HALFSPACE DEPTH: *GEOMETRY OF MULTIVARIATE QUANTILES*

Stanislav Nagy

Prague 2024

Charles University, Prague
Department of Probability and Mathematical Statistics

Outline

Statistical depth

Halfspace depth

Selected properties and problems

Floating bodies

Motivation: Grünbaum's inequality

(Dupin's) floating bodies

Convex floating bodies

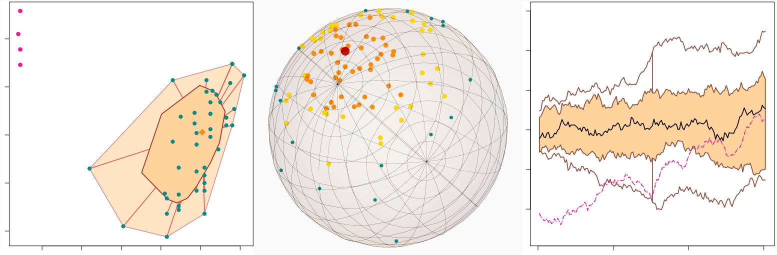
Simplicial depth and beyond

STATISTICAL DEPTH

Multivariate nonparametrics

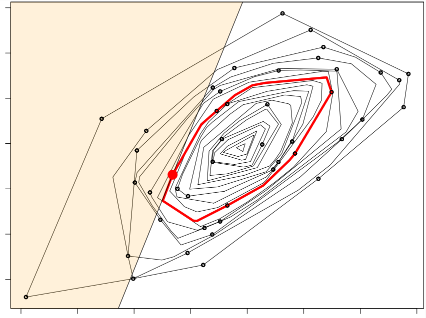
Nonparametric statistics:

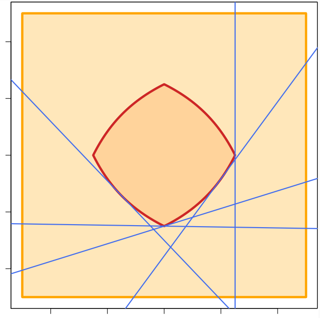
- Inference without assumptions — more flexible, but harder mathematically.
- On the real line using the ordering — median, quantiles, ranks...
- What are ranks or quantiles for multivariate (non-Euclidean) data?



Statistical depth

Statistical depth function: Ordering data in multivariate spaces.



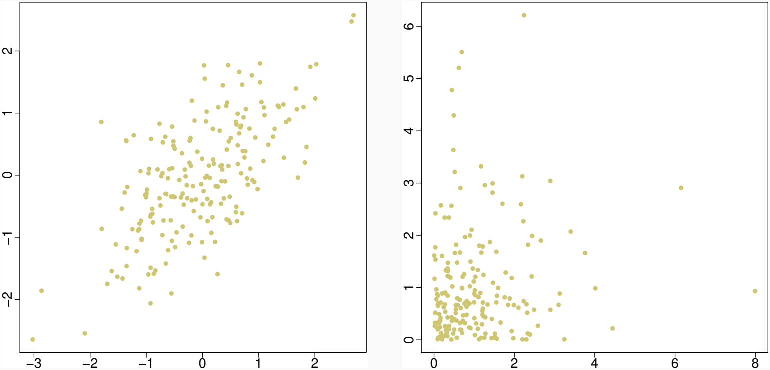


Introduced in 1975 (Tukey); studied intensively since the 1990s.

Statistical depth function

For $\mathcal{P}\left(\mathbb{R}^d\right)$ Borel probability measures on $\mathbb{R}^d$, consider the depth

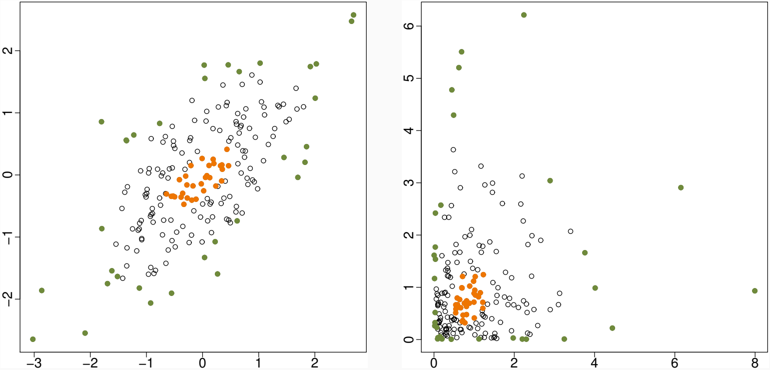
$$D\colon \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) \to [0,1]\colon (x, P) \mapsto D(x, P).$$



Statistical depth function

For $\mathcal{P}\left(\mathbb{R}^d\right)$ Borel probability measures on $\mathbb{R}^d$, consider the depth

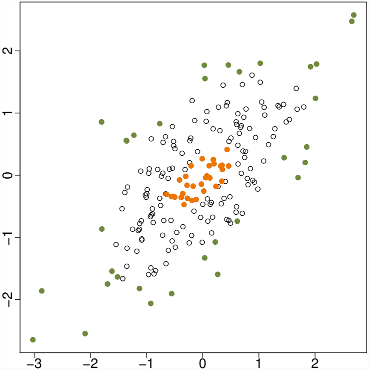
$$D\colon \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) \to [0,1]\colon (x, P) \mapsto D(x, P).$$

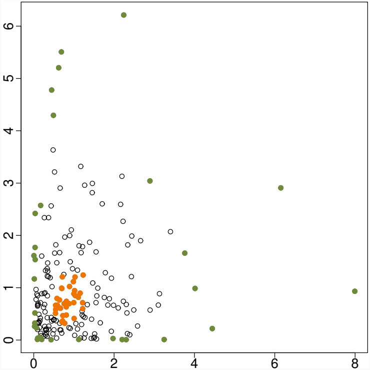


Halfspace depth

Halfspace depth (Tukey, 1975) of a point $x \in \mathbb{R}^d$ w.r.t. $P \in \mathcal{P}\left(\mathbb{R}^d\right)$

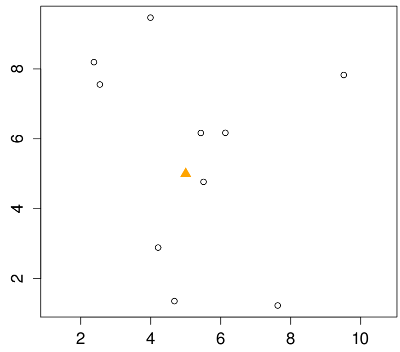
$$D(x;P) = \inf_{H \in \mathcal{H}(x)} P\left(H\right).$$





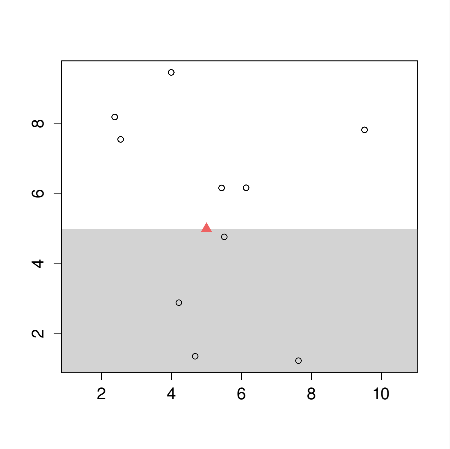
Halfspace depth

$$D(x; P_n) = \min \frac{\text{\# of observations in a halfspace that contains } x}{n}$$



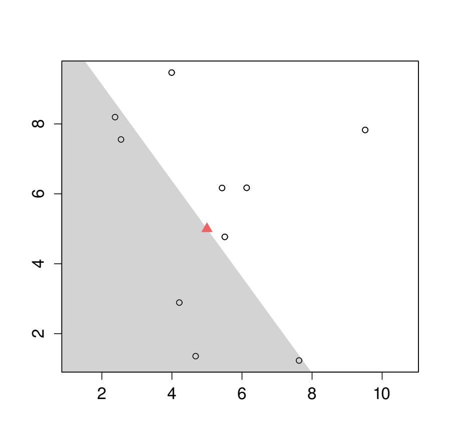
Halfspace depth

$$D(x; P_n) = \min \frac{\text{\# of observations in a halfspace that contains } x}{n}$$



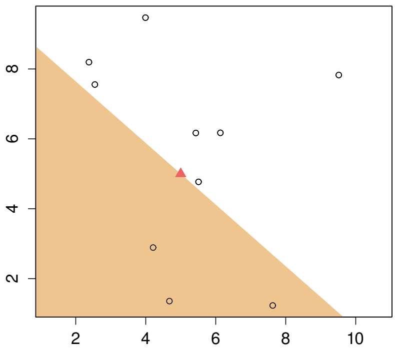
Halfspace depth

$$D(x; P_n) = \min \frac{\text{\# of observations in a halfspace that contains } x}{n}$$

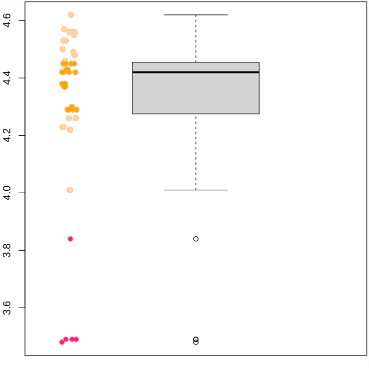


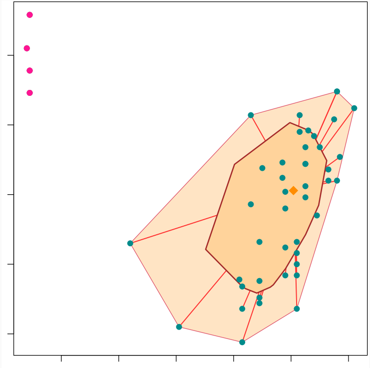
Halfspace depth

$$D(x; P_n) = \min \frac{\text{\# of observations in a halfspace that contains } x}{n}$$



Bagplot: A multivariate boxplot (Rousseeuw et al., 1999)

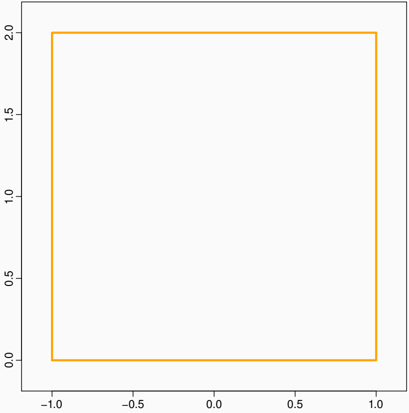




Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

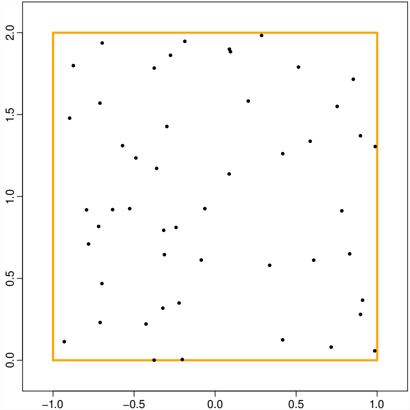
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

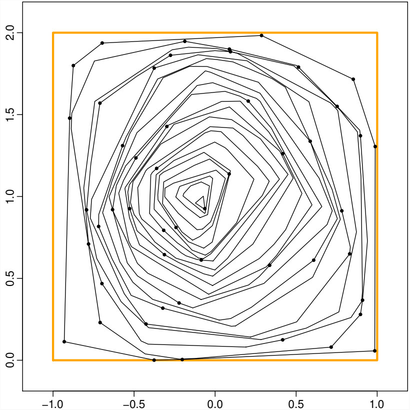
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

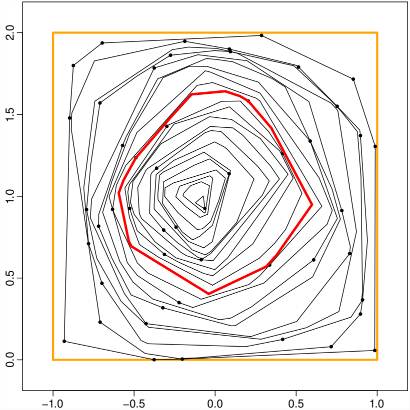
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

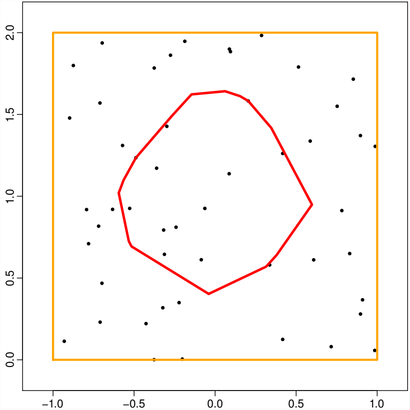
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

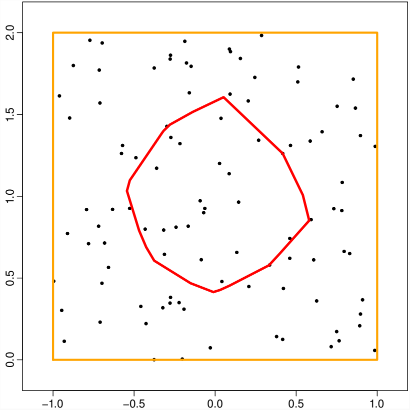
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

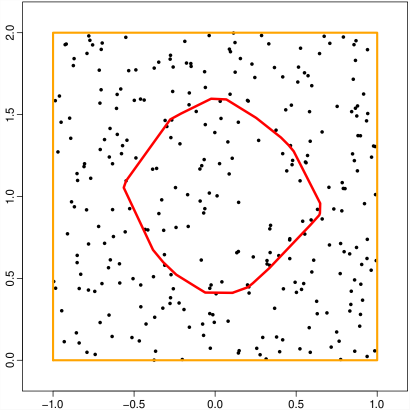
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

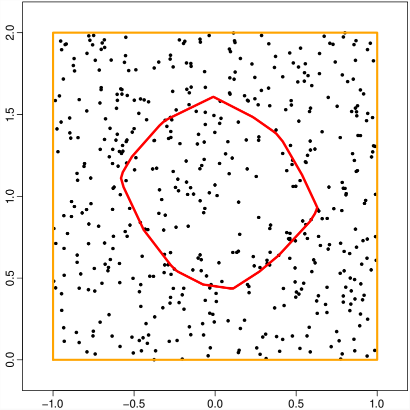
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

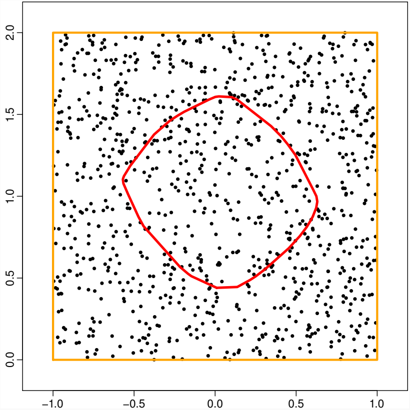
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

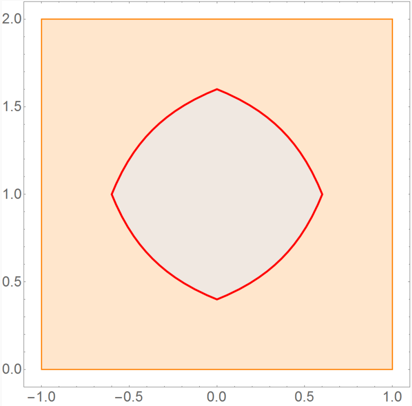
$$P_\delta = \left\{x \in \mathbb{R}^d : D(x; P) \geq \delta\right\} \text{ is convex}$$



Depth: Level sets

We can write (Rousseeuw and Struyf, 1999; Zuo and Serfling, 2000)

$$P_\delta = \left\{x \in \mathbb{R}^d : D(x; P) \geq \delta\right\} = \bigcap \{H \in \mathcal{H} : P(H) > 1 - \delta\} .$$

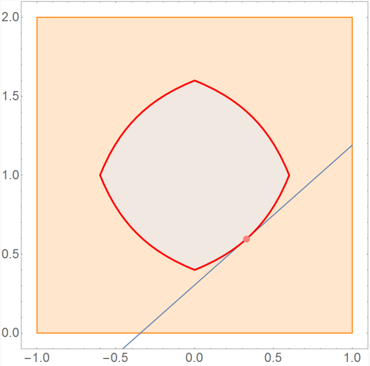


Depth: Asymptotic normality

Let $P_n \in \mathcal{P}\left(\mathbb{R}^d\right)$ be the empirical measure of n i.i.d. variables from P.

$\sqrt{n}\,(D(x;P_n) - D(x;P))$ is asymptotically normal

$\iff$ $D(x;P)$ is realised by a single halfspace $H \in \mathcal{H}(x)$ (Massé, 2004)

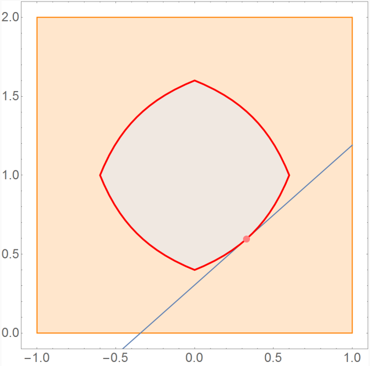


Depth: Asymptotic normality

Let $P_n \in \mathcal{P}\left(\mathbb{R}^d\right)$ be the empirical measure of n i.i.d. variables from P.

$\sqrt{n}\,(D(x; P_n) - D(x; P))$ is asymptotically normal

$$\iff \quad \text{the contour of } D(\cdot\,; P) \text{ is smooth at } x$$

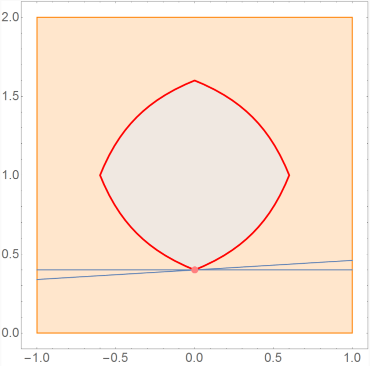


Depth: Asymptotic normality

Let $P_n \in \mathcal{P}\left(\mathbb{R}^d\right)$ be the empirical measure of n i.i.d. variables from P.

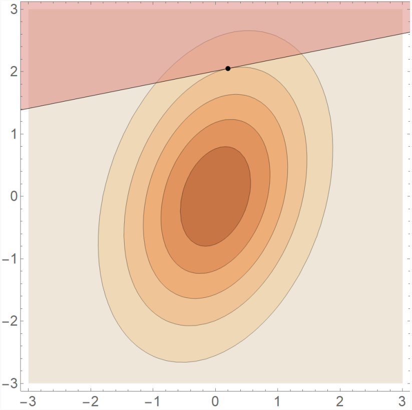
$\sqrt{n}\,(D(x;P_n) - D(x;P))$ is asymptotically normal

$$\iff \quad \text{the contour of } D(\cdot;P) \text{ is smooth at } x$$



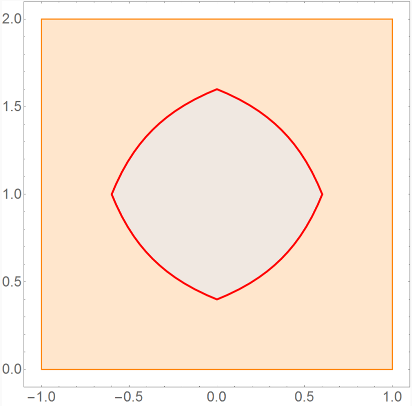
Problem: Smoothness of the depth

Elliptically symmetric distributions have smooth depth contours



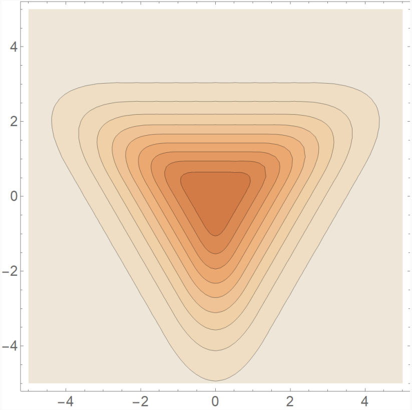
Problem: Smoothness of the depth

Many common distributions do not have smooth depth



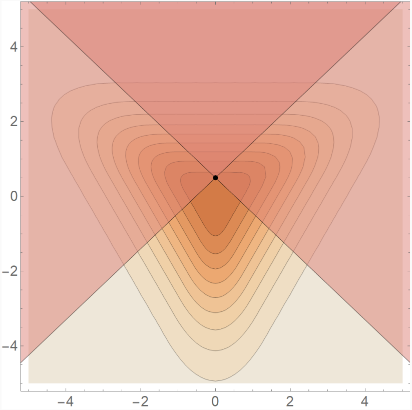
Problem: Smoothness of the depth

Smooth quasi-concave density is not sufficient for smooth depth



Problem: Smoothness of the depth

Smooth quasi-concave density is not sufficient for smooth depth



Problem: Smoothness of the depth

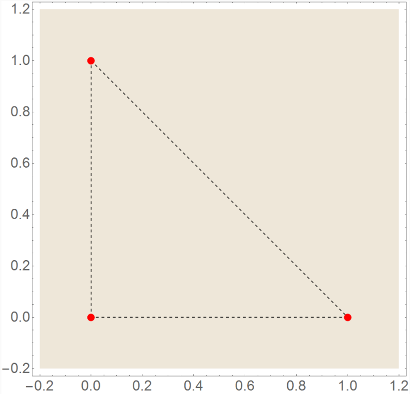
Problem (Massé and Theodorescu, 1994)

(P_1) Does there exist a non-elliptical distribution with smooth depth contours?

Problem: Depth of a median

For P in the vertices of a simplex in $\mathbb{R}^d$ (Donoho and Gasko, 1992)

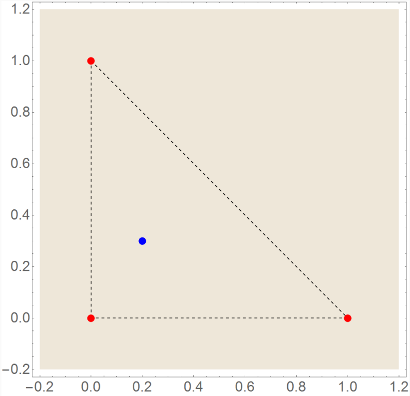
$$\sup_{x \in \mathbb{R}^d} D(x; P) = (d+1)^{-1} \xrightarrow[d \to \infty]{} 0$$



Problem: Depth of a median

For P in the vertices of a simplex in $\mathbb{R}^d$ (Donoho and Gasko, 1992)

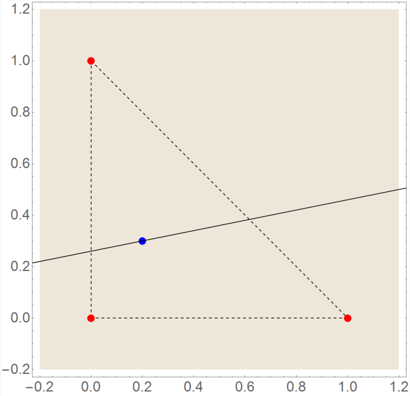
$$\sup_{x \in \mathbb{R}^d} D(x; P) = (d+1)^{-1} \xrightarrow[d \to \infty]{} 0$$



Problem: Depth of a median

For P in the vertices of a simplex in $\mathbb{R}^d$ (Donoho and Gasko, 1992)

$$\sup_{x \in \mathbb{R}^d} D(x; P) = (d+1)^{-1} \xrightarrow[d \to \infty]{} 0$$



Problem: Depth of a median

Problem (Donoho and Gasko, 1992)

(P_2) The maximum depth in $\mathbb{R}^d$ is at least $1/(d+1)$. Can we say more?

Problem: Characterization conjecture

Problem (Struyf and Rousseeuw, 1999)

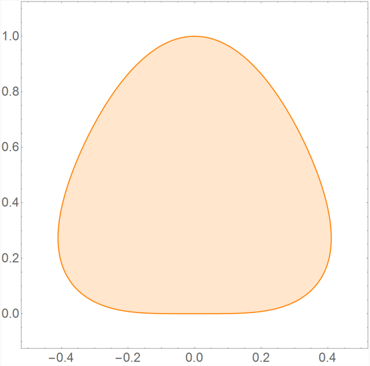
(P$_3$) Is it possible for two different distributions $P, Q \in \mathcal{P}\left(\mathbb{R}^d\right)$ to have the same depth at all $x \in \mathbb{R}^d$?

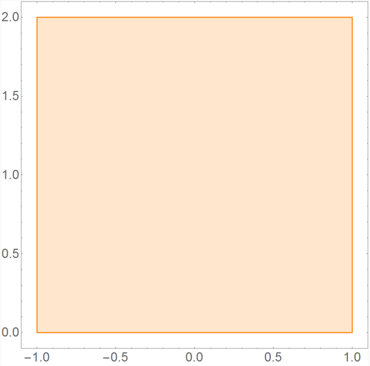
Partial answers:

- Certainly not for $d = 1$ (there depth $\sim$ distribution function).
- Not if P is atomic (Struyf and Rousseeuw, 1999; Koshevoy, 2002; Hassairi and Regaieg, 2007; Laketa and Nagy, 2021).
- Not if the contours of $D(\cdot; P)$ are smooth (Kong and Zuo, 2010).
- Long conjectured general negative answer (Koshevoy, 2003; Hassairi and Regaieg, 2008; Cuesta-Albertos and Nieto-Reyes, 2008).

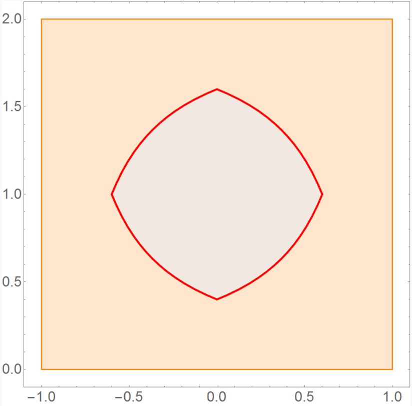
FLOATING BODIES

Convex body is a non-empty, compact and convex set $K \subset \mathbb{R}^d$ (Webster, 1994; Schneider, 2014).





Depth of a convex body K



Motivation: Grünbaum's inequality

Proposition (Grünbaum, 1960)

Let $K \subset \mathbb{R}^d$ *be a convex body,* $\mathrm{vol}(K) = 1$*, and* X *uniform on* K*. Then*

$$D(\mathsf{E}\, X; K) \geq \left(\frac{d}{d+1}\right)^d .$$

Motivation: Grünbaum's inequality

Proposition (Grünbaum, 1960)

Let $K \subset \mathbb{R}^d$ be a convex body, $\operatorname{vol}(K) = 1$, *and X uniform on K. Then*

$$D(\mathrm{E}\,X; K) \geq \left(\frac{d}{d+1}\right)^d.$$

- $\lim_{d\to\infty} \left(\frac{d}{d+1}\right)^d = \exp(-1) \approx 0.37.$

APPLICATIONS
DE GÉOMÉTRIE
ET
DE MÉCHANIQUE;

A LA MARINE, AUX PONTS ET CHAUSSÉES, ETC.,

POUR FAIRE SUITE

AUX DÉVELOPPEMENTS DE GÉOMÉTRIE,

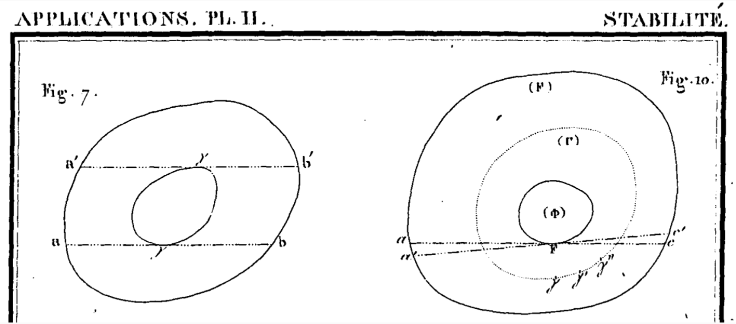
PAR CHARLES DUPIN,

Membre de l'Institut de France, Académie des Sciences; ancien Secrétaire de l'Académie Ionnienne, Associé étranger de l'Institut de Naples, Associé honoraire de l'Académie royale d'Irlande, et de la Société des Ingénieurs civils de la Grande-Bretagne, Membre des Académies royales des Sciences de Stockholm, de Turin, de Montpellier, etc., de la Société des Arts de Genève, de la Société d'Encouragement pour l'industrie française, Membre du Comité consultatif des Arts et Manufactures de France, Professeur de Méchanique au Conservatoire, Officier supérieur au corps du Génie Maritime, et Membre de la Légion-d'Honneur.

PARIS,

BACHELIER, SUCCESSEUR DE M^{me}. V^{e}. COURCIER, LIBRAIRE,
QUAI DES AUGUSTINS.

1822.

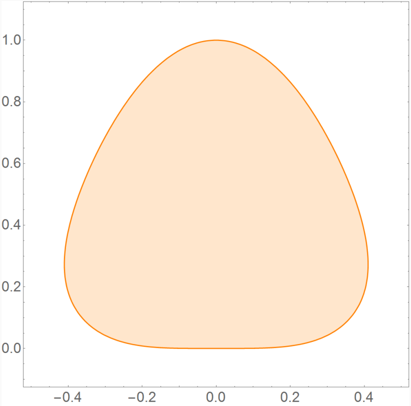


Definition (Dupin, 1822)

A convex body $K_{[\delta]}$ is called the Dupin floating body of a convex body $K \subset \mathbb{R}^d$ for $\delta \in [0, \text{vol}(K)/2]$ if each supporting hyperplane of $K_{[\delta]}$ cuts off a set of volume δ from K.

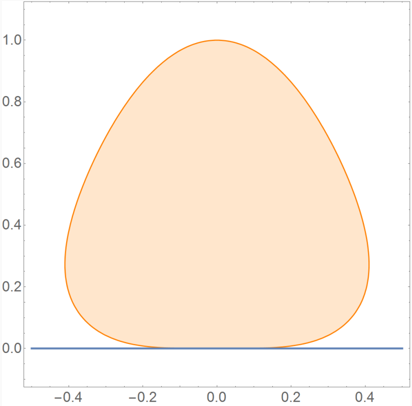
Floating body

Dupin's floating body of $K \subset \mathbb{R}^2$ for $\delta = 0.3$



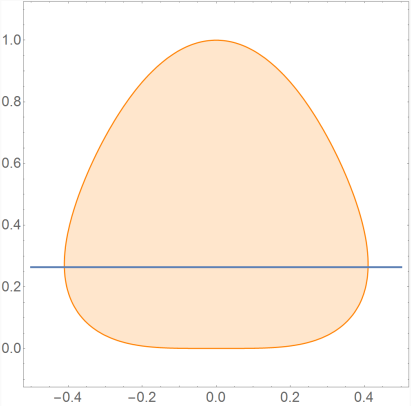
Floating body

Dupin's floating body of $K \subset \mathbb{R}^2$ for $\delta = 0.3$



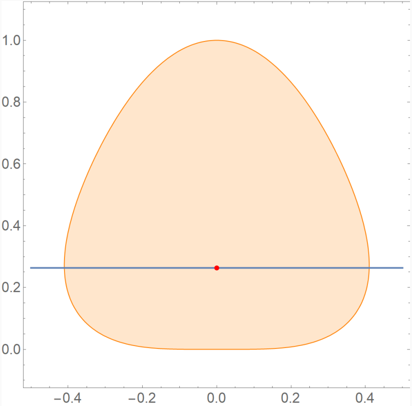
Floating body

Dupin's floating body of $K \subset \mathbb{R}^2$ for $\delta = 0.3$



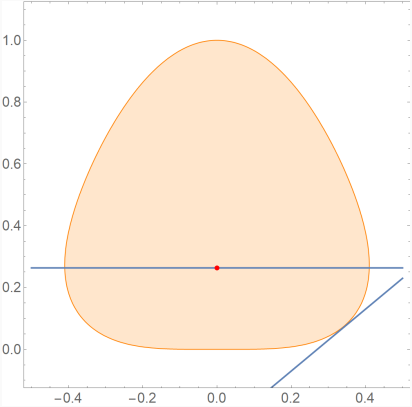
Floating body

Dupin's floating body of $K \subset \mathbb{R}^2$ for $\delta = 0.3$



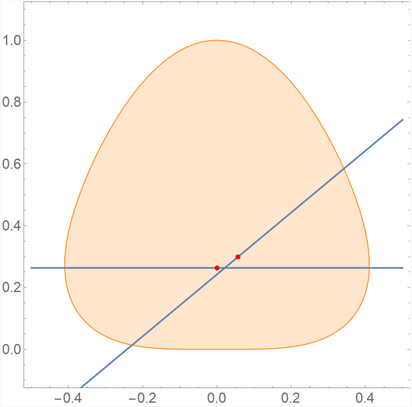
Floating body

Dupin's floating body of $K \subset \mathbb{R}^2$ for $\delta = 0.3$



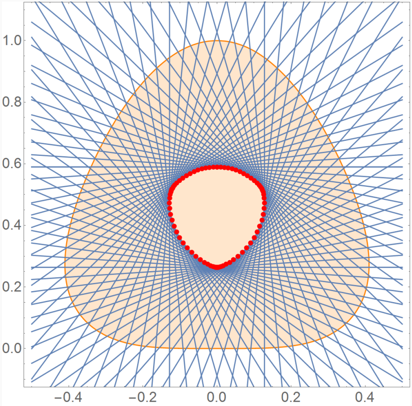
Floating body

Dupin's floating body of $K \subset \mathbb{R}^2$ for $\delta = 0.3$



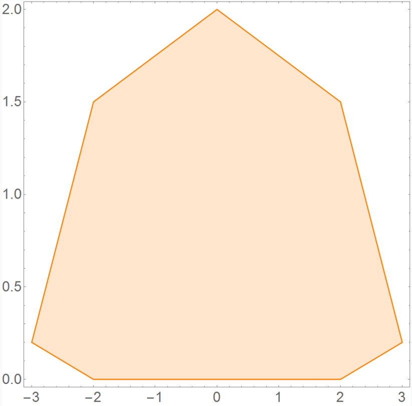
Floating body

Dupin's floating body of $K \subset \mathbb{R}^2$ for $\delta = 0.3$

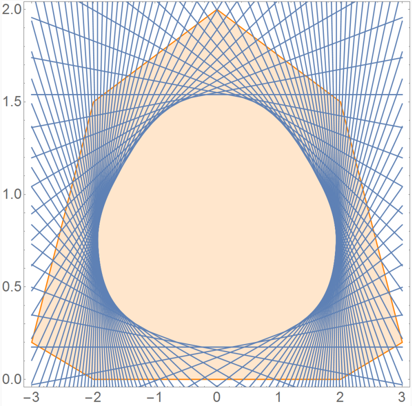


Floating body

Dupin's floating body of $K \subset \mathbb{R}^2$ for $\delta = 0.1$

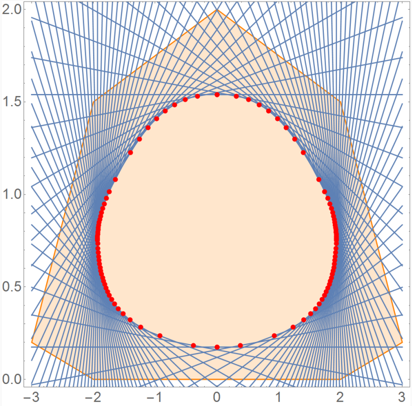


Dupin's floating body of $K \subset \mathbb{R}^2$ for $\delta = 0.1$



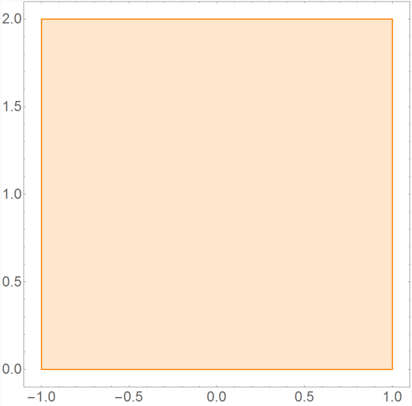
Floating body

Dupin's floating body of $K \subset \mathbb{R}^2$ for $\delta = 0.1$



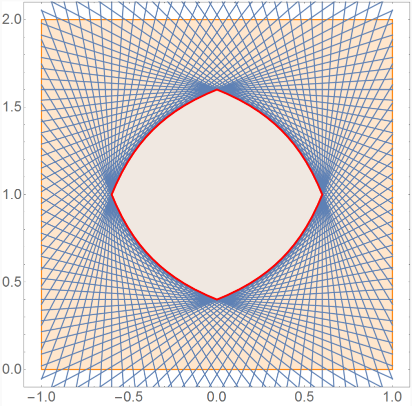
Floating body

Dupin's floating body of $K \subset \mathbb{R}^2$ for $\delta = 0.3$



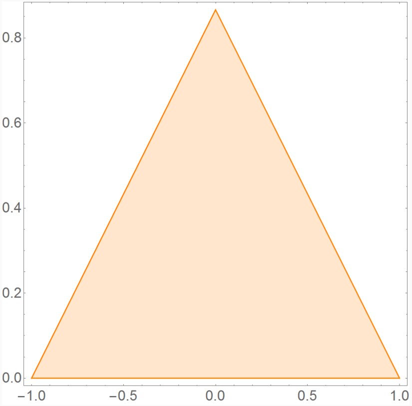
Floating body

Dupin's floating body of $K \subset \mathbb{R}^2$ for $\delta = 0.3$



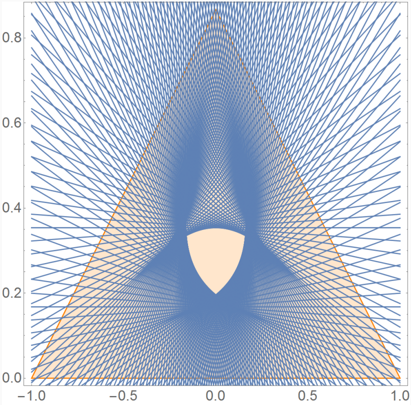
Floating body

Dupin's floating body of K does not have to exist



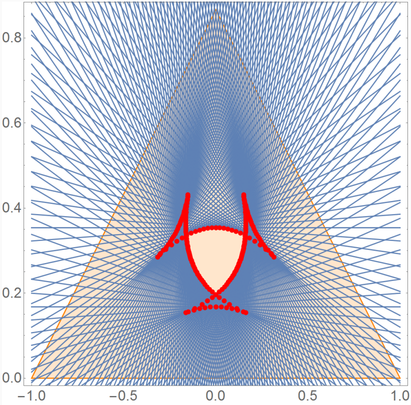
Floating body

Dupin's floating body of K does not have to exist



Floating body

Dupin's floating body of K does not have to exist



Definition (Schütt and Werner, 1990)

Let $K \subset \mathbb{R}^d$ be a convex body with $\mathrm{vol}\,(K) = 1$ and $\delta \in (0, 1/2)$.
The **convex floating body** of K associated with δ is given by

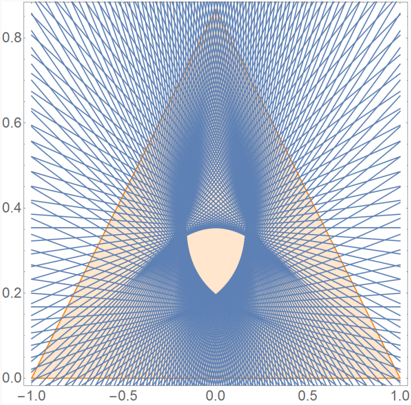
$$K_\delta = \bigcap \{H \in \mathcal{H} \colon \ \mathrm{vol}\,(K \cap H) \geq 1 - \delta\} \, .$$

Proposition (Schütt and Werner, 1990)

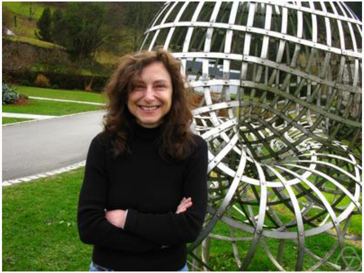
- *K_δ always exists.*
- *If $K_{[\delta]}$ exists, then $K_{[\delta]} = K_\delta$.*
- *Just as $K_{[\delta]}$, also K_δ has "nice" properties.*

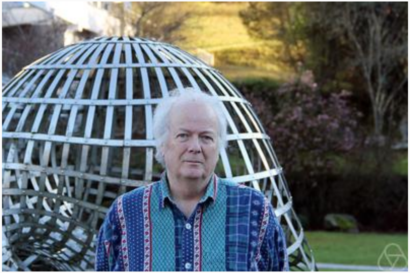
Convex floating body

Convex floating body of K always exists



Elisabeth Werner and Carsten Schütt





[1] Stanislav Nagy, Carsten Schütt, and Elisabeth M. Werner. Halfspace depth and floating body. *Statistics Surveys*, 13:52–118, 2019.

Grünbaum's inequality

- If K is a convex body, $D(\mathbb{E}X; K) \geq \exp(-1)$ (Grünbaum, 1960);
- Extensions to log-concave, κ-concave and quasi-concave measures and densities (Ball, 1986, 1988; Caplin and Nalebuff, 1991; Bobkov 2003, 2010);

$\Longrightarrow$ **(P_2)** The more concave density, the higher maximum depth.

Smoothness of floating bodies

Problem (Massé and Theodorescu, 1994)

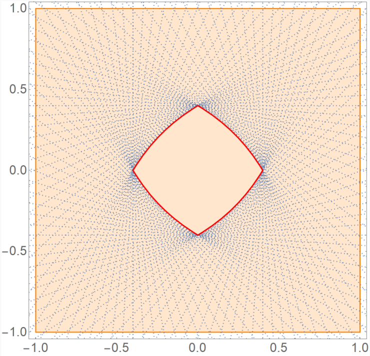
(P_1) Does there exist a non-elliptical distribution with smooth depth contours?

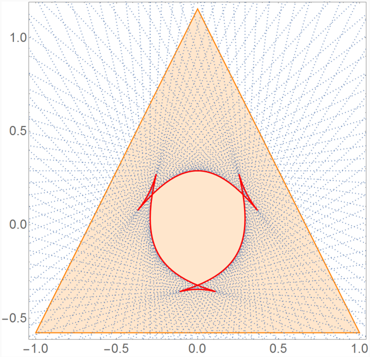
Proposition (Meyer and Reisner, 1991)

Uniform distributions on smooth, symmetric, strictly convex bodies have smooth depth.

Problem: Structure of floating bodies

For P uniform on a polytope K, describe the boundary structure of K_δ.





Depth characterization conjecture

Question: (Struyf and Rousseeuw, 1999)
Does for any $P \neq Q$ in $\mathcal{P}\left(\mathbb{R}^d\right)$ exist $x \in \mathbb{R}^d$ such that $D(x; P) \neq D(x; Q)$?

Positive answers for $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ such that:

- $d = 1$ (there depth $\sim$ distribution function).
- P is purely atomic, with finitely many atoms.
  (Struyf and Rousseeuw, 1999; Koshevoy, 2002; Laketa and Nagy, 2021)
- P is atomic. (Cuesta-Albertos and Nieto-Reyes, 2008)
- P is properly integrable. (Koshevoy, 2003)
- P has a smooth density. (Hassairi and Regaieg, 2008)
- all Dupin's floating bodies of P exist.
  (Kong and Zuo, 2010; Nagy, Schütt, Werner, 2019)

Conjectured positive answer.

(Cuesta-Albertos and Nieto-Reyes, 2008; Kong and Mizera, 2012)

Depth characterization conjecture

Question: (Struyf and Rousseeuw, 1999)

Does for any $P \neq Q$ in $\mathcal{P}\left(\mathbb{R}^d\right)$ exist $x \in \mathbb{R}^d$ such that $D(x; P) \neq D(x; Q)$?

Positive answers for $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ such that:

- $d = 1$ (there depth $\sim$ distribution function).
- P is purely atomic, with finitely many atoms.
  (Struyf and Rousseeuw, 1999; Koshevoy, 2002; Laketa and Nagy, 2021)
- ~~P is atomic.~~ (Cuesta-Albertos and Nieto-Reyes, 2008)
- ~~P is properly integrable.~~ (Koshevoy, 2003)
- ~~P has a smooth density.~~ (Hassairi and Regaieg, 2008)
- all Dupin's floating bodies of P exist.
  (Kong and Zuo, 2010; Nagy, Schütt, Werner, 2019)

Conjectured positive answer.

(Cuesta-Albertos and Nieto-Reyes, 2008; Kong and Mizera, 2012)

Characterization conjecture

Question: (Struyf and Rousseeuw, 1999)

Does for any $P \neq Q$ in $\mathcal{P}\left(\mathbb{R}^d\right)$ exist $x \in \mathbb{R}^d$ such that $D(x; P) \neq D(x; Q)$?

Not for $d > 1$.

[1] Stanislav Nagy. Halfspace depth does not characterize probability distributions. *Statistical Papers*, 62:1135–1139, 2021.

[2] Stanislav Nagy. The halfspace depth characterization problem. *Nonparametric Statistics*, 379–389. Springer International Publishing. 2020.

Depth characterization: Proof I

A measure $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ is called α-symmetric (Eaton, 1981) if

$$\psi(t) = \int_{\mathbb{R}^d} \exp\left(\mathrm{i}\,\langle t, x\rangle\right) \mathrm{d}\, P(x) = \xi\left(\|t\|_\alpha\right) \qquad \text{for all } t \in \mathbb{R}^d$$

for some $\xi \colon \mathbb{R} \to \mathbb{R}$. For $X = (X_1, \ldots, X_d) \sim P$, these measures satisfy

$$\langle X, u\rangle \overset{d}{=} \|u\|_\alpha X_1 \qquad \text{for all } u \in \mathbb{S}^{d-1}.$$

For the depth of α-symmetric P

$$D\left(x; P\right) = \inf_{u \in \mathbb{S}^{d-1}} \mathrm{P}\left(\langle X, u\rangle \leq \langle x, u\rangle\right) = \inf_{u \in \mathbb{S}^{d-1}} \mathrm{P}\left(\|u\|_\alpha X_1 \leq \langle x, u\rangle\right)$$
$$= \mathrm{P}\left(X_1 \leq \inf_{u \in \mathbb{S}^{d-1}} \langle x, u\rangle / \|u\|_\alpha\right) = F_1\left(-\|x\|_\beta\right)$$

for β the conjugate index to α, and F_1 the c.d.f. of X_1.

Depth characterization: Proof II

Fix $\gamma \in (0,1)$ and take $\psi_\alpha(t) = \exp\left(-\|t\|_\alpha^\gamma\right)$ for $\gamma \le \alpha \le 1$. Then

- Measure P_α with characteristic function ψ_α exists (Lévy, 1937);
- The conjugate index to $\alpha \le 1$ is $\beta = \infty$; and
- For the characteristic function of X_1 with $X \sim P_\alpha$ we have

$$\mathrm{E}\exp\left(\mathrm{i}\, t X_1\right) = \exp\left(-|t|^\gamma\right) \quad \text{for all } t \in \mathbb{R},$$

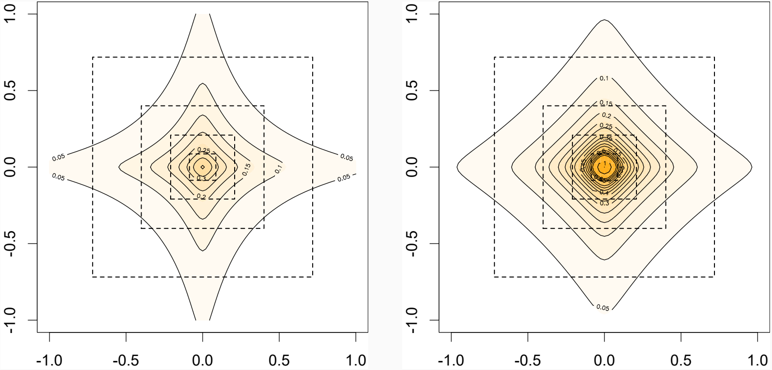
  i.e. F_1 does not depend on α.

All $P_\alpha \in \mathcal{P}\left(\mathbb{R}^d\right)$ have the same depth

$$D\left(x; P_\alpha\right) = F_1\left(-\|x\|_\infty\right) \quad \text{for all } x \in \mathbb{R}^d.$$

Depth characterization: Proof III

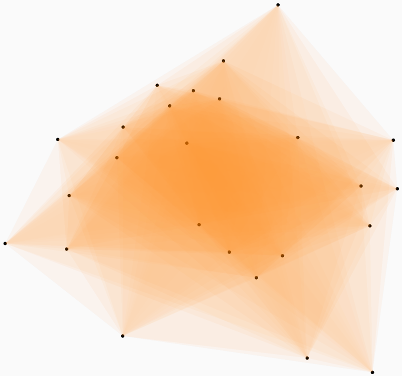
For $\gamma = 1/2$, the density of P_α with $\alpha = 1$ (left) and $\alpha = 1/2$ (right).



SIMPLICIAL DEPTH AND BEYOND

Simplicial depth (Liu, 1988) of $x \in \mathbb{R}^d$ w.r.t. $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ is

$$SD(x; P) = \mathrm{P}\left(x \in \triangle(X_1, \ldots, X_{d+1})\right).$$



How to compute simplicial depth?

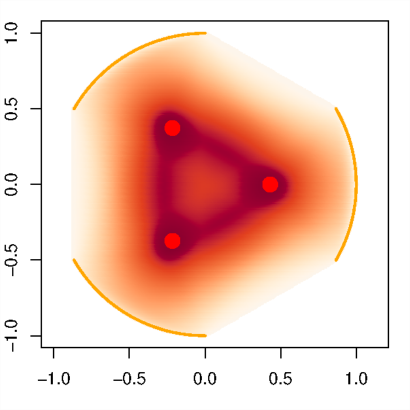
Simplicial depth (Liu, 1988) of $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ w.r.t. $P \in \mathcal{P}\left(\mathbb{R}^2\right)$ is

$$SD(x; P) = \mathrm{P}\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \triangle\left(\begin{pmatrix} X_{1,1} \\ X_{1,2} \end{pmatrix}, \begin{pmatrix} X_{2,1} \\ X_{2,2} \end{pmatrix}, \begin{pmatrix} X_{3,1} \\ X_{3,2} \end{pmatrix}\right)\right)$$
$$= \iint \iint \iint \mathbb{I}\left[x \in \triangle\right] \mathrm{d}\, P(x_{1,1}, x_{1,2}) \,\mathrm{d}\, P(x_{2,1}, x_{2,2}) \,\mathrm{d}\, P(x_{3,1}, x_{3,2}).$$

- $d \times (d+1)$ integrals in $\mathbb{R}^d$.
- Impossible to calculate already for Gaussian distributions in $\mathbb{R}^2$.
- How does the real (that is, population) simplicial depth of P even look like?

Population simplicial depth

The integrals are simpler if $P \in \mathcal{P}\left(\mathbb{R}^2\right)$ lives on a curve.

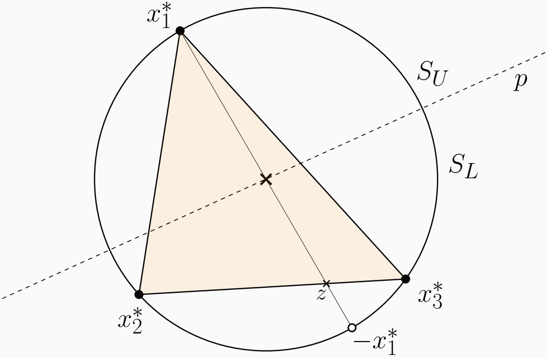


Three medians?

Computing simplicial depth exactly

We want to compute the simplicial depth in $\mathbb{R}^2$ exactly:

$$SD(x; P) = \mathrm{P}\left(x \in \triangle(X_1, X_2, X_3)\right).$$



Computing simplicial depth exactly

Proposition (Mendroš and Nagy, 2023+)

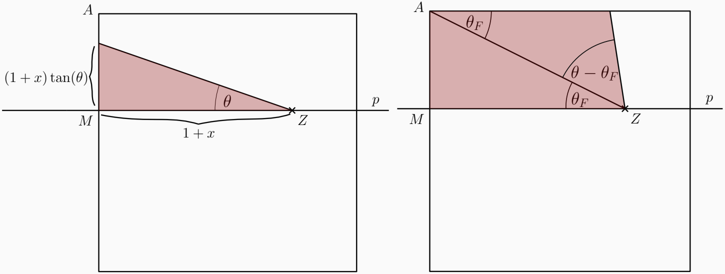
Let $a \in \mathbb{R}^2$ *be any non-zero vector,* $X \sim P \in \mathcal{P}\left(\mathbb{R}^2\right)$ *be absolutely continuous and let* $q = P(a^\top X > 0)$*. Then*

$$SD(0; P) = 6\, q \cdot (1-q)^2 \int_0^\pi G(\theta) \cdot (1 - G(\theta))\, dF(\theta) + 6\, q^2 \cdot (1-q) \int_0^\pi F(\theta) \cdot (1 - F(\theta))\, dG(\theta),$$

where F (or G) is the upper (or lower) circular distribution function of P.

Simplicial depth of a square

$P = \mathrm{Unif}([-1,1]^2)$: Finding the circular distribution function $F(\theta)$.



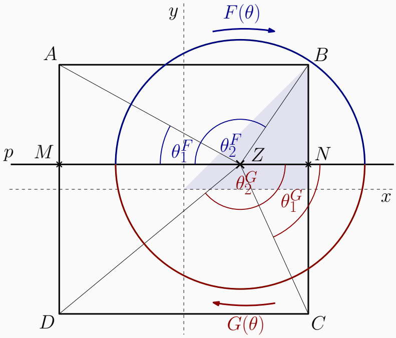
Simplicial depth of a square

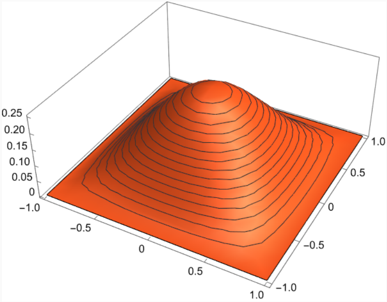
For $0 \le x \le 1$ and $0 \le y \le x$ we have

$$SD((x,y);P) =$$
$$\frac{(x-1)^2}{32}\left[-\frac{2\left(3y^4(-x^2+3x+3)-y^2(7x^2+18x+9)+6x^2+9x+4\right)}{(x+1)(y^2-1)}\right.$$
$$\left.+3(y^2(3x-1)+x+1)\log\left(\frac{1+x}{1-x}\right)+3y(y^2(x-1)+3x+1)\log\left(\frac{1-y}{1+y}\right)\right].$$

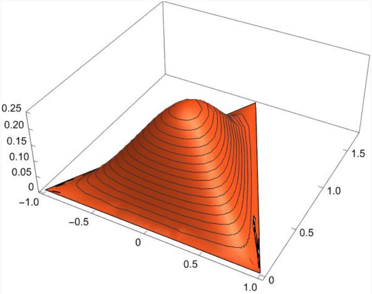
In other parts of $[-1,1]^2$ symmetrically.

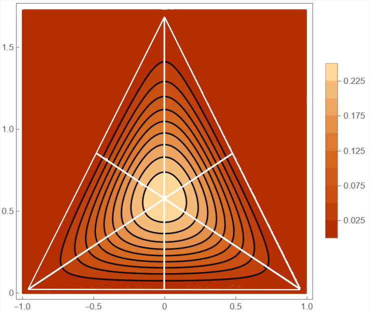
Simplicial depth of a square



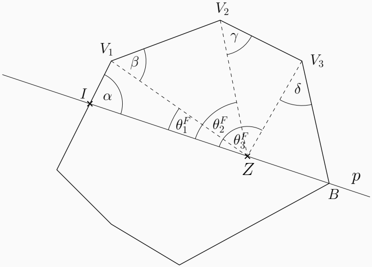


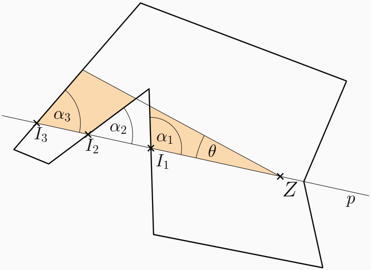
Simplicial depth of a triangle





Simplicial depth of polygons





Simplicial depth in mathematics

Simplicial depth (Liu, 1988) of $x \in \mathbb{R}^d$ w.r.t. $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ is

$$SD(x; P) = \mathrm{P}\left(x \in \triangle(X_1, \ldots, X_{d+1})\right).$$

- Studied since the 1950s in geometry.
- First selection lemma: $\max_{x \in \mathbb{R}^d} SD(x; P) \geq c_d > 0$, with $c_1 = 1/2$, $c_2 = 2/9$, $c_d = (d!)(d+1)^{-d}$ (conjectured).
- Applications to breakdown point (BP) of the simplicial median: The simplicial median is robust, but its BP decreases fast with d.

[1] Stanislav Nagy. Simplicial depth and its median: Selected properties and limitations. (2023) *Statistical Analysis and Data Mining* 16(4), 374–390.

Conclusion

Quantiles and multivariate data:

- Many different approaches; inherently geometric.
- Halfspace depth and the floating body are the same concept.
- Halfspace depth does not characterize distributions.
- Simplicial depth in $\mathbb{R}^2$ can be evaluated (sometimes).

What we do not know:

- When are floating bodies smooth?
- When does halfspace depth characterize distributions?
- Is the triangle characterized by its halfspace depth?
- How to evaluate simplicial depth in $\mathbb{R}^d$, $d > 2$?

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] David L. Donoho and Miriam Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.*, 20(4):1803–1827, 1992.

[3] Branko Grünbaum. Partitions of mass-distributions and of convex bodies by hyperplanes. *Pacific J. Math.*, 10:1257–1261, 1960.

[4] Regina Y. Liu. On a notion of data depth based on random simplices. *Ann. Statist.*, 18(1):405–414, 1990.

[5] M. Meyer and S. Reisner. A geometric property of the boundary of symmetric convex bodies and convexity of flotation surfaces. *Geom. Dedicata*, 37(3):327–337, 1991.

[6] Carsten Schütt and Elisabeth M. Werner. The convex floating body. *Math. Scand.*, 66(2):275–290, 1990.

[7] Anja Struyf and Peter J. Rousseeuw. Halfspace depth and regression depth characterize the empirical distribution. *J. Multivariate Anal.*, 69(1):135–153, 1999.

[8] 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.

More at

GeMS.karlin.mff.cuni.cz

and in

[1] Stanislav Nagy, Carsten Schütt, and Elisabeth M. Werner. Halfspace depth and floating body. *Statistics Surveys*, 13:52–118, 2019.

[2] Stanislav Nagy. Halfspace depth does not characterize probability distributions. *Statistical Papers*, 62:1135–1139, 2021.

[3] Stanislav Nagy. The halfspace depth characterization problem. *Nonparametric Statistics*, 379–389. Springer International Publishing. 2020.

[4] Stanislav Nagy, and Jiří Dvořák. Illumination depth. *Journal of Computational and Graphical Statistics*, 30(1): 78–90, 2021.

[5] Stanislav Nagy. Simplicial depth and its median: Selected properties and limitations. *Statistical Analysis and Data Mining* 16(4): 374–390, 2023.