# Statistical depth:
# Part I: The depth function

Stanislav Nagy

Charles University
Department of Probability and Mathematical Statistics

Warsaw 2021

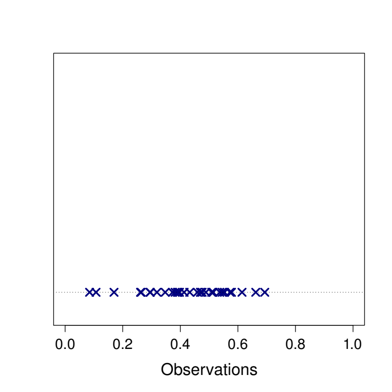
Co-funded by the
Erasmus+ Programme
of the European Union

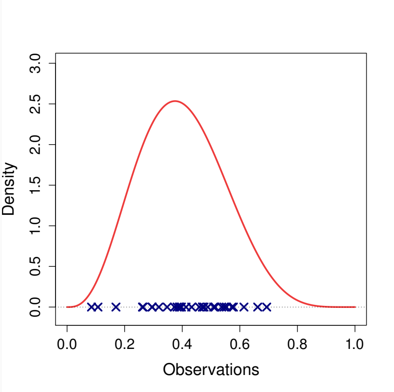
# Part I: Statistical data depth (Introduction)

- Motivation: Orderings and quantiles
  - Point estimation
  - Data visualisation
  - L-estimation and testing
- Halfspace depth: Multivariate quantiles
  - Halfspace depth and its properties
  - Applications: Non-parametric statistics in Euclidean spaces
  - Difficulties and open problems
- General depth and local depths

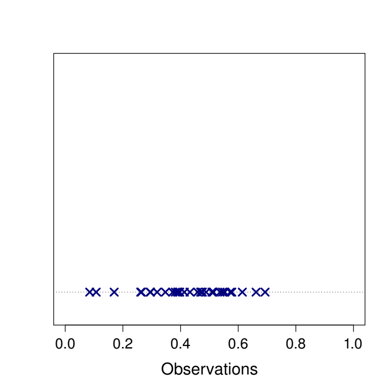
# Motivation: Orderings and quantiles

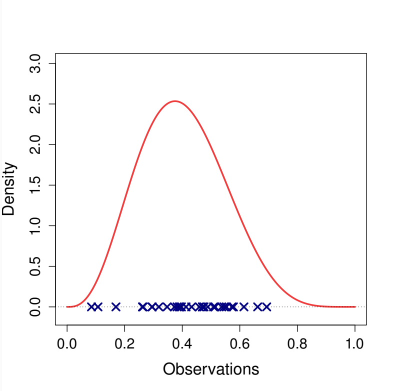
A random sample $X_1, \ldots, X_n$ of **univariate** observations ($\times$)



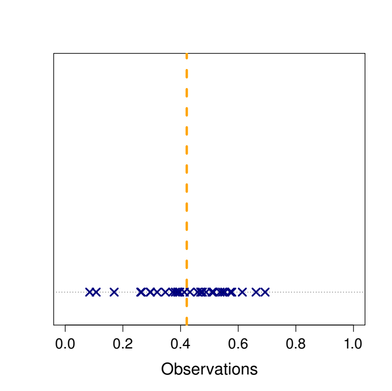



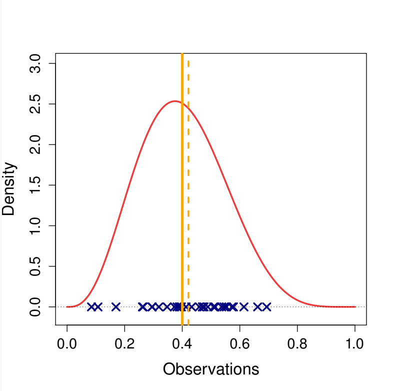
$X_1, \ldots, X_n \sim P \in \mathcal{P}(\mathbb{R})$ with a density



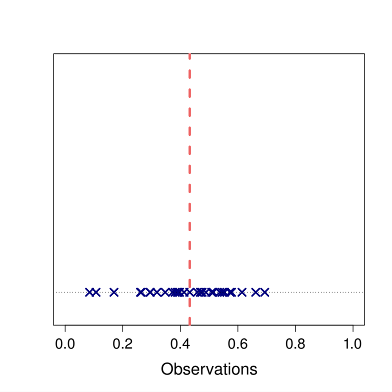



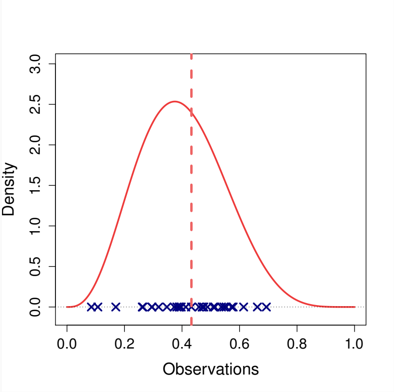
Mean $\mathrm{E}\,X_1 = \int_{\mathbb{R}} x \,\mathrm{d}\,P(x)$ estimated by $\sum_{i=1}^{n} X_i/n$



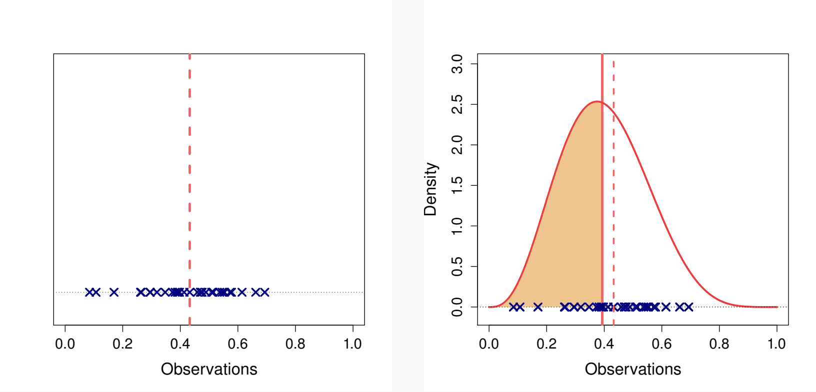



Sample median: the middle-most observation $X_{(n/2)}$

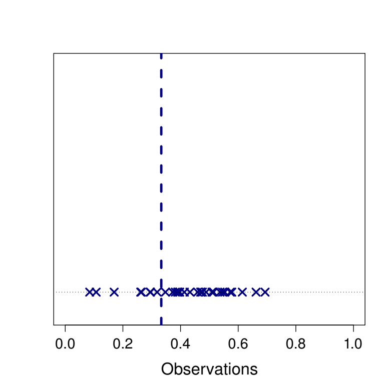


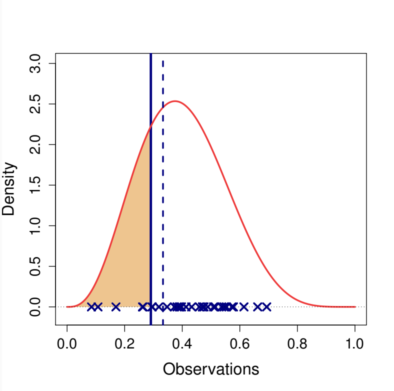



$$q(0.5) = \sup\{x \in \mathbb{R} \colon P((-\infty, x]) \leq 0.5\}$$

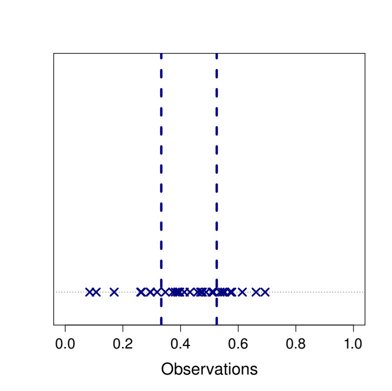


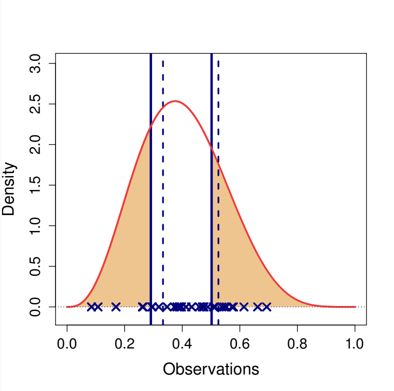
$$q(0.25) = \sup\{x \in \mathbb{R} : P((-\infty, x]) \leq 0.25\}$$



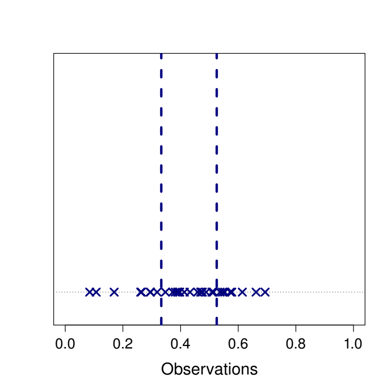



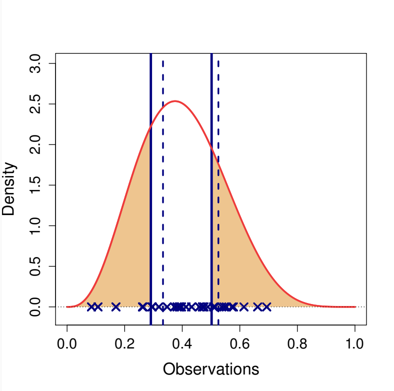
$$q(0.75) = \sup\{x \in \mathbb{R} : P((-\infty, x]) \leq 0.75\}$$



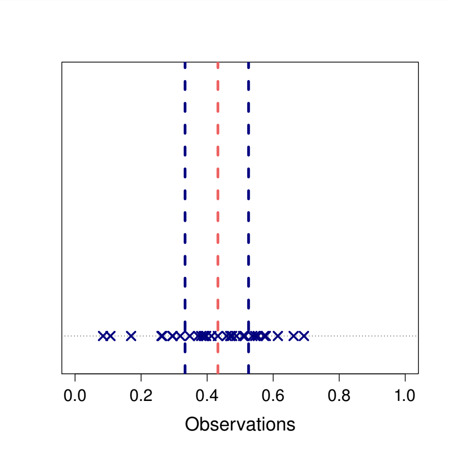



$$IQR = q(0.75) - q(0.25)$$

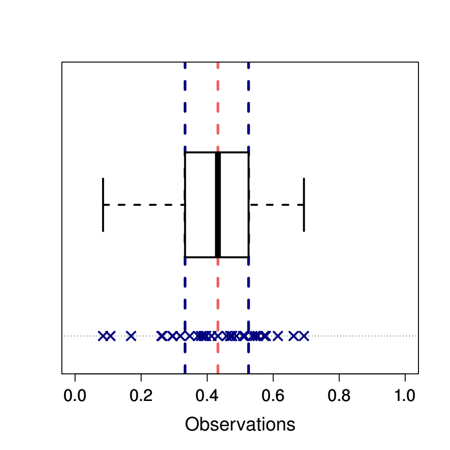


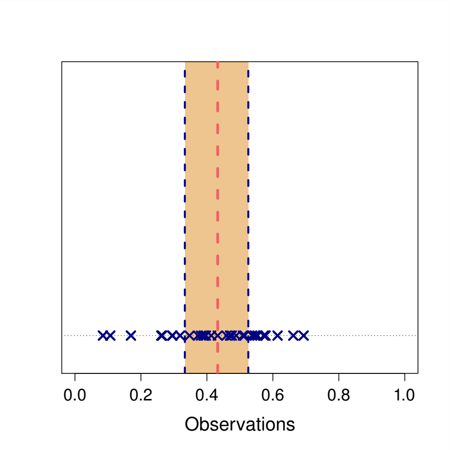
Quantile-based visualisation tool (Tukey, 1969)



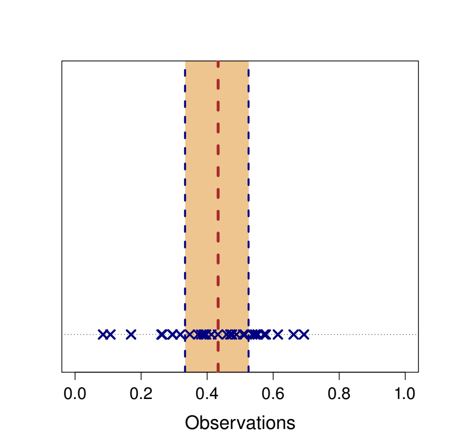
Quantile-based visualisation tool (Tukey, 1969)



Central part of the data

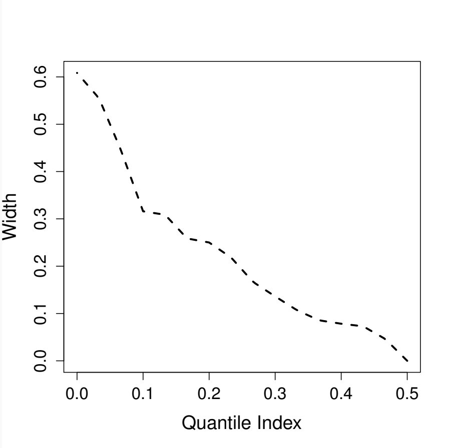


L-statistics: Functions of the order statistics (e.g. the trimmed mean)

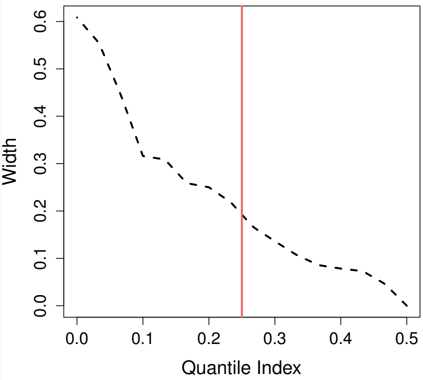


# Scale curve

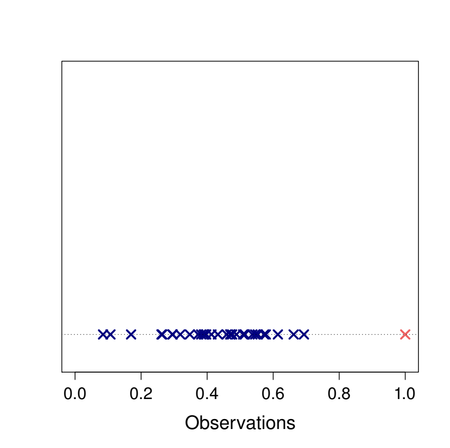
$$s\colon [0,1/2] \to [0,\infty)\colon t \mapsto q(1-t) - q(t)$$



$$s\colon [0, 1/2] \to [0, \infty)\colon t \mapsto q(1-t) - q(t)$$

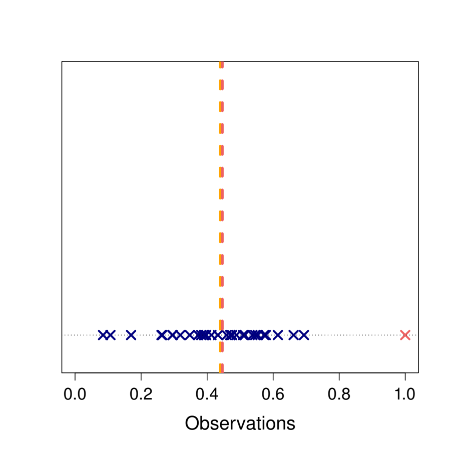


Contaminate the dataset with an error $X_{n+1} = 1$



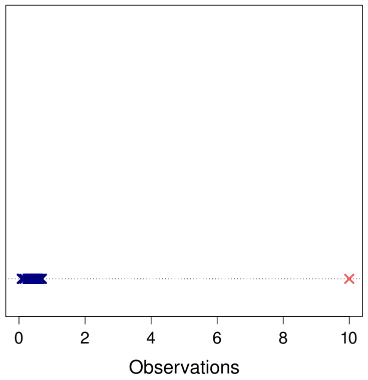
# An outlier

Mean and median of the contaminated data

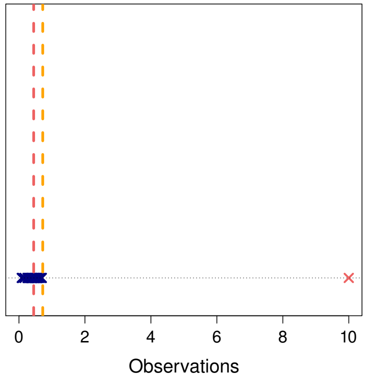


# A severe outlier

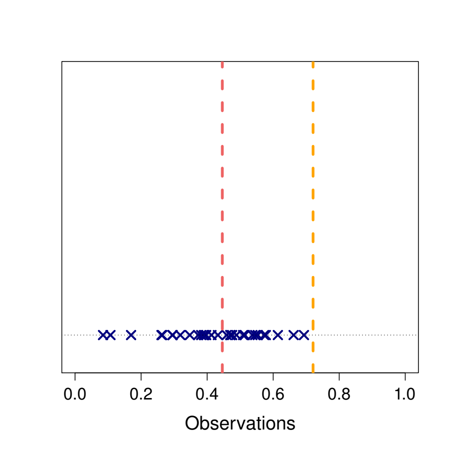
Contaminate with $X_{n+1} = 10$



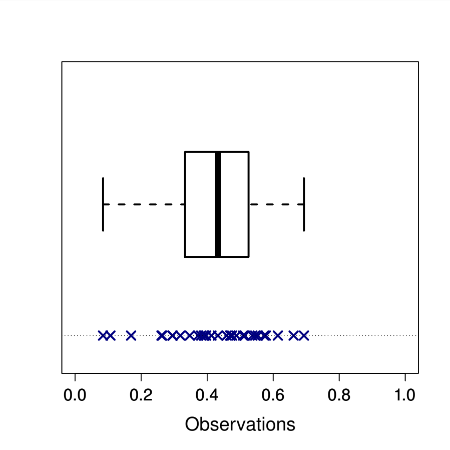
Mean and median of the contaminated data



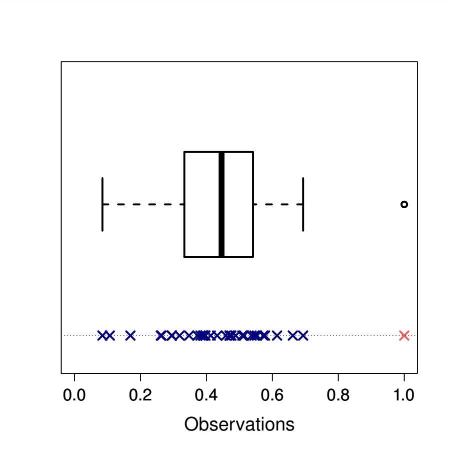
Mean and median of the contaminated data



Boxplot of the original data



Boxplot of the contaminated data



Let $X_1, \ldots, X_n \sim P$ and $Y_1, \ldots, Y_m \sim Q$ be independent univariate random samples (no ties are assumed). Test

$$H_0\colon P = Q \quad \text{against} \quad H_1\colon P \neq Q.$$

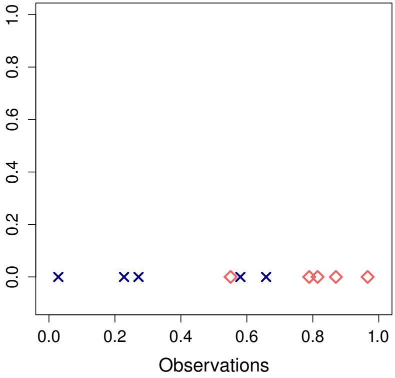
Wilcoxon's rank sum test (Wilcoxon, 1945):

- Pool the two samples into $Z_1, \ldots, Z_{n+m}$ and rank these observations (1 through $n+m$).
- Sum up the ranks of those observations which came from the sample from $P$. Denote by $R$.
- Reject $H_0$ if $R$ is either too small, or too large.

# Wilcoxon's rank sum test: Illustration (beta distributions)

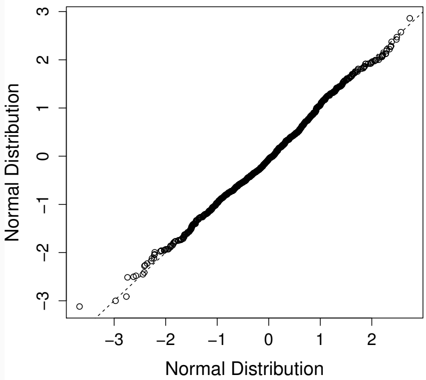
$$X \sim B(1,2), Y \sim B(2,1), n = m = 5$$

$R = 17$ (range from 15 to 40), p-value 0.03



Quantile-versus-quantile plot (Gnanadesikan and Wilk, 1968)

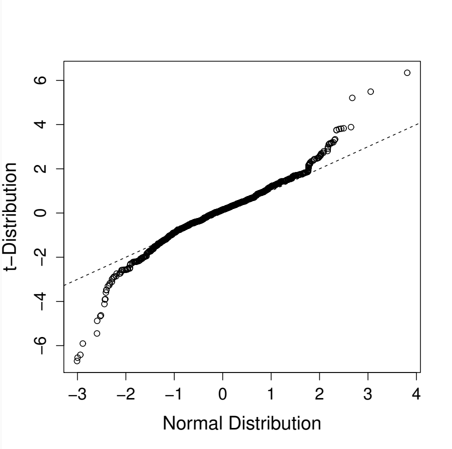
$$t \mapsto (q_X(t), q_Y(t))$$



# Q-Q PLOT

Quantile-versus-quantile plot (Gnanadesikan and Wilk, 1968)

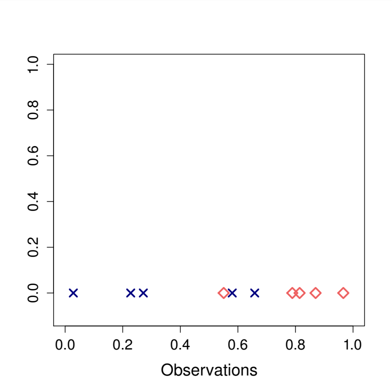
$$t \mapsto (q_X(t), q_Y(t))$$

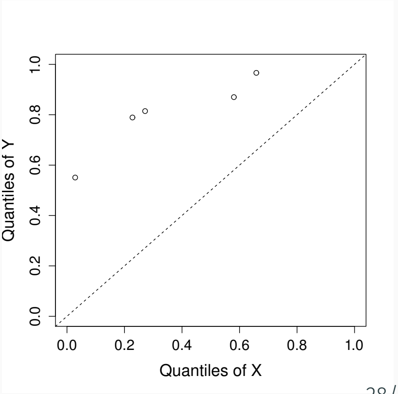


# Wilcoxon's rank sum test: Illustration

$$X \sim B(1,2), Y \sim B(2,1), n = m = 5$$

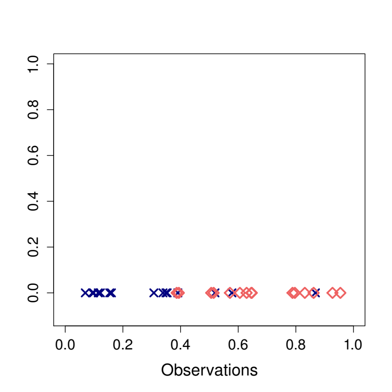
$R = 17$ (range from 15 to 40), p-value 0.03

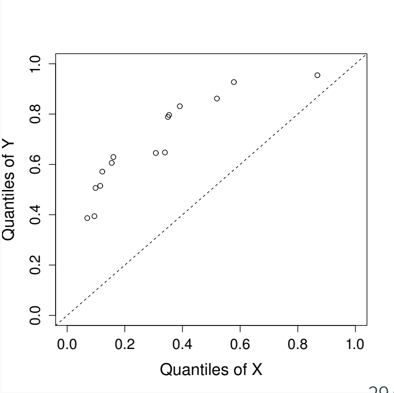


$$X \sim B(1, 2), Y \sim B(2, 1), n = m = 15$$

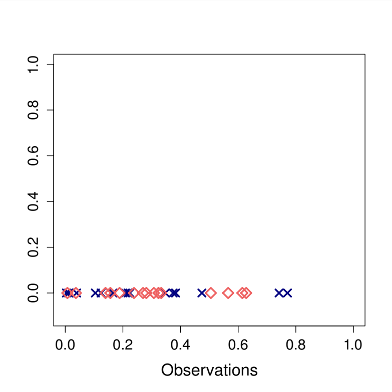
$R = 143$ (range from 120 to 345), p-value 0.00

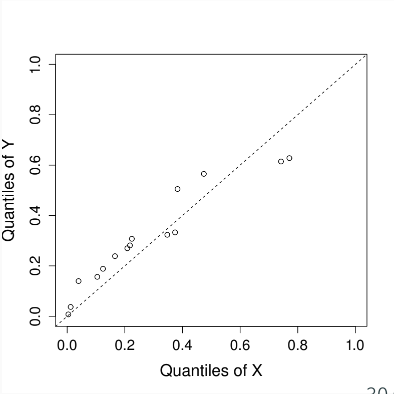


$$X \sim B(1,2), Y \sim B(1,2), n = m = 15$$

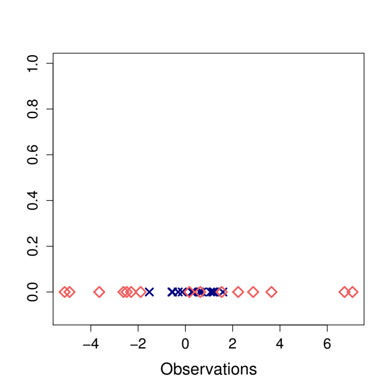
$R = 220$ (range from 120 to 345), p-value 0.62

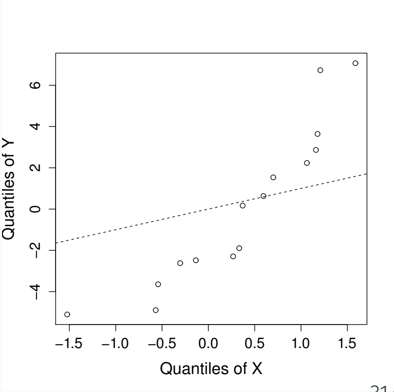


$$X \sim N(0,1), Y \sim N(0,16), n = m = 15$$

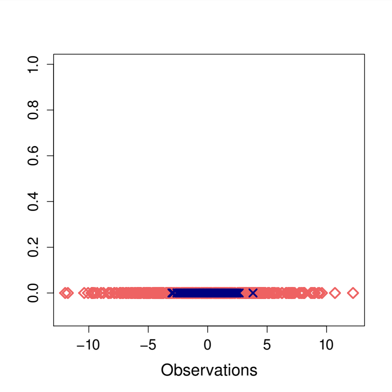
$R = 242$ (range from 120 to 345), p-value 0.71

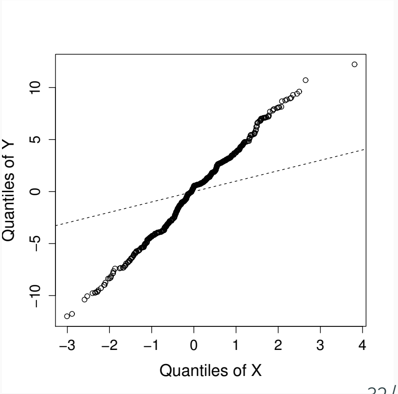


$$X \sim N(0,1), Y \sim N(0,16), n = m = 500$$

p-value 0.30



## Summary: Ranks and orders

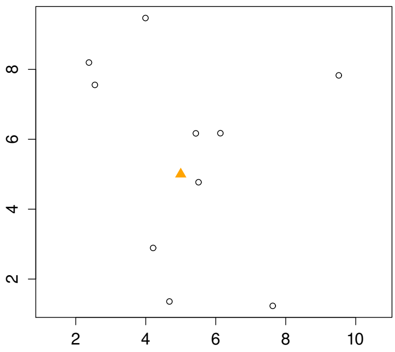
In $\mathbb{R}$, ranks and order statistics enable:

- effective data visualisation (Q-Q plot);
- outlier detection (boxplot);
- construction of robust estimators (L-statistics);
- non-parametric data analysis (rank tests).

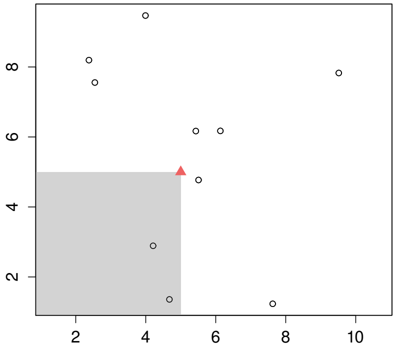
All thanks to the **linear ordering** on the sample space.

# Halfspace depth: Multivariate quantiles

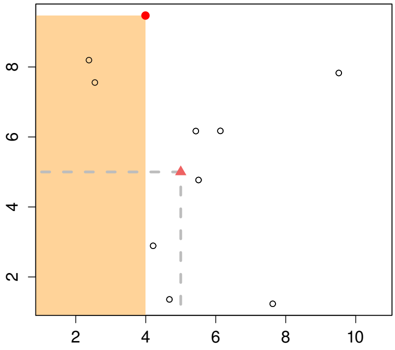
How to order **multivariate data**?



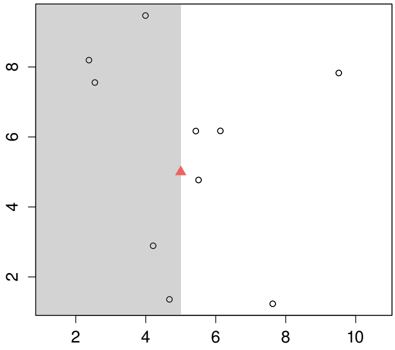
How to order **multivariate data**?



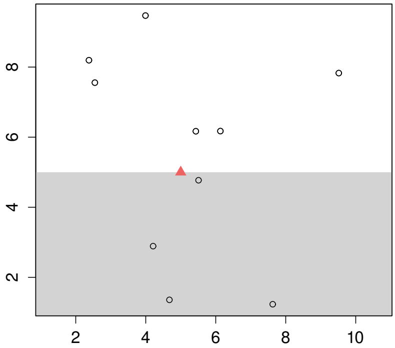
How to order **multivariate data**?



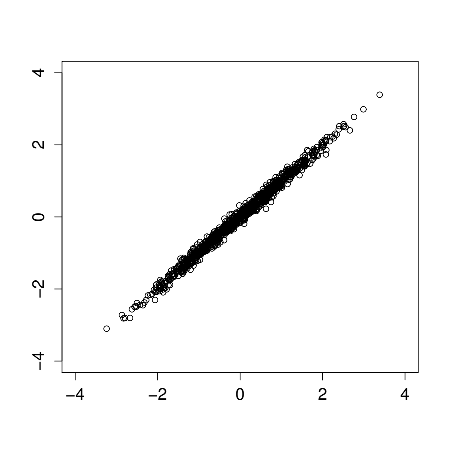
How to order **multivariate data**?



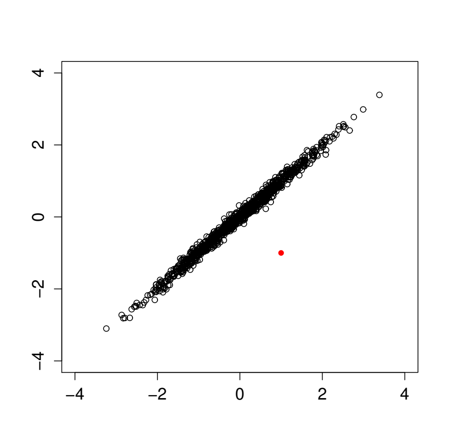
How to order **multivariate data**?



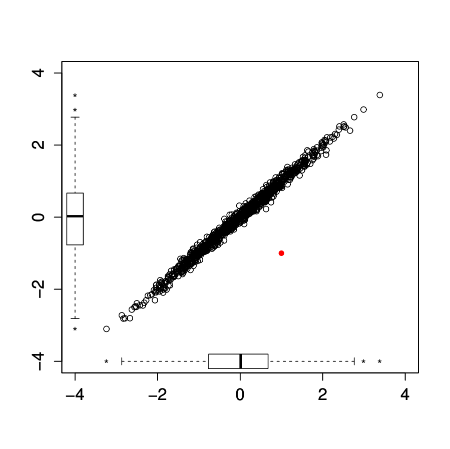
How to order **multivariate data**?



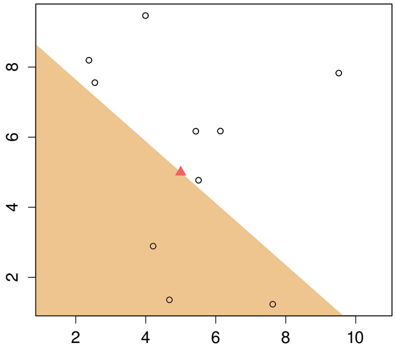
How to order **multivariate data**?



How to order **multivariate data**?

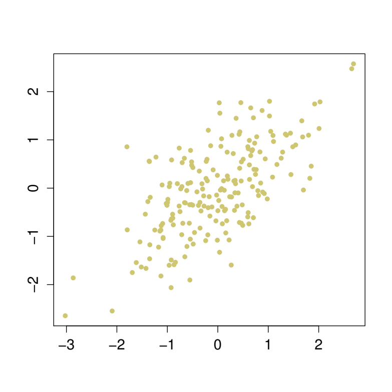


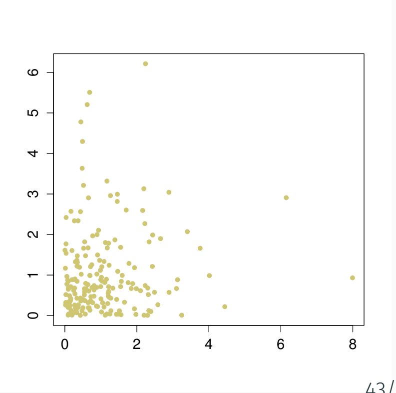
How to order **multivariate data**?



For a random variable $X \sim P \in \mathcal{P}(\mathbb{R}^d)$, consider the depth of $x \in \mathbb{R}^d$ w.r.t. $P$

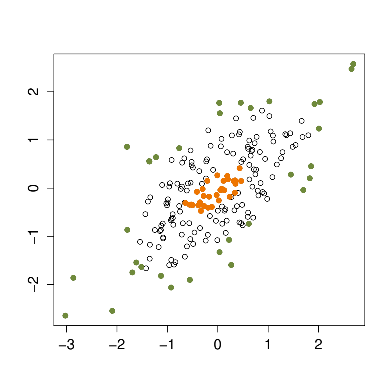
$$D\colon \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) \to [0,1]\colon (x,P) \mapsto D(x,P).$$

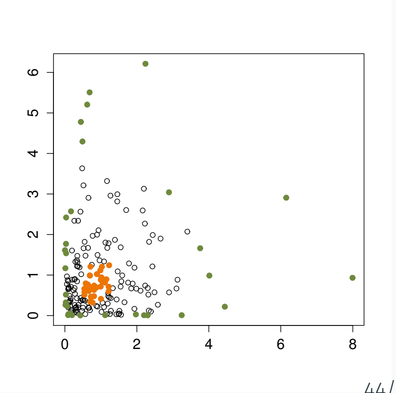


For a random variable $X \sim P \in \mathcal{P}\left(\mathbb{R}^d\right)$, consider the depth of $x \in \mathbb{R}^d$ w.r.t. $P$

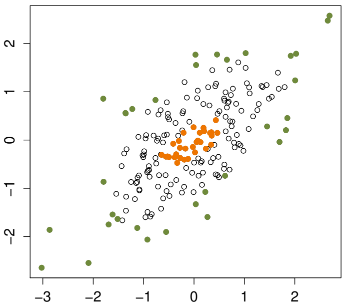
$$D\colon \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) \to [0,1]\colon (x,P) \mapsto D(x,P).$$

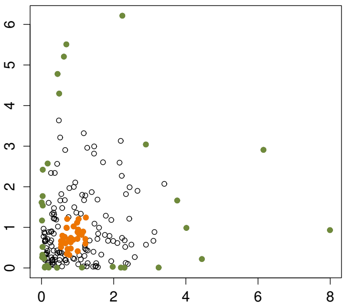


Halfspace depth (Tukey, 1975) of an observation in $\mathbb{R}^d$

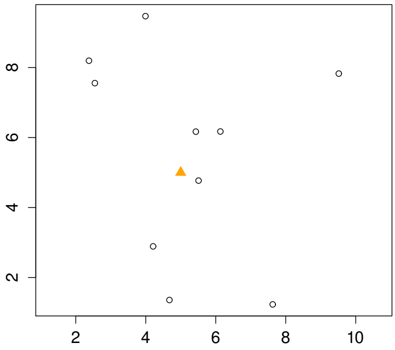
$$hD(x; P) = \inf_{H \in \mathcal{H}(x)} P(H).$$



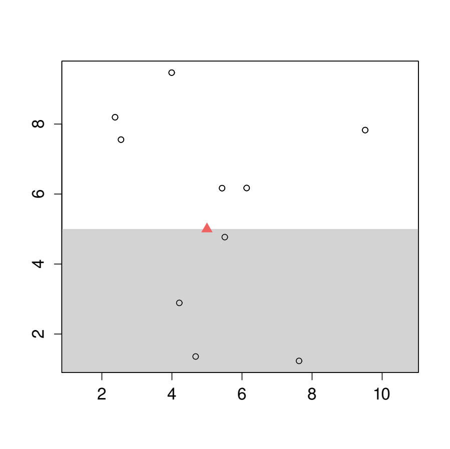
# Halfspace depth

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

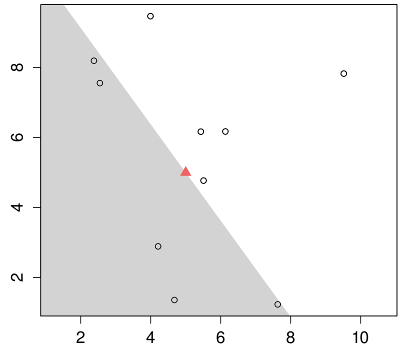


# Halfspace depth

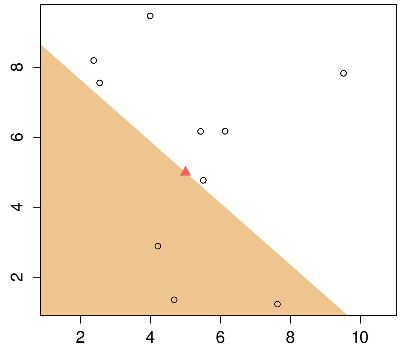
$$hD(x; P_n) = \min \frac{\#\text{ of observations in a halfspace that contains } x}{n}$$



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



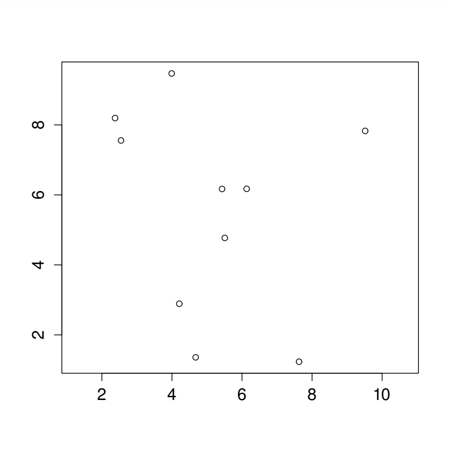
$$hD\left(x; P_n\right) = \min \frac{\text{\# of observations in a halfspace that contains } x}{n}$$



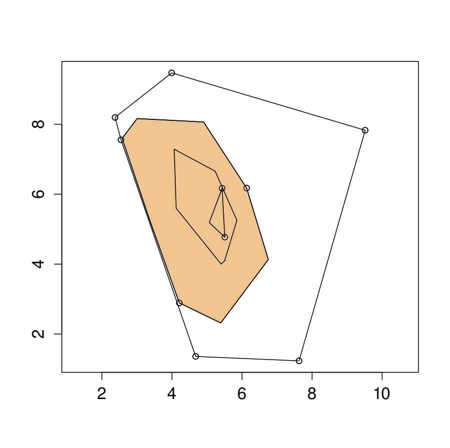
# A brief history of *hD* (in statistics)

- 1955 The idea with minimal halfspaces first used by Hodges;
- 1975 Tukey proposes *hD* as a visualisation tool;
- 1982 Donoho studies *hD* in his Ph.D. thesis;
- 1992 depth introduced in the AoS (Donoho and Gasko, 1992);
- 1999 Rousseeuw and Ruts study *hD* in full generality;
- 2000 Zuo and Serfling provide a general framework for the depth.

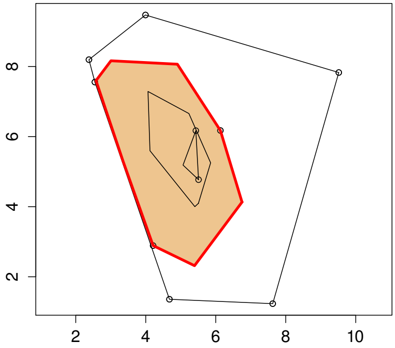
$$hD_\delta(P) = \{x \in \mathbb{R}^d \colon hD(x; P) \geq \delta\}$$



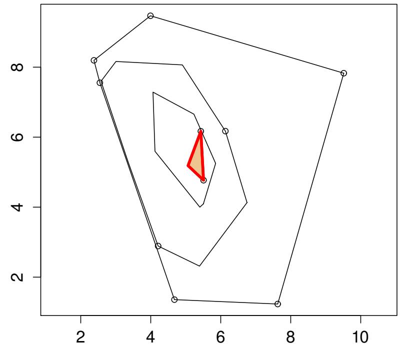
$$hD_\delta(P) = \{x \in \mathbb{R}^d \colon hD(x; P) \geq \delta\}$$



Topological boundary of $hD_\delta(P)$



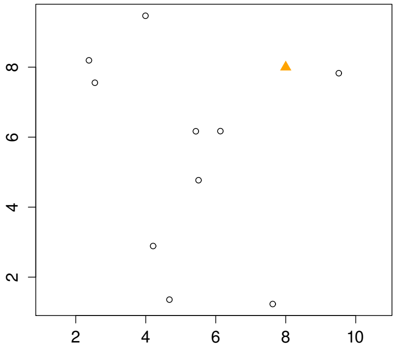
Point(s) at which the depth $hD(\cdot; P)$ is maximized over $\mathbb{R}^d$



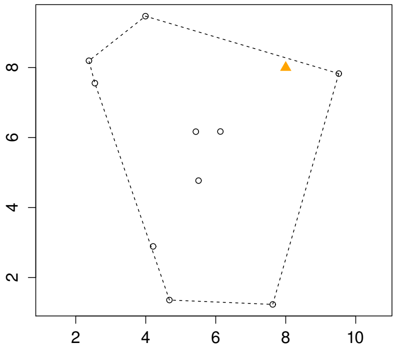
It holds true that

- $hD(x;P)$ is well defined for any $x \in \mathbb{R}^d$ and $P \in \mathcal{P}\left(\mathbb{R}^d\right)$;
- $hD(x;P) \in [0,1]$;
- a halfspace median always exists;
- $hD(x;P) \leq \delta$ iff $\forall \varepsilon > \delta \; \exists H \in \mathcal{H}(x)\colon P(H) \leq \varepsilon$;
- $hD(x;P) = \inf_{u \in \mathbb{S}^{d-1}} hD(\langle x, u \rangle ; P_{\langle X, u \rangle})$.

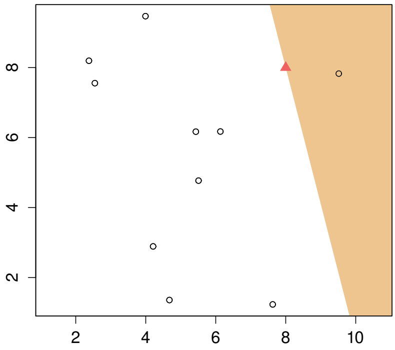
Minimizing halfspace at $x$ is $H \in \mathcal{H}(x)$ such that $P(H) = hD(x; P)$.



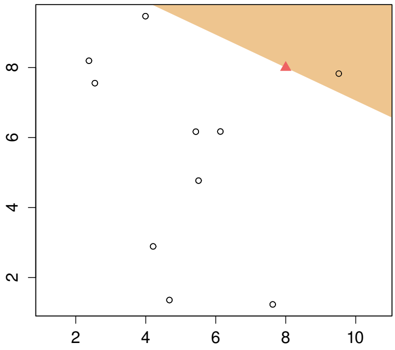
Minimizing halfspace is not always unique



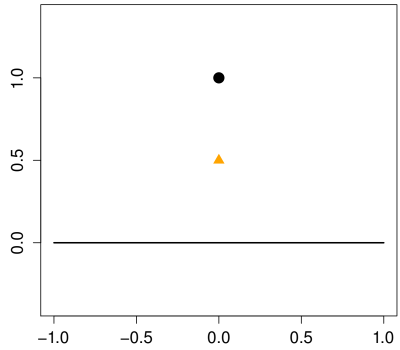
Minimizing halfspace is not always unique



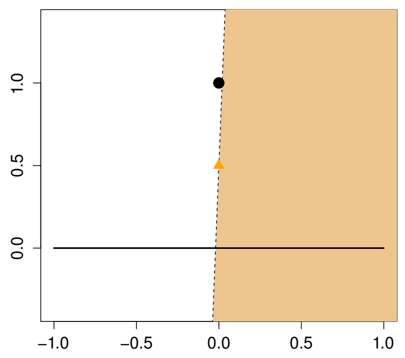
Minimizing halfspace is not always unique



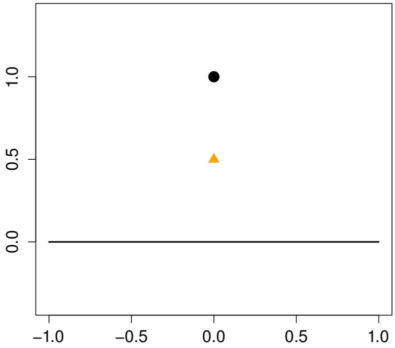
Minimizing halfspace does not always exist



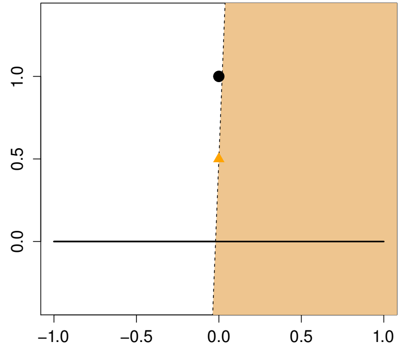
Minimizing halfspace does not always exist



$P(\partial H) = 0$ for each halfspace $H$



Minimizing halfspace does not always exist

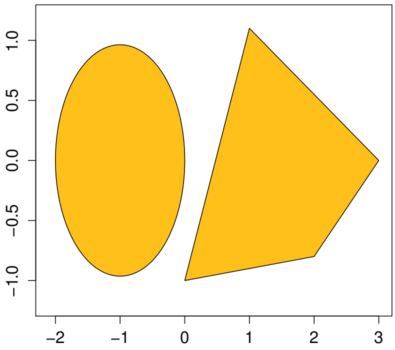



There always is such a flag halfspace (Pokorný et al., 2021+)

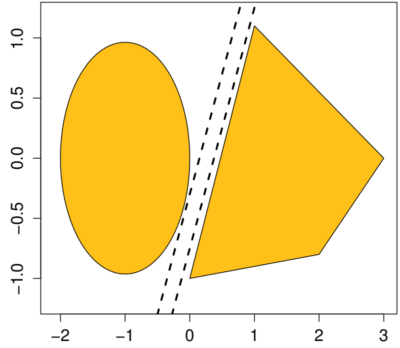
# Assumption 2: Contiguous support (C)

The mass of $P$ cannot be divided by a slab of zero probability

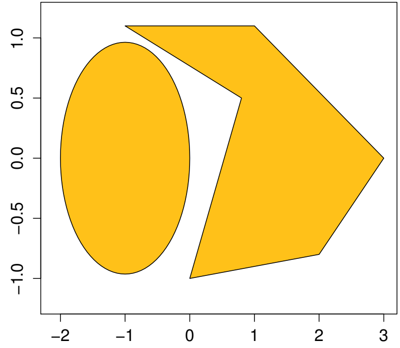
(Mizera and Volauf, 2002)



The mass of $P$ cannot be divided by a slab of zero probability (Mizera and Volauf, 2002)



The mass of $P$ cannot be divided by a slab of zero probability (Mizera and Volauf, 2002)



For $P$ that satisfies (S)

- $hD(x; P) \in [0, 1/2]$;
- a minimizing halfspace exists at any $x \in \mathbb{R}^d$;
- if (C) is also true, we have unique halfspace median (Mizera and Volauf, 2002).

# Further properties

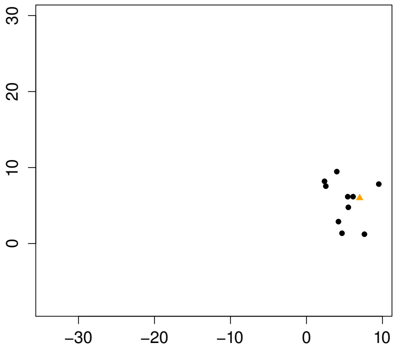
For $P$ that satisfies (S)

- $hD(x; P) \in [0, 1/2]$;
- a minimizing halfspace exists at any $x \in \mathbb{R}^d$;
- ~~if (C) is also true, we have unique halfspace median~~ (Mizera and Volauf, 2002).

Conditions under which the halfspace median is unique are more complicated (Part II).

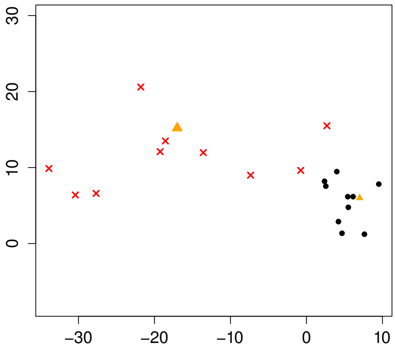
For any $A \in \mathbb{R}^{d \times d}$ non-singular and $b \in \mathbb{R}^d$

$$hD(x; P_X) = hD(Ax + b; P_{AX+b}).$$



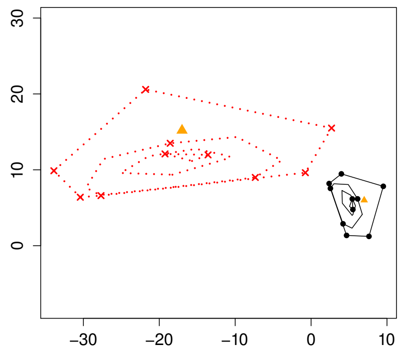
For any $A \in \mathbb{R}^{d \times d}$ non-singular and $b \in \mathbb{R}^d$

$$hD(x; P_X) = hD(Ax + b; P_{AX+b}).$$



For any $A \in \mathbb{R}^{d \times d}$ non-singular and $b \in \mathbb{R}^d$

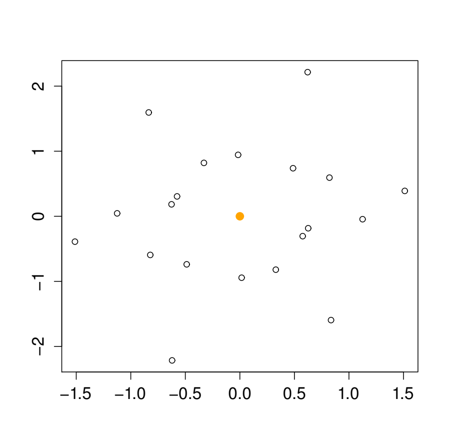
$$hD(x; P_X) = hD(Ax + b; P_{AX+b}).$$



# Maximality

If $X$ is symmetric (i.e. $P_X = P_{-X}$), then

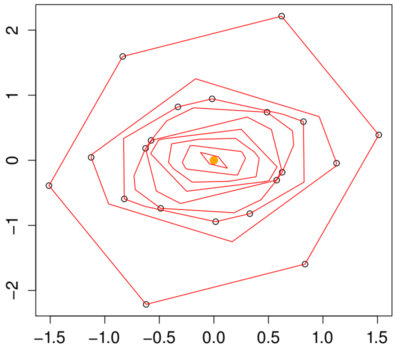
$$hD(0;P) = \sup_{x \in \mathbb{R}^d} hD(x;P).$$



# Maximality

If $X$ is symmetric (i.e. $P_X = P_{-X}$), then

$$hD(0; P) = \sup_{x \in \mathbb{R}^d} hD(x; P).$$



**Theorem (Mizera and Volauf, 2002)**
*For any $x_\nu \to x$ in $\mathbb{R}^d$ and $P_\nu \xrightarrow[\nu\to\infty]{w} P$ in $\mathcal{P}\left(\mathbb{R}^d\right)$*

$$\limsup_{\nu\to\infty} hD(x_\nu; P_\nu) \leq hD(x; P).$$

*In particular,*

$$\limsup_{\nu\to\infty} hD(x_\nu; P) \leq hD(x; P).$$

*If $P$ satisfies* (S) *then also*

$$\lim_{\nu\to\infty} hD(x_\nu; P_\nu) = hD(x; P).$$

**Proof:** Portmanteau theorem: $P_\nu \xrightarrow[\nu\to\infty]{w} P$ if and only if $\limsup_{\nu\to\infty} P_\nu(F) \leq P(F)$ for all $F$ closed. Now $\mathcal{H}(x)$ is a collection of closed sets.

Mizera and Volauf (2002):

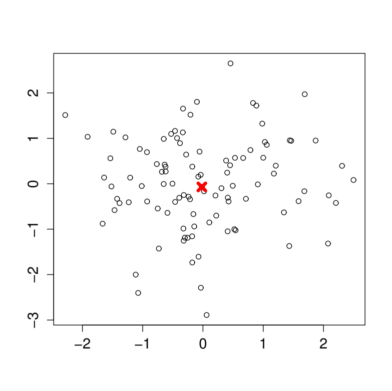
$$\limsup_{\nu\to\infty} hD(x_\nu; P) \le hD(x; P).$$

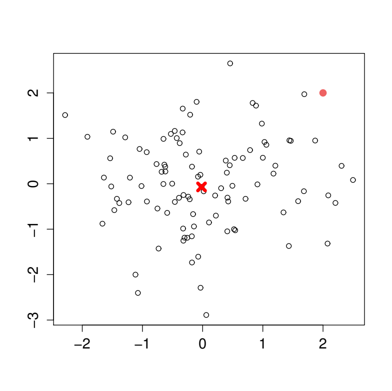
Let $\{x_\nu\}_{\nu=1}^{\infty} \subset hD_\delta(P)$, $x_\nu \xrightarrow[\nu\to\infty]{} x$. Then

$$hD(x; P) \ge \limsup_{\nu\to\infty} hD(x_\nu; P) \ge \delta.$$

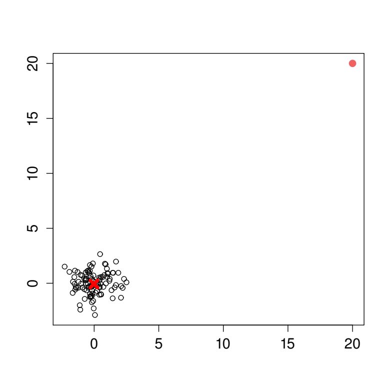
- The central regions $hD_\delta(P)$ are always closed.
- The halfspace median set is non-empty and compact.

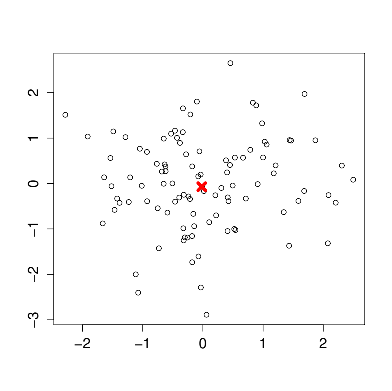
Halfspace median is a robust estimator



Halfspace median is a robust estimator



Consider the breakdown point of an estimator $T \colon \mathcal{P}\left(\mathbb{R}^d\right) \to \mathbb{R}^d$ at a dataset $\mathbb{X}_n = \{X_1, \ldots, X_n\} \subset \mathbb{R}^d$

$$\varepsilon(T, \mathbb{X}_n) = \min_{1 \leq m \leq n} \left\{ \frac{m}{m+n} \colon \sup_{\mathbb{Y}_m} \|T(\mathbb{X}_n) - T(\mathbb{X}_n \cup \mathbb{Y}_m)\| = \infty \right\}.$$

The asymptotic breakdown point $\varepsilon^*(T)$ of $T$ is the (almost sure) limit of $\varepsilon(T, \mathbb{X}_n)$ as $n \to \infty$ and $X_1, X_2, \cdots \sim P$ independent.

- For $T$ the mean we have $\varepsilon^*(T) = 0$.
- For $T$ the univariate median we have $\varepsilon^*(T) = 1/2$.

The best possible value is $\varepsilon^*(T) = 1/2$, of course.

# Halfspace median: Breakdown point and optimality

For $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ elliptically symmetric is the halfspace median $T$

- highly robust also in high dimensions (Donoho and Gasko, 1992):

$$\varepsilon^*(T) = \begin{cases} 1/2 & \text{for } d = 2, \\ 1/3 & \text{for } d > 2. \end{cases}$$

- a **minimax optimal** estimator in Huber's contamination model (Chen et al., 2018);
- converges with the same rate even with $\mathcal{O}(\sqrt{n\,d})$ contaminating points as $n \to \infty$.
  (only $\mathcal{O}(\sqrt{n})$ for the coordinatewise median)

Arguably, **the best robust affine equivariant estimator** we have.

**Theorem (Donoho and Gasko, 1992)**
*For any $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ almost surely*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}^d} |hD(x;P_n) - hD(x;P)| = 0.$$

**Proof:** We have

$$\begin{aligned}\sup_{x\in\mathbb{R}^d} |hD(x;P_n) - hD(x;P)| &= \sup_{x\in\mathbb{R}^d} \left| \inf_{H\in\mathcal{H}(x)} P_n(H) - \inf_{H\in\mathcal{H}(x)} P(H) \right| \\ &\leq \sup_{x\in\mathbb{R}^d} \sup_{H\in\mathcal{H}(x)} |P_n(H) - P(H)| = \sup_{H\in\mathcal{H}} |P_n(H) - P(H)|\,.\end{aligned}$$

The last expression is known to vanish almost surely as $n \to \infty$ due to the Glivenko-Cantelli theory (e.g. Dudley, 1999).

**Theorem (Donoho and Gasko, 1992)**
*For any* $P \in \mathcal{P}\left(\mathbb{R}^d\right)$

$$\lim_{\|x\| \to \infty} hD(x; P) = 0.$$

**Proof:** Easy.

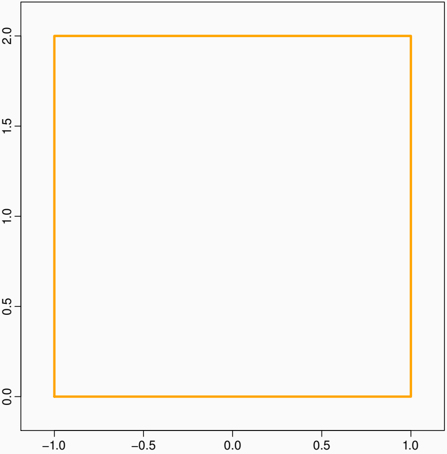
For each $\delta > 0$ it holds true that (Rousseeuw and Ruts, 1999)

- $hD_\delta(P) = \bigcap\{H \in \mathcal{H} \colon P(H) > 1 - \delta\}$;
- $hD_\delta(P)$ is closed;
- $hD_\delta(P)$ is bounded;
- $hD_\delta(P)$ is convex.

$hD(\cdot; P)$ is a quasi-concave function for any $P$.

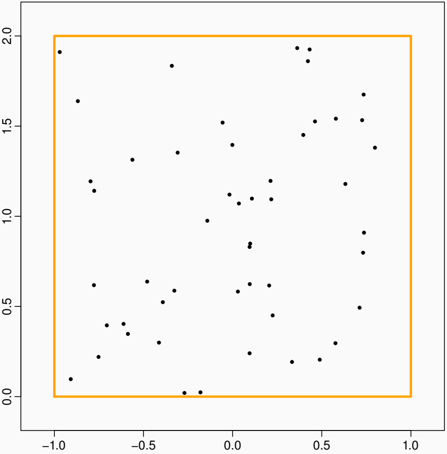
$hD$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

$$\left\{x \in \mathbb{R}^d : hD(x; P) \geq \delta\right\} \text{ is a convex set}$$



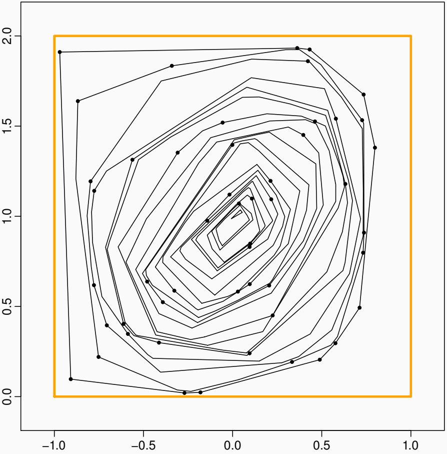
$hD$ is always quasi-concave, i.e. for each $\delta \in [0,1]$

$$\left\{x \in \mathbb{R}^d : hD(x;P) \geq \delta\right\} \text{ is a convex set}$$



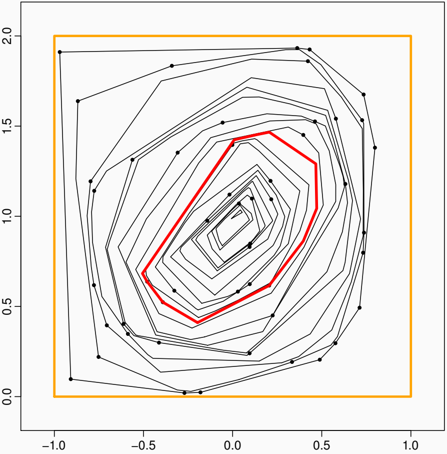
$hD$ is always quasi-concave, i.e. for each $\delta \in [0,1]$

$$\left\{x \in \mathbb{R}^d : hD(x;P) \geq \delta\right\} \text{ is a convex set}$$



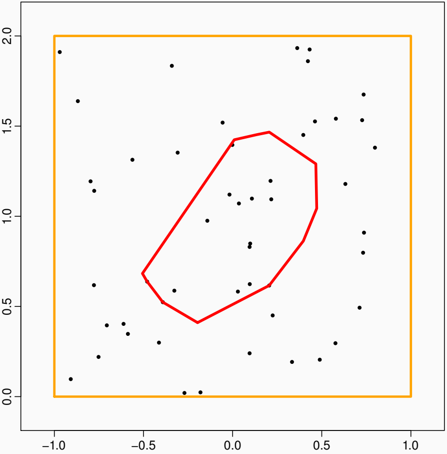
$hD$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

$$\left\{x \in \mathbb{R}^d : hD(x; P) \geq \delta\right\} \text{ is a convex set}$$



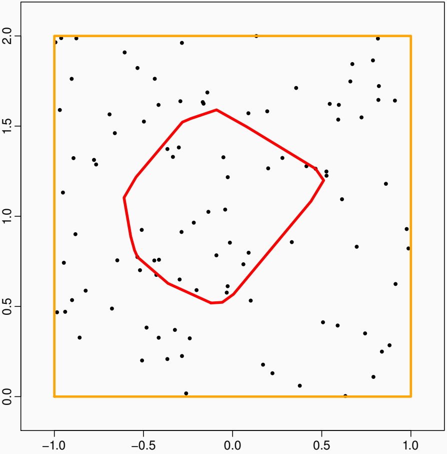
Consider the set-valued mapping

$$\delta \mapsto \left\{x \in \mathbb{R}^d \colon hD(x; P) \geq \delta\right\}$$



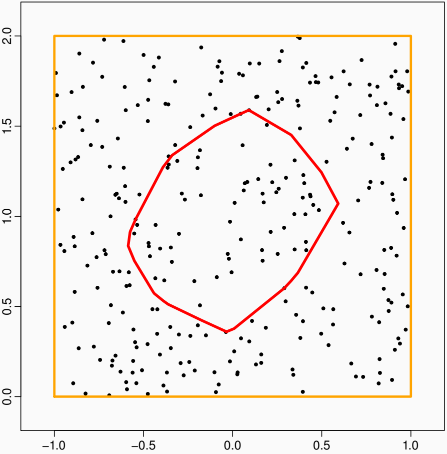
Consider the set-valued mapping

$$\delta \mapsto \left\{ x \in \mathbb{R}^d \colon hD(x; P) \geq \delta \right\}$$



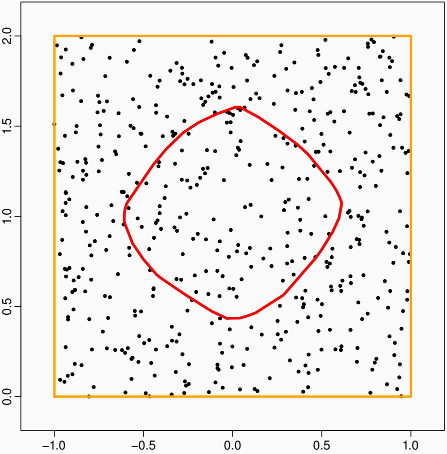
Consider the set-valued mapping

$$\delta \mapsto \left\{ x \in \mathbb{R}^d \colon hD(x; P) \geq \delta \right\}$$



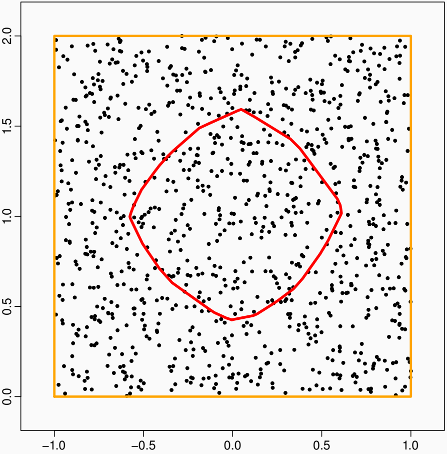
Consider the set-valued mapping

$$\delta \mapsto \left\{ x \in \mathbb{R}^d \colon hD(x; P) \geq \delta \right\}$$



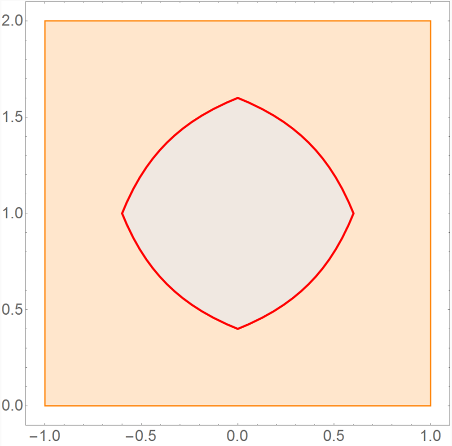
Consider the set-valued mapping

$$\delta \mapsto \left\{ x \in \mathbb{R}^d \colon hD(x;P) \geq \delta \right\}$$



Consider the set-valued mapping

$$\delta \mapsto \left\{x \in \mathbb{R}^d \colon hD(x;P) \geq \delta\right\}$$



Convex sets are equipped with the Hausdorff distance $d_H$

$$d_H(K_1, K_2) = \max\left\{\sup_{x\in K_1}\inf_{y\in K_2}\|x-y\|, \sup_{x\in K_2}\inf_{y\in K_1}\|x-y\|\right\}$$

for $K_1, K_2$ convex compact in $\mathbb{R}^d$.

**Theorem (Dyckerhoff, 2017+; Laketa and Nagy, 2021)**
*Let* **(S)** *and* **(C)** *be true for $P$. Then the mapping $\delta \mapsto hD_\delta(P)$ is continuous. Further, for any $\delta$*

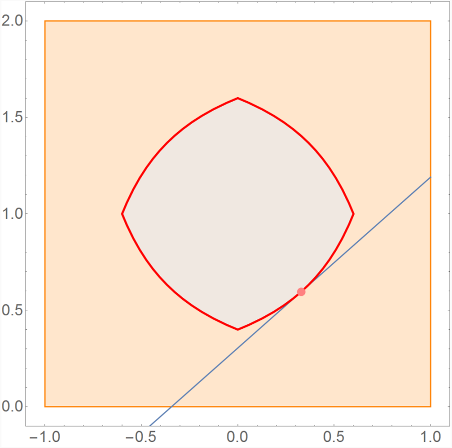
$$d_H(hD_\delta(P_n), hD_\delta(P)) \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

**Proof:** Set convergence and its properties.

# Asymptotic normality

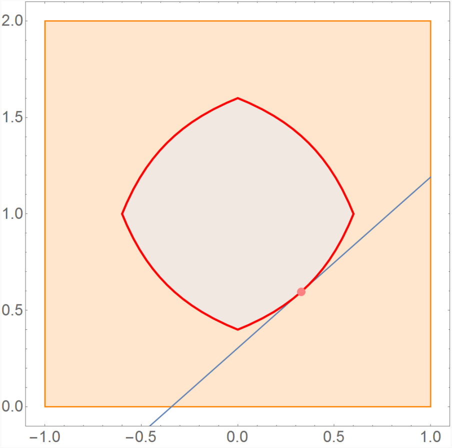
$\sqrt{n}\,(hD(x; P_n) - hD(x; P))$ is asymptotically normal (Massé, 2004)

$\iff$ there is a unique minimizing halfspace $H \in \mathcal{H}(x)$ of $P$ at $x$



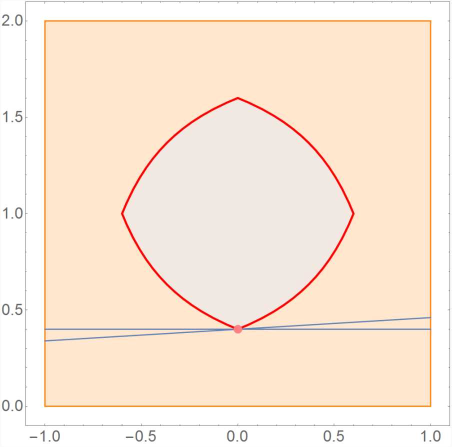
$\sqrt{n}\,(hD(x; P_n) - hD(x; P))$ is asymptotically normal

$\iff$ the contour of $hD(\cdot; P)$ is smooth at $x$

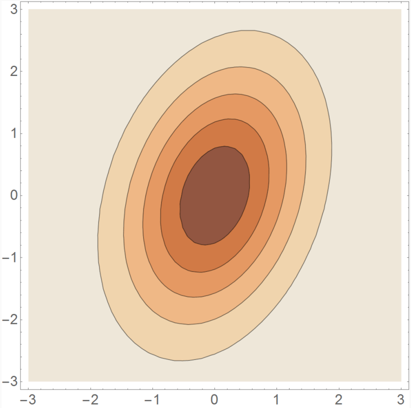


$\sqrt{n}\,(hD(x; P_n) - hD(x; P))$ is asymptotically normal

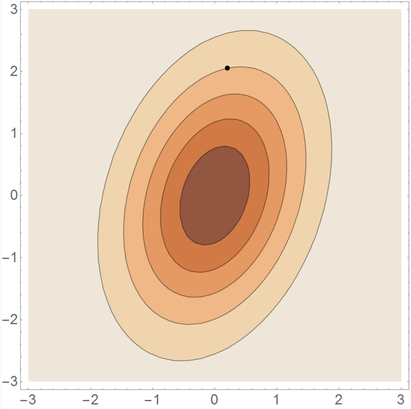
$\iff$ the contour of $hD(\cdot; P)$ is smooth at $x$



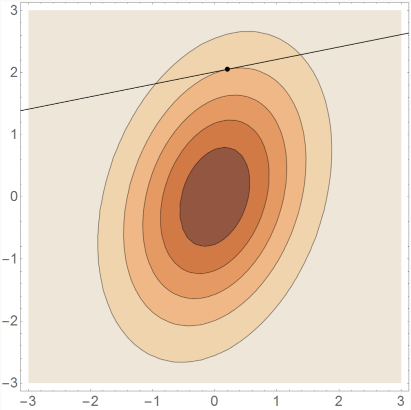
Elliptically symmetric distributions have elliptical depth contours



Elliptically symmetric distributions have elliptical depth contours



Elliptically symmetric distributions have elliptical depth contours



A measure $P \in \mathcal{P}(\mathbb{R}^d)$ is called $\alpha$-symmetric (Eaton, 1981) if its characteristic function takes the form

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

for some $\xi \colon \mathbb{R} \to \mathbb{R}$, where $\|\cdot\|_\alpha$ is the $\mathcal{L}_\alpha$-norm on $\mathbb{R}^d$.

For $X = (X_1, \ldots, X_d) \sim P$, these measures satisfy (Fang et al., 1990)

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

Examples:

- for $\alpha = 2$ we obtain the spherically symmetric distributions;
- for $\alpha = 1$ and $\xi(t) = \exp(-t)$ we get a multivariate Cauchy;
- for no other $\alpha \in (0, 2]$ there exists an explicit form for the density of $P$.

**Theorem (Massé and Theodorescu, 1994)**
*Let $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ be $\alpha$-symmetric. Set the conjugate exponent to $\alpha$*

$$\beta = \begin{cases} \alpha/(\alpha-1) & \text{if } \alpha > 1, \\ \infty & \text{if } \alpha \leq 1. \end{cases}$$

*Then the depth regions $hD_\delta(P)$ are the level sets of the norm $\|\cdot\|_\beta$.*

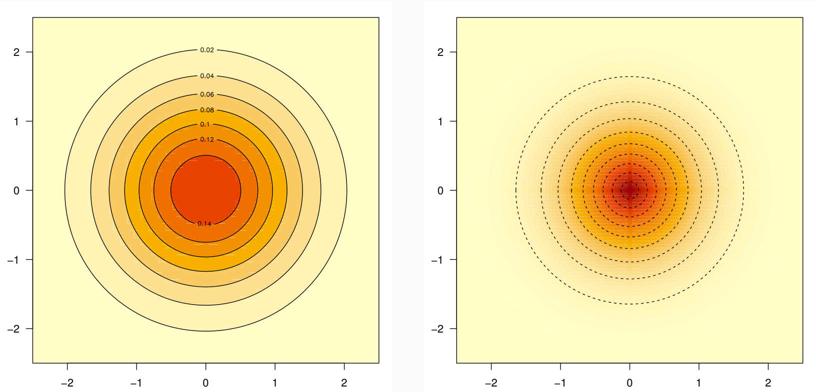
**Proof:** General proof in part II of the lectures. For $\alpha = 2$ and $X = (X_1, \dots, X_d) \sim P$ we can write

$$hD\left(x; P\right) = \inf_{u \in \mathbb{S}^{d-1}} \mathsf{P}\left(\langle X, u\rangle \leq \langle x, u\rangle\right) = \inf_{u \in \mathbb{S}^{d-1}} \mathsf{P}\left(\|u\| X_1 \leq \langle x, u\rangle\right)$$

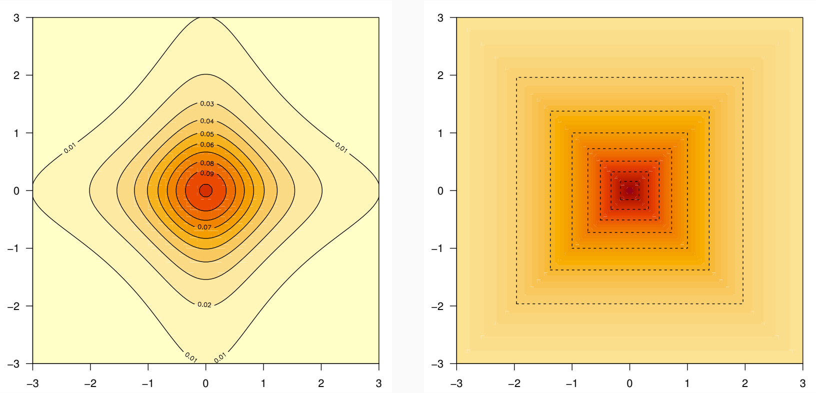
$$= \mathsf{P}\left(X_1 \leq \inf_{u \in \mathbb{S}^{d-1}} \langle x, u\rangle / \|u\|\right) = F_1\left(-\|x\|\right)$$

for $F_1$ the c.d.f. of $X_1$.

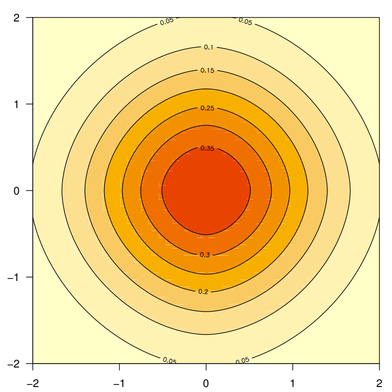
Multivariate normal distribution ($\alpha = 2$)

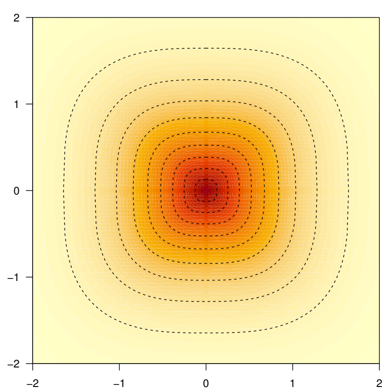


Multivariate Cauchy distribution ($\alpha = 1$)

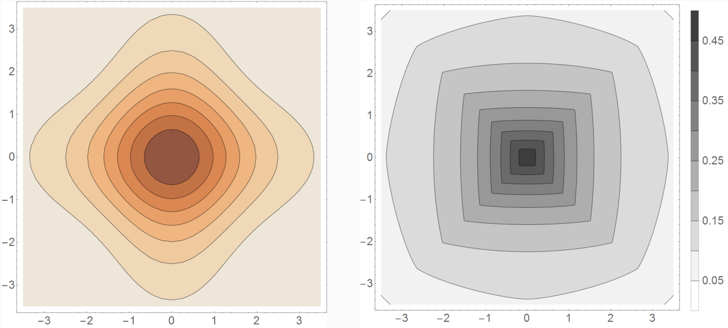


Multivariate 1.5-symmetric distribution ($\beta = 3$)

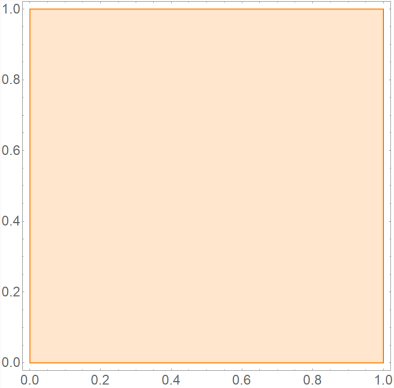


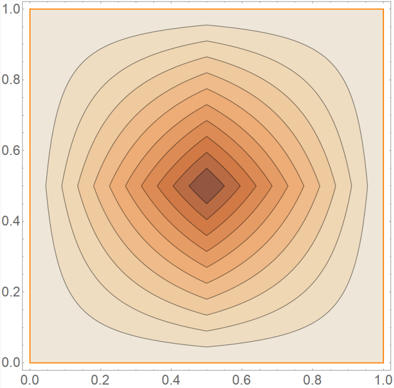



Mixture of two bivariate normal distributions (Gijbels and Nagy, 2016)



Uniform distribution on a simple convex body

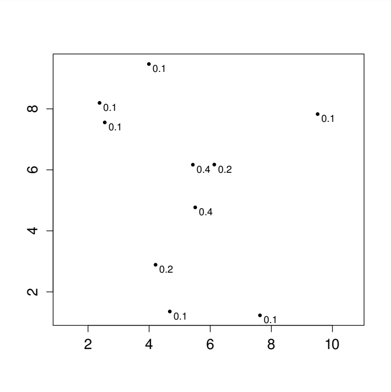


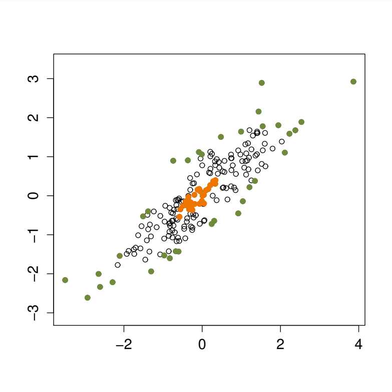



**Problem (Massé and Theodorescu, 1994)**

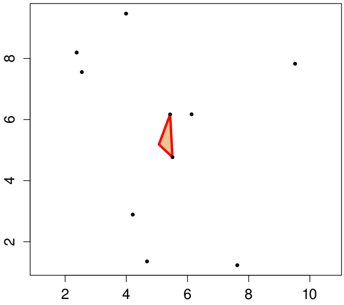
Is there any non-$\alpha$-symmetric distribution with smooth depth contours?

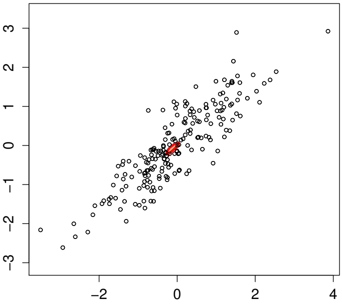
Depth induces a centre — outward ordering of points



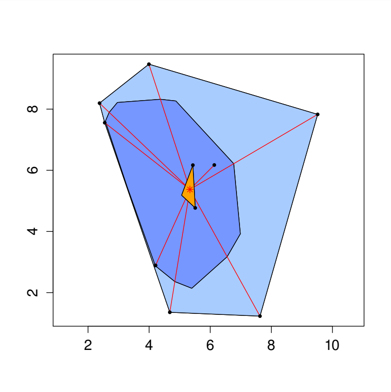



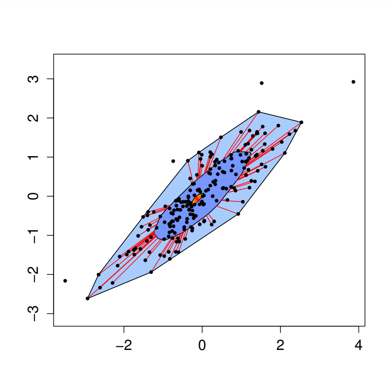
Point(s) that maximize the depth over $\mathbb{R}^d$



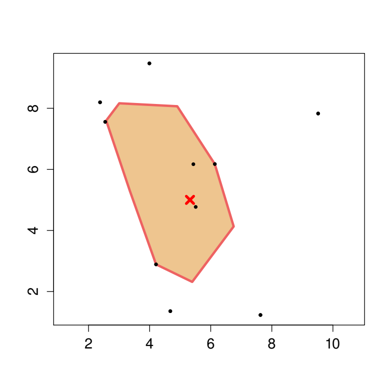
Central bag: 50 % deepest observations (Rousseeuw et al., 1999)

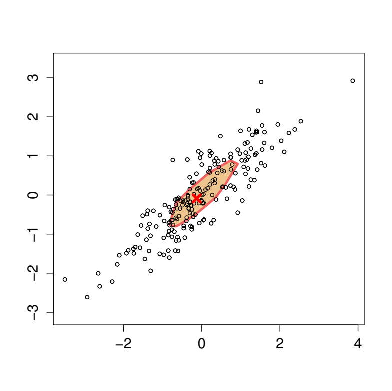


## Depth-trimmed mean (Fraiman and Meloche, 1999)

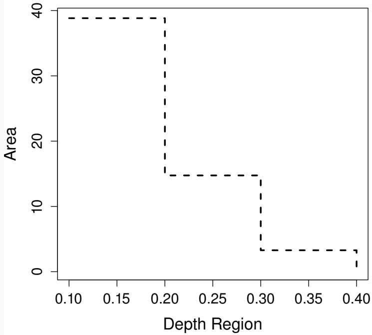
$$\sum_{i=1}^{n} X_i \,\mathbb{I}\left(hD(X_i; P_n) \geq \delta\right) \Big/ \sum_{i=1}^{n} \mathbb{I}\left(hD(X_i; P_n) \geq \delta\right)$$

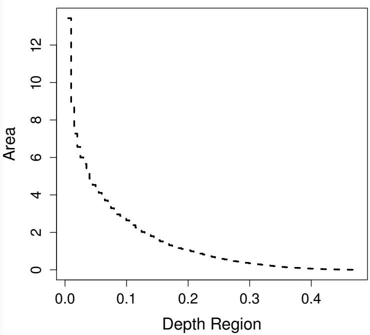


Volume of the depth region (Liu et al., 1999)

$$s\colon [0,1] \to [0,\infty)\colon \delta \mapsto \lambda\left(hD_\delta(P)\right)$$



Let $X_1, \ldots, X_n \sim P$ and $Y_1, \ldots, Y_m \sim Q$ be independent **multivariate** random samples. Test

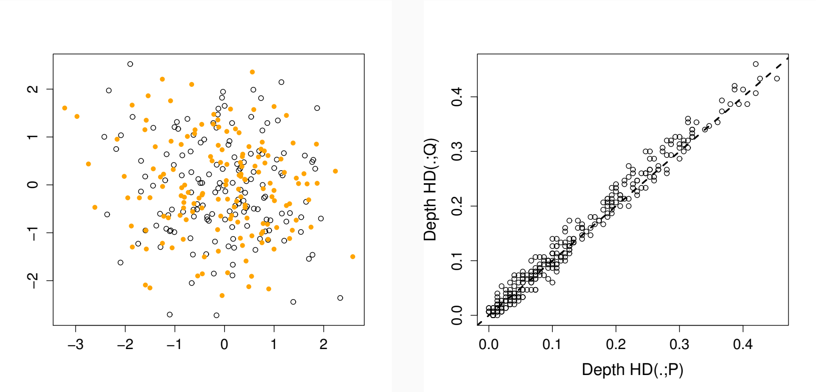
$$H_0\colon P = Q \quad \text{against} \quad H_1\colon P \neq Q.$$

Wilcoxon's rank sum test (Liu and Singh, 1993):

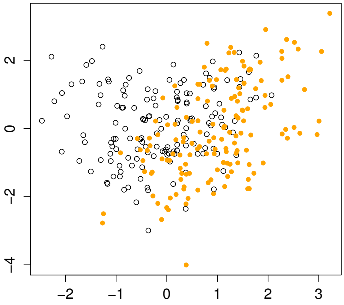
- Pool the two samples into $Z_1, \ldots, Z_{n+m}$ and rank these observations by their depth (1 through $n + m$).
- Add up the ranks of those observations which came from the sample from $P$. Denote by $R$.
- Reject $H_0$ if $R$ is either too small, or too large.

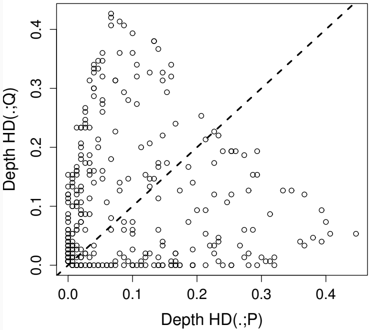
**Question:** Is this test a reasonable multivariate analogue to the Wilcoxon test from $\mathbb{R}$?

Replace quantiles by depth in Q-Q plots (Liu et al., 1999)

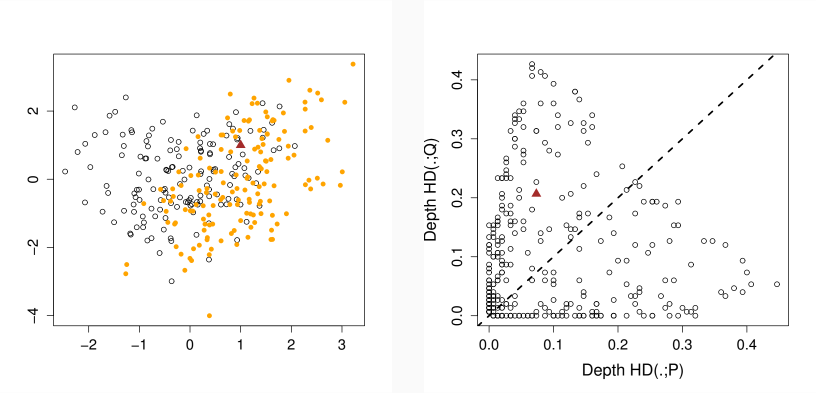


Replace quantiles by depth in Q-Q plots (Liu et al., 1999)

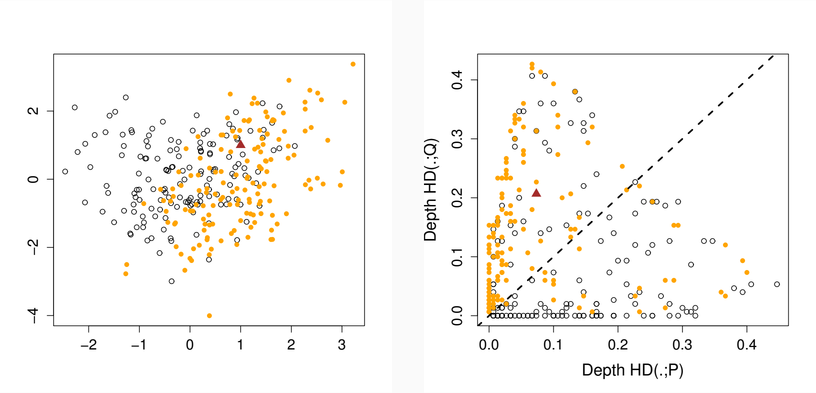


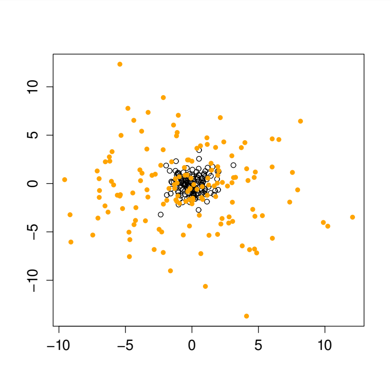
Classify a new observation into one of the groups (Li et al., 2012)

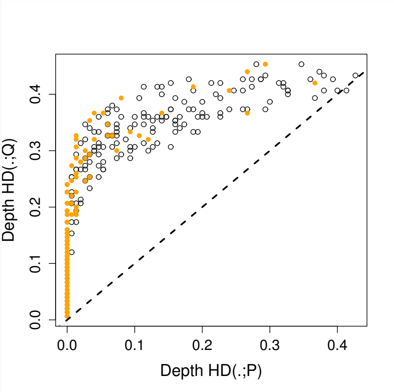


Classify a new observation into one of the groups (Li et al., 2012)



D-D plots with unequal scatters



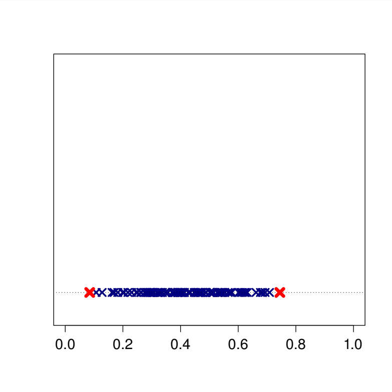



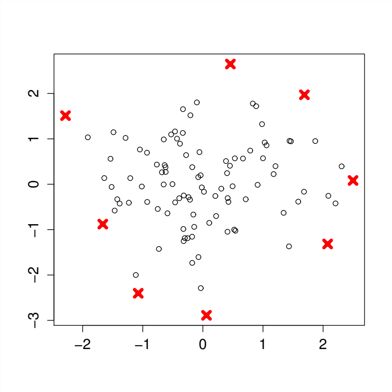
Computational cost of *hD*:

- Computing a single depth $hD(x; P)$ for $x \in \mathbb{R}^d$ is NP-hard in general (Johnson and Preparata, 1978);
- best known exact algorithms of complexity $\mathcal{O}(\log(n)n^{d-1})$ (Rousseeuw and Struyf, 1998);
- feasible exact computation available $n \leq 1000$ and $d \leq 5$ (Dyckerhoff and Mozharovskyi, 2016);
- very fast approximation algorithms exist (Dyckerhoff, 2004; Chen et al., 2013; Dyckerhoff et al., 2021);
- fast computation of central regions / halfspace median (Liu et al., 2019).

Implemented in R packages `depth` (Genest et al., 2008), `ddalpha` (Pokotylo et al., 2013), or `TukeyRegion` (Barber and Mozharovskyi, 2017).

With increasing $d$ the number of depth-ties increases



# Some open problems

Little is known about

- uniform distributional asymptotics;
- higher order asymptotics;
- detection of rough points;
- finite/large sample properties of depth-based tests and estimators;
- population depth and its properties.

**Conjecture (Struyf and Rousseeuw, 1998)**
For any $P, Q \in \mathcal{P}\left(\mathbb{R}^d\right)$, $P \neq Q$ there exists $x \in \mathbb{R}^d$ such that $hD(x; P) \neq hD(x; Q)$.

**Conjecture (Struyf and Rousseeuw, 1998)**
For any $P, Q \in \mathcal{P}\left(\mathbb{R}^d\right)$, $P \neq Q$ there exists $x \in \mathbb{R}^d$ such that $hD(x; P) \neq hD(x; Q)$.

Partial positive answers: This is true if

- $P$ and $Q$ are absolutely continuous with a compact support (Koshevoy, 2001);
- $P$ and $Q$ are empirical (Koshevoy, 2002);
- $P$ is atomic (Cuesta-Albertos and Nieto-Reyes, 2008);
- $P$ and $Q$ have smooth densities (Hassairi and Regaieg, 2008);
- $P$ and $Q$ have smooth depth contours (Kong and Zuo, 2010).

**Conjecture (Struyf and Rousseeuw, 1998)**
For any $P, Q \in \mathcal{P}\left(\mathbb{R}^d\right)$, $P \neq Q$ there exists $x \in \mathbb{R}^d$ such that $hD(x; P) \neq hD(x; Q)$.

Partial positive answers: This is true if

- ~~$P$ and $Q$ are absolutely continuous with a compact support (Koshevoy, 2001);~~
- $P$ and $Q$ are empirical (Koshevoy, 2002);
- ~~$P$ is atomic (Cuesta-Albertos and Nieto-Reyes, 2008);~~
- ~~$P$ and $Q$ have smooth densities (Hassairi and Regaieg, 2008);~~
- $P$ and $Q$ have smooth depth contours (Kong and Zuo, 2010).

# General depth and local depths

According to Zuo and Serfling (2000), statistical depth is a function

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

that satisfies

1. affine invariance;
2. maximality at the centre of symmetry for $P$ symmetric;
3. monotonicity on rays from the depth median;
4. vanishing at infinity.

According to Zuo and Serfling (2000), statistical depth is a function

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

that satisfies

1. affine invariance;
2. maximality at the centre of symmetry for $P$ symmetric;
3. monotonicity on rays from the depth median;
4. vanishing at infinity.

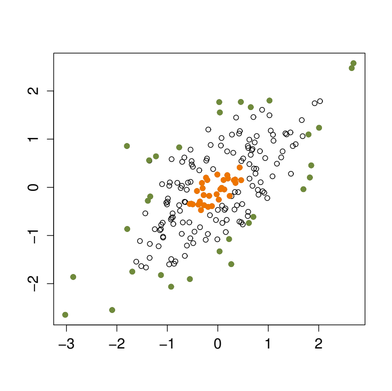
Serfling (2006) requires in addition also

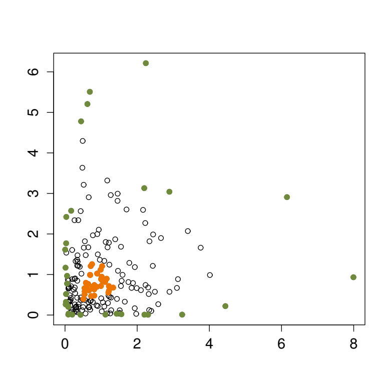
5. upper semi-continuity as a function of $x$;
6. continuity as a functional of $P$;
7. quasi-concavity in $x$.

# Simplicial depth

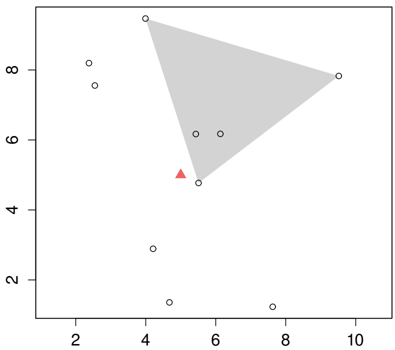
Simplicial depth (Liu, 1988) of $x \in \mathbb{R}^d$

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



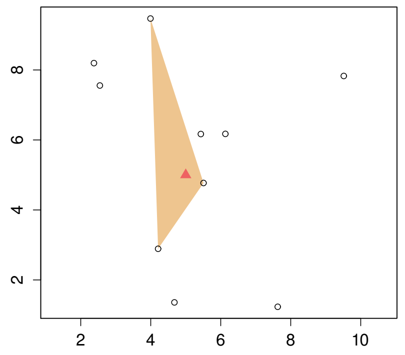



$$sD(x; P_n) = \binom{n}{d+1}^{-1} \sum_{1 \leq X_{i_1} < \cdots < X_{i_{d+1}} \leq n} \mathbb{I}\left(x \in \mathbb{S}(X_{i_1}, \ldots, X_{i_{d+1}})\right).$$



# Simplicial depth

$$sD(x; P_n) = \binom{n}{d+1}^{-1} \sum_{1 \leq X_{i_1} < \cdots < X_{i_{d+1}} \leq n} \mathbb{I}\left(x \in \mathbb{S}(X_{i_1}, \ldots, X_{i_{d+1}})\right).$$



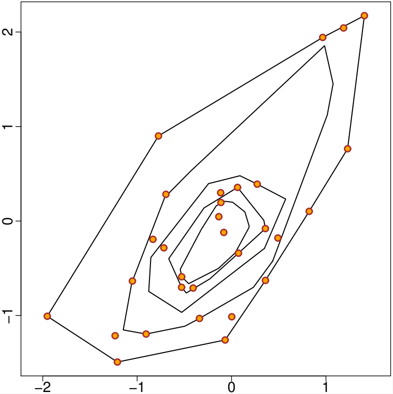
Advantages of simplicial depth:

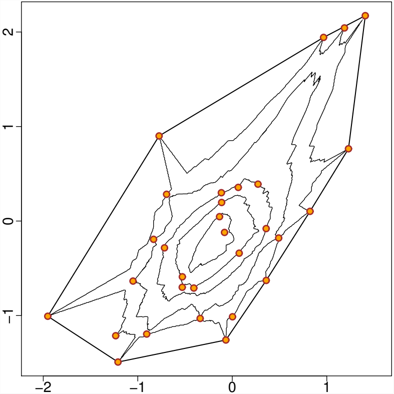
- affine invariant;
- U-statistic (i.e. nice statistical properties);
- upper semi-continuous in $x$;
- induces a robust median;
- vanishes at infinity.

But:

- not quasi-concave or monotonically decreasing;
- computationally expensive;
- population version difficult to study theoretically.

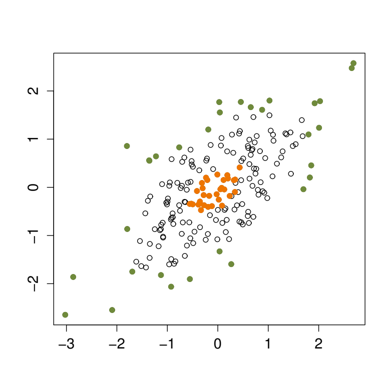
Halfspace (left) and simplicial (right) depth contours

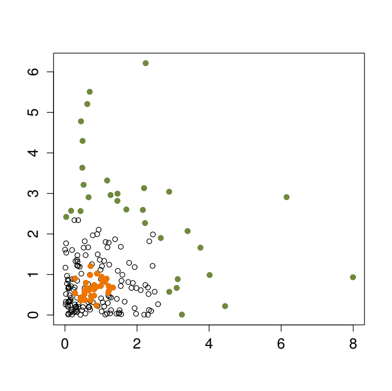


Simplicial volume depth (Oja, 1983) of $x \in \mathbb{R}^d$

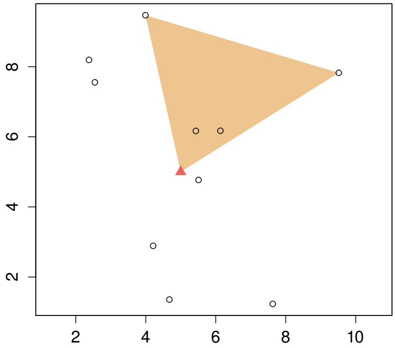
$$svD(x; P) = \left(1 + \mathrm{E}\,\lambda\left(\mathbb{S}\left(x, X_1, \ldots, X_d\right)\right)\right)^{-1}.$$



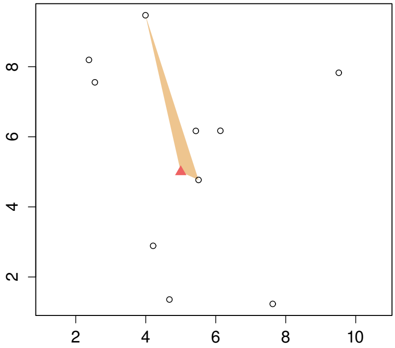



# Simplicial volume depth (Oja's depth)

$$svD(x; P_n) = \left(1 + \binom{n}{d}^{-1} \sum_i \lambda\left(\mathbb{S}\left(x, X_{i_1}, \ldots, X_{i_d}\right)\right)\right)^{-1}$$

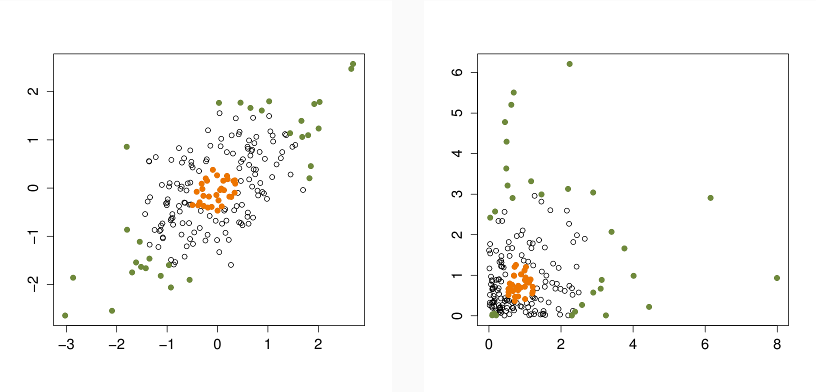


$$svD\left(x;P_n\right)=\left(1+\binom{n}{d}^{-1}\sum_i \lambda\left(\mathbb{S}\left(x,X_{i_1},\ldots,X_{i_d}\right)\right)\right)^{-1}$$



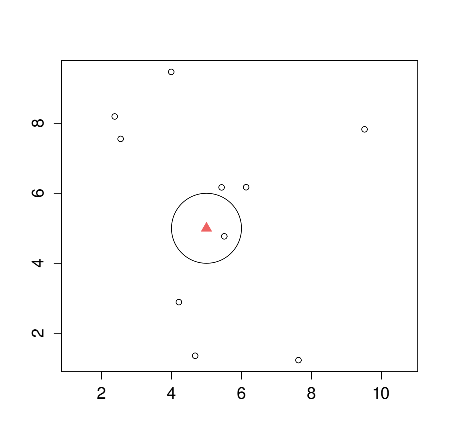
Spatial depth (Chaudhuri, 1996) of $x \in \mathbb{R}^d$

$$spD(x, P) = 1 - \left\| \mathrm{E} \frac{x - X}{\|x - X\|} \right\| .$$

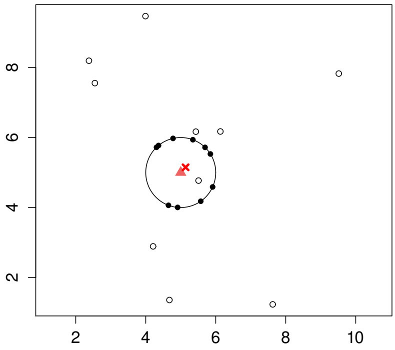


# Spatial depth

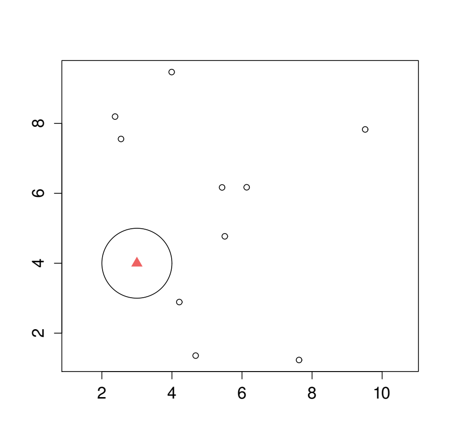
$$spD(x;P) = 1 - \left\| \mathbb{E} \frac{x - X}{\|x - X\|} \right\|$$



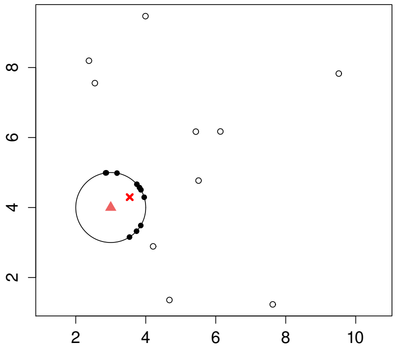
$$spD(x; P) = 1 - \left\| \mathrm{E} \frac{x - X}{\|x - X\|} \right\|$$



$$spD(x;P) = 1 - \left\| \mathbb{E}\, \frac{x - X}{\|x - X\|} \right\|$$



$$spD(x;P) = 1 - \left\| \mathrm{E} \frac{x - X}{\|x - X\|} \right\|$$



# Spatial depth: Properties

Advantages:

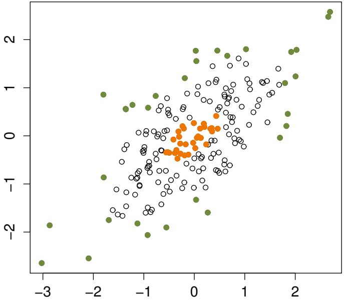
- rotation invariant;
- maximized at the spatial median, i.e. a point $x$ that minimizes
$$\mathrm{E}\,\|X - x\|\,;$$
- robust median;
- vanishes at infinity;
- very fast computation ($\mathcal{O}(n)$);
- works also in high-dimensional spaces.

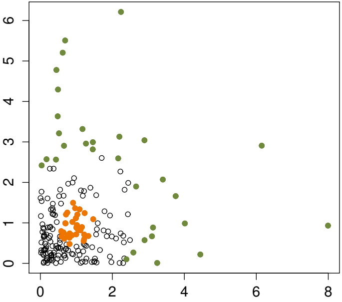
But:

- not affine invariant;
- not quasi-concave or monotonically decreasing.

Mahalanobis depth (Mahalanobis, 1936) of $x \in \mathbb{R}^d$

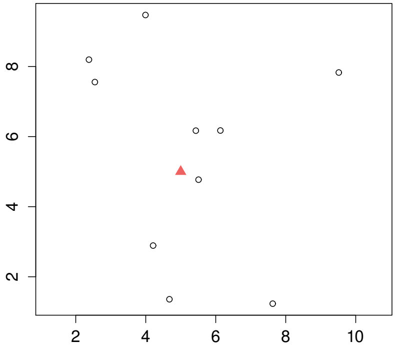
$$mD(x; P) = \left(1 + (x - \operatorname{E} X)^{\top} (\operatorname{Var} X)^{-1} (x - \operatorname{E} X)\right)^{-1}.$$



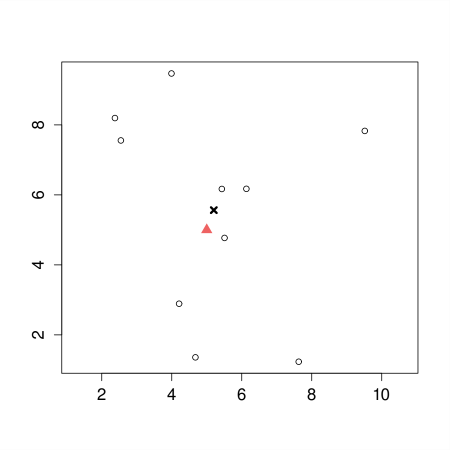



# Mahalanobis depth

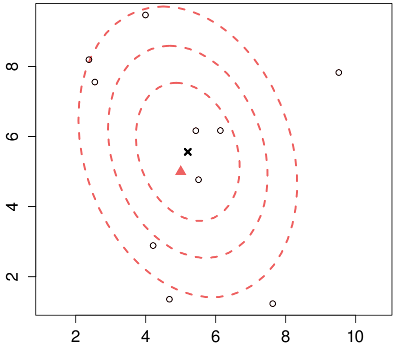
$mD(x; P) \sim$ Mahalanobis distance from $\mathrm{E}\,X$



$mD(x; P) \sim$ Mahalanobis distance from $\mathrm{E}\,X$

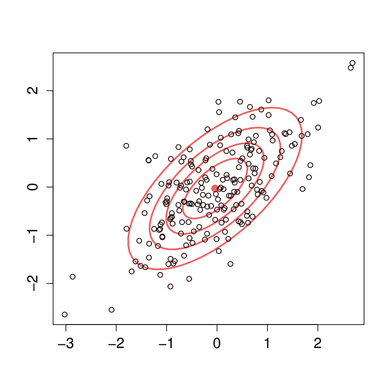


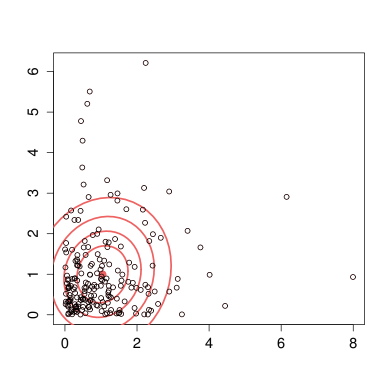
$mD(x; P) \sim$ Mahalanobis distance from $\mathrm{E}\,X$



Mahalanobis depth (Mahalanobis, 1936) of $x \in \mathbb{R}^d$

$$mD(x; P) = \left(1 + (x - \mathrm{E}\,X)^{\mathsf{T}} (\mathrm{Var}\,X)^{-1} (x - \mathrm{E}\,X)\right)^{-1}.$$



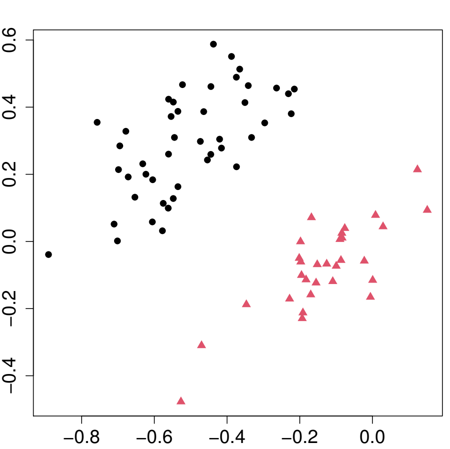



Disadvantages:

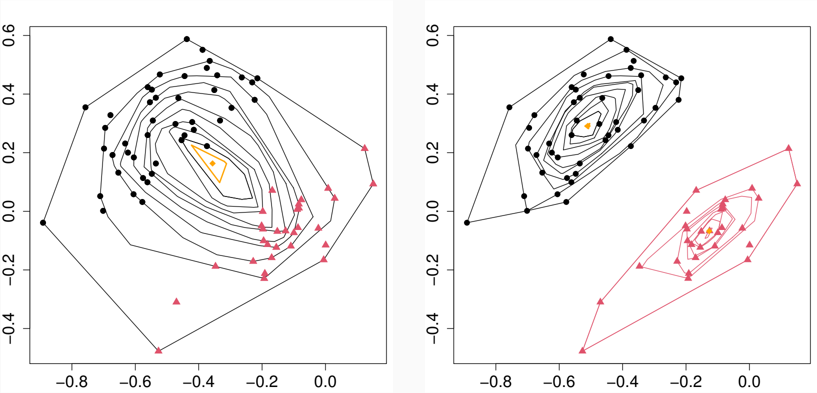
- not always defined (i.e. not entirely non-parametric);
- maximized at the mean ($\implies$ not robust);
- rigid contours (concentric ellipses of the same shape).

Not really a depth.

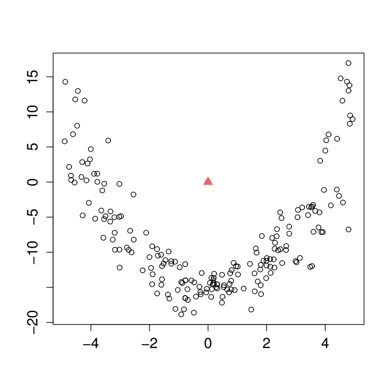
The depth suits well only for analysing unimodal distributions

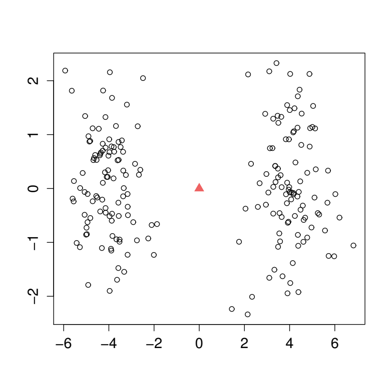


The depth suits well only for analysing unimodal distributions

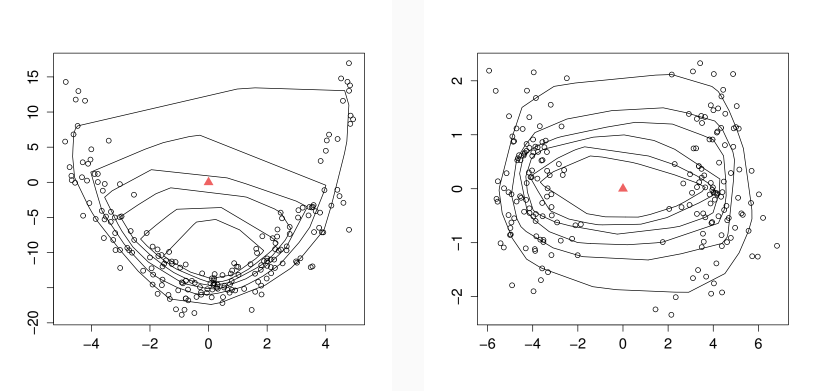


A proper depth is intended to be unimodal and quasi-concave

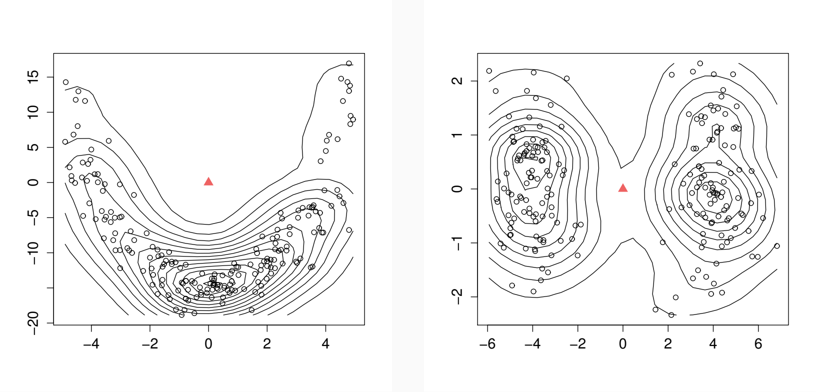


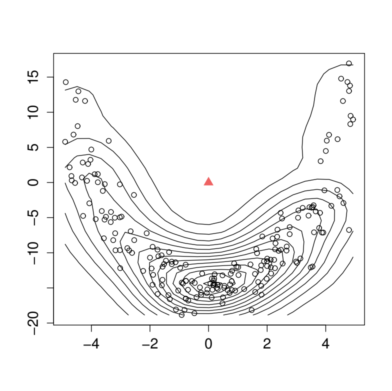
A proper depth is intended to be unimodal and quasi-concave

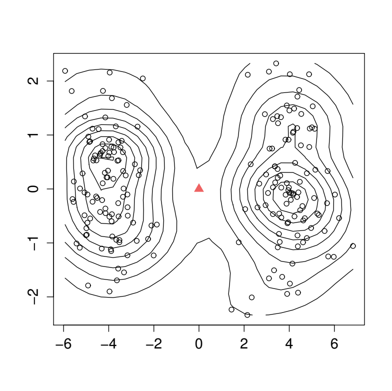


Relaxation of unimodality leads to local depths

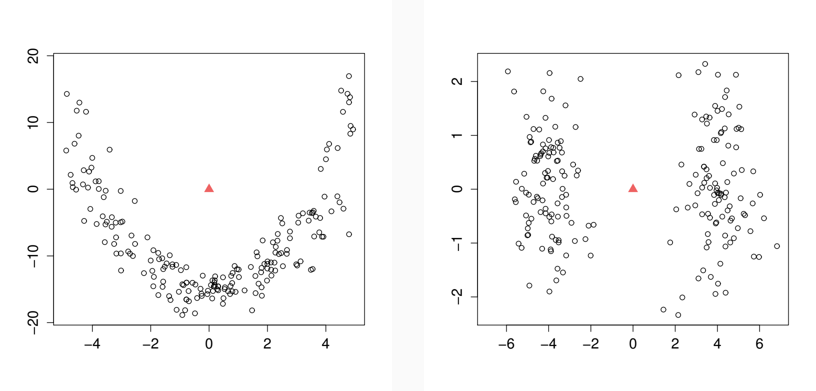


Multivariate density estimator (Fraiman and Meloche, 1999)

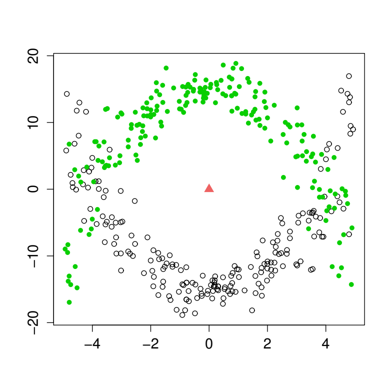


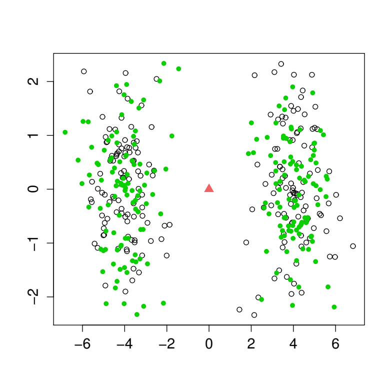



Localization of $hD$ (Paindaveine and Van Bever, 2013)

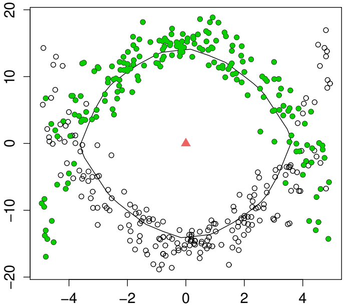


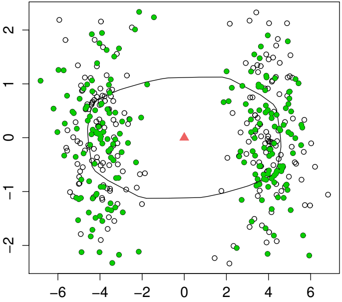
Localization of *hD* (Paindaveine and Van Bever, 2013)



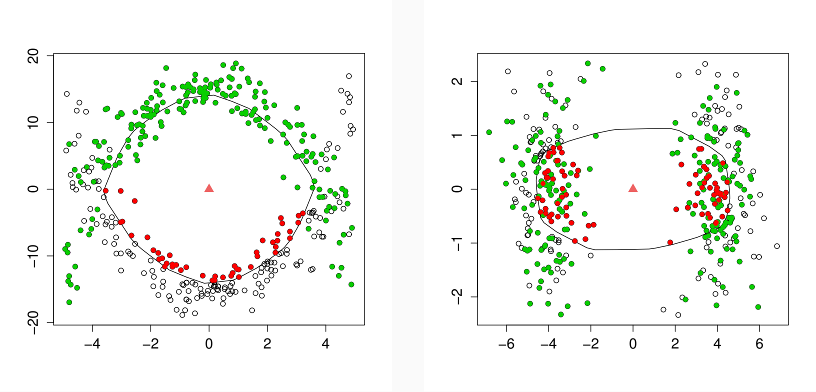



Localization of *hD* (Paindaveine and Van Bever, 2013)

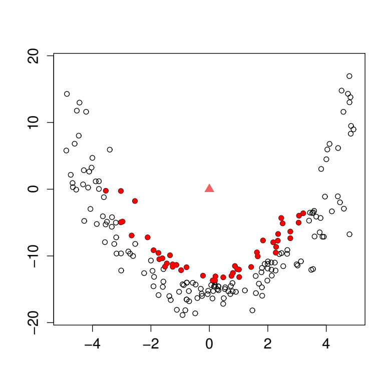


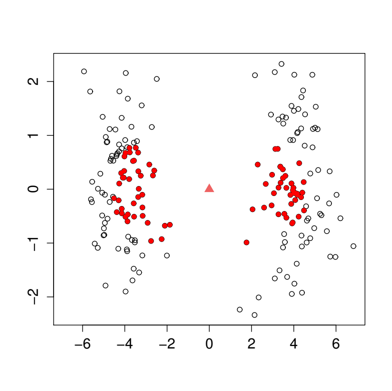



Localization of *hD* (Paindaveine and Van Bever, 2013)

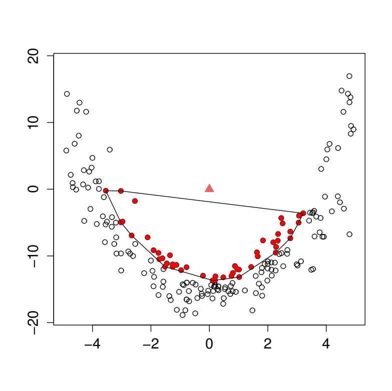


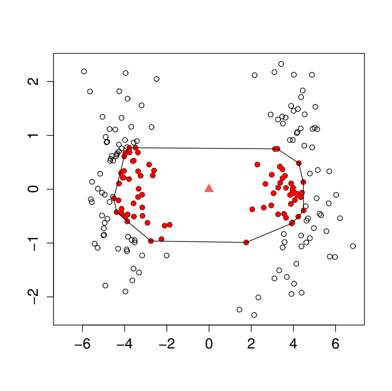
Localization of *hD* (Paindaveine and Van Bever, 2013)



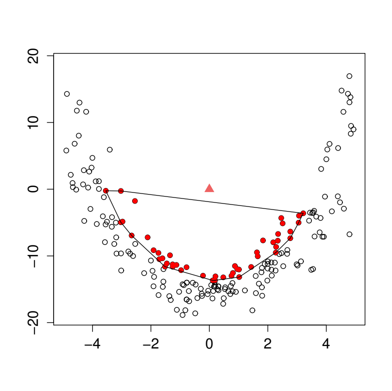



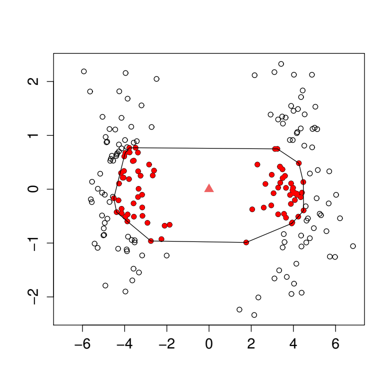
Localization of *hD* (Paindaveine and Van Bever, 2013)



Other approaches exist (Kotík and Hlubinka, 2017)



# Further extensions

Depths for more exotic data — variants of the halfspace and simplicial depth:

- for directional data (data in $\mathbb{S}^{d-1}$) (Liu and Singh, 1992);
- for data on images/graphs/networks (Small, 1997);
- for infinite-dimensional (functional) data (Fraiman and Muniz, 2001; López-Pintado and Romo, 2009);
- for general metric spaces (Carrizosa, 1996);
- in regression problems (Rousseeuw and Hubert, 1999);
- ...

Many proposals, many tests, many simulations.
No sufficient comprehensive theory.

# Conclusions

Statistical depth is

- easy to understand (i.e. extremely popular);
- promises many applications; but also
- computationally intensive;
- with isolated and underdeveloped theory.

# Selected Literature


- D. L. Donoho and M. Gasko, Breakdown properties of location estimates based on halfspace depth and projected outlyingness, Ann. Statist., 20 (1992), pp. 1803–1827.
- R. Y. Liu, On a notion of simplicial depth, Proc. Natl. Acad. Sci. U.S.A., 85 (1988), pp. 1732–1734.
- R. Y. Liu, J. M. Parelius, and K. Singh, Multivariate analysis by data depth: descriptive statistics, graphics and inference, Ann. Statist., 27 (1999), pp. 783–858.
- P. J. Rousseeuw and I. Ruts, The depth function of a population distribution, Metrika, 49 (1999), pp. 213–244.
- J. W. Tukey, Mathematics and the picturing of data, in Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2, Canad. Math. Congress, Montreal, Que., 1975, pp. 523–531.
- Y. Zuo and R. Serfling, General notions of statistical depth function, Ann. Statist., 28 (2000), pp. 461–482.