FUNCTIONAL DEPTH: RECENT PROGRESS AND PERSPECTIVES

Stanislav Nagy

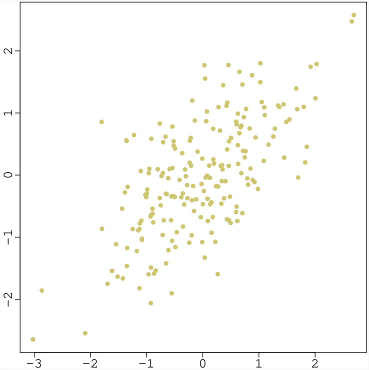
Valladolid 2021

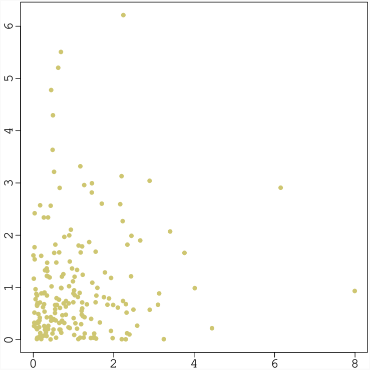
Charles University
Department of Probability and Mathematical Statistics

Statistical depth

For Borel probability measures $\mathcal{P}\left(\mathbb{R}^d\right)$ the statistical depth is a mapping

$$D\colon \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) \to [0,1]\colon (x,P) \mapsto D(x;P).$$

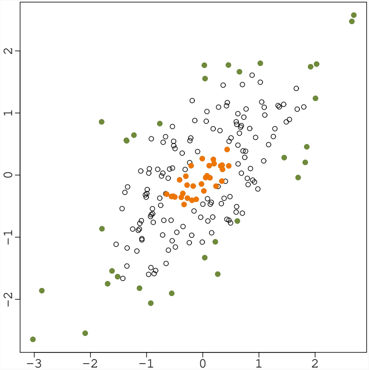


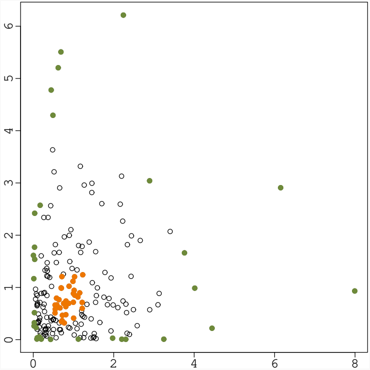


Statistical depth

For Borel probability measures $\mathcal{P}\left(\mathbb{R}^d\right)$ the statistical depth is a mapping

$$D\colon \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) \to [0,1]\colon (x,P) \mapsto D(x;P).$$

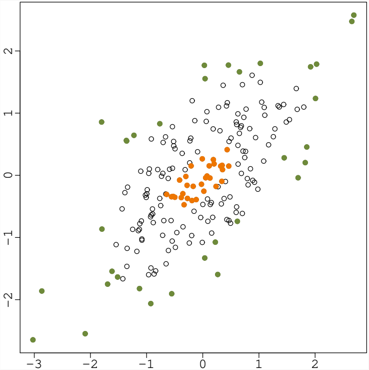


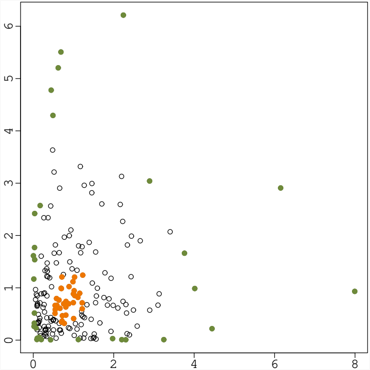


Halfspace depth

Halfspace depth or Tukey depth (Tukey, 1975) of $x \in \mathbb{R}^d$

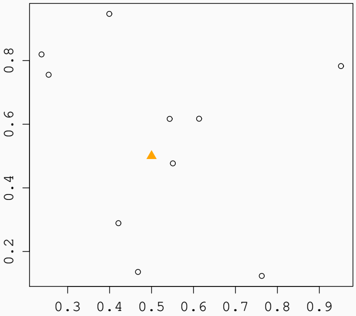
$$hD(x; P) = \inf_{H \in \mathcal{H}(x)} P(H).$$





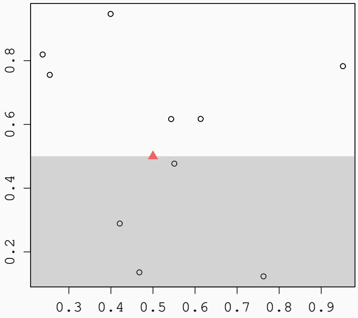
Halfspace depth

$$hD\left(x; \{X_1, \ldots, X_n\}\right) = \frac{\text{\# of observations in a halfspace that contains } x}{n}$$



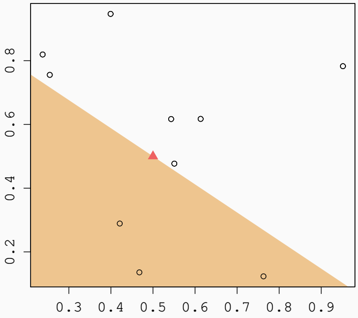
Halfspace depth

$$hD\left(x; \{X_1, \ldots, X_n\}\right) = \frac{\text{\# of observations in a halfspace that contains } x}{n}$$



Halfspace depth

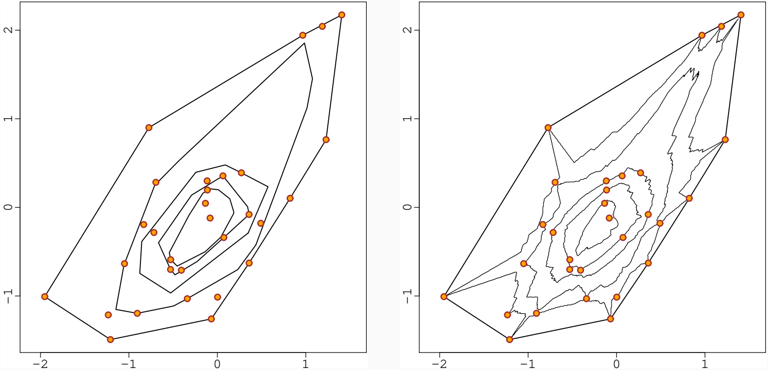
$$hD\left(x; \{X_1, \ldots, X_n\}\right) = \frac{\text{\# of observations in a halfspace that contains } x}{n}$$



Halfspace/Simplicial depth central regions

Halfspace depth contours (left) and simplicial depth contours (right)

$$hD(x; P) = \inf_{H \in \mathcal{H}(x)} P(H), \qquad sD(x; P) = \mathsf{P}(x \in \mathbf{S}[X_1, \ldots, X_{d+1}]).$$



Depth: Desired properties (informally)

A depth $D\colon \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to [0,1]\colon\ (x,P) \mapsto D(x;P)$ should be (Zuo and Serfling, 2000; Serfling, 2006):

(P1) *Invariant for affine transforms*;

(P2) *Maximal at the center of symmetry of P*;

(P3) *Decreasing along rays from the center*;

(P4) *Vanishing as x goes to infinity*;

(P5) *Semi-continuous in x*;

(P6) *Continuous in P*;

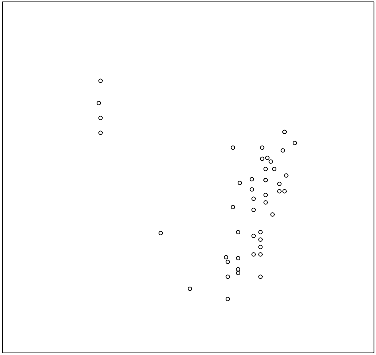
and sometimes also

(P3') *Quasi-concave in x*: All upper level sets of $D(\cdot;P)$ are convex.

The depth then ranks the data reasonably well.

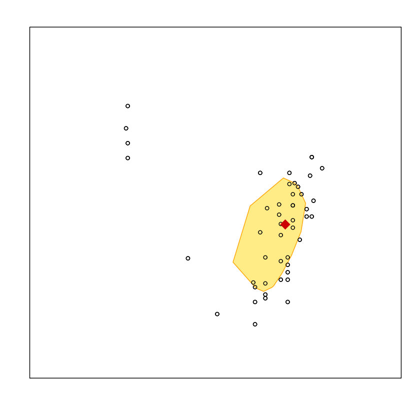
Application: Bagplot

Bagplot — depth-based boxplot in $\mathbb{R}^d$ (Rousseeuw et al., 1999)



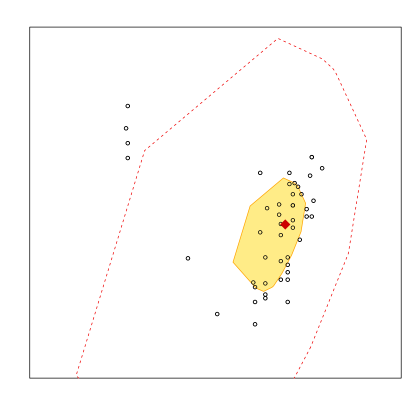
Application: Bagplot

Bagplot — depth-based boxplot in $\mathbb{R}^d$ (Rousseeuw et al., 1999)



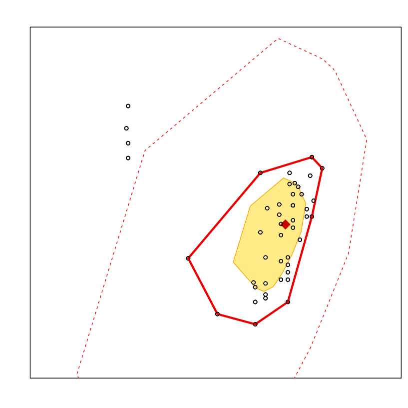
Application: Bagplot

Bagplot — depth-based boxplot in $\mathbb{R}^d$ (Rousseeuw et al., 1999)



Application: Bagplot

Bagplot — depth-based boxplot in $\mathbb{R}^d$ (Rousseeuw et al., 1999)



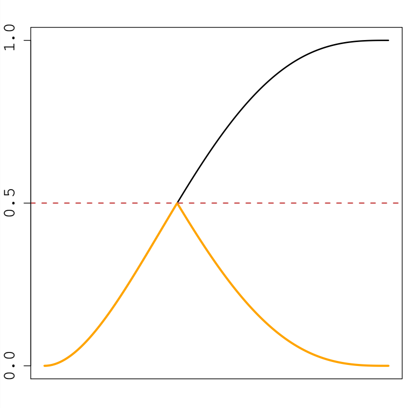
Functional data

$X \sim P \in \mathcal{P}(\mathcal{F})$ and $X_1, \ldots, X_n$ i.i.d. from P. Consider the depth of functional observations w.r.t. P (or P_n the empirical measure of $X_1, \ldots, X_n$)

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

Halfspace depth in $\mathbb{R}$

$$hD_1(u; Q) = \min\{\mathrm{F}_Q(u), 1 - \mathrm{F}_Q(u-)\} \approx 1/2 - |1/2 - \mathrm{F}_Q(u)|$$



Depth in function spaces

For $\mathcal{F}$ a Banach space and $X \sim P \in \mathcal{P}(\mathcal{F})$, what is the depth of $x \in \mathcal{F}$?

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

- For the halfspace depth, only the linear structure of $\mathbb{R}^d$ is needed:

$$\begin{aligned} hD(x; P) &= \inf_{u \in \mathbb{R}^d} P\left(\left\{y \in \mathbb{R}^d \colon \langle y, u \rangle \leq \langle x, u \rangle\right\}\right) \\ &= \inf_{u \in \mathbb{R}^d} hD_1\left(\langle x, u \rangle ; P_{\langle X, u \rangle}\right). \end{aligned}$$

- The simplicial depth in $\mathbb{R}^d$ depends on d, the dimension of the space.

Depth in function spaces

For $\mathcal{F}$ a Banach space and $X \sim P \in \mathcal{P}(\mathcal{F})$, what is the depth of $x \in \mathcal{F}$?

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

- Functional halfspace depth: for $\mathcal{F}^*$ the dual space of $\mathcal{F}$
  (Dutta et al., 2011)

$$\begin{aligned} hD(x;P) &= \inf_{\varphi \in \mathcal{F}^*} P\left(\{y \in \mathcal{F} \colon \varphi(y) \leq \varphi(x)\}\right) \\ &= \inf_{\varphi \in \mathcal{F}^*} hD_1\left(\varphi(x); P_{\varphi(X)}\right). \end{aligned}$$

- The simplicial depth does not work directly in function spaces.

Note: An extension of the simplicial depth to functional data is the band depth.
(López-Pintado and Romo, 2009)

Functional halfspace depth in $L^2(\mathcal{T})$

Each functional datum lives in its own dimension:

Observation
For a random sample $X_1, \ldots, X_n$ of truly infinite-dimensional functional data, X_n lies outside of the convex hull of $X_1, \ldots, X_{n-1}$, almost surely.

The Hahn-Banach theorem then implies that the sample functional halfspace depth is constant zero, P-almost everywhere.

Observation
The functional halfspace depth degenerates.

Halfspace depth degenerates

For (certain) Gaussian processes $P \in \mathcal{P}(\mathcal{F})$ we have that (Chakraborty and Chaudhuri, 2013)

$$hD(x; P) = \inf_{\varphi \in \mathcal{F}^*} hD_1\left(\varphi(x); P_{\varphi(X)}\right) = 0 \text{ for } P\text{-almost all } x \in \mathcal{F}.$$

Also other functional depths, e.g. the projection depths (Zuo and Serfling, 2000), degenerate too.

Condition 0. Depth should not degenerate. That is, it is not allowed that for some $P \in \mathcal{P}(\mathcal{F})$ we have $D(x; P) = 0$ for P-almost all $x \in \mathcal{F}$.

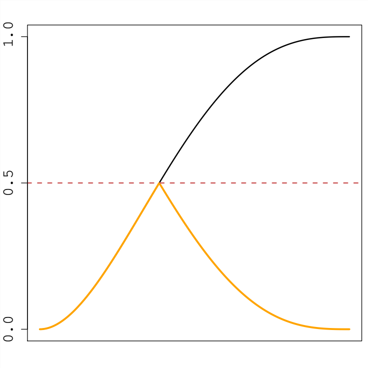
➔ Restrict the set of projections in hD from the dual $\mathcal{F}^*$ to a smaller, but still representative and well interpretable subset.

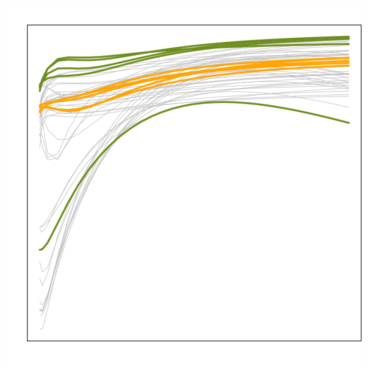
Integrated depths

Average depth of a functional value

(Fraiman and Muniz, 2001; Cuevas and Fraiman, 2009; López-Pintado and Romo, 2009)

$$FD\,(x;P) = \int_{\mathcal{T}} D_1(x(t),P_t)\,\mathrm{d}\,t, \qquad D_1(u;Q) = 1/2 - |1/2 - \mathrm{F}_Q(u)|\,.$$

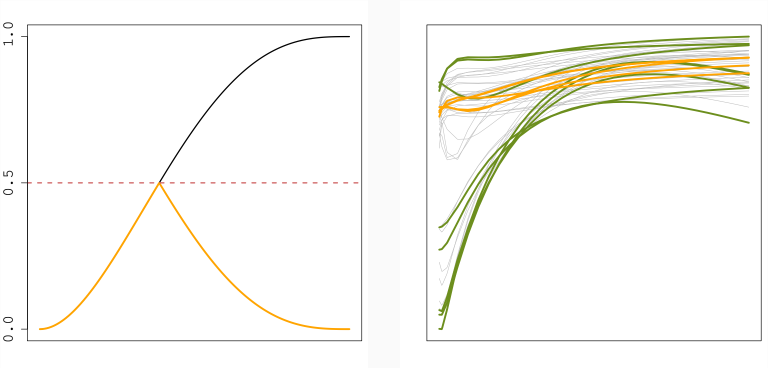




Infimal depths

Smallest depth of a functional value (Mosler, 2013; Narisetty and Nair, 2016)

$$ID\,(x;P) = \inf_{t\in\mathcal{T}} D_1\,(x(t);P_t)\,, \qquad D_1(u;Q) = 1/2 - |1/2 - \mathrm{F}_Q(u)|\,.$$



Functional depth

Basic types of depth for functional data:

- integrated depth

$$FD\,(x;P) = \int_{\mathcal{T}} D_1(x(t), P_t)\,\mathrm{d}\,t,$$

- infimal depth

$$ID\,(x;P) = \inf_{t \in \mathcal{T}} D_1\,(x(t);P_t)\,.$$

General functional depth

For a Banach space $\mathcal{F}$, $P \in \mathcal{P}(\mathcal{F})$, $\Phi \subset \mathcal{F}^*$, and λ a measure on Φ:

- integrated depth

$$FD(x; P) = \int_{\Phi} D_1(\varphi(x), P_{\varphi(X)}) \,\mathrm{d}\,\lambda(\varphi),$$

- infimal depth

$$ID(x; P) = \inf_{\varphi \in \Phi} D_1\left(\varphi(x), P_{\varphi(X)}\right).$$

The set $\Phi \subset B^*$ is typically the collection of evaluation functionals

$$\{\varphi_t \colon x \mapsto x(t) \colon t \in \mathcal{T}\},$$

but not necessarily so. λ can be the Lebesgue measure on $\mathcal{T}$.

Degeneracy problem

Condition 0. Depth should not degenerate. That is, it is not allowed that $D(x; P) = 0$ for P-almost all $x \in \mathcal{F}$ for any $P \in \mathcal{P}(\mathcal{F})$.

The integrated depth does not degenerate, but the infimal depth *almost* does.

Example: Consider $X \sim P \in \mathcal{P}(\mathcal{C}([0, 1]))$ given as the linear interpolant of

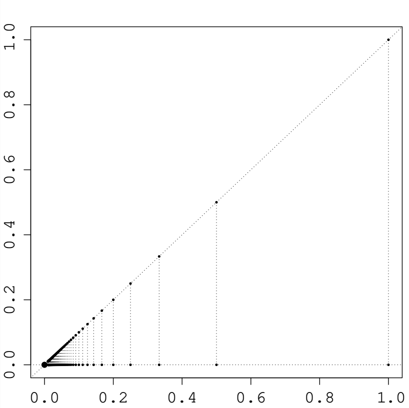
- $X(0) = 0$, and
- $X(1/m) = \mathrm{Bernoulli}(1/2)/m$ independent for $m = 1, 2, \ldots$.

Then $ID(x; P_n) = 0$ for P-almost all $x \in \mathcal{C}([0, 1])$, almost surely.

Infimal depths: Degeneracy problem

For $X \sim P$ our randomly jumping function and any $x \in \mathcal{C}([0,1])$

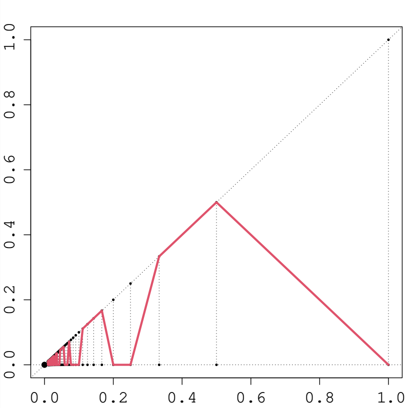
$$ID(x; P) \geq 1/4 \times \mathbb{I}\{0 \leq x(t) \leq t \text{ for all } t \in [0,1]\}$$



Infimal depths: Degeneracy problem

For $X_1, \ldots, X_n$ a random sample from P with empirical measure P_n

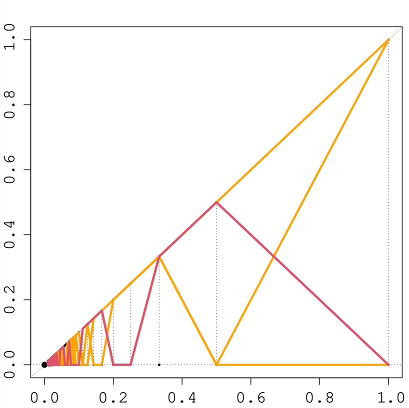
$$ID(x; P_n) = 0 \text{ for } P\text{-almost all } x \in \mathcal{C}([0,1]).$$



Infimal depths: Degeneracy problem

For $X_1, \ldots, X_n$ a random sample from P with empirical measure P_n

$$ID(x; P_n) = 0 \text{ for } P\text{-almost all } x \in \mathcal{C}([0,1]).$$



Infimal depths: A new degeneracy problem

In our example, for P-almost any $x \in \mathcal{C}([0,1])$ with $0 \leq x(t) \leq t$ for all $t \in [0,1]$ we have

$$ID(x; P) \geq 1/4,$$

but

$$ID(x; P_n) = 0 \text{ for all } n = 1, 2, \dots, \text{ almost surely.}$$

➔ The estimator of the depth does not work, i.e.

$$\lim_{n \to \infty} ID(x; P_n) \neq ID(x; P).$$

Condition 1. The depth must possess a consistent sample version, at least for *reasonable* distributions.

Depth distribution

Consider the depth distribution of $x \in L^2(\mathcal{T})$, that is the law of

$$D_x^P \colon \left(\mathcal{T}, \mathcal{B}(\mathcal{T}), \lambda\right) \to [0,1] \colon t \mapsto hD\left(x(t); P_t\right)$$

being a random variable on $\mathcal{T}$.

- The integrated depth is the mean of D_x^P

$$FD(x; P) = \int_{\mathcal{T}} hD\left(x(t); P_t\right) \,\mathrm{d}\,\lambda(t) = \mathsf{E}\, D_x^P.$$

- The infimal depth is the (essential) infimum of D_x

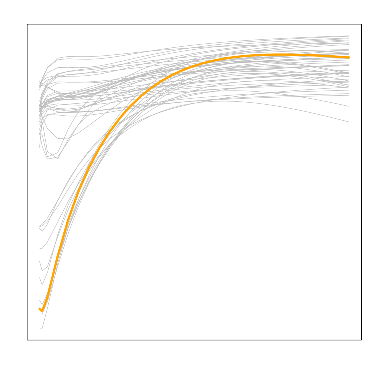
$$ID(x; P) = \inf_{t \in \mathcal{T}} hD\left(x(t); P_t\right),$$

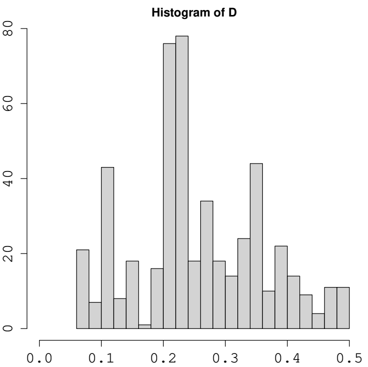
  that is the lower end-point of the support of D_x^P.

Depth distribution

The depth distribution of $x \in L^2(\mathcal{T})$ w.r.t. the random sample

$$D_x^P : \left(\mathcal{T}, \mathcal{B}(\mathcal{T}), \lambda\right) \to [0,1] : t \mapsto hD\left(x(t); P_t\right)$$

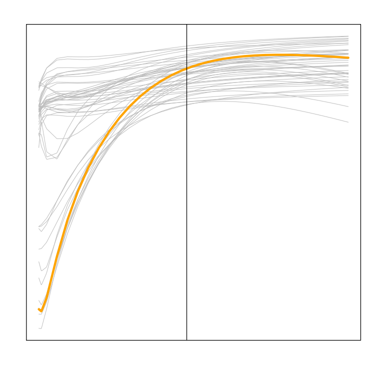


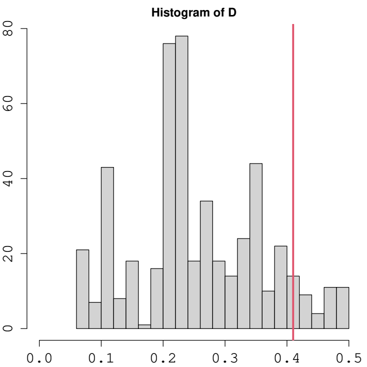


Depth distribution

The depth distribution of $x \in L^2(\mathcal{T})$ w.r.t. the random sample

$$D_x^P : \left(\mathcal{T}, \mathcal{B}(\mathcal{T}), \lambda\right) \to [0,1] : t \mapsto hD\left(x(t); P_t\right)$$

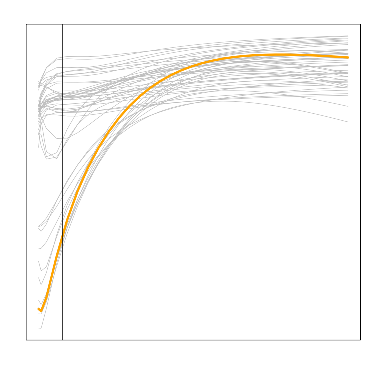


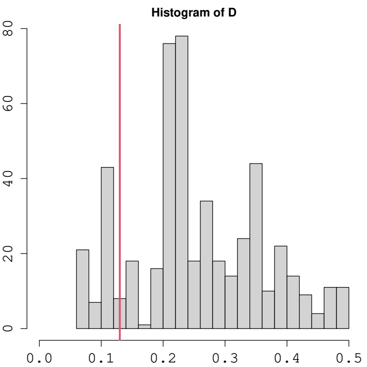


Depth distribution

The depth distribution of $x \in L^2(\mathcal{T})$ w.r.t. the random sample

$$D_x^P\colon\ (\mathcal{T}, \mathcal{B}(\mathcal{T}), \lambda) \to [0,1]\colon t \mapsto hD\left(x(t); P_t\right)$$





Adaptive feature choice: Depth distribution

The k-integrated depth with $k \in \mathbb{R} \setminus \{0\}$

$$D^k(x;P) = \left(\int_{\mathcal{T}} \left(hD(x(t);P_t) + 1/2\right)^k \,\mathrm{d}\,\lambda(t) \right)^{1/k} - 1/2$$
$$= \left(\mathsf{E}\left(D_x^P + 1/2\right)^k \right)^{1/k} - 1/2$$

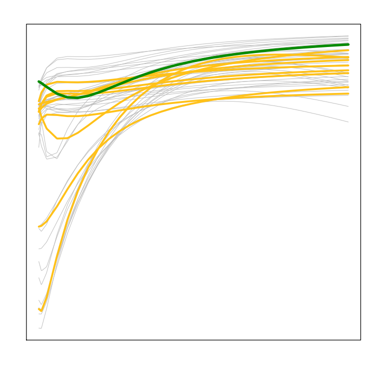
is, basically, the k-th moment of the depth distribution of x. We obtain a family of depths

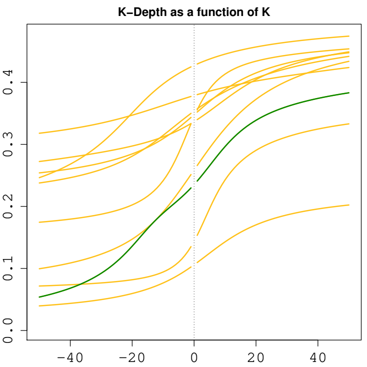
- for $k = 1$ the usual integrated depth;
- as $k \to -\infty$ a version of the infimal depth;
- choice of k allows us to fine tune in practice.

(Nagy, Helander, Van Bever, Viitasaari, and Ilmonen, 2021)

Trajectories of the k-integrated depths

The trajectories $k \mapsto D^k(x;P) = \left(\mathrm{E}\left(D_x^P + 1/2\right)^k\right)^{1/k} - 1/2$





General functional depths?

One can choose any (location) parameter L of the depth distribution

$$D_L(x; P) = L(D_x^P)$$

to obtain a custom tailored depth functional. Examples are

- quantiles,
- trimmed means,
- M-estimators...

Or integrated quantiles for $q \in (0, 1)$ and $F_{x,P}$ the c.d.f. of D_x^P

$$L(D_x^P) = \int_0^q F_{x,P}^{-1}(u) \,\mathrm{d}\, u.$$

(Work in progress with López-Pintado, 2021+)

The resulting depths possess quite different properties.

Case in point: Sample version consistency and Condition 1.

A theoretical issue: Consistency

Let $P_n \in \mathcal{P}(\mathcal{F})$ be the (random) empirical measure of a random sample $X_1, \ldots, X_n$ from P.

A depth D on space $\mathcal{F}$ is

- consistent if

$$D(x; P_n) \xrightarrow[n\to\infty]{\text{a.s.}} D(x; P) \quad \text{for all } x \in \mathcal{F};$$

- uniformly consistent if

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

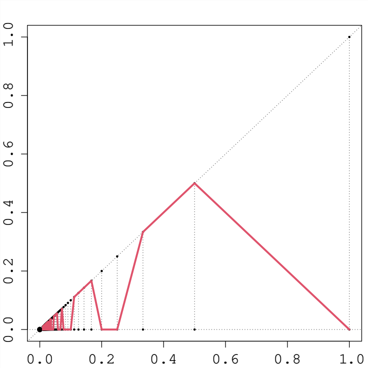
➔ In $\mathcal{F} = \mathbb{R}^d$, the halfspace / simplicial depth is uniformly consistent (empirical processes and the Vapnik-Červonenkis theory).

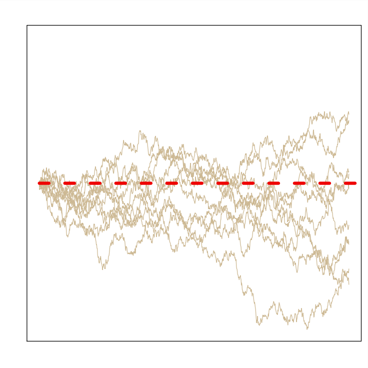
➔ In function spaces uniform consistency requires new theories.

➔ Functional depths are often not consistent uniformly.

Infimal (quantile) depths are not consistent

ID is not consistent for, e.g., *P* the Wiener measure.





ID can be shown to be consistent under more restrictive conditions.
(Gijbels and Nagy, 2015)

Uniform consistency of general depths

Theorem (Nagy and López-Pintado, 2021+)

Suppose that

- $\mathcal{F} = \mathcal{C}(\mathcal{T})$, *and*
- *the functional* $L\colon \mathcal{P}(\mathcal{T}) \to [0,1]$ *is uniformly continuous for the weak topology on* $\mathcal{P}(\mathcal{T})$.

Then the general functional depth based on L *is uniformly consistent.*

Corollary

- *All (k-)integrated depths are uniformly consistent over* $\mathcal{F}$, *for any* $P \in \mathcal{P}(\mathcal{F})$, *for both* $\mathcal{F} = L^2(\mathcal{T})$ *and* $\mathcal{F} = \mathcal{C}(\mathcal{T})$. *(Nagy et al., 2016; 2021)*
- *All integrated quantile depths are uniformly consistent over* $\mathcal{C}(\mathcal{T})$, *for* $P \in \mathcal{P}(\mathcal{C}(\mathcal{T}))$ *with smooth marginals. (Nagy and López-Pintado, 2021+)*

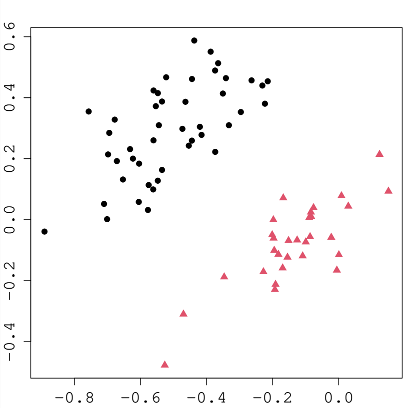
General functional depth: Properties

Theoretical properties of general functional depths, under appropriate assumptions on $L\colon \mathcal{P}(\mathcal{T}) \to [0,1]$:

- Non-degeneracy (**Condition 0**), including quantitative versions;
- Maximality at the coordinate-wise median, reachable by continuous functions;
- Invariance and monotonicity properties;
- (Semi-)Continuity in both $x \in \mathcal{F}$ and $P \in \mathcal{P}(\mathcal{F})$;
- Uniform consistency also for imperfectly observed functional data, multivariate functional data, image and video data.

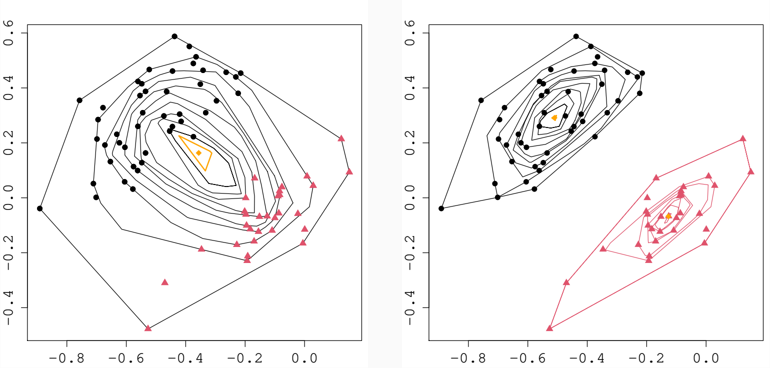
All this is true for both (k-)integrated depths and quantile integrated depths. (Nagy et al., 2016, 2021; Nagy and López-Pintado, 2021+)

The depth suits well only for analyzing unimodal distributions



Depth (in $\mathbb{R}^d$) is not for mixtures

The depth suits well only for analyzing unimodal distributions



Local/density-like depths

One further depth for functional data:

- ***h*-depth** (Cuevas et al., 2006)

$$D_\kappa(x; P) = \mathrm{E}_{X\sim P}\left[\kappa(\|x - X\|)\right]$$

  estimated by

$$D_\kappa(x; P_n) = n^{-1} \sum_{i=1}^{n} \kappa(\|x - X_i\|).$$

  Here, $\kappa\colon [0, \infty) \to [0, 1]$ is a continuous, non-increasing function with $\lim_{u\to\infty} \kappa(u) = 0$.

➔ A **density-like depth** allowing for multiple "modes" in P.

h-depth in machine learning

Observation (Wynne and Nagy, 2021)
For "typical" choices of κ, the h-depth is equivalent with a special ***kernel mean embedding*** *in an appropriate RKHS.*

Consequences:

- Uniform consistency including rates of convergence;
- Consistency/rates of convergence of the induced deepest function;
- Uniform distributional asymptotics;
- All this also for imperfectly observed, or dependent data.

h-depth: Further consequences

The characterization property:
For any $P \neq Q \in \mathcal{P}(\mathcal{F})$ there exists $x \in \mathcal{F}$ with $D_\kappa(x; P) \neq D_\kappa(x; Q)$.

(Random) functional depths:
Some **random functional depths** (Cuevas et al., 2007, Cuesta-Albertos and Nieto-Reyes, 2008) admit explicit forms, i.e. they do not need to be approximated.

Conclusions

What we know:

- Functional depth is a very active field of FDA,
- with many potential applications,
- and many depths have been proposed.
- The selection of a depth is crucial, and must be problem-specific.
- Theoretical properties of the depth must be observed.

Open problems:

➔ Desiderata for the depth
(affine invariance? convexity in function spaces?)

➔ Statistical properties (finer asymptotics, bootstrap).

➔ Applications to analysis?

➔ How to choose a depth?

Selected literature

[1] Antonio Cuevas, Manuel Febrero, and Ricardo Fraiman. On the use of the bootstrap for estimating functions with functional data. *Comput. Statist. Data Anal.*, 51(2):1063–1074, 2006.

[2] Ricardo Fraiman and Graciela Muniz. Trimmed means for functional data. *Test*, 10(2):419–440, 2001.

[3] Sara López-Pintado and Juan Romo. On the concept of depth for functional data. *J. Amer. Statist. Assoc.*, 104(486):718–734, 2009.

[4] Stanislav Nagy, Irène Gijbels, Marek Omelka, and Daniel Hlubinka. Integrated depth for functional data: Statistical properties and consistency. *ESAIM Probab. Stat.*, 20:95–130, 2016.

[5] Stanislav Nagy, Sami Helander, Germain Van Bever, Lauri Viitasaari, and Pauliina Ilmonen. Flexible integrated functional depths. *Bernoulli*, 27(1):673–701, 2021.

[6] Stanislav Nagy and Sara López-Pintado. Depth for functional data: Theoretical properties. *in preparation*, 2021.

[7] John W. Tukey. Mathematics and the picturing of data. In *Proc. ICM, 1974, Vol. 2*, pages 523–531. Canad. Math. Congress, Montreal, Que., 1975.

[8] George Wynne and Stanislav Nagy. Statistical depth meets machine learning: Kernel mean embeddings and depth in functional data analysis. *arXiv preprint arXiv:2105.12778*, 2021.

[9] Yijun Zuo and Robert Serfling. General notions of statistical depth function. *Ann. Statist.*, 28(2):461–482, 2000.