from Agostinelli and Romanazzi (2011) in the context of the Boston

data set in Section 2.2, (ii) an argument proving that $P^0_{\mathbf{x}}$ (see (4.1)) does not exist for $\mathbf{x} \in \text{Supp}(f)$, (iii) an example showing that a point $\mathbf{x}$ on the boundary of $\text{Supp}(f)$ can exhibit any limiting local depth as $\beta \to 0$, and (iv) a proof of Theorem 4.2.

References

Agostinelli, C. and Romanazzi, M. (2011), "Local depth," *J. Statist. Plann. Inference*, 141, 817–830.

Chaudhuri, P. (1996), "On a Geometric Notion of Quantiles for Multivariate Data," *J. Amer. Statist. Assoc.*, 91, 862–872.

Chen, Y., Dang, X., Peng, H., and Bart, H. L. J. (2009), "Outlier detection with the kernelized spatial depth function," *IEEE Trans. Patt. Anal Mach. Int.*, 31, 288–305.

Dümbgen, L. (1992), "Limit theorems for the simplicial depth," *Statist. Probab. Lett.*, 14, 119–128.

Dutta, S. and Ghosh, A. K. (2012), "On robust classification using projection depth," *Ann. Inst. Statist. Math.*, 64, 657–676.

Dutta, S., Ghosh, A. K., and Chaudhuri, P. (2011), "Some intriguing properties of Tukey's half-space depth," *Bernoulli*, 17, 1420–1434.

Ghosh, A. K. and Chaudhuri, P. (2005), "On maximum depth and related classifiers," *Scand. J. Statist.*, 32, 327–350.

Hallin, M., Paindaveine, D., and Šiman, M. (2010), "Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth (with discussion)," *Ann. Statist.*, 38, 635–669.

Härdle, W. (1991), *Smoothing techniques*, Springer Series in Statistics, New York:

Springer-Verlag.

Harrison, D. and Rubinfeld, D. (1978), "Hedonic prices and the demand for clean air," *J. Environ. Econ. Manage.*, 5, 81–102.

He, X. and Wang, G. (1997), "Convergence of depth contours for multivariate datasets," *Ann. Statist.*, 25, 495–504.

Hlubinka, D., Kotík, L., and Vencálek, O. (2010), "Weighted halfspace depth," *Kybernetika*, 46, 125–148.

Izem, R., Rafalin, E., and Souvaine, D. L. (2008), "Describing multivariate distributions with nonlinear variation using data depth," Technical Report TR-2006-2, Tufts University, Department of Computer Science.

Kong, L. and Zuo, Y. (2010), "Smooth depth contours characterize the underlying distribution," *J. Multivariate Anal.*, 101, 2222–2226.

Koshevoy, G. and Mosler, K. (1997), "Zonoid trimming for multivariate distributions," *Ann. Statist.*, 25, 1998–2017.

Li, J., Cuesta-Albertos, J., and Liu, R. Y. (2012), "DD-Classifier: Nonparametric classification procedures based on DD-plots." *J. Amer. Statist. Assoc.*, 107, 737–753.

Li, J. and Liu, R. Y. (2004), "New nonparametric tests of multivariate locations and scales using data depth," *Statist. Sci.*, 19, 686–696.

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

Liu, R. Y., Parelius, J. M., and Singh, K. (1999), "Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion)," *Ann. Statist.*, 27, 783–858.

Liu, X. and Zuo, Y. (2011a), "Computing halfspace depth and regression depth," *Communications in Statistics, Simulations and Computation*.

— (2011b), "Computing projection depth and its associated estimators," *Statistics and Computing*.

Liu, X., Zuo, Y., and Wang, Z. (2011), "Exactly computing bivariate projection depth contours and median," *Comput. Statist. Data Anal.*

Lok, W. and Lee, S. M. S. (2011), "A new statistical depth function with applications to multimodal data," *J. Nonparametr. Stat.*, 23, 617–631.

McLachlan, G. and Peel, D. (2000), *Finite mixture models*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York.

McLachlan, G. J. and Basford, K. E. (1988), *Mixture models*, vol. 84 of *Statistics: Textbooks and Monographs*, New York: Marcel Dekker Inc.

Mizera, I. and Volauf, M. (2002), "Continuity of halfspace depth contours and maximum depth estimators: diagnostics of depth-related methods," *J. Multivariate Anal.*, 83, 365–388.

Paindaveine, D. and Van Bever, G. (2012), "Nonparametrically consistent depth-based classifiers," Unpublished.

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

Rousseeuw, P. J. and Struyf, A. (2002), "A depth test for symmetry," in *Goodness-of-fit tests and model validity (Paris, 2000)*, Boston, MA: Birkhäuser Boston, Stat. Ind. Technol., pp. 401–412.

— (2004), "Characterizing angular symmetry and regression symmetry," *J. Statist. Plann. Inference*, 122, 161–173.

Serfling, R. (2004), "Nonparametric multivariate descriptive measures based on spatial quantiles," *J. Statist. Plann. Inference*, 123, 259–278.

— (2010), "Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization," *J. Nonparametr. Stat.*, 22, 915–926.

Tukey, J. W. (1975), "Mathematics and the picturing of data," in *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*, Canad. Math. Congress, Montreal, Que., pp. 523–531.

Wei, Y. (2008), "An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts," *J. Amer. Statist. Assoc.*, 103.

Zuo, Y. (1998), *Contributions to the theory and applications of statistical depth functions*, ProQuest LLC, Ann Arbor, MI, thesis (Ph.D.)–The University of Texas at Dallas.

— (2003), "Projection-based depth functions and associated medians," *Ann. Statist.*, 31, 1460–1490.

Zuo, Y. and Serfling, R. (2000a), "General notions of statistical depth function," *Ann. Statist.*, 28, 461–482.

— (2000b), "On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry," *J. Statist. Plann. Inference*, 84, 55–79.

— (2000c), "Structural properties and convergence results for contours of sample statistical depth functions," *Ann. Statist.*, 28, 483–499.