Supplementary Materials for *From Depth to Local Depth : A Focus on Centrality*

Davy Paindaveine* and Germain Van Bever

Université libre de Bruxelles, Brussels, Belgium

Abstract

These Supplementary Materials report (i) a comparison with the local half-space and simplicial depths from Agostinelli and Romanazzi (2011) in the context of the Boston data set treated in Section 2.2 of the manuscript, (ii) an argument proving that $P_{\mathbf{x}}^{0}$ (see (4.1) in the manuscript) does not exist for $\mathbf{x} \in \mathrm{Supp}(f)$, (iii) an example showing that a point $\mathbf{x}$ on the boundary of $\mathrm{Supp}(f)$ can exhibit any limiting local depth as $\beta \to 0$, and (iv) a proof of Theorem 4.2.

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 More on Boston data

We consider again the *Boston* data set treated in Section 2.2 of the manuscript, and still focus on the same 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 upper panels of Figure 1 below coincide with the two panels of Figure 2 in the manuscript : the upper left panel shows a scatter plot of the 506 bivariate data points, along with the four particular locations considered in Section 2.2, and the upper right panel shows the plots of the proposed local (halfspace and simplicial) depths $\beta \mapsto LD_i^\beta(\mathbf{x}, P^{(n)})$, $i = H, S$, where $\mathbf{x}$ is any of those four locations.

For the sake of comparison, the lower left panel of Figure 1 provides scaled versions of the Agostinelli and Romanazzi (2011) local halfspace and simplicial depths of the four locations, as a function of the locality level τ in some appropriate range. More precisely,

- for halfspace depth, the maximum value $\tau_{\max}$ of τ was chosen as the minimal τ-value for which the τ-local depths of the four locations all coincide with the corresponding global halfspace depths ;
- for simplicial depth, the τ-values at which the τ-local depths were evaluated are, as suggested in Agostinelli and Romanazzi (2011), the percentiles of the volumes of the $\binom{n}{3}$ data-based simplices, which also ensures that global depth is obtained for the largest τ considered.

Scaling was performed in such a way that, at any fixed τ, the largest τ-local (halfspace and simplicial) depths considered are equal to one (this still allows to investigate, for any fixed τ, the corresponding—halfspace and simplicial—depth rankings, and was done because those local depths are hardly comparable for different τ-values).

It is seen that the resulting local depth rankings depend much more on the choice of (halfspace or simplicial) depth than for the proposed local depths (particularly so for the green location). For halfspace depth, the red point remains the deepest for most τ-values ; it actually is so for all τ-values in the lower right panel of Figure 1, that reports the corresponding local depths after a unit change expressing the DIS variable in yards (this consists in multiplying DIS by 1760, but the results are similar for much smaller factors).

The particular τ-indexing used for local simplicial depth makes it affine-invariant, but local halfspace depth fails to be affine-invariant, irrespective of the τ-indexing used ; the unit change considered, unpleasantly, has then a strong impact on the local halfspace depth from Agostinelli and Romanazzi (2011) (the τ-local halfspace depth of the green location now dramatically decreases for small τ-values, and, as mentioned above, the red location remains, for all τ, the halfspace deepest point among the four locations). In contrast, our local depths are affine-invariant, hence are not affected by any unit change.

2 More on extreme localization

In this section, we provide further results on extreme localization. We first prove that, for $\mathbf{x} \in \mathrm{Supp}(f)$, the limiting distribution $P^0_{\mathbf{x}}$ in (4.1) of the manuscript does not exist (Section 2.1). Then we provide an example showing that a point $\mathbf{x}$ on the boundary of $\mathrm{Supp}(f)$ can exhibit any limiting local depth as $\beta \to 0$ (Section 2.2). Finally, we prove Theorem 4.2 (Section 2.3).

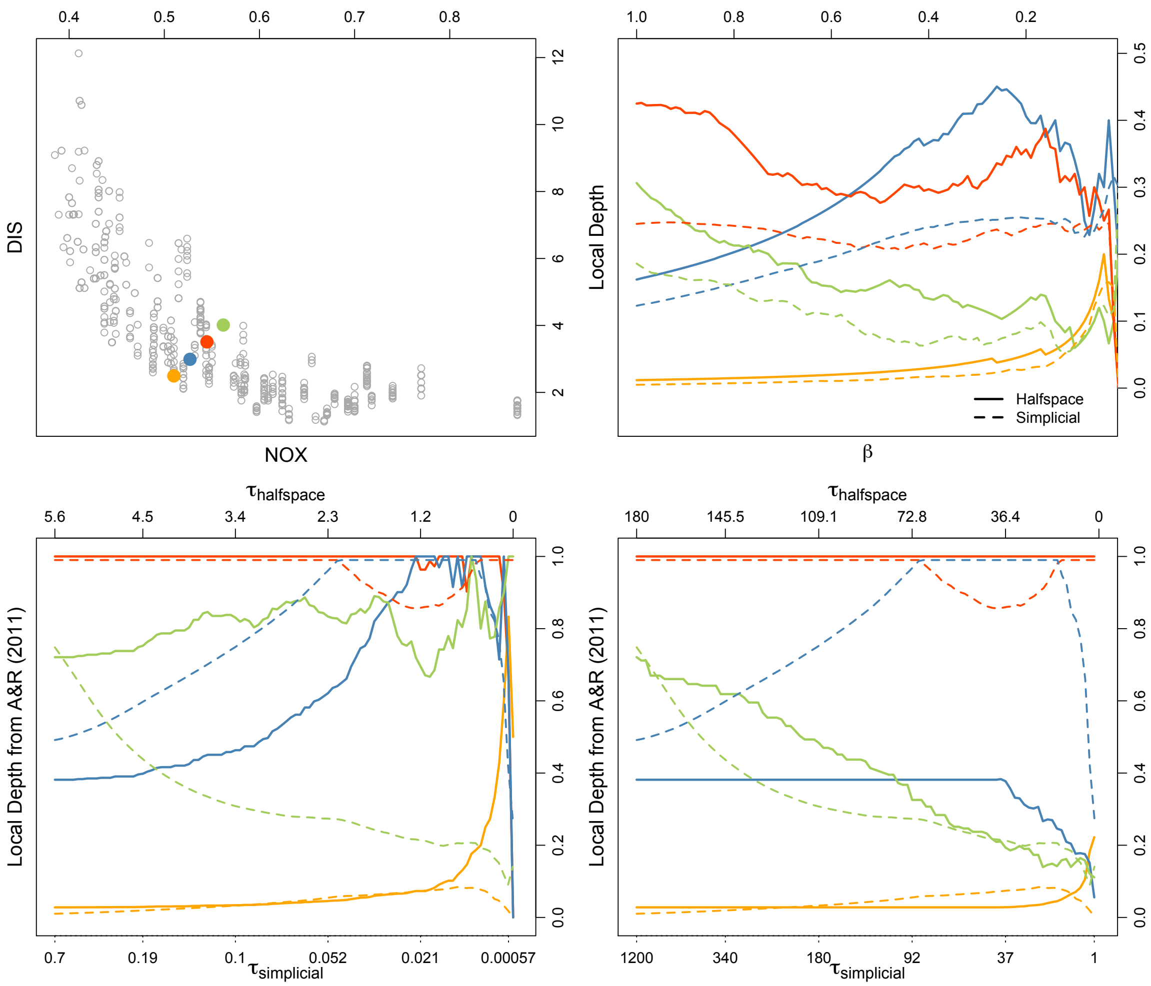

Figure 1: (Upper left:) Scatterplot of the NOX and DIS variables from the Boston data set, with four particular locations. (Upper right:) Plots, as a function of the locality level β, of the proposed local halfspace (solid curves) and simplicial (dashed curves) depths of these locations. (Lower left:) scaled versions of the corresponding Agostinelli and Romanazzi (2011) local depths; see Section 1 for details. (Lower right:) The same curves as in the lower left panel, when expressing the DIS variable in yards (such unit changes do not affect the local depths in the upper right panel).

2.1 Non-existence of $P^0_{\mathbf{x}}$ for $\mathbf{x} \in \mathbf{Supp}(f)$

To prove the non-existence of $P^0_{\mathbf{x}}$, let us reach a contradiction by assuming that it does exist. Note first that, from Lemma 4.1(i), it is clear that any open halfspace that does not contain $\mathbf{x}$ needs to have $P^0_{\mathbf{x}}$-probability zero, which implies that an open halfspace H having $\mathbf{x}$ on its boundary should also receive $P^0_{\mathbf{x}}$-probability zero. However, a direct computation—along the same lines as in the proof of Theorem 4.1 of the manuscript—rather provides that

$$P^0_{\mathbf{x}}[H] = \lim_{\beta \to 0} P^\beta_{\mathbf{x}}[H] = \lim_{\beta \to 0} \frac{\mathrm{Vol}(R^\beta_{\mathbf{x}}(P) \cap H)}{\mathrm{Vol}(R^\beta_{\mathbf{x}}(P))} = \frac{1}{2},$$

a contradiction.

2.2 Local depth on the boundary of the support

We here provide a bivariate example showing that a point $\mathbf{x}$ on 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 of the depth D (as in Section 4 of the manuscript, we assume that (Q1)-(Q3) hold).

For any $\eta \in (0, \pi)$, let $P = P_\eta$ be the uniform distribution on the unit disk $B_{\mathbf{0}}(1)$ (centered at the origin $\mathbf{0} = \binom{0}{0}$) deprived from a sector with radius $1/2$ and angle η, that is, more precisely, the uniform distribution on the set

$$C_\eta := B_{\mathbf{0}}(1) \setminus \left\{ \mathbf{x} = \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix} \in B_{\mathbf{0}}(1/2) : r < 1/2 \text{ and } \varphi \in [-\pi, \pi] \setminus [-\eta/2, \eta/2] \right\};$$

see Figure 2. For any $\eta \in (0, \pi)$, the origin lies on the boundary of the support of P_η, and one can show that $\ell_\eta := \lim_{\beta \to 0} LD^\beta(\mathbf{0}, P_\eta)$ ranges from $\ell_0 = \lim_{\eta \overset{>}{\to} 0} \ell_\eta = c_D$ to $\ell_\pi = \lim_{\eta \overset{<}{\to} \pi} \ell_\eta = 0$. This confirms that points on the boundary of the support, for extreme localization, may receive arbitrarily small or arbitrarily large local depths. This is far from being undesirable, though, and, in this example, is perfectly translating the obvious fact that (extreme) local centrality of the origin is a decreasing function of η. Note that the global depths $D(\mathbf{0}, P_\eta)$, $\eta \in (0, \pi)$, remain bounded away from zero.

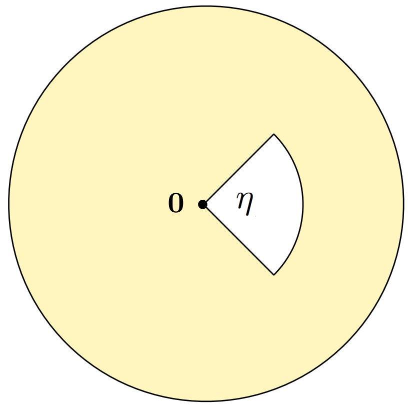

Figure 2: Support C_η (in yellow) of the uniform distribution considered in Section 2.2.

In contrast, the univariate local (halfspace or simplicial) depths typically do not allow for arbitrary values on the boundary of the support. Indeed, it can easily be shown that, in the univariate case, $\lim_{\beta\to 0} LD_H^\beta(x, P) = 0 = \lim_{\beta\to 0} LD_S^\beta(x, P)$ as soon as, for some $\varepsilon > 0$, f is continuous in $(x - \varepsilon, x + \varepsilon) \setminus \{x\}$.

2.3 Proof of Theorem 4.2

Finally, we prove Theorem 4.2 of the manuscript, which requires Lemma 1 below. Let X be an absolutely continuous random variable with cdf F and pdf f, and put $g(\beta) := (F(x) - F(x_\beta))/\beta$ and $h(\beta) := (F(2x - x_\beta) - F(x))/\beta$, where x_β was defined in the statement of Proposition 5.1.

Lemma 1 *Fix $x \in \mathrm{Supp}_+(f)$. (i) If f is continuous in a neighborhood of x, then (i) $\lim_{\beta\to 0} g(\beta) = \frac{1}{2} = \lim_{\beta\to 0} h(\beta)$; (ii) if f admits a continuous derivative f' in a neighborhood of x, then $\lim_{\beta\to 0} g'(\beta) = -\frac{f'(x)}{8f^2(x)}$ and $\lim_{\beta\to 0} h'(\beta) = \frac{f'(x)}{8f^2(x)}$; if f admits a continuous second derivative f'' in a neighborhood of x, then (iii) $\lim_{\beta\to 0} h''(\beta) = 0$ and $\lim_{\beta\to 0} g''(\beta) = 0$.*

Proof of Lemma 1. First note that x_β is the $(1-\beta)/2$-quantile of the symmetrized distribution of X about x, that is, $x_\beta = (F^Y)^{-1}(\frac{1-\beta}{2})$, where F^Y is the cdf $\frac{1}{2}F + \frac{1}{2}F^{2x-X}$. Below, the corresponding pdf will be denoted f^Y.

(i) Absolute continuity implies that $\lim_{\beta\to 0} x_\beta = x$. Therefore, L'Hôpital's rule can be applied and, together with the continuity of f in a neighborhood of x and the expression for x_β given above, we obtain

$$\lim_{\beta\to 0} g(\beta) = \lim_{\beta\to 0}\Big(-f(x_\beta)\,\frac{dx_\beta}{d\beta}\Big) = \lim_{\beta\to 0}\frac{f(x_\beta)}{2f^Y(x_\beta)}.$$

Since $f^Y(x) = f(x)$, the result follows for g. Computations for $h(\beta)$ are extremely similar, hence will be omitted here (as well as in the proof of (ii)-(iii) below).

(ii) Straightforward calculus shows that $g'(\beta) = (\frac{f(x_\beta)}{2f^Y(x_\beta)} - g(\beta))/\beta$. Taking the limit of $g'(\beta)$ and applying L'Hôpital's rule gives

$$\lim_{\beta\to 0} g'(\beta) = \lim_{\beta\to 0}\frac{d}{d\beta}\left[\frac{f(x_\beta)}{2f^Y(x_\beta)}\right] - \lim_{\beta\to 0} g'(\beta).$$

The result then follows after some calculations using that $(f^Y)'$ is continuous in a neighborhood of x and takes value zero at x.

(iii) Tedious calculations show that

$$g''(\beta) = \frac{1}{\beta}\left\{\frac{-f'(x_\beta) + f(x_\beta)(f^Y)'(x_\beta)\left(f^Y(x_\beta)\right)^{-1}}{4\left(f^Y(x_\beta)\right)^2} - 2g'(\beta)\right\}.$$

L'Hôpital's rule then establishes the result, after some derivations using that $(f^Y)''$ is continuous in a neighborhood of x and takes value $f''(x)$ at x. $\square$

We can now prove Theorem 4.2 of the manuscript.

Proof of Theorem 4.2. Under the assumptions considered, it is clearly sufficient to prove that, as $\beta \to 0$,

(a) $LD_H^\beta(x,P) \to \frac{1}{2}$ and $\frac{\partial}{\partial\beta}LD_H^\beta(x,P) \to -\frac{|f'(x)|}{8f^2(x)}$;

(b) $LD_S^\beta(x,P) \to \frac{1}{2}$, $\frac{\partial}{\partial\beta}LD_S^\beta(x,P) \to 0$, and $\frac{\partial^2}{\partial\beta^2}LD_S^\beta(x,P) \to -\frac{(f'(x))^2}{16f^4(x)}$.

Proposition 5.1 shows that $LD_H^\beta = \min(g(\beta), h(\beta))$. A simple Taylor expansion then yields

$$\lim_{\beta \to 0} LD_H^\beta(x, P) = \begin{cases} \lim_{\beta \to 0} g'(\beta) & \text{if } f'(x) > 0 \\ \lim_{\beta \to 0} h'(\beta) & \text{if } f'(x) < 0 \\ \lim_{\beta \to 0} \min(g'(\beta), h'(\beta)) & \text{if } f'(x) = 0 \end{cases}$$

Lemma 1 then directly establishes (a). In order to prove (b), note that Proposition 5.1 states that $LD_S^\beta(x, P) = 2g(\beta)h(\beta)$, which yields $\frac{\partial}{\partial\beta} LD_S^\beta(x, P) = 2g'(\beta)h(\beta) + 2g(\beta)h'(\beta)$ and $\frac{\partial^2}{\partial\beta^2} LD_S^\beta(x, P) = 2g''(\beta)h(\beta) + 4g'(\beta)h'(\beta) + 2g(\beta)h''(\beta)$. The limits in (b) then follow from Lemma 1. $\square$

References

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