Data Depth

Stanislav Nagy

Department of Probability and Mathematical Statistics
Charles University in Prague

6.12.2012

This work was supported by Czech Science Foundation (grant P402/12/G097)

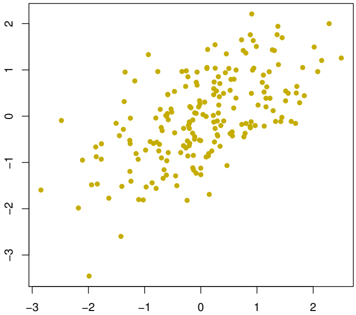
1. Depth Measure and its Smoothness for Multivariate Data
   - Smoothness of Halfspace Depth Contours
2. Functional Data Depth: Theory
   - Functional Band Depths
   - Consistency
   - Counterexample
   - Fixing the Continuousness
   - Integral and Vector Depths
3. Functional Data Depth: Practice
   - Problem of Functional Data Classification
   - Using Depth for Classification
   - Simulation Study
4. Conclusions

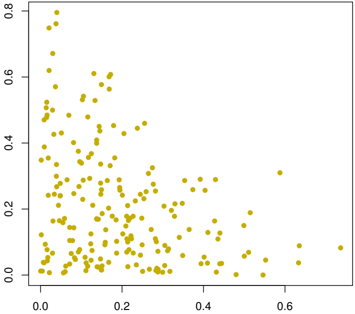
Outline

1. Depth Measure and its Smoothness for Multivariate Data
   - Smoothness of Halfspace Depth Contours
2. Functional Data Depth: Theory
   - Functional Band Depths
   - Consistency
   - Counterexample
   - Fixing the Continuousness
   - Integral and Vector Depths
3. Functional Data Depth: Practice
   - Problem of Functional Data Classification
   - Using Depth for Classification
   - Simulation Study
4. Conclusions

Data Depth

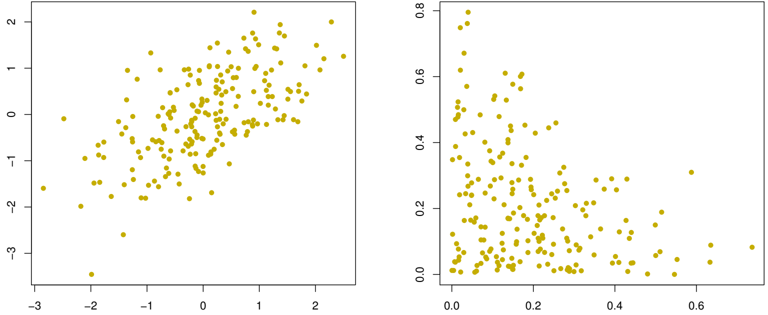
Consider a random variable $X \sim P \in \mathcal{P}\left(\mathbb{R}^d\right)$.





Data Depth

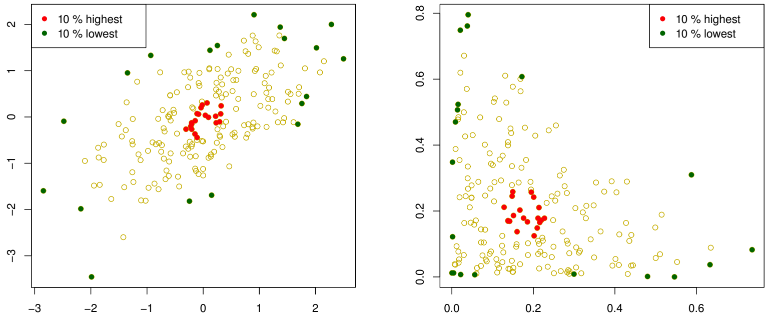
Consider a random variable $X \sim P \in \mathcal{P}\left(\mathbb{R}^d\right)$.



How to **define ordering** of these data?

Data Depth

Consider a random variable $X \sim P \in \mathcal{P}\left(\mathbb{R}^d\right)$.



Using **data depth!**

Depth Generally

According to Zuo and Serfling [13], **Statistical depth** is a function possessing:

- affine transformation **invariance**
- **maximality at the center of symmetry** of the distribution for the class of symmetric distributions
- **monotonicity** relative to the point with the highest depth
- **vanishing** at infinity

Depth Generally

According to Zuo and Serfling [13], **Statistical depth** is a function possessing:

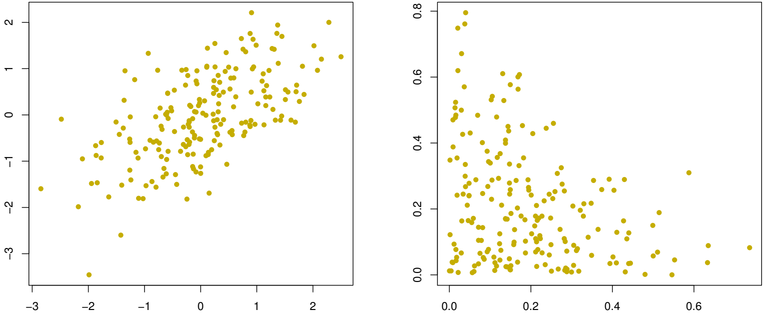
- affine transformation **invariance**
- **maximality at the center of symmetry** of the distribution for the class of symmetric distributions
- **monotonicity** relative to the point with the highest depth
- **vanishing** at infinity

We obtain a function recognizing “typical” and “outlier” observations, **a generalization of quantiles** for multivariate data.

Halfspace Depth

Halfspace depth (Tukey [11]) *HD* of an observation from $\mathbb{R}^d$

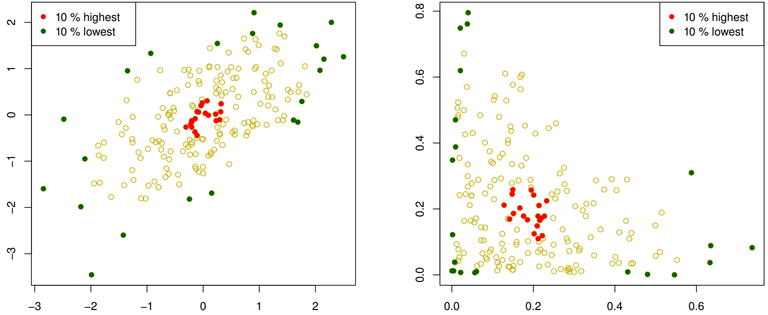
$$HD(x;P) = \inf_{H \in \mathcal{H}(x)} P(X \in H)$$



Halfspace Depth

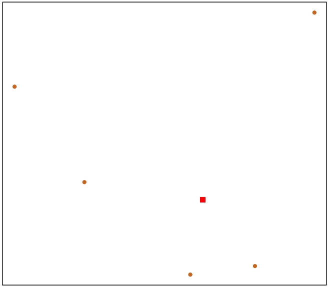
Halfspace depth (Tukey [11]) *HD* of an observation from $\mathbb{R}^d$

$$HD(x;P) = \inf_{H \in \mathcal{H}(x)} P(X \in H)$$



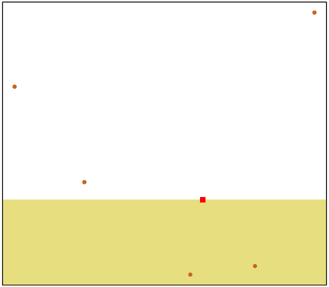
Halfspace Depth

$HD(x; X_1, \ldots, X_n) =$ least ratio of observations in a halfspace containing x



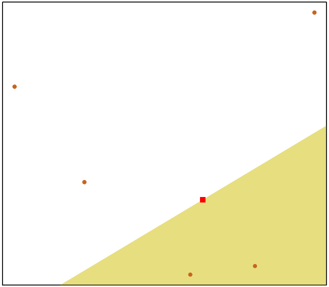
Halfspace Depth

$HD(x; X_1, \ldots, X_n) =$ least ratio of observations in a halfspace containing x



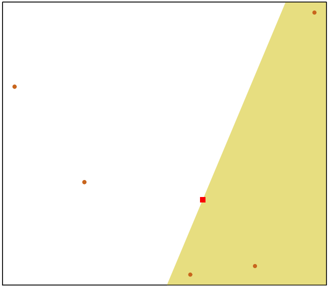
Halfspace Depth

$HD(x; X_1, \ldots, X_n) =$ least ratio of observations in a halfspace containing x



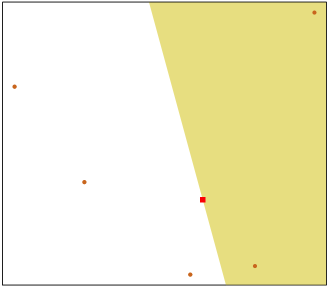
Halfspace Depth

$HD(x; X_1, \ldots, X_n) =$ least ratio of observations in a halfspace containing x



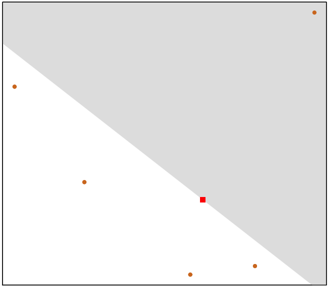
Halfspace Depth

$HD(x; X_1, \ldots, X_n) =$ least ratio of observations in a halfspace containing x



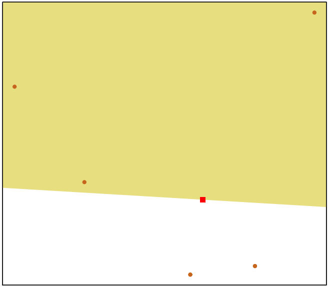
Halfspace Depth

$HD(x; X_1, \ldots, X_n) =$ least ratio of observations in a halfspace containing x



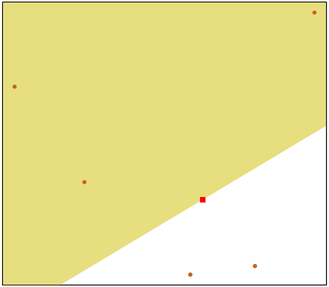
Halfspace Depth

$HD(x; X_1, \ldots, X_n) =$ least ratio of observations in a halfspace containing x



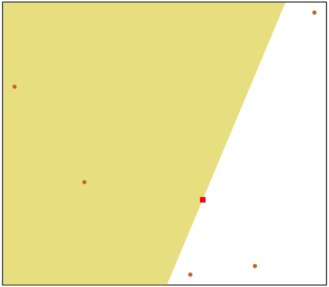
Halfspace Depth

$HD(x; X_1, \ldots, X_n) =$ least ratio of observations in a halfspace containing x



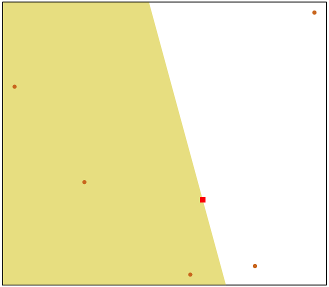
Halfspace Depth

$HD(x; X_1, \ldots, X_n) =$ least ratio of observations in a halfspace containing x



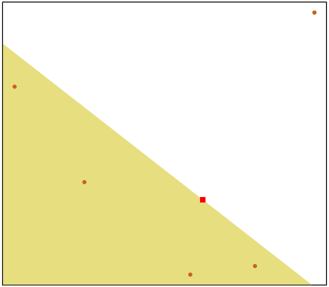
Halfspace Depth

$HD(x; X_1, \ldots, X_n) =$ least ratio of observations in a halfspace containing x



Halfspace Depth

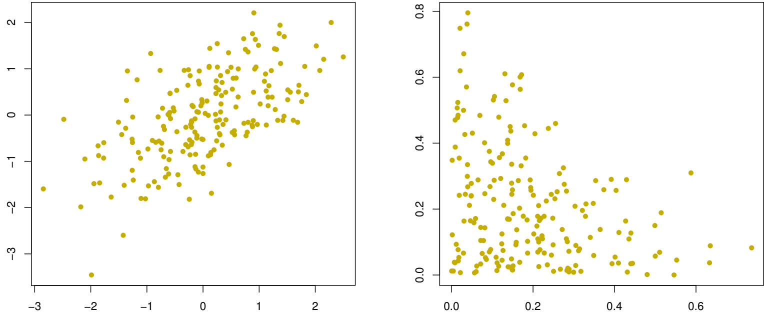
$HD(x; X_1, \ldots, X_n)$ = least ratio of observations in a halfspace containing x



Simplicial Depth

Simplicial depth (Liu [7]) *SD* of an observation from $\mathbb{R}^d$

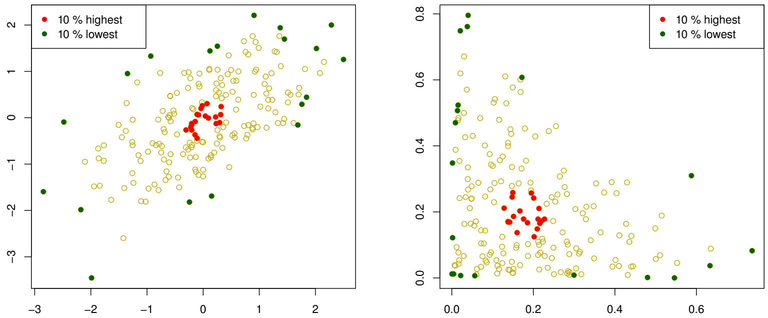
$$SD(x;P) = P\left(x \in \mathbb{S}_{X_1,\ldots,X_{d+1}}\right)$$



Simplicial Depth

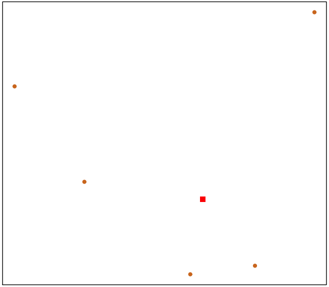
Simplicial depth (Liu [7]) *SD* of an observation from $\mathbb{R}^d$

$$SD(x;P) = P\left(x \in \mathbb{S}_{X_1,\ldots,X_{d+1}}\right)$$



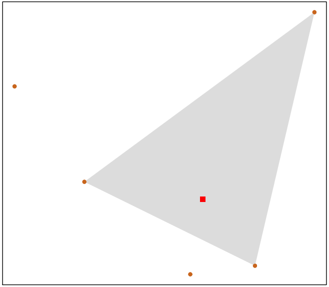
Simplicial Depth

$$SD\left(x;X_1,\ldots,X_n\right)=\frac{\text{number of simplices containing } x}{\text{total number of simplices}}$$



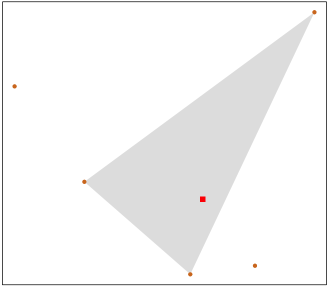
Simplicial Depth

$$SD\left(x; X_1, \ldots, X_n\right) = \frac{\text{number of simplices containing } x}{\text{total number of simplices}}$$



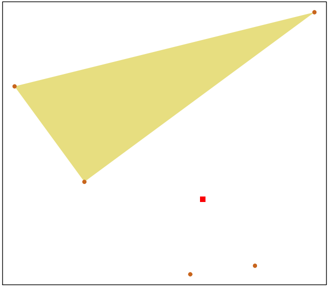
Simplicial Depth

$$SD\left(x; X_1, \ldots, X_n\right) = \frac{\text{number of simplices containing } x}{\text{total number of simplices}}$$



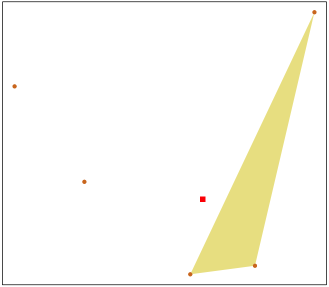
Simplicial Depth

$$SD(x; X_1, \ldots, X_n) = \frac{\text{number of simplices containing } x}{\text{total number of simplices}}$$



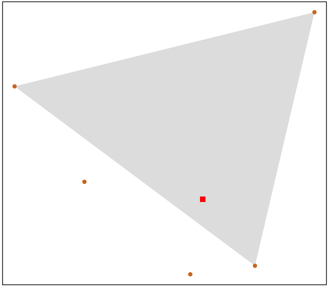
Simplicial Depth

$$SD\left(x;X_1,\ldots,X_n\right)=\frac{\text{number of simplices containing } x}{\text{total number of simplices}}$$



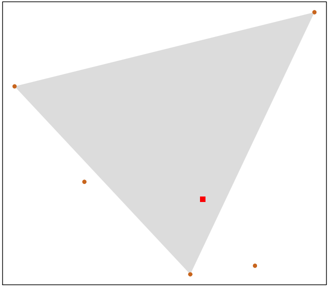
Simplicial Depth

$$SD\left(x; X_1, \ldots, X_n\right) = \frac{\text{number of simplices containing } x}{\text{total number of simplices}}$$



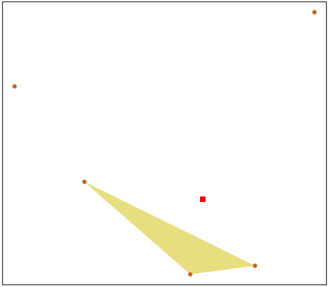
Simplicial Depth

$$SD(x; X_1, \ldots, X_n) = \frac{\text{number of simplices containing } x}{\text{total number of simplices}}$$



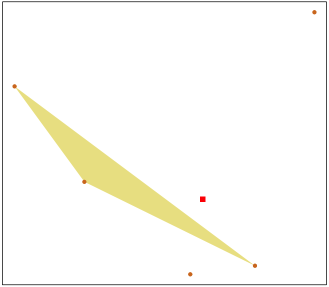
Simplicial Depth

$$SD\left(x; X_1, \ldots, X_n\right) = \frac{\text{number of simplices containing } x}{\text{total number of simplices}}$$



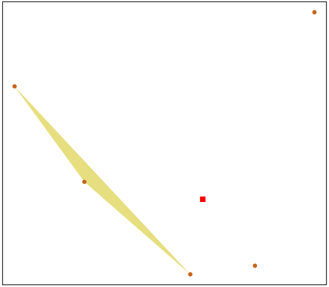
Simplicial Depth

$$SD\left(x; X_1, \ldots, X_n\right) = \frac{\text{number of simplices containing } x}{\text{total number of simplices}}$$



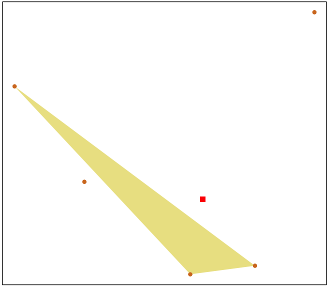
Simplicial Depth

$$SD(x; X_1, \ldots, X_n) = \frac{\text{number of simplices containing } x}{\text{total number of simplices}}$$



Simplicial Depth

$$SD\left(x; X_1, \ldots, X_n\right) = \frac{\text{number of simplices containing } x}{\text{total number of simplices}}$$



Characterization of Distribution

Is $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ characterized by $\left\{HD\left(x;P\right) \middle| x \in \mathbb{R}^d\right\}$?

Characterization of Distribution

Is $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ characterized by $\left\{HD\left(x;P\right) \middle| x \in \mathbb{R}^d\right\}$?

Yes, if

- P is an **empirical measure** (Struyf and Rousseeuw 1999),
- P is **a.c. with a compact support** (Koshevoy 2001),
- P is **atomic** (Koshevoy 2002)
- P has a $\mathcal{C}^{(2)}$ **density** (Hassairi and Regaieg 2008),
- the HD **contours are smooth** (Kong and Zuo 2010).

Characterization of Distribution

Is $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ characterized by $\left\{HD\left(x;P\right)\middle|x \in \mathbb{R}^d\right\}$?

Yes, if

- P is an **empirical measure** (Struyf and Rousseeuw 1999),
- P is **a.c. with a compact support** (Koshevoy 2001),
- P is **atomic** (Koshevoy 2002)
- P has a $\mathcal{C}^{(2)}$ **density** (Hassairi and Regaieg 2008),
- the *HD* **contours are smooth** (Kong and Zuo 2010).

When are *HD* contours smooth?

When are Halfspace Depth Contours Smooth?

Theorem:

Let $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ be contiguous and $x \in \mathbb{R}^d$. Then the halfspace depth contours are smooth at x if and only if there exists a unique halfspace $H \in \mathcal{H}(x)$ such that

$$HD\left(x;P\right) = P\left(X \in H\right).$$

When are Halfspace Depth Contours Smooth?

Theorem:

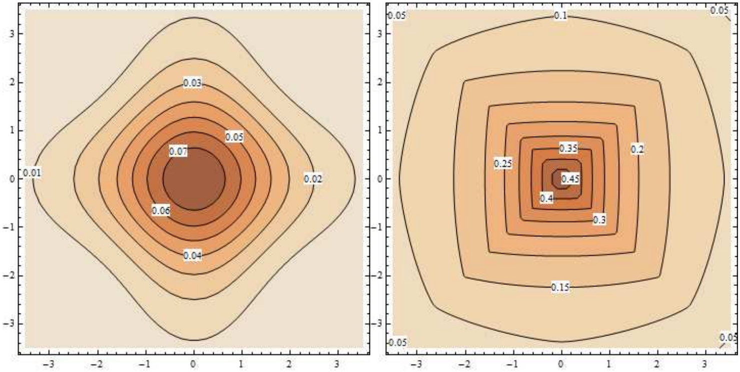
Let $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ be contiguous and $x \in \mathbb{R}^d$. Then the halfspace depth contours are smooth at x if and only if there exists a unique halfspace $H \in \mathcal{H}(x)$ such that

$$HD\left(x;P\right) = P\left(X \in H\right).$$

As a corollary, a point x from the hyperspace of reflectional symmetry R of P is depth regular (depth contours at at x are smooth) if and only if ***HD* is attained only at a halfspace orthogonal to *R***.

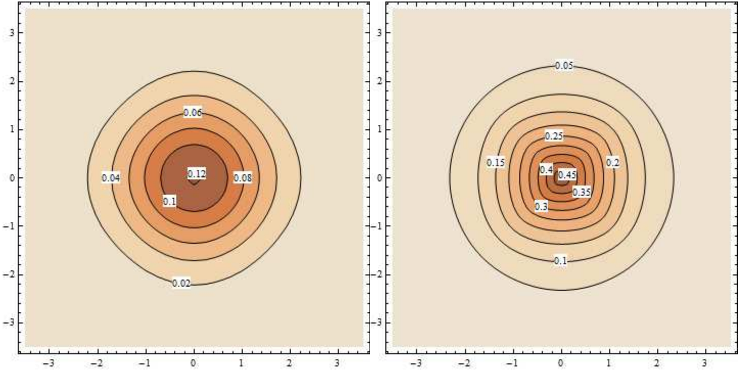
Example 1: Gaussian Distributions Mixture

A **strictly unimodal** distribution and non-smooth *HD* contours.



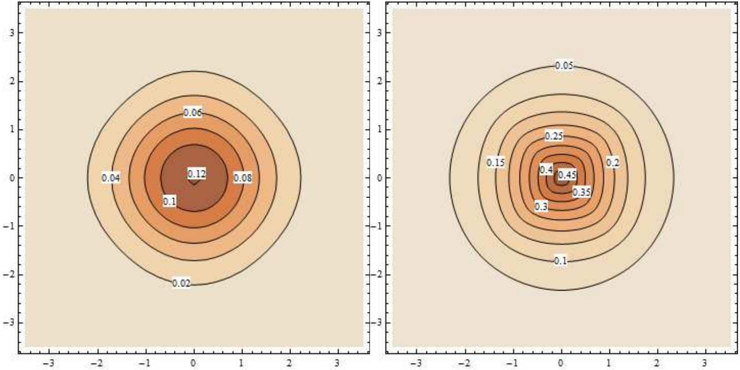
Example 2: Gaussian Distributions Mixture

Another **strictly unimodal** distribution.



Example 2: Gaussian Distributions Mixture

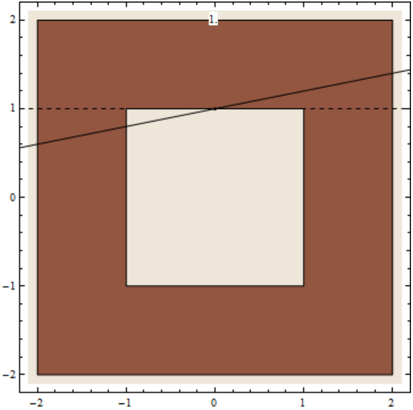
Another **strictly unimodal** distribution.



Is non-smooth **only for** $\sigma_1^2 > \sigma_2^2(2+\sqrt{3})$.

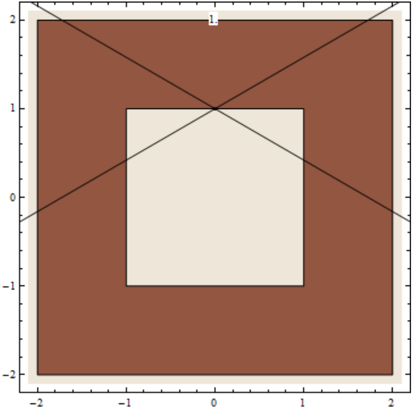
Example 3: Rectangle

A distribution with non-smooth *HD* contours.



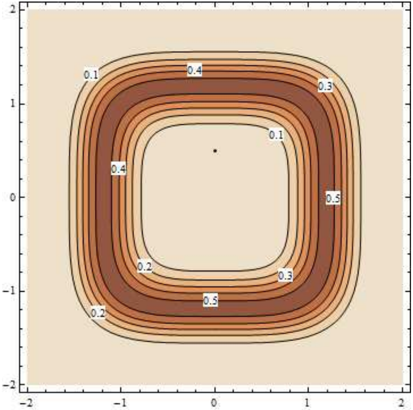
Example 3: Rectangle

A distribution with non-smooth *HD* contours.



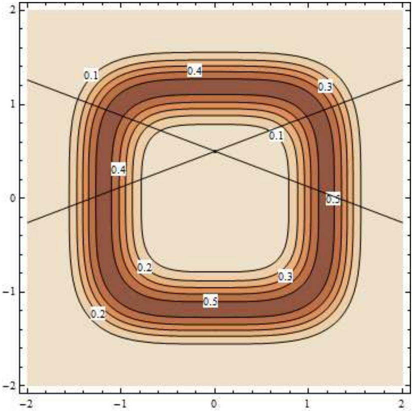
Example 4: L^4 symmetrical distribution

An L^4 **symmetrical** distribution with non-smooth *HD* contours.



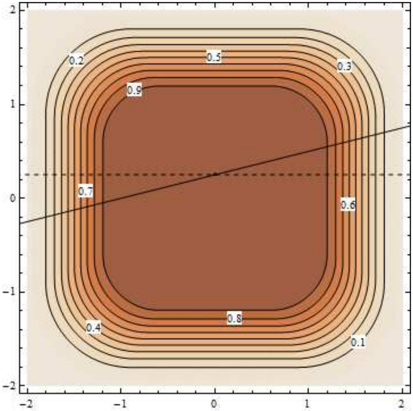
Example 4: L^4 symmetrical distribution

An L^4 **symmetrical** distribution with non-smooth *HD* contours.



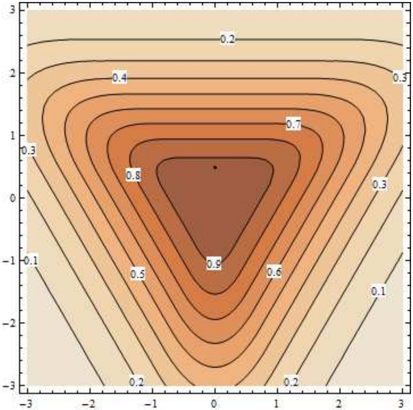
Example 5: quasi-concave distribution 1

A **quasi-concave** distribution with non-smooth *HD* contours.



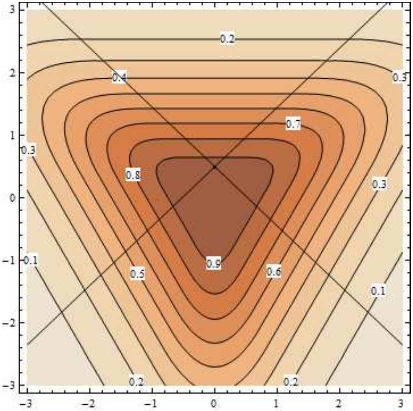
Example 6: quasi-concave distribution 2

A **strictly quasi-concave** distribution with non-smooth *HD* contours.



Example 6: quasi-concave distribution 2

A **strictly quasi-concave** distribution with non-smooth *HD* contours.



Smooth Halfspace Depth Contours: Conclusions

Conclusion

Not even the density smoothness, strict quasi-concavity and reflectional symmetry suffices for the halfspace depth contours to be smooth at every point of $\mathbb{R}^d$.

Smooth Halfspace Depth Contours: Conclusions

Conclusion

Not even the density smoothness, strict quasi-concavity and reflectional symmetry suffices for the halfspace depth contours to be smooth at every point of $\mathbb{R}^d$.

Can this be guaranteed at least for even **smaller classes of distributions**?

- **angularly symmetrical** and strictly quasi-concave, or merely
- L^p **symmetrical** and strictly quasi-concave?

For further discussion, see Nagy [9].

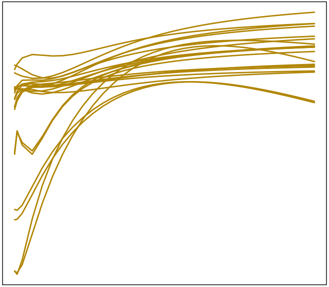
Outline

1. Depth Measure and its Smoothness for Multivariate Data
   - Smoothness of Halfspace Depth Contours
2. Functional Data Depth: Theory
   - Functional Band Depths
   - Consistency
   - Counterexample
   - Fixing the Continuousness
   - Integral and Vector Depths
3. Functional Data Depth: Practice
   - Problem of Functional Data Classification
   - Using Depth for Classification
   - Simulation Study
4. Conclusions

Functional Data

$X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ and $X_1, \ldots, X_n$ a r.s. from P. Consider the depth of functional observations w.r.t. P (or P_n)

$$D \colon \mathcal{C}([0,1]) \times \mathcal{P}(\mathcal{C}([0,1])) \to [0,1].$$

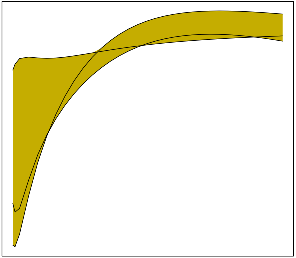


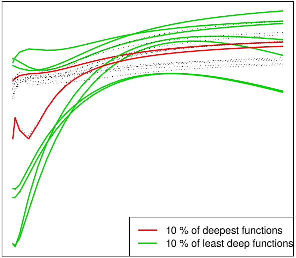
Band Depth

López-Pintado and Romo [8] for $J = 2, 3, \ldots$

$$BD^{J)}(x; P) = \frac{1}{J-1} \sum_{j=2}^{J} P\left[G(x) \subset B(X_1, X_2, \ldots, X_j)\right],$$

where $G(x)$ is **the graph of a function** x and $B(x_1, x_2, \ldots, x_j)$ is **a band of functions** $x_1, x_2, \ldots, x_j$

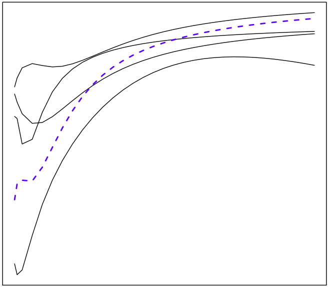




Band Depth

The sample version is **a U-statistic of order J**.

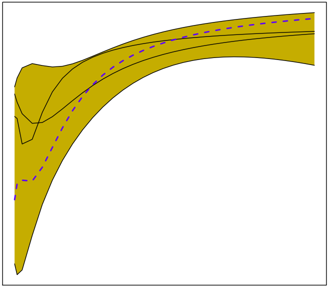
$$BD^{J)}(x;P_n) = \frac{1}{J-1}\sum_{j=2}^{J}\binom{n}{j}^{-1}\sum_{1\leq i_1<i_2<\ldots<i_j\leq n}\mathbb{I}\left[G(x)\subset B\left(X_{i_1},X_{i_2},\ldots,X_{i_j}\right)\right].$$



Band Depth

The sample version is **a U-statistic of order J**.

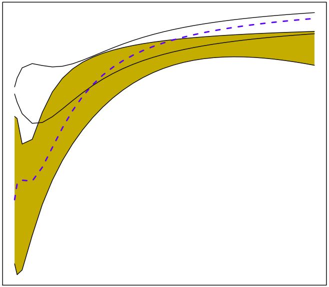
$$BD^{J)}(x;P_n) = \frac{1}{J-1}\sum_{j=2}^{J}\binom{n}{j}^{-1}\sum_{1\leq i_1<i_2<\ldots<i_j\leq n}\mathbb{I}\left[G(x)\subset B\left(X_{i_1},X_{i_2},\ldots,X_{i_j}\right)\right].$$



Band Depth

The sample version is **a U-statistic of order J**.

$$BD^{J)}(x;P_n) = \frac{1}{J-1}\sum_{j=2}^{J}\binom{n}{j}^{-1}\sum_{1\leq i_1<i_2<\ldots<i_j\leq n}\mathbb{I}\left[G(x)\subset B\left(X_{i_1},X_{i_2},\ldots,X_{i_j}\right)\right].$$



Strong Consistency

Depth D is on a set $S \subset \mathcal{C}([0,1])$ **consistent**

- **pointwise** if

$$D(x; P_n) - D(x; P) \xrightarrow[n \to \infty]{a.s.} 0 \text{ for all } x \in S,$$

Strong Consistency

Depth D is on a set $S \subset \mathcal{C}([0,1])$ **consistent**

- **pointwise** if

$$D(x; P_n) - D(x; P) \xrightarrow[n\to\infty]{a.s.} 0 \text{ for all } x \in S,$$

- **uniformly** if

$$\sup_{x \in S} |D(x; P_n) - D(x; P)| \xrightarrow[n\to\infty]{a.s.} 0,$$

Strong Consistency

Depth D is on a set $S \subset \mathcal{C}([0,1])$ **consistent**

- **pointwise** if

$$D(x; P_n) - D(x; P) \xrightarrow[n\to\infty]{a.s.} 0 \text{ for all } x \in S,$$

- **uniformly** if

$$\sup_{x\in S} |D(x; P_n) - D(x; P)| \xrightarrow[n\to\infty]{a.s.} 0,$$

- **universally** if

$$\sup_{x\in S} |D(x; P_n) - D(x; P)| \xrightarrow[n\to\infty]{a.s.} 0 \text{ for all } P \in \mathcal{P}(\mathcal{C}([0,1])),$$

Strong Consistency

Depth D is on a set $S \subset \mathcal{C}([0,1])$ **consistent**

- **pointwise** if

$$D(x;P_n) - D(x;P) \xrightarrow[n\to\infty]{a.s.} 0 \text{ for all } x \in S,$$

- **uniformly** if

$$\sup_{x\in S} |D(x;P_n) - D(x;P)| \xrightarrow[n\to\infty]{a.s.} 0,$$

- **universally** if

$$\sup_{x\in S} |D(x;P_n) - D(x;P)| \xrightarrow[n\to\infty]{a.s.} 0 \text{ for all } P \in \mathcal{P}(\mathcal{C}([0,1])),$$

- $\mathcal{P}$**-uniformly** if

$$\sup_{P\in\mathcal{P}(\mathcal{C}([0,1]))} \sup_{x\in S} |D(x;P_n) - D(x;P)| \xrightarrow[n\to\infty]{a.s.} 0.$$

Band Depth

Band Depth (L-P López-Pintado, R Romo):

- L-P, R: Depth-based classification for functional data (DIMACS 2006)
- L-P, R: Depth-based inference for functional data (CSDA 2007)
- L-P, Jornsten: Functional analysis via extensions of the band depth (IMS Lecture Notes, 2007)
- **L-P, R: On the Concept of Depth for Functional Data (JASA 2009)**
- L-P, R: Robust depth-based tools for the analysis of gene expression data (Biostatistics 2010)
- L-P, R: A half-region depth for functional data (CSDA 2011)
- ...

Band Depth Consistency

López-Pintado and Romo [8, Thm 4]

Theorem:

Let $P \in \mathcal{P}(\mathcal{C}([0,1]))$ with a.c. marginals. Then $BD^{J)}$ is uniformly consistent on every equi-continuous set S, i.e.

$$\sup_{x \in S} \left| BD^{J)}(x; P_n) - BD^{J)}(x; P) \right| \xrightarrow[n \to \infty]{a.s.} 0.$$

Band Depth Consistency: Proof

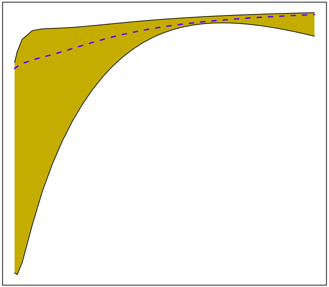
Proof: As $\lim_{\|x\|\to\infty} BD^{J)}(x;P) = 0$, consider only $\{\|x\| < M\}$ for $M > 0$. According to Arzéla-Ascoli's Theorem, a uniformly bounded set of equi-continuous functions is totally bounded.
Because $BD^{J)}(.;P)$ is for P with a.c. marginals **a continuous functional, it is enough to prove** for $N \in \mathbb{N}$ fixed

$$\max_{\{x_i\}_{i=1}^{N} \subset S} \left| BD^{J)}(x_i;P_n) - BD^{J)}(x_i;P) \right| \xrightarrow[n\to\infty]{a.s.} 0.$$

This holds, since $BD^{J)}(.;P_n)$ is a bounded U-statistic. □

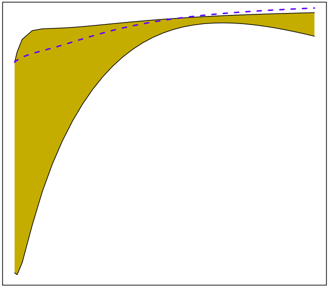
Why the Proof Does Not Work?

$BD^{J)}(.;P)$ is continuous, but $BD^{J)}(.;P_n)$ **is not**!



Why the Proof Does Not Work?

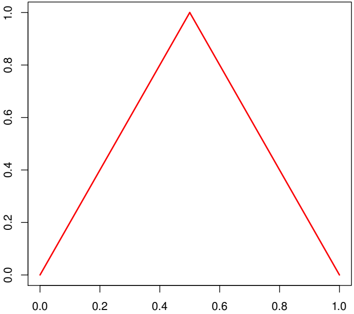
$BD^{J)}(.;P)$ is continuous, but $BD^{J)}(.;P_n)$ **is not**!



Why the Proof Does Not Work?

$$\max_{\{x_i\}_{i=1}^{N} \subset S} \left| BD^{J)}(x_i; P_n) - BD^{J)}(x_i; P) \right| \xrightarrow[n\to\infty]{a.s.} 0$$

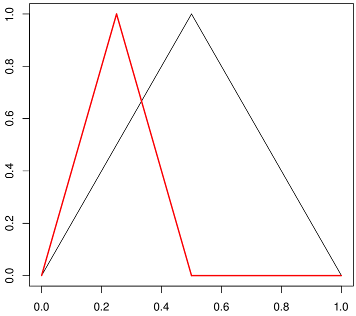
does not give uniform convergence!



Why the Proof Does Not Work?

$$\max_{\{x_i\}_{i=1}^N \subset S} \left| BD^{J)}(x_i; P_n) - BD^{J)}(x_i; P) \right| \xrightarrow[n\to\infty]{a.s.} 0$$

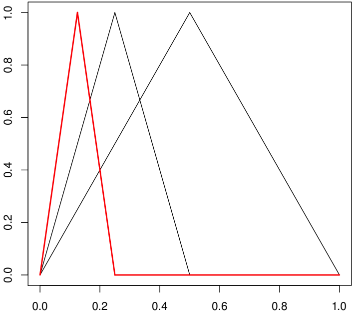
does not give uniform convergence!



Why the Proof Does Not Work?

$$\max_{\{x_i\}_{i=1}^N \subset S} \left| BD^{J)}(x_i; P_n) - BD^{J)}(x_i; P) \right| \xrightarrow[n\to\infty]{a.s.} 0$$

does not give uniform convergence!



Is Band Depth Consistent?

Starting from the theory of **empirical processes** (for $J = 2$):

- The validity of

$$\dim_{\mathrm{VC}} \{(x_1, x_2) | G(x) \subset B(x_1, x_2)\}_{x \in S} = \infty$$

  for $S \subset \mathcal{C}([0,1])$ compact suggests, that the depth **is not $\mathcal{P}$-uniformly consistent** (Assouad's Thm - [3, Thm 6.4.5]).
- The existence of **boolean σ-independent sequence** of functions in the class

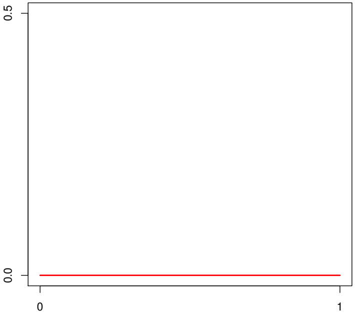
$$\{(x_1, x_2) | G(x) \subset B(x_1, x_2)\}_{x \in S}$$

  suggest, that the depth **is not universally consistent** (van Handel's Thm - [12, Thm 1.3]).

Band Depth Consistence: Counterexample

Define $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ as follows:

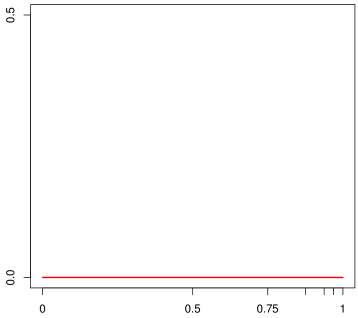
- $P(X(t) = 0 \text{ for all } t \in [0,1]) = 0.5.$



Band Depth Consistence: Counterexample

Define $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ as follows:

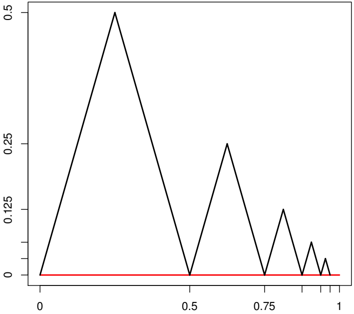
- Divide the interval $[0,1]$ "diadically" into disjoint subintervals I_j of lengths $\{2^{-j}\}_{j\in\mathbb{N}}$.



Band Depth Consistence: Counterexample

Define $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ as follows:

- If $X \not\equiv 0$, set X zero on every I_j with probability 0.5 or have a jump with probability 0.5. The jumps occur independently.



Band Depth Consistence: Counterexample

Let x_j be a function with **a single jump** on the interval I_j, 0 otherwise.

Then:

- $BD^{2)}(x_j; P) = 0.25$ for all $j \in \mathbb{N}$

Band Depth Consistence: Counterexample

Let x_j be a function with **a single jump** on the interval I_j, 0 otherwise. Then:

- $BD^{2)}(x_j;P) = 0.25$ for all $j \in \mathbb{N}$
- Let n be even. If there exists $j_n \in \mathbb{N}$ such that exactly $n/2$ functions have a jump on I_{j_n} and $n/2$ functions is zero at $[0,1]$, then x_{j_n} lies in

$$\binom{n}{2} - 2\binom{n/2}{2} = \frac{n^2}{4}$$

bands.

Band Depth Consistence: Counterexample

Let x_j be a function with **a single jump** on the interval I_j, 0 otherwise. Then:

- $BD^{2)}(x_j; P) = 0.25$ for all $j \in \mathbb{N}$
- Let n be even. If there exists $j_n \in \mathbb{N}$ such that exactly $n/2$ functions have a jump on I_{j_n} and $n/2$ functions is zero at $[0,1]$, then x_{j_n} lies in

$$\binom{n}{2} - 2\binom{n/2}{2} = \frac{n^2}{4}$$

  bands.
- For such a j_n we have

$$BD^{J)}(x_{j_n}; P_n) = \frac{\frac{n^2}{4}}{\binom{n}{2}} = \frac{n}{2(n-1)} \xrightarrow[n\to\infty]{} 0.5.$$

Band Depth Consistence: Counterexample

Let x_j be a function with **a single jump** on the interval I_j, 0 otherwise. Then:

- $BD^{2)}(x_j; P) = 0.25$ for all $j \in \mathbb{N}$
- Let n be even. If there exists $j_n \in \mathbb{N}$ such that exactly $n/2$ functions have a jump on I_{j_n} and $n/2$ functions is zero at $[0,1]$, then x_{j_n} lies in
$$\binom{n}{2} - 2\binom{n/2}{2} = \frac{n^2}{4}$$
bands.
- For such a j_n we have
$$BD^{J)}(x_{j_n}; P_n) = \frac{\frac{n^2}{4}}{\binom{n}{2}} = \frac{n}{2(n-1)} \xrightarrow[n\to\infty]{} 0.5.$$

Band Depth Consistence: Counterexample

Let x_j be a function with **a single jump** on the interval I_j, 0 otherwise. Then:

- $BD^{2)}(x_j; P) = 0.25$ for all $j \in \mathbb{N}$
- Let n be even. If there exists $j_n \in \mathbb{N}$ such that exactly $n/2$ functions have a jump on I_{j_n} and $n/2$ functions is zero at $[0,1]$, then x_{j_n} lies in

$$\binom{n}{2} - 2\binom{n/2}{2} = \frac{n^2}{4}$$

  bands.
- For such a j_n we have

$$BD^{J)}(x_{j_n}; P_n) = \frac{\frac{n^2}{4}}{\binom{n}{2}} = \frac{n}{2(n-1)} \xrightarrow[n\to\infty]{} 0.5.$$

But does exist **infinitely many** of such couples (n, j_n)?

Band Depth Consistence: Counterexample

But does exist **infinitely many** of such couples (n, j_n)? **Yes!**

- Almost surely there is infinitely many n such that exactly $n/2$ functions if zero on $[0,1]$ **(state 0 is permanent in a symmetric random walk)**.

Band Depth Consistence: Counterexample

But does exist **infinitely many** of such couples (n, j_n)? **Yes!**

- Almost surely there is infinitely many n such that exactly $n/2$ functions if zero on $[0, 1]$ **(state 0 is permanent in a symmetric random walk)**.
- For every such n a.s. exists j_n such that all the $n/2$ functions with jumps have on the interval I_{j_n} a jump **(Borel-Cantelli)**.

Band Depth Consistence: Counterexample

But does exist **infinitely many** of such couples (n, j_n)? **Yes!**

- Almost surely there is infinitely many n such that exactly $n/2$ functions if zero on $[0,1]$ **(state 0 is permanent in a symmetric random walk)**.
- For every such n a.s. exists j_n such that all the $n/2$ functions with jumps have on the interval I_{j_n} a jump **(Borel-Cantelli)**.
- We get infinitely many n such that for each of them there exists a function $\{x_j\}_{j\in\mathbb{N}}$ for which $BD^{2)}(x_{j_n}; P_n) \approx 0.5$. Hence, for every $\varepsilon > 0$ and infinitely many $n \in \mathbb{N}$ a.s. holds

$$\sup_{j\in\mathbb{N}} \left| BD^{2)}(x_j; P_n) - BD^{2)}(x_j; P) \right| > 0.25 - \varepsilon$$

Band Depth Consistence: Counterexample

But does exist **infinitely many** of such couples (n, j_n)? **Yes!**

- Almost surely there is infinitely many n such that exactly $n/2$ functions if zero on $[0,1]$ **(state 0 is permanent in a symmetric random walk)**.
- For every such n a.s. exists j_n such that all the $n/2$ functions with jumps have on the interval I_{j_n} a jump **(Borel-Cantelli)**.
- We get infinitely many n such that for each of them there exists a function $\{x_j\}_{j\in\mathbb{N}}$ for which $BD^{2)}(x_{j_n}; P_n) \approx 0.5$. Hence, for every $\varepsilon > 0$ and infinitely many $n \in \mathbb{N}$ a.s. holds

$$\sup_{j\in\mathbb{N}} \left| BD^{2)}(x_j; P_n) - BD^{2)}(x_j; P) \right| > 0.25 - \varepsilon$$

Band Depth Consistence: Counterexample

But does exist **infinitely many** of such couples (n, j_n)? **Yes!**

- Almost surely there is infinitely many n such that exactly $n/2$ functions if zero on $[0,1]$ **(state 0 is permanent in a symmetric random walk)**.
- For every such n a.s. exists j_n such that all the $n/2$ functions with jumps have on the interval I_{j_n} a jump **(Borel-Cantelli)**.
- We get infinitely many n such that for each of them there exists a function $\{x_j\}_{j\in\mathbb{N}}$ for which $BD^{2)}(x_{j_n}; P_n) \approx 0.5$. Hence, for every $\varepsilon > 0$ and infinitely many $n \in \mathbb{N}$ a.s. holds

$$\sup_{j\in\mathbb{N}} \left| BD^{2)}(x_j; P_n) - BD^{2)}(x_j; P) \right| > 0.25 - \varepsilon$$

Thus, $BD^{2)}$ **is not uniformly consistent** w.r.t. P.

Fixing the Continuousness

The problem of López-Pinado and Romo's proof was that the depth $BD^{J)}(.;P_n)$ was not (uniformly) continuous. Instead of measuring the outlyingness of a function from a band by an indicator, let's measure **distance from a band**, i.e. for a metric d on $\mathcal{C}([0,1])$ use

$$\mathrm{E}\left[1 - w\left(d\left(x; B\left(X_1, X_2\right)\right)\right)\right]$$

instead of

$$P\left[G(x) \subset B\left(X_1, X_2\right)\right] = \mathrm{E}\left[\mathbb{I}\left[G(x) \subset B\left(X_1, X_2\right)\right]\right],$$

where $w\colon [0,\infty) \to [0,1]$, $w(0) = 1$, $\lim_{t\to\infty} w(t) = 0$ is equi-continuous **smoothing function**, e.g. e^{-t}.
Consider supremum and L_1 metric for simplicity.

Fixing the Continuousness

Theorem:

Let w be a smoothing function and $S \subset \mathcal{C}([0,1])$ relatively compact. Then the band depths smoothed by w

$$BD^{J)}(.;.,w,d) : \mathcal{C}([0,1]) \times \mathcal{P}(\mathcal{C}([0,1])) \to [0,1]$$

are for supremum norm, as well as for L_1 norm $\mathcal{P}$-uniformly consistent on S.

Proof: A strengthened version of López-Pintado and Romo's proof is used. It is proved that the class

$$\left\{ BD^{J)}(x;P,w,d) \middle| x \in \mathcal{C}([0,1]), P \in \mathcal{P}(\mathcal{C}([0,1])) \right\}$$

is uniformly continuous and the properties of U-statistics are utilized (Borovskich a Koroljuk [6, Thm 2.1.4]).

Fraiman-Muniz Type of Depth

Fraiman and Muniz [4]

$$ID(x;P) = \int_0^1 D(x(t);P_t)\,\mathrm{d}t,$$

where D is univariate “depth” like

- **halfspace depth**

$$D(x(t);P_t) = \min\{F_t(x(t)), 1 - F_t(x(t))\},$$

- **simplicial depth**

$$D(x(t);P_t) = F_t(x(t))(1 - F_t(x(t))).$$

Generalization of Fraiman-Muniz Type of Depth

The idea of Fraiman and Muniz may be easily generalized to **vector-valued functions**

$$ID(x;P) = \int_0^1 D(x(t);P_t)\,\mathrm{d}t,$$

where D is usual multivariate depth,

$$x = (x_1, \ldots, x_K), \text{ where } x_k : [0,1] \to \mathbb{R}$$

and $P \in \mathcal{P}\left(\mathcal{C}([0,1])^K\right)$.

This is how we define **K-vector depths** and by application to differentiable functions also **K-derivatives depths** (Hlubinka and Nagy [5]).

Integral Depths Consistency

Theorem:

Let the sample version of a depth $D : \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) \to [0,1]$ *have a form of a U-statistic and be universally consistent. Then the depth for vector-valued functions*

$$ID\left(x;P\right) = \int_0^1 D\left(x\left(t\right);P_t\right)\,dt$$

is universally consistent on $\mathcal{C}([0,1])^d$*, under some measurability assumptions.*

Proof: Utilizing Lebesgue dominated convergence Theorem we obtain weak universal consistency, which is for U-processes equivalent to (strong) universal consistency (cf. de la Peña a Giné [2, p.227]). □

The Theorem can be applied for example for **simplicial depth** as D.

Other Properties of Integral Depths

A range of **other properties** of integral depth for vector-valued functions can be proved (Nagy and Hlubinka [10]):

- **measurability** as a functional on $\mathcal{C}([0,1])^K \times \mathcal{P}\left(\mathcal{C}([0,1])^K\right)$,

Other Properties of Integral Depths

A range of **other properties** of integral depth for vector-valued functions can be proved (Nagy and Hlubinka [10]):

- **measurability** as a functional on $\mathcal{C}([0,1])^K \times \mathcal{P}\left(\mathcal{C}([0,1])^K\right)$,
- functional version of **affine invariance** for *ID* and *dID*,

Other Properties of Integral Depths

A range of **other properties** of integral depth for vector-valued functions can be proved (Nagy and Hlubinka [10]):

- **measurability** as a functional on $\mathcal{C}([0,1])^K \times \mathcal{P}\left(\mathcal{C}([0,1])^K\right)$,
- functional version of **affine invariance** for *ID* and *dID*,
- **monotonicity relative to the deepest point**,

Other Properties of Integral Depths

A range of **other properties** of integral depth for vector-valued functions can be proved (Nagy and Hlubinka [10]):

- **measurability** as a functional on $\mathcal{C}([0,1])^K \times \mathcal{P}\left(\mathcal{C}([0,1])^K\right)$,
- functional version of **affine invariance** for *ID* and *dID*,
- **monotonicity relative to the deepest point**,
- **continuity** (or **semicontinuity**) as functional of $x \in \mathcal{C}([0,1])^K$,

Other Properties of Integral Depths

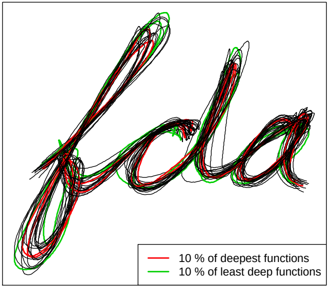
A range of **other properties** of integral depth for vector-valued functions can be proved (Nagy and Hlubinka [10]):

- **measurability** as a functional on $\mathcal{C}([0,1])^K \times \mathcal{P}\left(\mathcal{C}([0,1])^K\right)$,
- functional version of **affine invariance** for *ID* and *dID*,
- **monotonicity relative to the deepest point**,
- **continuity** (or **semicontinuity**) as functional of $x \in \mathcal{C}([0,1])^K$,
- **qualitative robustness**, i.e. continuity as a functional of $P \in \mathcal{P}\left(\mathcal{C}([0,1])^K\right)$ in the weak convergence sense.

K-Vector Depth

Integral depths for vector functions

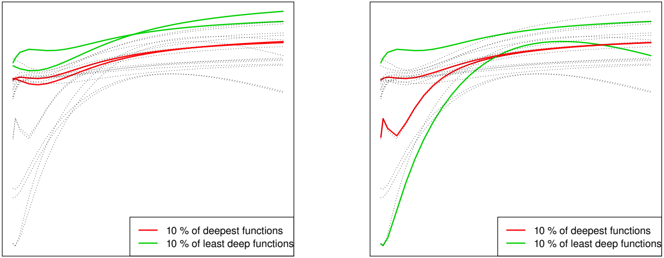
$$ID(x;P) = \int_0^1 D\left(\left(x_1(t), x_2(t)\right); \left(P_{1,t}, P_{2,t}\right)\right) \mathrm{d}t,$$



K-Derivatives Depth

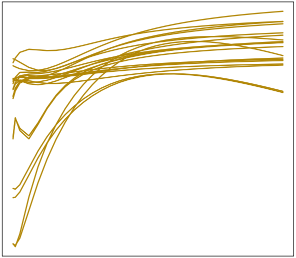
Integral depths for differentiable functions

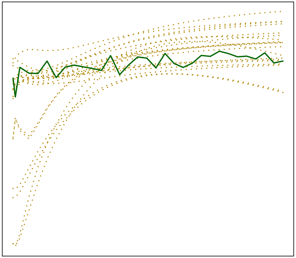
$$dID\left(x;P\right)=\int_0^1 D\left(\left(x(t),x'(t)\right);\left(P_t,P'_t\right)\right)\,\mathrm{d}t,$$



Contaminated Dataset

Consider now the **contaminated functional dataset**. Does the depth recognize the outlier?

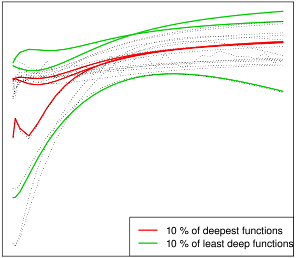


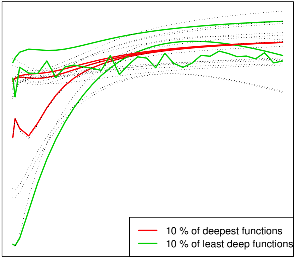


K-Derivatives Depth Again

Integral depths for differentiable functions

$$dID\left(x;P\right) = \int_0^1 D\left(\left(x(t), x'(t)\right); \left(P_t, P_t'\right)\right) \mathrm{d}t,$$





Future Challenges

- Generalization of van Handel's (Assouad's) Theorem for U-processes.
- **$\mathcal{P}$-uniform consistency of integral depths.**

$$P_\gamma - \dim \{\lambda [t | x(t) \in B(x_1(t), x_2(t))]\}_{x \in S} = \infty \quad \forall \gamma > 0$$

for $S \subset \mathcal{C}([0,1])$ compact suggests, that the depth **is not $\mathcal{P}$-uniformly consistent** (Alon's Thm) [1, Thm 2.2]).

Outline

1 Depth Measure and its Smoothness for Multivariate Data
- Smoothness of Halfspace Depth Contours

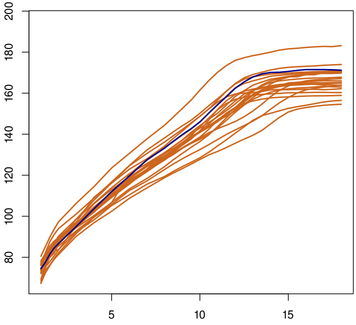
2 Functional Data Depth: Theory
- Functional Band Depths
- Consistency
- Counterexample
- Fixing the Continuousness
- Integral and Vector Depths

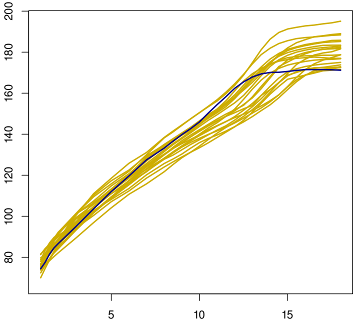
3 Functional Data Depth: Practice
- Problem of Functional Data Classification
- Using Depth for Classification
- Simulation Study

4 Conclusions

Determining the Distribution: Children's Growth Data

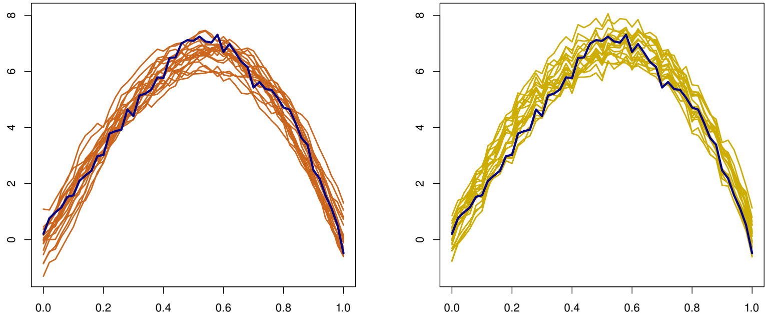
Let $P_1, P_2 \in \mathcal{P}(\mathcal{C}([0,1]))$ and $X \sim P_m$, $m \in \{1,2\}$ is unknown. What is the distribution of X?





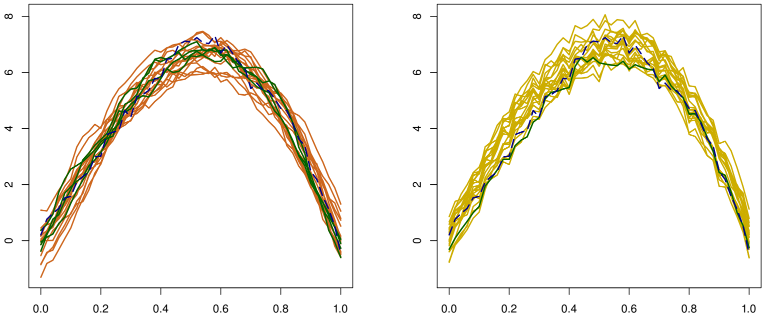
Determining the Distribution: a More Difficult Example

Let $P_1, P_2 \in \mathcal{P}(\mathcal{C}([0,1]))$ and $X \sim P_m$, $m \in \{1,2\}$ is unknown. What is the distribution of X?



Nearest Neighbor Rule

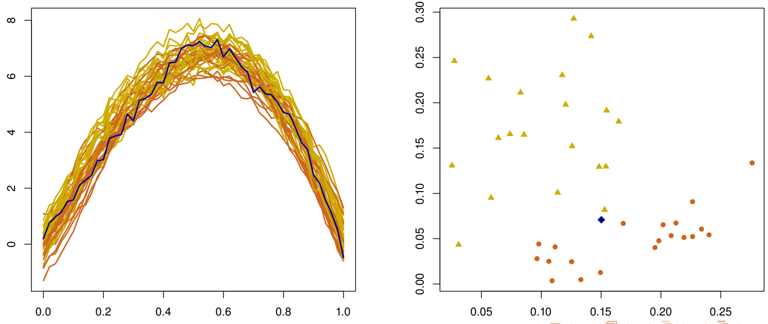
The k-nearest neighbor rule *KNN* with respect to a particular metric on space $\mathcal{C}([0,1])$ (e.g. L_2, $k = 5$):



Classification – DD-plot

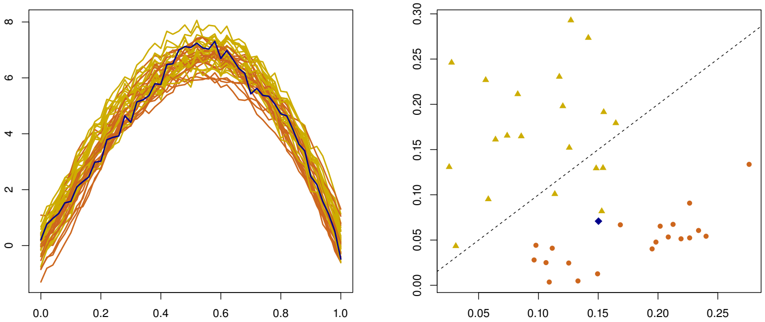
For given training samples $\mathbb{X}_1$, $\mathbb{X}_2$ and depth D, the **DD-transformation** of data can be computed as

$$DD\colon \mathcal{C}([0,1]) \to \mathbb{R}^2\colon x \mapsto \left(D(x;\mathbb{X}_1), D(x;\mathbb{X}_2)\right)^T$$



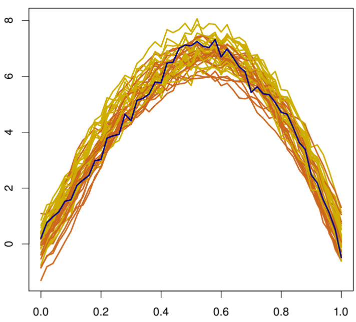
DD-plot and Highest Depth Rule

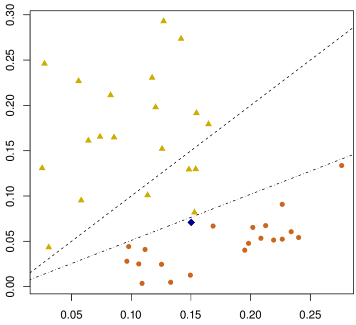
The function is assigned to the sample with **highest depth value** $\arg\max_{i=1,2} D(x; \mathbb{X}_i)$ (Cuevas et al. 2007)



DD-plot and Li's Rule

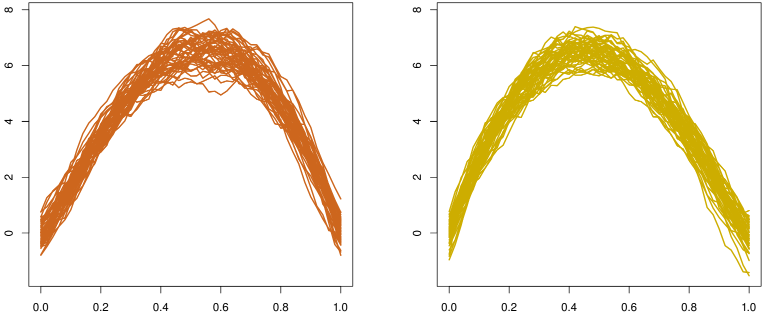
An increasing **best separating** function (linear, or polynomial) is utilized to classify the DD-transformations (Li et al. 2010)





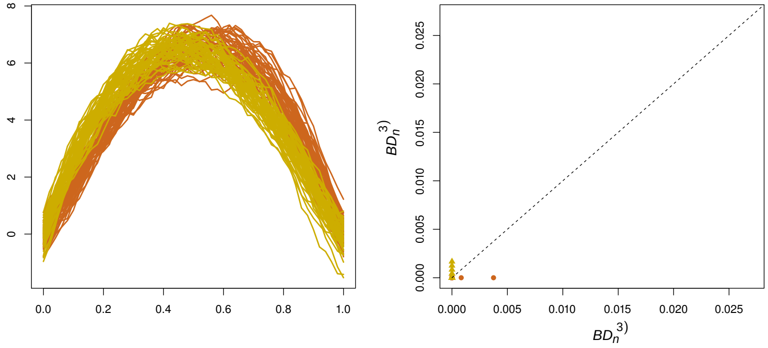
Location-shifted Model: Functions

$$m_1(t) = 30(1-t)t^{1.2},\ m_2(t) = 30t(1-t)^{1.2}$$

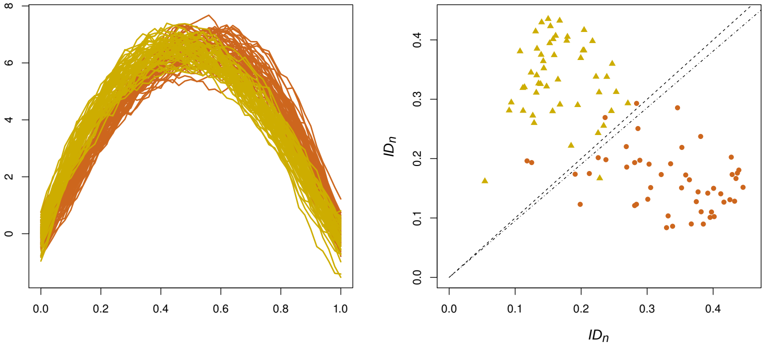


$$R_1(s,t) = 0.2\exp\left(-\frac{|s-t|}{0.3}\right),\ R_2(s,t) = 0.2\exp\left(-\frac{|s-t|}{0.3}\right)$$

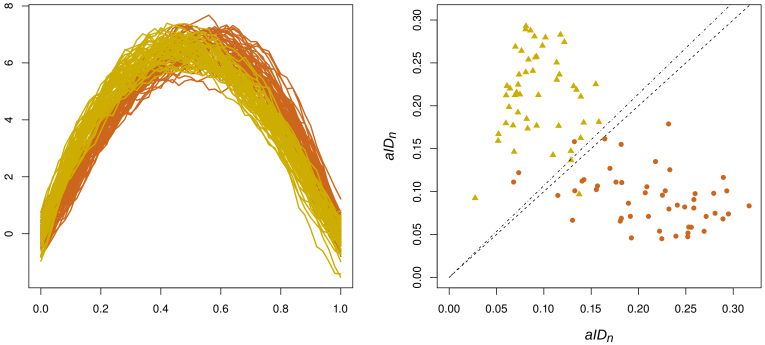
Location-shifted Model: $BD_n^{3)}$



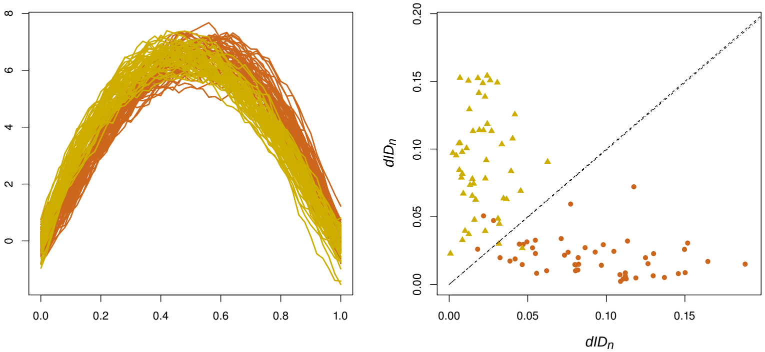
Location-shifted Model: ID_n



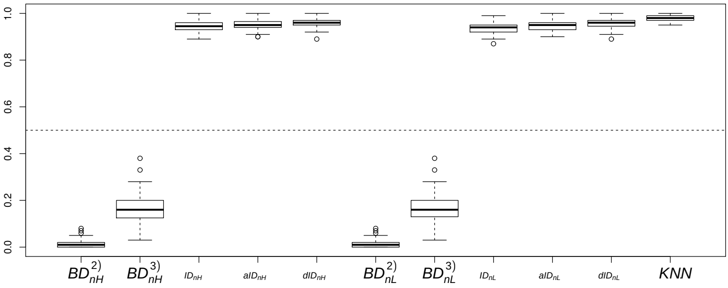
Location-shifted Model: aID_n



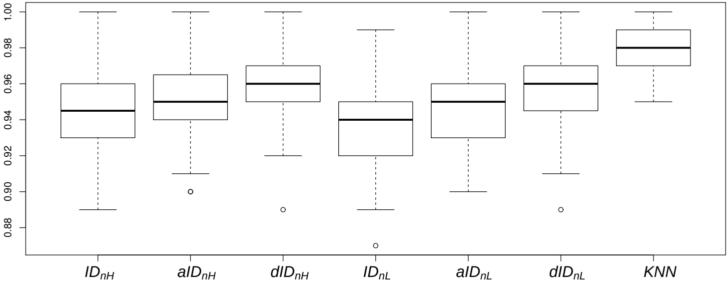
Location-shifted Model: dID_n



Location-shifted Model: Results 1

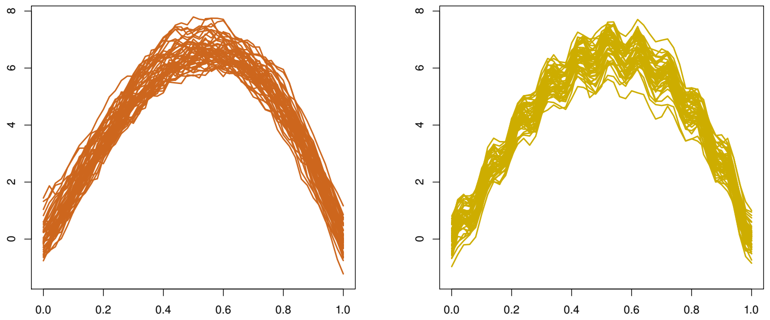


Location-shifted Model: Results 2



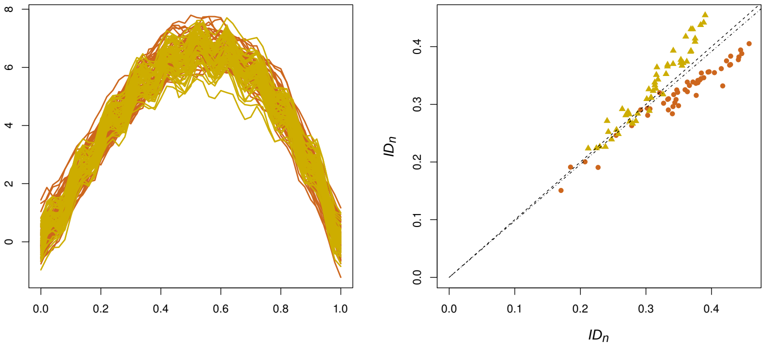
Shape-shifted Model: Functions

$$m_1(t) = 30(1-t)t^{1.2},\ m_2(t) = 30(1-t)t^{1.2} + \frac{\sin(20\pi t)}{3}$$

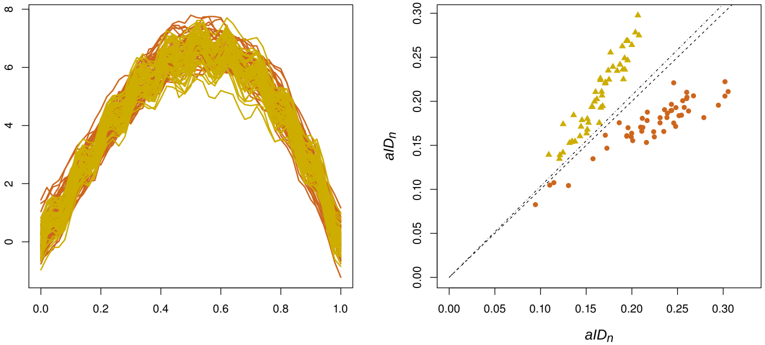


$$R_1(s,t) = 0.2\exp\left(-\frac{|s-t|}{0.3}\right),\ R_2(s,t) = 0.2\exp\left(-\frac{|s-t|}{0.3}\right)$$

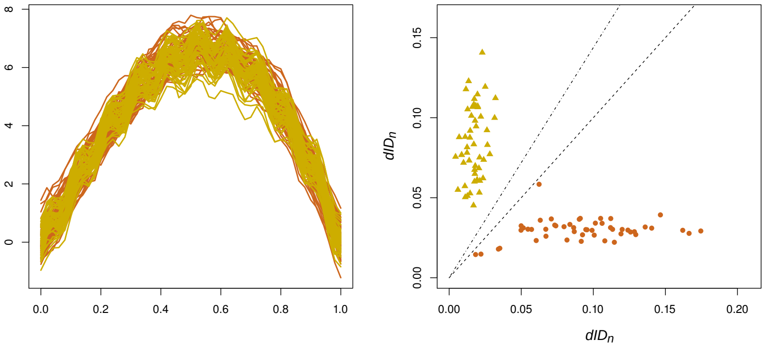
Shape-shifted Model: ID_n



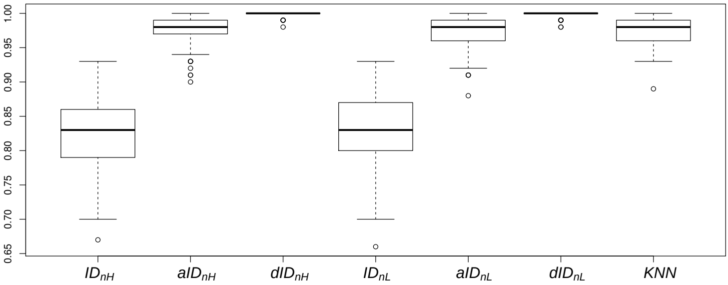
Shape-shifted Model: aID_n



Shape-shifted Model: dID_n

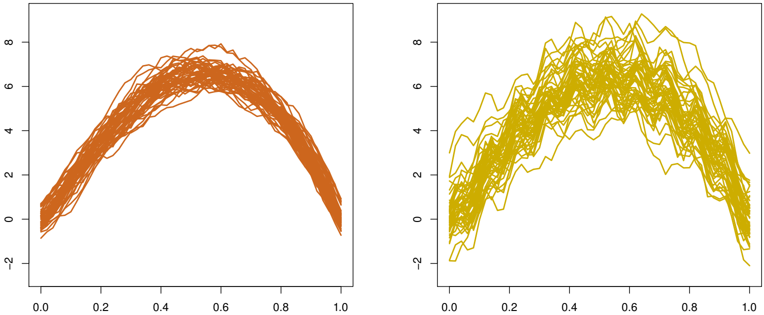


Shape-shifted Model: Results



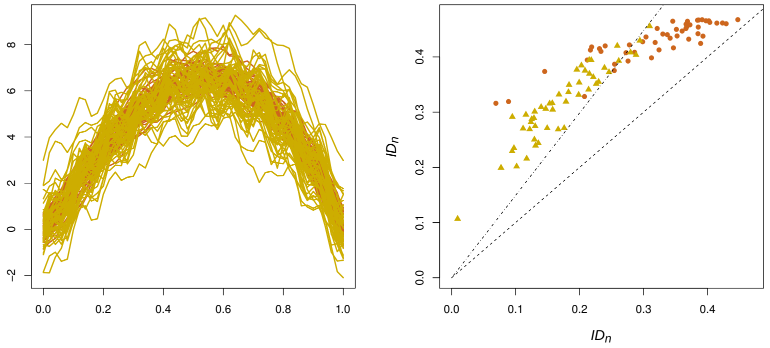
Variance Difference Model: Functions

$$m_1(t) = 30(1-t)t^{1.2},\ m_2(t) = 30(1-t)t^{1.2} + \frac{\sin(20\pi t)}{3}$$

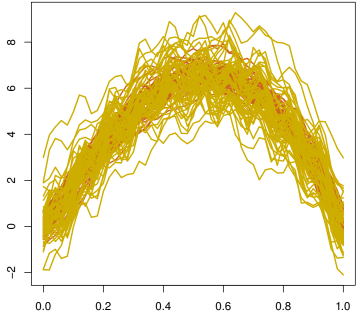


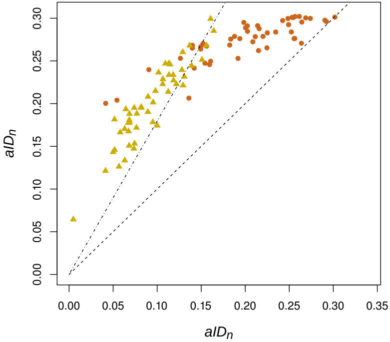
$$R_1(s,t) = 0.2\exp\left(-\frac{|s-t|}{0.3}\right),\ R_2(s,t) = \exp\left(-\frac{|s-t|}{0.3}\right)$$

Variance Difference Model: ID_n

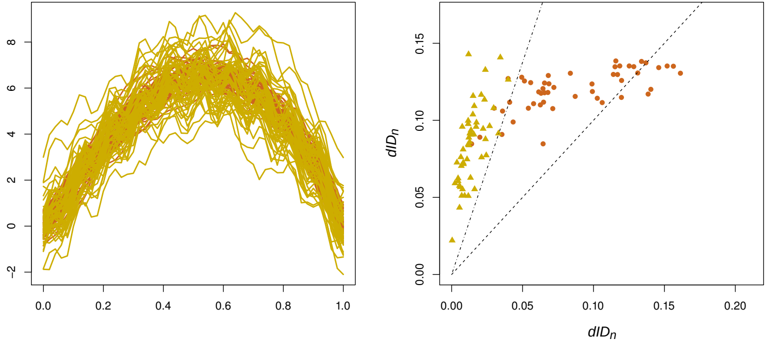


Variance Difference Model: aID_n

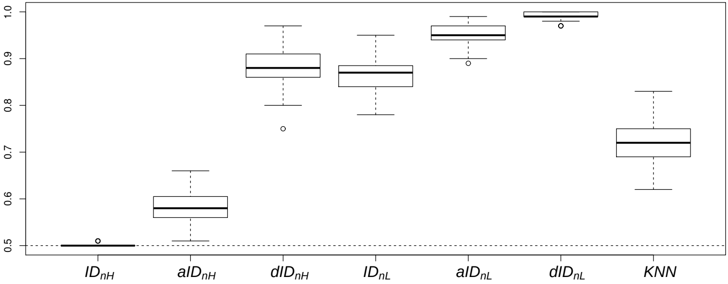




Variance Difference Model: dID_n



Variance Difference Model: Results



Outline

1. Depth Measure and its Smoothness for Multivariate Data
   - Smoothness of Halfspace Depth Contours
2. Functional Data Depth: Theory
   - Functional Band Depths
   - Consistency
   - Counterexample
   - Fixing the Continuousness
   - Integral and Vector Depths
3. Functional Data Depth: Practice
   - Problem of Functional Data Classification
   - Using Depth for Classification
   - Simulation Study
4. Conclusions

Depth-based Classification

How to choose a depth?

Depth-based Classification

How to choose a depth?

- **Band depths fail** in the case of noisy observations.

Depth-based Classification

How to choose a depth?

- **Band depths fail** in the case of noisy observations.
- **Fraiman-Muniz depths** are reasonable in the location-shifted model, but **fail in the case of shape shifted models**.

Depth-based Classification

How to choose a depth?

- **Band depths fail** in the case of noisy observations.
- **Fraiman-Muniz depths** are reasonable in the location-shifted model, but **fail in the case of shape shifted models**.
- **Depths including derivatives** with proper weights provide **good pattern recognition** for the location-shifted models as well as for the shape-shifted models.

Depth-based Classification

How to choose a depth?

- **Band depths fail** in the case of noisy observations.
- **Fraiman-Muniz depths** are reasonable in the location-shifted model, but **fail in the case of shape shifted models**.
- **Depths including derivatives** with proper weights provide **good pattern recognition** for the location-shifted models as well as for the shape-shifted models.

Depth-based Classification

How to choose a depth?

- **Band depths fail** in the case of noisy observations.
- **Fraiman-Muniz depths** are reasonable in the location-shifted model, but **fail in the case of shape shifted models**.
- **Depths including derivatives** with proper weights provide **good pattern recognition** for the location-shifted models as well as for the shape-shifted models.

In most of the non-trivial examples the K-derivative depths classify **better than the nearest neighbor methods**.

Depth-based Classification

How to choose a DD-plot analysis method?

Depth-based Classification

How to choose a DD-plot analysis method?

- **Highest depth rule** is reliable if the difference is caused by the mean function, but **fails in the variance difference setup**.

Depth-based Classification

How to choose a DD-plot analysis method?

- **Highest depth rule** is reliable if the difference is caused by the mean function, but **fails in the variance difference setup**.
- **Li's rules** identify the location and shape difference (if a proper depth is used) as well as the variance structure difference.

Depth-based Classification

How to choose a DD-plot analysis method?

- **Highest depth rule** is reliable if the difference is caused by the mean function, but **fails in the variance difference setup**.
- **Li's rules** identify the location and shape difference (if a proper depth is used) as well as the variance structure difference.

Depth-based Classification

How to choose a DD-plot analysis method?

- **Highest depth rule** is reliable if the difference is caused by the mean function, but **fails in the variance difference setup**.
- **Li's rules** identify the location and shape difference (if a proper depth is used) as well as the variance structure difference.

The nearest neighbor rule appears to be weak in comparison with Li's rules, mainly in the variance difference models.

Conclusions: Band Depths

As far as band depths are concerned, we have seen that:

- they provide **bad results in applications**,
- **are hard to be counted** ($\mathcal{O}\left(n^J\right)$ against $\mathcal{O}\left(n\right)$ for integral depths),
- **need not to be uniformly consistent**.

Conclusion

Avoid using band depths, aim for integral alternatives!

References I

- Noga Alon, Shai Ben-David, Nicolò Cesa-Bianchi, and David Haussler. Scale-sensitive dimensions, uniform convergence, and learnability. *J. ACM*, 44(4):615–631, 1997.
- Víctor de la Peña and Evarist Giné. *Decoupling. From dependence to independence. Randomly stopped processes, U-statistics and processes, martingales and beyond.* New York, NY: Springer. xv, 392 p., 1998.
- R. M. Dudley. *Uniform central limit theorems*, volume 63 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999.
- Ricardo Fraiman and Graciela Muniz. Trimmed means for functional data. *Test*, 10(2):419–440, 2001.
- Daniel Hlubinka and Stanislav Nagy. Functional data depth and classification. *submitted to Comput. Stat.*
- V. S. Koroljuk and Yu. V. Borovskich. *Theory of U-statistics*, volume 273 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994. Translated from the 1989 Russian original by P. V. Malyshev and D. V. Malyshev and revised by the authors.
- Regina Y. Liu. On a notion of data depth based on random simplices. *Ann. Statist.*, 18(1):405–414, 1990.

References II

- Sara López-Pintado and Juan Romo. On the concept of depth for functional data. *J. Amer. Statist. Assoc.*, 104(486):718–734, 2009.
- Stanislav Nagy. Distributions with Smooth Density and Non-Smooth Tukey Depth Contours. *preprint.*
- Stanislav Nagy and Daniel Hlubinka. Depth for Vector-Valued Functions and Some of its Basic Properties. *preprint.*
- John W. Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*, pages 523–531. Canad. Math. Congress, Montreal, Que., 1975.
- Ramon van Handel. The universal Glivenko-Cantelli property. *Probab. Th. Rel. Fields, to appear.*
- Yijun Zuo and Robert Serfling. General notions of statistical depth function. *Ann. Statist.*, 28(2):461–482, 2000.