Data Depth

Stanislav Nagy

Department of Probability and Mathematical Statistics
Charles University in Prague


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

How to define ordering of these data?
Consider a random variable $X \sim P \in \mathcal{P}(\mathbb{R}^d)$.

Using data depth!
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
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 \((\text{Tukey [11]})\, HD\) of an observation from \(\mathbb{R}^d\)

\[
HD(x; P) = \inf_{H \in \mathcal{H}(x)} P(X \in H)
\]
Halfspace depth (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

\[ HD(x; X_1, \ldots, X_n) = \text{least ratio of observations in a halfspace containing } x \]
Halfspace Depth

\[ HD(x; X_1, \ldots, X_n) = \text{least ratio of observations in a halfspace containing } x \]
Halfspace Depth

\[ HD(x; X_1, \ldots, X_n) = \text{least ratio of observations in a halfspace containing } x \]
Halfspace Depth

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

\[ HD(x; X_1, \ldots, X_n) = \text{least ratio of observations in a halfspace containing } x \]
Halfspace Depth

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

$HD(x; X_1, \ldots, X_n) = \text{least ratio of observations in a halfspace containing } x$
Depth Measure and its Smoothness for Multivariate Data
Functional Data Depth: Theory
Functional Data Depth: Practice
Conclusions

**Halfspace Depth**

\[ HD(x; X_1, \ldots, X_n) = \text{least ratio of observations in a halfspace containing } x \]
Halfspace Depth

\[ HD(x; X_1, \ldots, X_n) = \text{least ratio of observations in a halfspace containing } x \]
Halfspace Depth

\[ HD(x; X_1, \ldots, X_n) = \text{least ratio of observations in a halfspace containing } x \]
Halfspace Depth

\[ HD(x; X_1, \ldots, X_n) = \text{least ratio of observations in a halfspace containing } x \]
Simplicial depth \((\text{Liu} \ [7])\) \(SD\) 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 \( \mathbb{R}^d \)

\[
SD(x; P) = P \left( x \in S_{X_1, \ldots, X_{d+1}} \right)
\]
Simplicial Depth

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

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

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

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

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

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

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

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

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

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

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

Is $P \in \mathcal{P}(\mathbb{R}^d)$ characterized by $\{HD(x; P) \mid x \in \mathbb{R}^d\}$?
Is $P \in \mathcal{P}(\mathbb{R}^d)$ characterized by $\{HD(x; P) \mid x \in \mathbb{R}^d\}$? **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 $C^{(2)}$ **density** (Hassairi and Regaieg 2008),
- the **HD contours are smooth** (Kong and Zuo 2010).
Is $P \in \mathcal{P}(\mathbb{R}^d)$ characterized by $\{HD(x; P) \mid x \in \mathbb{R}^d\}$?

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 $C^{(2)}$ density (Hassairi and Regaieig 2008),
- the HD contours are smooth (Kong and Zuo 2010).

When are HD contours smooth?
Theorem:

Let \( P \in \mathcal{P}(\mathbb{R}^d) \) 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

\[
HD(x; P) = P(X \in H).
\]
When are Halfspace Depth Contours Smooth?

**Theorem:**

Let \( P \in \mathcal{P}(\mathbb{R}^d) \) 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

\[
HD(x; P) = P(X \in H).
\]

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

$$D : C([0,1]) \times \mathcal{P}(C([0,1])) \to [0,1].$$
López-Pintado and Romo [8] for $J = 2, 3, \ldots$

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

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

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

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

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

Depth $D$ is on a set $S \subset C([0, 1])$ consistent pointwise if

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

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

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

- uniformly if
  
  $$\sup_{x \in S} |D(x; P_n) - D(x; P)| \xrightarrow[n \to \infty]{} 0,$$
Depth $D$ is on a set $S \subset C([0, 1])$ **consistent**

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

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

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

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

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

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

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

- **$\mathcal{P}$-uniformly** if
  \[ \sup_{P \in \mathcal{P}(C([0,1]))} \sup_{x \in S} |D(x; P_n) - D(x; P)| \xrightarrow{a.s. \ n \to \infty} 0. \]
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)
- ...
López-Pintado and Romo [8, Thm 4]

**Theorem:**

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

$$\sup_{x \in S} \left| BD^J(x; P_n) - BD^J(x; P) \right| \xrightarrow{a.s.} 0.$$
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 \( BD_J(\cdot; P) \) is for \( P \) with a.c. marginals an **continuous** functional, it is enough to prove for \( N \in \mathbb{N} \) fixed

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

This holds, since \( BD_J(\cdot; P_n) \) is a bounded U-statistic. \( \square \)
Why the Proof Does Not Work?

$BD^J(\cdot; P)$ is continuous, but $BD^J(\cdot; P_n)$ 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?

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

does not give uniform convergence!
Why the Proof Does Not Work?

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

does not give uniform convergence!
Why the Proof Does Not Work?

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

does not give uniform convergence!

![Graph showing the behavior of the depth measure and its convergence](image-url)
Is Band Depth Consistent?

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

- The validity of
  \[ \dim_{VC} \{(x_1, x_2) | G(x) \subset B(x_1, x_2)\}_{x \in S} = \infty \]
  for $S \subset C([0, 1])$ compact suggests, that the depth is not $\mathcal{P}$-uniformly consistent (Assouad’s Thm - [3, Thm 6.4.5]).

- The existence of boolean $\sigma$-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]).
Define $X \sim P \in \mathcal{P}(C([0, 1]))$ as follows:

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

![Graph showing a horizontal line at 0.5]
Define $X \sim P \in \mathcal{P}(C([0, 1]))$ as follows:

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

```
0.0 0.5
0 0.5 0.75 1
```

Stanislav Nagy
Band Depth Consistence: Counterexample

Define $X \sim P \in \mathcal{P}(C([0, 1]))$ 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.
Let $x_j$ be a function with a single jump on the interval $I_j$, 0 otherwise. Then:

\[ BD^2(x_j; P) = 0.25 \text{ for all } j \in \mathbb{N} \]
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

$$\left( \begin{array}{c} n \\ 2 \end{array} \right) - 2 \left( \begin{array}{c} n/2 \\ 2 \end{array} \right) = \frac{n^2}{4}$$

bands.
- For such a $j_n$ we have

$$BD^{J}(x_{j_n}; P_n) = \frac{n^2}{\left( \begin{array}{c} n \\ 2 \end{array} \right)} = \frac{n}{2(n-1)} \xrightarrow{n \to \infty} 0.5.$$
Band Depth Consistence: Counterexample

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

- $BD^{2j} (x_j; P) = 0.25$ 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

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

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

But does exist infinitely many of such couples $(n, j_n)$?
Band Depth Consistence: Counterexample

But does exist \textit{infinitely many} of such couples \((n,j_n)\)? \textbf{Yes!}

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

Stanislav Nagy
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 \(l_{j_n}\) 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

\[
\sup_{j \in \mathbb{N}} \left| BD^2) (x_j; P_n) - BD^2) (x_j; P) \right| > 0.25 - \varepsilon
\]
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 $BD^J(\cdot; P_n)$ 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 $C([0,1])$ use

$$E[1 - w(d(x; B(X_1, X_2)))]$$

instead of

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

where $w: [0, \infty) \rightarrow [0, 1], w(0) = 1, \lim_{t \rightarrow \infty} w(t) = 0$ 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 $S \subset C([0,1])$ relatively compact. Then the band depths smoothed by $w$

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

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

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

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]

\[ ID(x; P) = \int_0^1 D(x(t); P_t) \, dt, \]

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**

\[ D(x(t); P_t) = F_t(x(t)) (1 - F_t(x(t))). \]
The idea of Fraiman and Muniz may be easily generalized to vector-valued functions

\[ ID(x; P) = \int_0^1 D(x(t); P_t) \, dt, \]

where \( D \) is usual multivariate depth,

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

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

This is how we define \textbf{K-vector depths} and by application to differentiable functions also \textbf{K-derivatives depths} (Hlubinka and Nagy [5]).
Integral Depths Consistency

Theorem:

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

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

is universally consistent on $C([0, 1])^d$, 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 $D$. Stanislav Nagy
A range of other properties of integral depth for vector-valued functions can be proved (Nagy and Hlubinka [10]):

- **measurability** as a functional on \( C([0,1])^K \times \mathcal{P}(C([0,1])^K) \).
A range of other properties of integral depth for vector-valued functions can be proved (Nagy and Hlubinka [10]):

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

Stanislav Nagy

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

- **measurability** as a functional on $C([0,1])^K \times \mathcal{P}(C([0,1])^K)$,
- functional version of **affine invariance** for $ID$ and $dID$,
- **monotonicity relative to the deepest point**,
A range of other properties of integral depth for vector-valued functions can be proved (Nagy and Hlubinka [10]):

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

Stanislav Nagy

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

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

Integral depths for vector functions

\[ ID(x; P) = \int_0^1 D((x_1(t), x_2(t)); (P_{1,t}, P_{2,t})) \, dt, \]

10% of deepest functions
10% of least deep functions
K-Derivatives Depth

Integral depths for differentiable functions

\[ d\text{ID}(x; P) = \int_0^1 D\left((x(t), x'(t)) ; (P_t, P'_t)\right) \, dt, \]
Consider now the **contaminated functional dataset**. Does the depth recognize the outlier?
K-Derivatives Depth Again

Integral depths for differentiable functions

\[ d\text{ID}(x; P) = \int_0^1 D \left( \left( x(t), x'(t) \right); (P_t, P'_t) \right) \, dt, \]
Future Challenges

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

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

for $S \subset C([0, 1])$ 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
Let $P_1, P_2 \in \mathcal{P}(C([0,1]))$ and $X \sim P_m$, $m \in \{1,2\}$ is unknown. What is the distribution of $X$?
Let $P_1, P_2 \in \mathcal{P}(C([0,1]))$ 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 $C([0, 1])$ (e.g. $L_2$, $k = 5$):

![Graphs showing nearest neighbor rule](image)
For given training samples $X_1, X_2$ and depth $D$, the **DD-transformation** of data can be computed as

$$DD: C([0, 1]) \rightarrow \mathbb{R}^2: x \mapsto (D(x; X_1), D(x; X_2))^T$$
The function is assigned to the sample with highest depth value 
\[ \text{arg } \max_{i=1,2} D(x; \overline{X}_i) \] (Cuevas et al. 2007)
An increasing **best separating** function (linear, or polynomial) is utilized to classify the DD-transformations (Li et al. 2010)
Location-shifted Model: Functions

\[ m_1(t) = 30 (1 - t) t^{1.2}, \quad m_2(t) = 30t (1 - t)^{1.2} \]

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

- $BD_{nH}^{2}$
- $BD_{nH}^{3}$
- $ID_{nH}$
- $aID_{nH}$
- $dID_{nH}$
- $BD_{nL}^{2}$
- $BD_{nL}^{3}$
- $ID_{nL}$
- $aID_{nL}$
- $dID_{nL}$
- KNN

Box plots illustrating the results for different methods.
Location-shifted Model: Results 2

![Box plots comparing different methods]

- $ID_{nH}$
- $aID_{nH}$
- $dID_{nH}$
- $ID_{nL}$
- $aID_{nL}$
- $dID_{nL}$
- KNN

Stanislav Nagy  
Depth
Shape-shifted Model: Functions

\[ m_1(t) = 30 \left( 1 - t \right) t^{1.2}, \quad m_2(t) = 30 \left( 1 - t \right) t^{1.2} + \frac{\sin(20\pi t)}{3} \]

\[ R_1(s, t) = 0.2 \exp \left( -\frac{|s-t|}{0.3} \right), \quad R_2(s, t) = 0.2 \exp \left( -\frac{|s-t|}{0.3} \right) \]
Shape-shifted Model: $ID_n$
Problem of Functional Data Classification
Using Depth for Classification
Simulation Study

Shape-shifted Model: $aID_n$
Shape-shifted Model: \( dID_n \)
Shape-shifted Model: Results
Variance Difference Model: Functions

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

\[ R_1(s, t) = 0.2 \exp\left(-\frac{|s-t|}{0.3}\right), \quad R_2(s, t) = \exp\left(-\frac{|s-t|}{0.3}\right) \]
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.
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.
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.
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.
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.
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.
As far as band depths are concerned, we have seen that:

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

**Conclusion**

*Avoid using band depths, aim for integral alternatives!*
References I


