Data Depth and its Place in Modern Mathematics
Part I: Statistical Depth Function

Stanislav Nagy

PRIMUS/17/SCI/03: Advanced Geometric Methods in Statistics

KPMS Praha 2017

Data Depth and Its Place in Modern Mathematics

PRIMUS/17/SCI/03

Advanced Geometric Methods in Statistics

2018–2020

- Multivariate statistics and convex geometry

KPMS	doc. Hlubinka, dr. Dvořák, S. Nagy;
KAM	doc. Valtr;
MÚ	prof. Rataj;
MFF	and students.

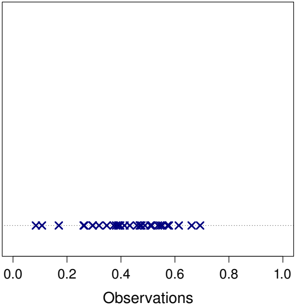
- 3 – 4 presentations on Mondays, October 2017 – January 2018

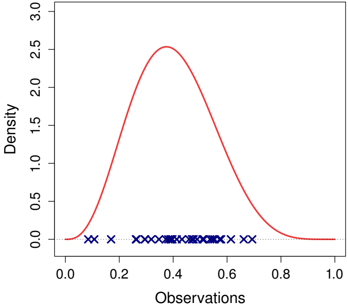
Part I: Statistical Data Depth (Introduction)

1. Motivation: Order Statistics, Quantiles and Ranks
   - Point estimation
   - Data visualisation
   - L-estimation and testing
2. Halfspace Depth: Quantiles for Multivariate Data
   - The depth and its properties
   - Applications: non-parametric statistics in Euclidean spaces
   - Difficulties and open problems
3. General Data Depth
   - Other depth measures
   - Local depths

Univariate Statistical Model

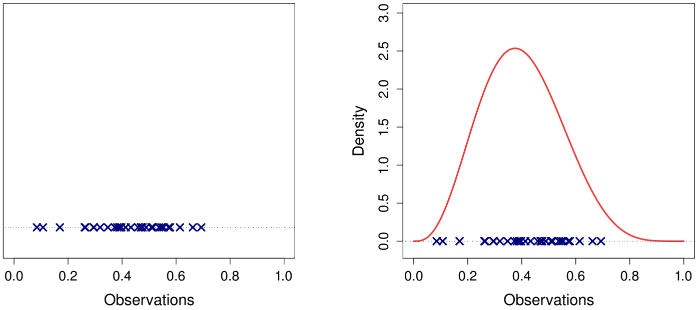
A random sample $X_1, \dots, X_n$ of **univariate** observations (**×**)





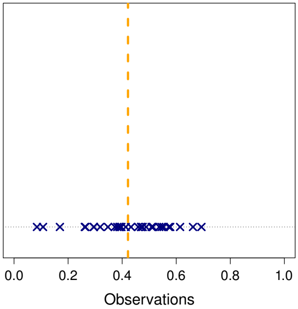
Univariate Statistical Model

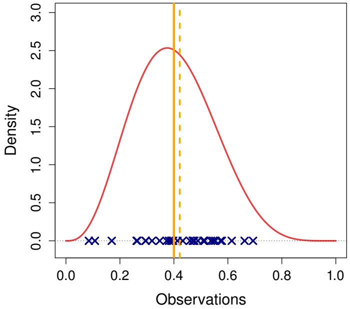
$$X_1, \ldots, X_n \sim P \in \mathcal{P}(\mathbb{R}) \text{ with a } \textbf{density}$$



Location Estimation: Mean

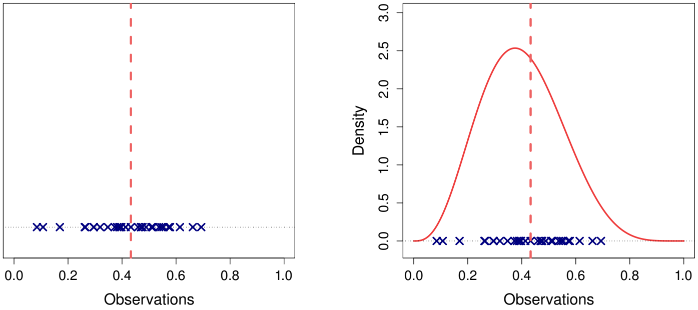
Mean $\mathrm{E}\, X_1 = \int_{\mathbb{R}} x \,\mathrm{d}P(x)$ estimated by $1/n \sum_{i=1}^{n} X_i$





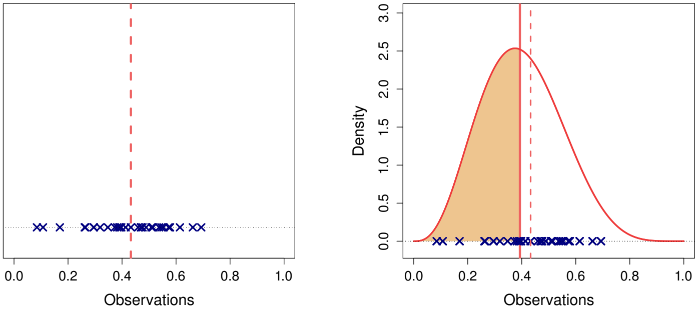
Location Estimation: Median

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



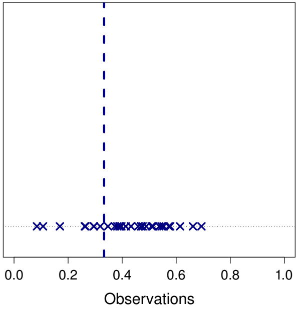
Quantiles for Univariate Data

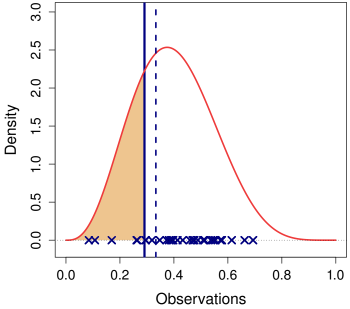
$$q(0.5) = \sup\{x \in \mathbb{R} \colon P((-\infty, x]) \leq 0.5\}$$



Quantiles for Univariate Data

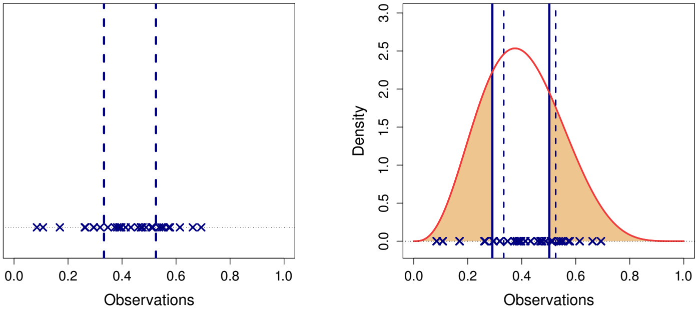
$$q(0.25) = \sup\{x \in \mathbb{R} \colon P((-\infty, x]) \leq 0.25\}$$





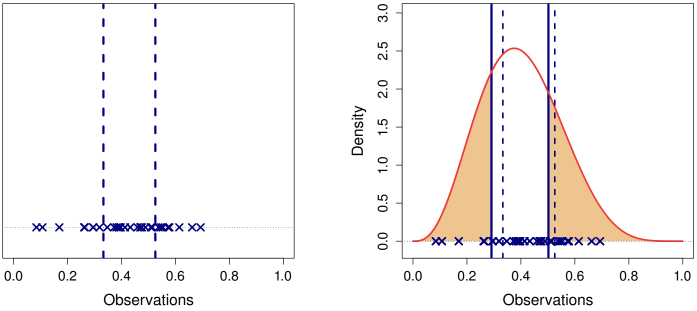
Quantiles for Univariate Data

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



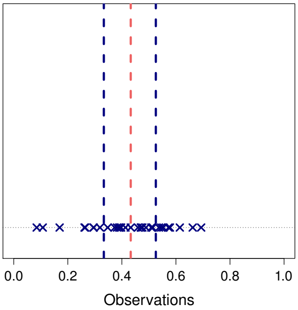
Quantiles for Univariate Data

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



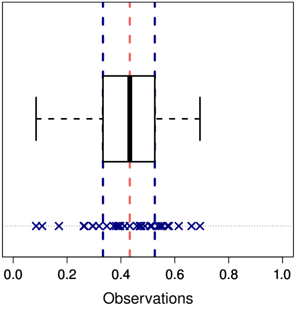
Boxplot

Quantile-based **visualisation tool** (Tukey, 1969)



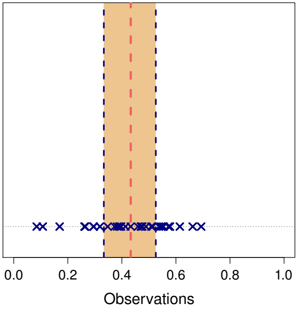
Boxplot

Quantile-based **visualisation tool** (Tukey, 1969)



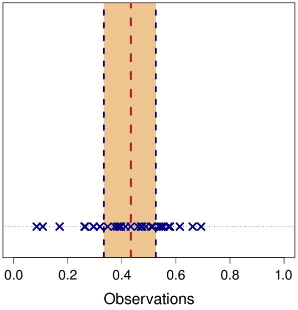
L-estimators

Central part of the data



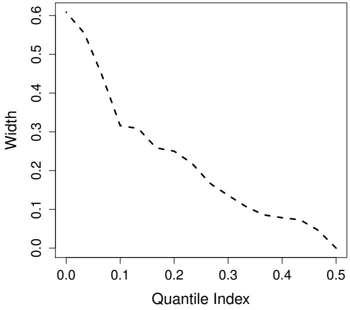
L-estimators

L-statistics: Functions of **order statistics** (trimmed mean)



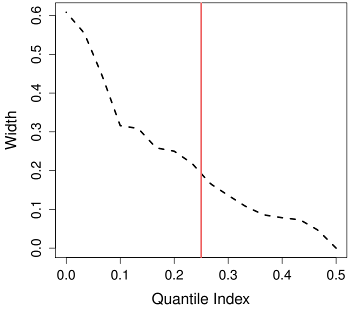
Scale Curve

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



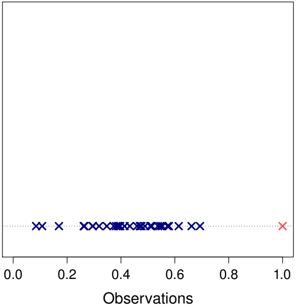
Scale Curve

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



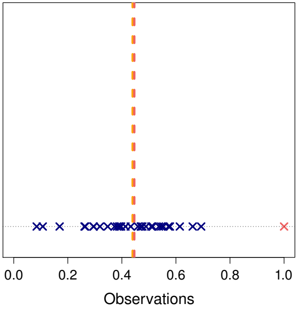
Outlier

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



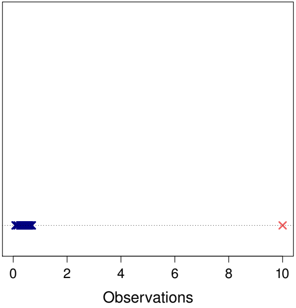
Outlier

Mean and **median** of the contaminated data



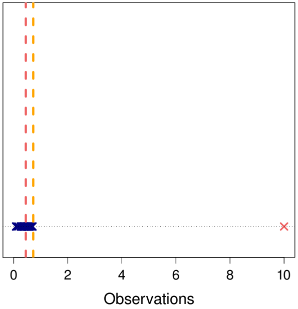
Severe Outlier

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



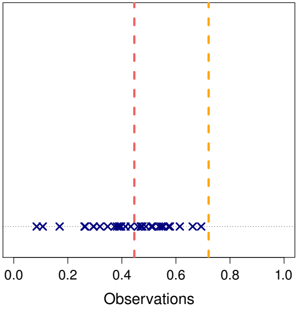
Severe Outlier

Mean and median of the contaminated data



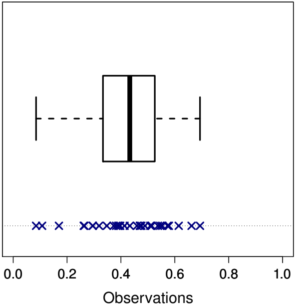
Severe Outlier

Mean and median of the contaminated data



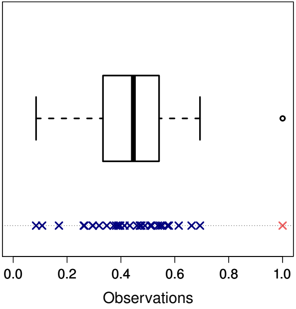
Boxplots

Boxplot of the **original data**



Boxplots

Boxplot of the **contaminated data**



Rank Tests: Two Sample Problem

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

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

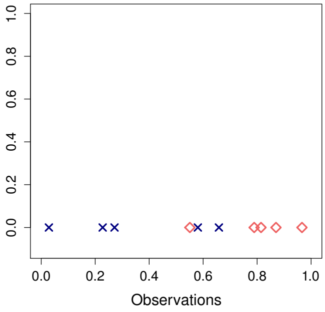
Wilcoxon's rank sum test (Wilcoxon, 1945):

- Pool the two samples into $Z_1, \ldots, Z_{n+m}$ and **rank** these observations (1 through $n+m$).
- 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.

Wilcoxon's Rank Sum Test: Illustration

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

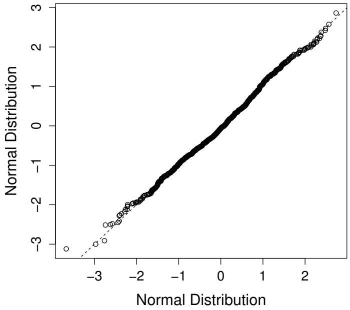
$R = 17$ (range from 15 to 40), p-value 0.03



Q-Q Plot

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

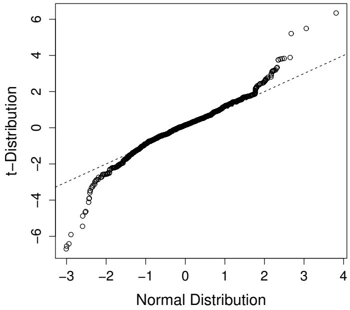
$$t \mapsto (q_X(t), q_Y(t))$$



Q-Q Plot

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

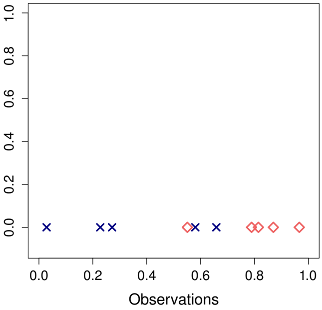
$$t \mapsto (q_X(t), q_Y(t))$$

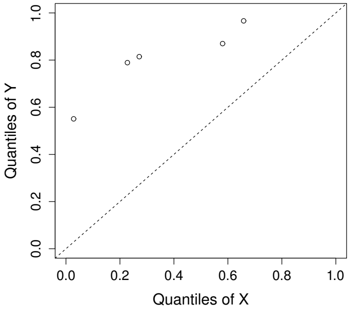


Wilcoxon's Rank Sum Test: Illustration

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

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

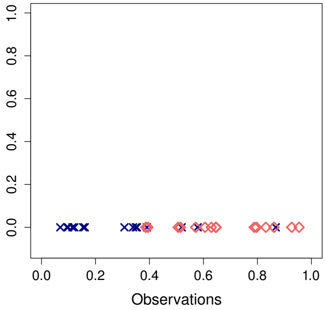


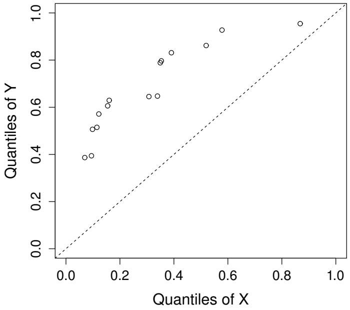


Wilcoxon's Rank Sum Test: Illustration

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

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

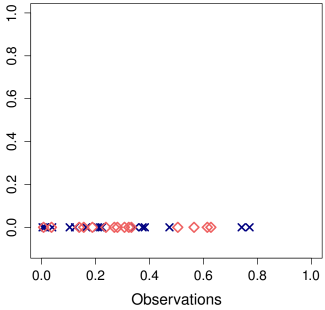


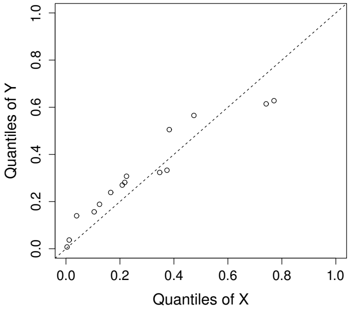


Wilcoxon’s Rank Sum Test: Illustration

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

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

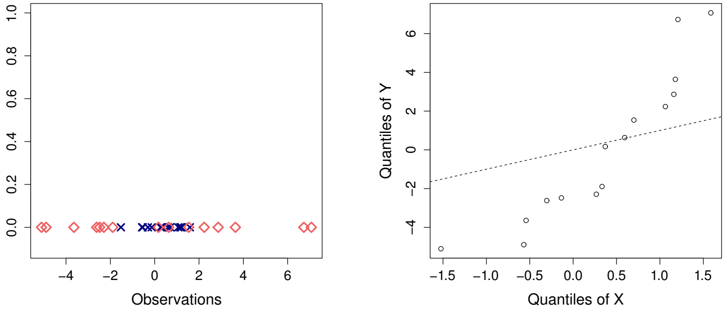




Wilcoxon's Rank Sum Test: Illustration

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

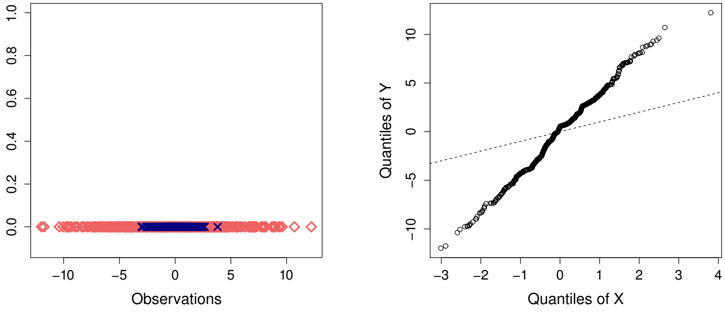
$R = 242$ (range from 120 to 345), p-value 0.71



Wilcoxon's Rank Sum Test: Illustration

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

p-value 0.30



Summary: Ranks and Orders

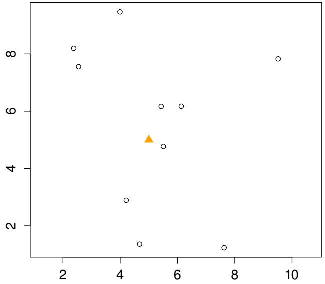
In $\mathbb{R}$, **rank** and **order statistics** enable:

- 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.

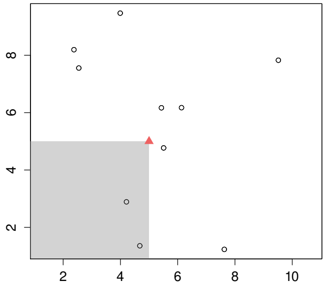
Multivariate Data

How to order **multivariate data**?



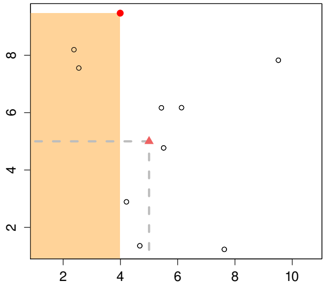
Multivariate Data

How to order **multivariate data**?



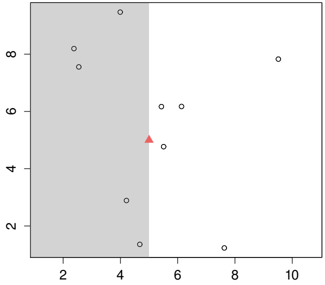
Multivariate Data

How to order **multivariate data**?



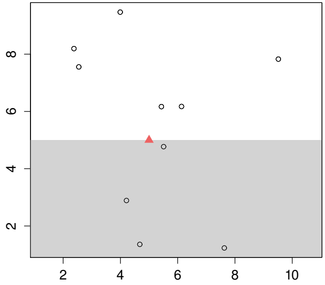
Multivariate Data

How to order **multivariate data**?



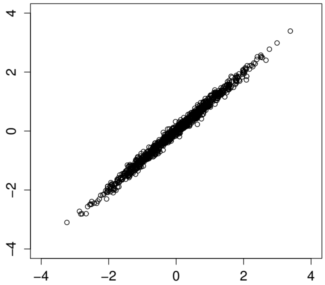
Multivariate Data

How to order **multivariate data**?



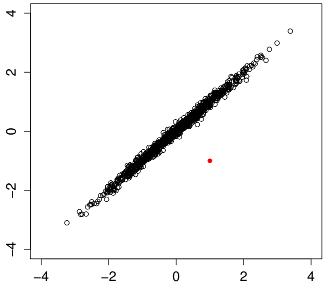
Multivariate Data

How to order **multivariate data**?



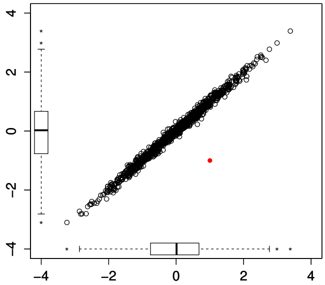
Multivariate Data

How to order **multivariate data**?



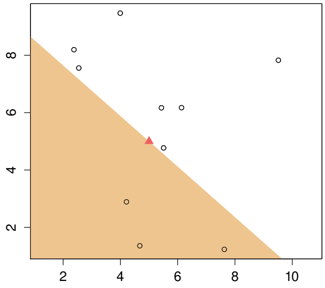
Multivariate Data

How to order **multivariate data**?



Multivariate Data

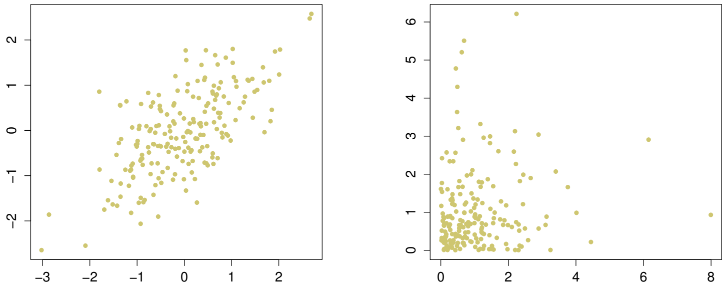
How to order **multivariate data**?



Data Depth

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

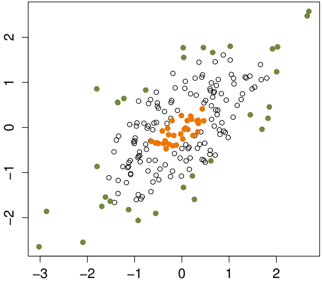
$$D\colon \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) \to [0,1]\colon (x,P) \mapsto D(x,P).$$

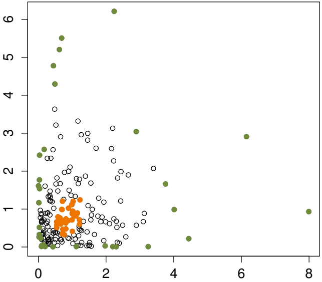


Data Depth

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

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

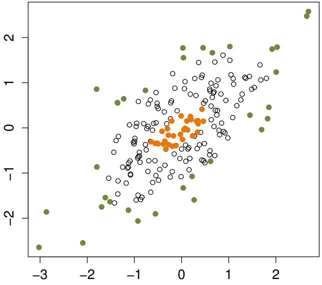


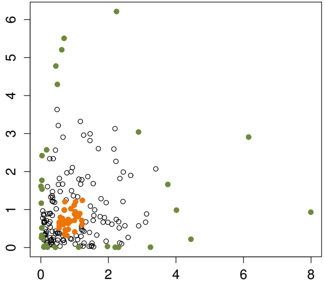


Halfspace Depth

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

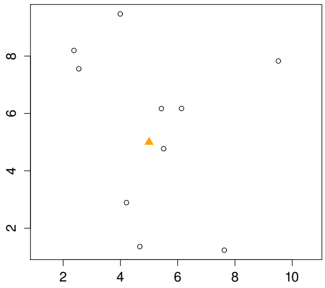
$$hD(x;P) = \inf_{H \in \mathcal{H}(x)} P(H).$$





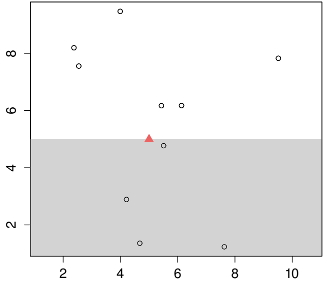
Halfspace Depth

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



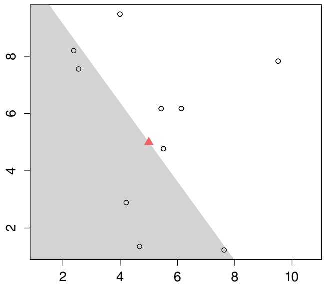
Halfspace Depth

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



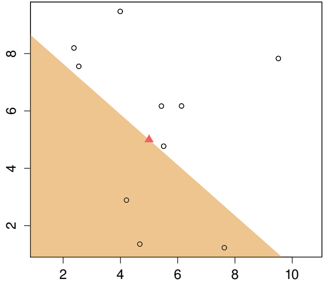
Halfspace Depth

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



Halfspace Depth

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



Brief History of *hD* (in Statistics)

1955 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;

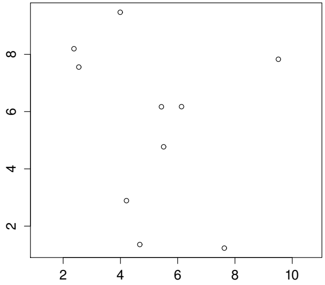
1992 depth introduced in AoS (Donoho and Gasko, 1992);

1999 Rousseeuw and Ruts study *hD* in full generality;

2000 Zuo and Serfling provide a general framework for the depth.

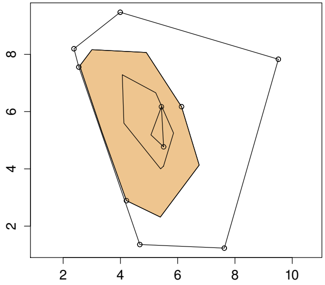
Depth Region

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



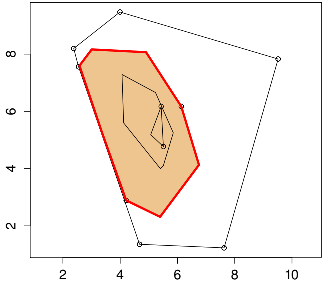
Depth Region

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



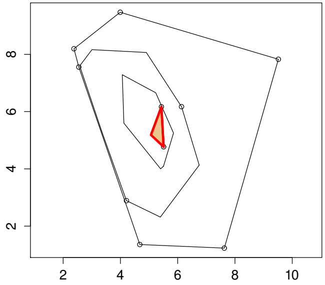
Depth Contour

Topological boundary of $hD_\alpha(P)$



Halfspace Median

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



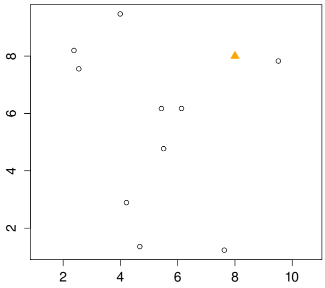
Elementary Properties

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**;
- $hD(x,P) \leq \alpha$ iff $\forall \beta > \alpha\ \exists H \in \mathcal{H}(x)\colon P(H) \leq \beta$;
- $hD(x;P) = \inf_{u \in \mathbb{S}^{d-1}} hD(\langle x,u\rangle ; P_{\langle X,u\rangle})$.

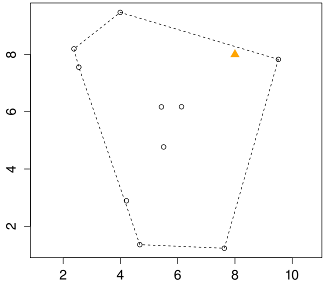
Minimizing Halfspace

The minimizing halfspace may **not be unique**



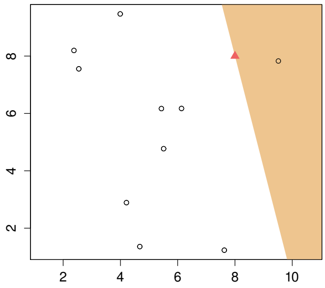
Minimizing Halfspace

The minimizing halfspace may **not be unique**



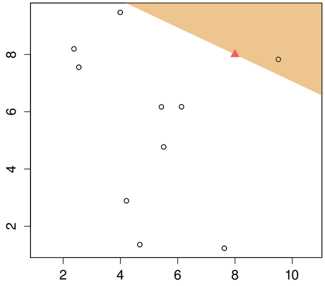
Minimizing Halfspace

The minimizing halfspace may **not be unique**



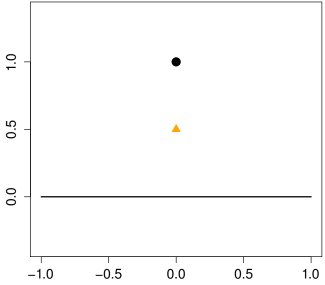
Minimizing Halfspace

The minimizing halfspace may **not be unique**



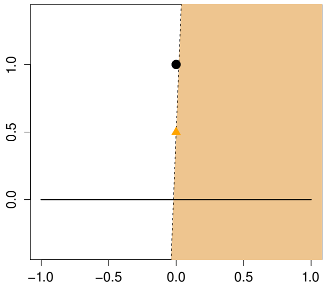
Minimizing Halfspace

The minimizing halfspace may **not exist**



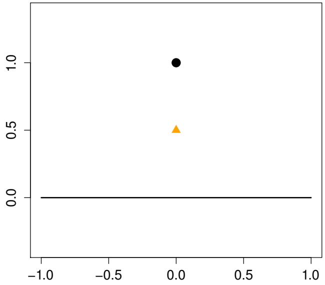
Minimizing Halfspace

The minimizing halfspace may **not exist**



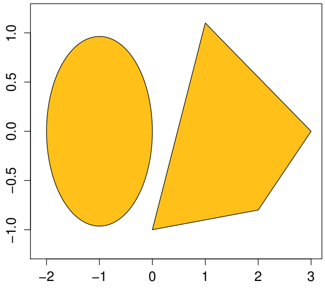
Assumption 1: Smoothness (S)

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



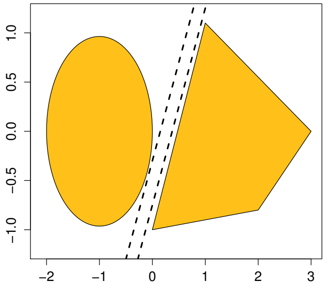
Assumption 2: Contiguous Support (C)

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



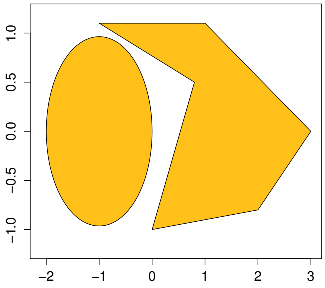
Assumption 2: Contiguous Support (C)

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



Assumption 2: Contiguous Support (C)

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



Further Properties

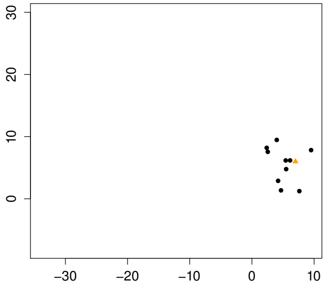
For P that **satisfies (S)**

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

Affine Invariance

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

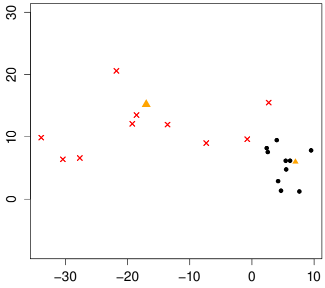
$$hD(x; P_X) = hD(Ax + b; P_{AX+b}).$$



Affine Invariance

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

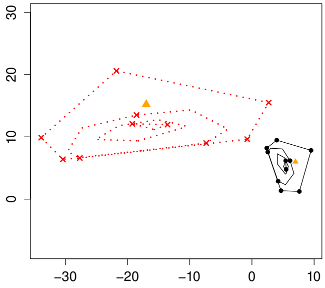
$$hD(x; P_X) = hD(Ax + b; P_{AX+b}).$$



Affine Invariance

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

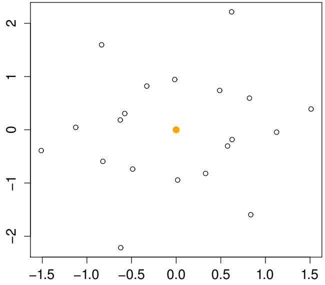
$$hD(x; P_X) = hD(Ax + b; P_{AX+b}).$$



Maximality

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

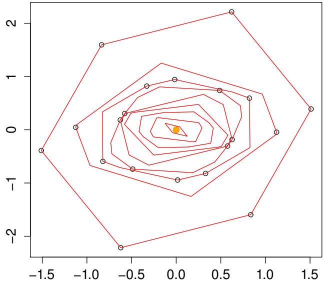
$$hD(0;P) = \sup_{x \in \mathbb{R}^d} hD(x;P).$$



Maximality

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

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



(Semi-)Continuity

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,

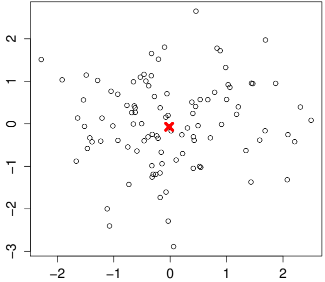
$$\limsup_{\nu\to\infty} hD(x_\nu;P) \leq hD(x;P).$$

If P *satisfies (S) then also*

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

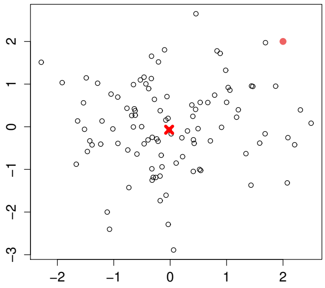
Robustness

Halfspace median is a **robust estimator**



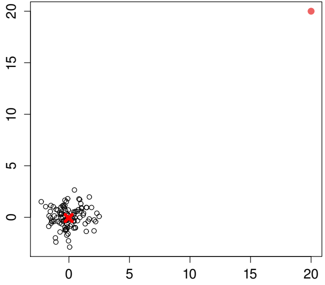
Robustness

Halfspace median is a **robust estimator**



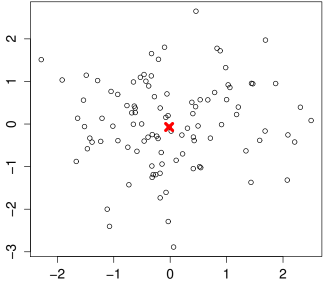
Robustness

Halfspace median is a **robust estimator**



Robustness

Halfspace median is a **robust estimator**



Sample Version Consistency

Theorem (Donoho and Gasko, 1992)

For any $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ almost surely

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

Vanishing at Infinity

Theorem (Donoho and Gasko, 1992)

For any $P \in \mathcal{P}\left(\mathbb{R}^d\right)$

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

Properties of Depth Regions

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

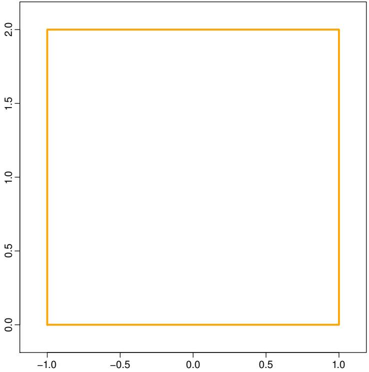
- $hD_{\alpha}(P) = \bigcap\{H \in \mathcal{H} : P(H) > 1 - \alpha\}$;
- $hD_{\alpha}(P)$ is **closed**;
- $hD_{\alpha}(P)$ is **bounded**;
- $hD_{\alpha}(P)$ is **convex**.

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

Quasi-Concavity

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

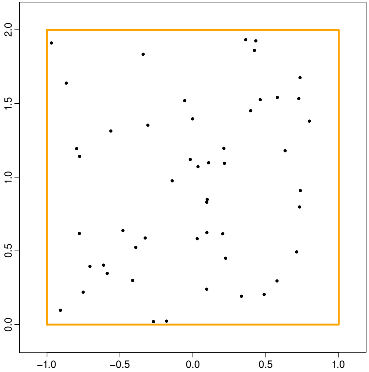
$$\left\{x \in \mathbb{R}^d : hD(x;P) \geq \alpha\right\} \text{ is a convex set}$$



Quasi-Concavity

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

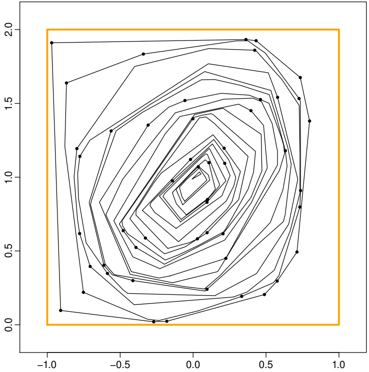
$$\left\{x \in \mathbb{R}^d : hD(x;P) \geq \alpha\right\} \text{ is a convex set}$$



Quasi-Concavity

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

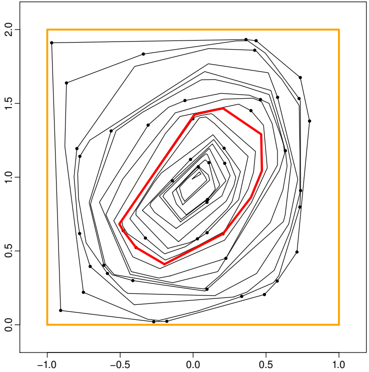
$$\{x \in \mathbb{R}^d : hD(x;P) \geq \alpha\} \text{ is a convex set}$$



Quasi-Concavity

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

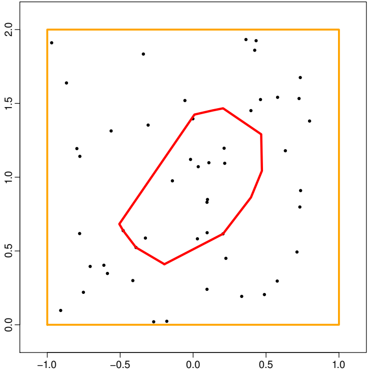
$$\left\{x \in \mathbb{R}^d : hD(x;P) \geq \alpha\right\} \text{ is a convex set}$$



Consistency of Depth Regions

Consider the mapping

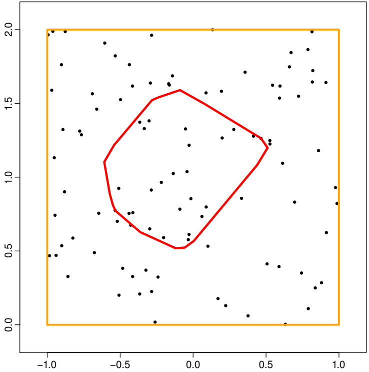
$$\alpha \mapsto \{x \in \mathbb{R}^d : hD(x;P) \geq \alpha\}$$



Consistency of Depth Regions

Consider the mapping

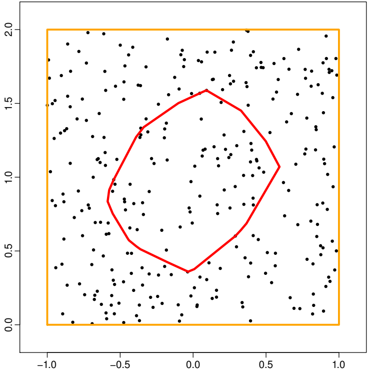
$$\alpha \mapsto \{x \in \mathbb{R}^d : hD(x;P) \geq \alpha\}$$



Consistency of Depth Regions

Consider the mapping

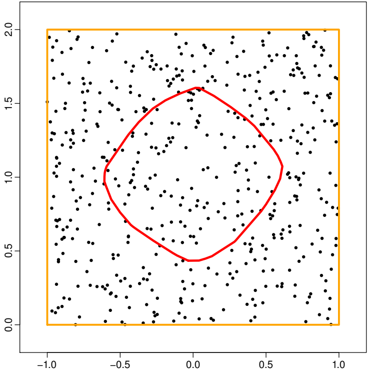
$$\alpha \mapsto \{x \in \mathbb{R}^d : hD(x;P) \geq \alpha\}$$



Consistency of Depth Regions

Consider the mapping

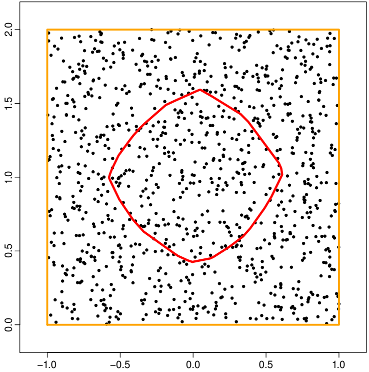
$$\alpha \mapsto \{x \in \mathbb{R}^d : hD(x; P) \geq \alpha\}$$



Consistency of Depth Regions

Consider the mapping

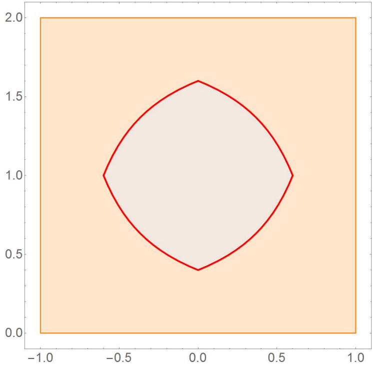
$$\alpha \mapsto \left\{ x \in \mathbb{R}^d : hD(x;P) \geq \alpha \right\}$$



Consistency of Depth Regions

Consider the mapping

$$\alpha \mapsto \{x \in \mathbb{R}^d : hD(x;P) \geq \alpha\}$$



Properties of Depth Regions

Convex sets are equipped with the Hausdorff distance d_H.

Theorem (Dyckerhoff, 2017+)

Let (S) and (C) be true for P. Then the mapping

$$\alpha \mapsto hD_\alpha(P)$$

is continuous. Further, for any α

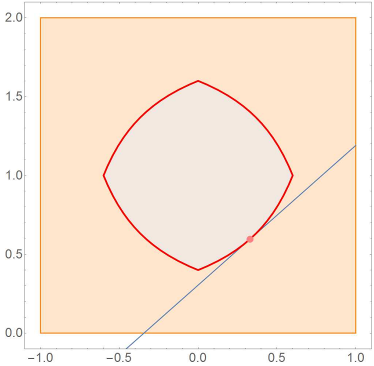
$$d_H\left(hD_\alpha(P_n), hD_\alpha(P)\right) \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

The previous results of Zuo and Serfling (2000b) are **not correct!**

Asymptotic Normality

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

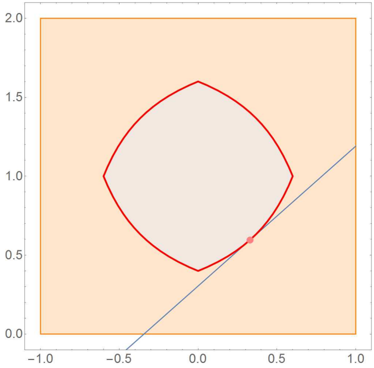
$\Longleftrightarrow$ $hD(x;P)$ is realised by a **single halfspace** $H \in \mathcal{H}$ (Massé, 2004)



Asymptotic Normality

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

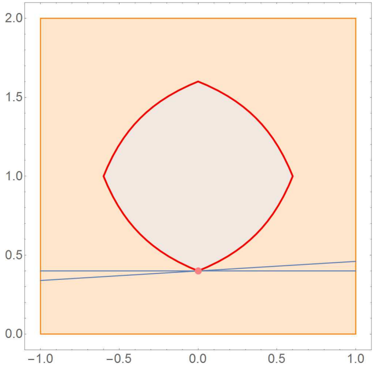
$\iff$ the contour of $hD(\cdot; P)$ is **smooth** at x (Gijbels and Nagy, 2016)



Asymptotic Normality

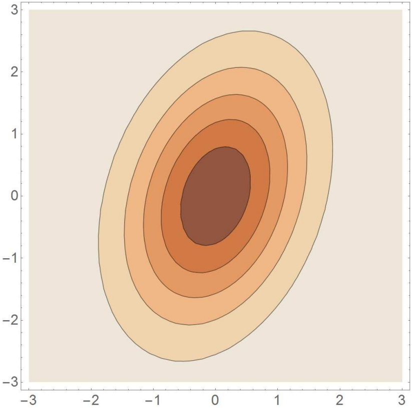
$\sqrt{n}\, hD(x; P_n)$ is **asymptotically normal**

$\iff$ the contour of $hD(\cdot; P)$ is **smooth** at x (Gijbels and Nagy, 2016)



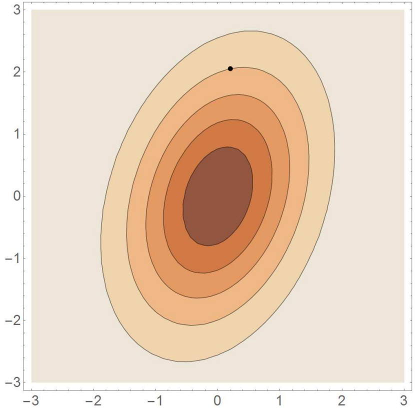
Population Depth: Elliptically Symmetric Distributions

Elliptically symmetric distributions have **elliptic depth contours**



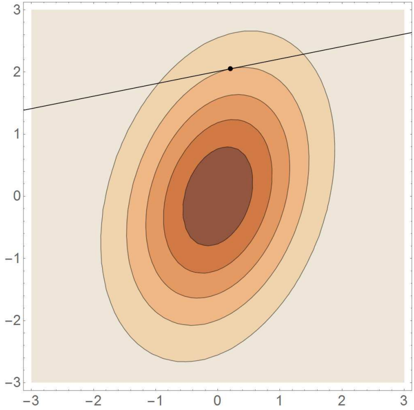
Population Depth: Elliptically Symmetric Distributions

Elliptically symmetric distributions have **elliptic depth contours**



Population Depth: Elliptically Symmetric Distributions

Elliptically symmetric distributions have **elliptic depth contours**



Population Depth: Multivariate Stable Distributions

For $p \in (0,2]$, $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ is a ***p*-stable distribution** if $\left(X_1,\ldots,X_d\right) \sim P$ has independent components and for any $u_1,\ldots,u_d \in \mathbb{R}$ it holds that

$$\sum_{i=1}^{d} u_i X_i \sim \|u\|_p X_1.$$

- for $p = 2$ we obtain the standard **multivariate normal** distribution;
- for $p = 1$ we obtain the standard **multivariate Cauchy** distribution;
- for other $p \in (0,2]$ there is **no explicit form** for the density of P.

Population Depth: Multivariate Stable Distributions

Theorem (Massé and Theodorescu, 1994)

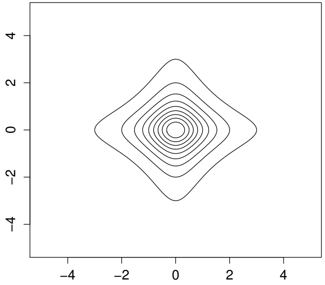
Let P be p-stable. Set

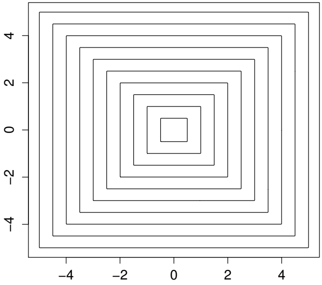
$$q = \begin{cases} p/(p-1) & \text{if } p > 1, \\ \infty & \text{if } p \leq 1. \end{cases}$$

Then the depth regions $hD_\alpha(P)$ are the level sets of the norm $\|\cdot\|_q$.

Population Depth: Multivariate Stable Distributions

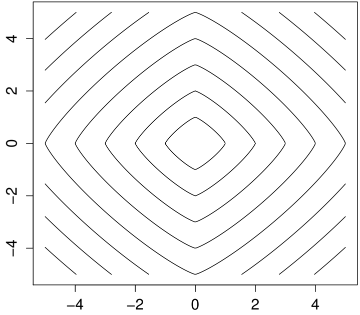
Multivariate Cauchy distribution ($p = 1$)





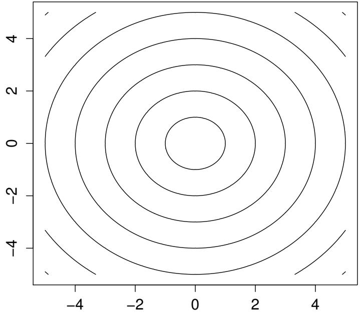
Population Depth: Multivariate Stable Distributions

Multivariate stable distribution ($p = 1.2$)



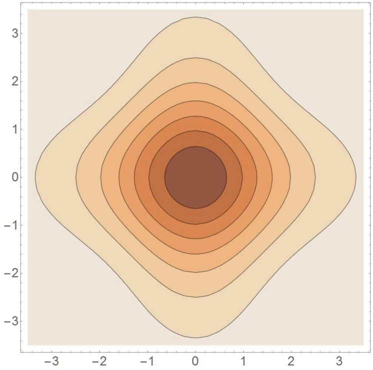
Population Depth: Multivariate Stable Distributions

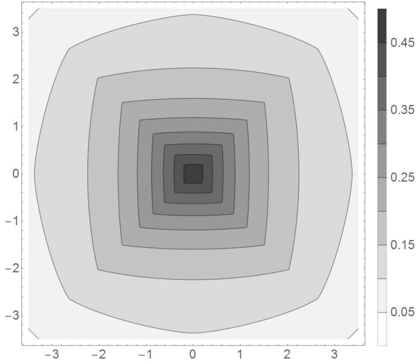
Multivariate normal distribution ($p = 2$)



Population Depth: Mixture of Normals

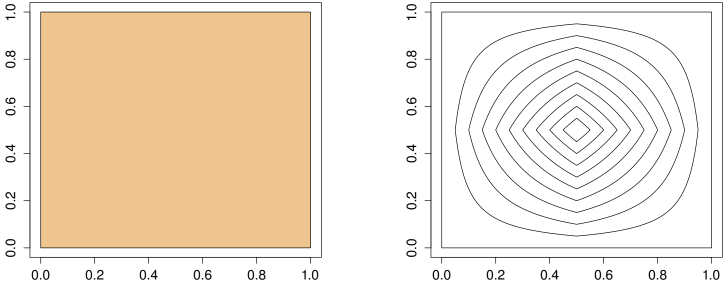
Mixture of two bivariate normal distributions





Population Depth: Uniform Distribution on a Square

Uniform distribution on a simple convex body



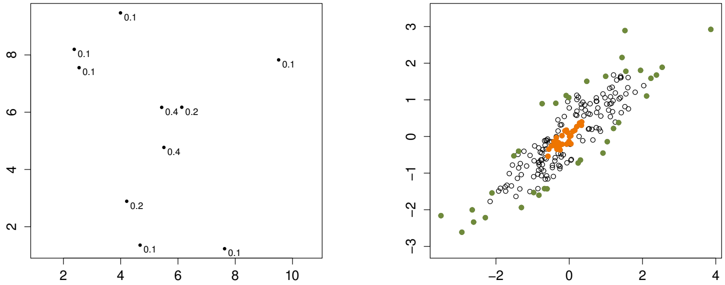
Problem: Smoothness of Depth Contours

Problem (Massé and Theodorescu, 1994)

Is there any non-elliptical distribution with smooth depth contours?

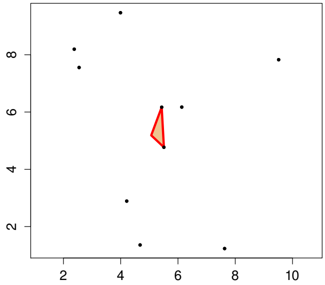
Data Ordering

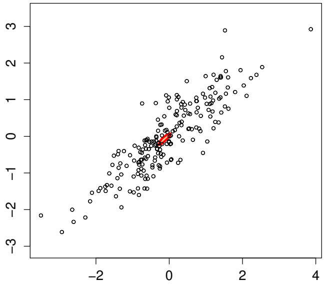
Depth induces a **centre - outward ordering** of points



Halfspace Median

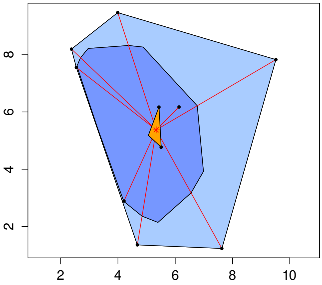
Point(s) that maximize the depth over $\mathbb{R}^d$

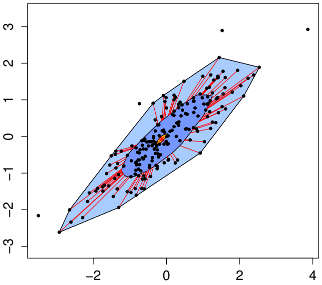




Bagplot: A Multivariate Boxplot

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

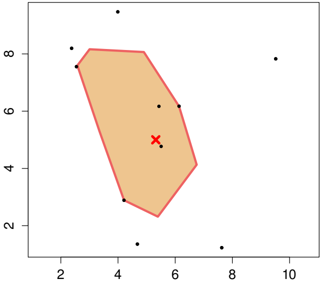


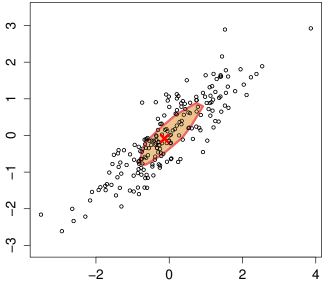


Multivariate L-statistics

Depth-trimmed mean (Fraiman and Meloche, 1999)

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

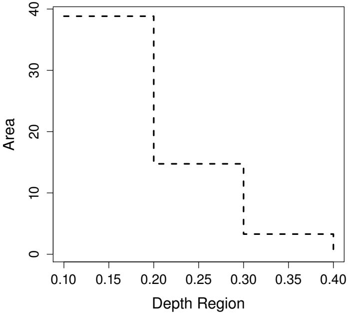


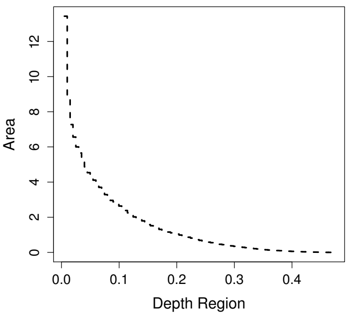


Scale Curve

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

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





Multivariate Rank Tests: Two Sample Problem

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

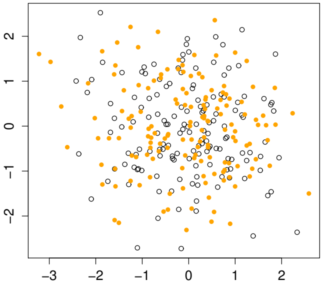
$$H_0 : P = Q \quad \text{against} \quad H_1 : P \neq Q.$$

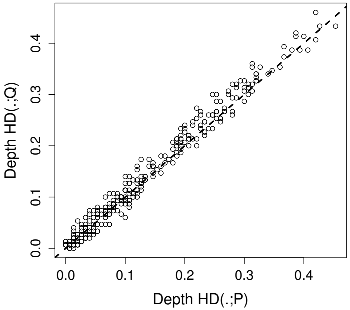
Wilcoxon's rank sum test (Liu and Singh, 1993):

- 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.

D-D Plots: Multivariate Q-Q Plots

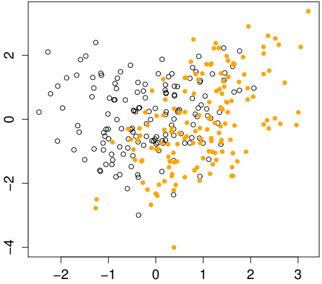
Replace quantiles by **depth in Q-Q plots** (Liu et al., 1999)

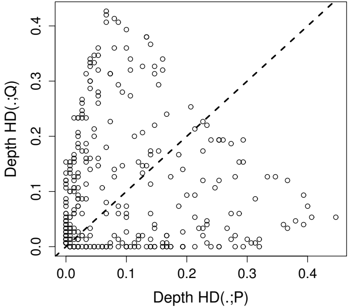




D-D Plots: Multivariate Q-Q Plots

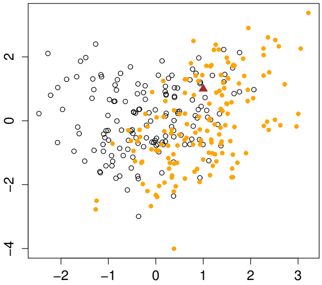
Replace quantiles by **depth in Q-Q plots** (Liu et al., 1999)

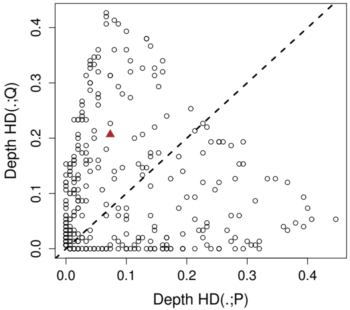




Classification

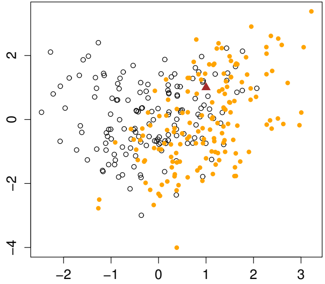
Classify a **new observation** into one of the groups (Li et al., 2012)

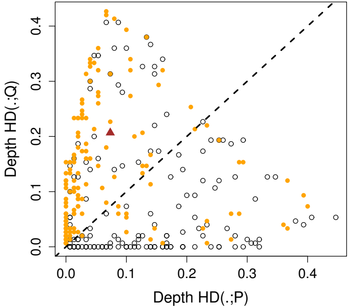




Classification

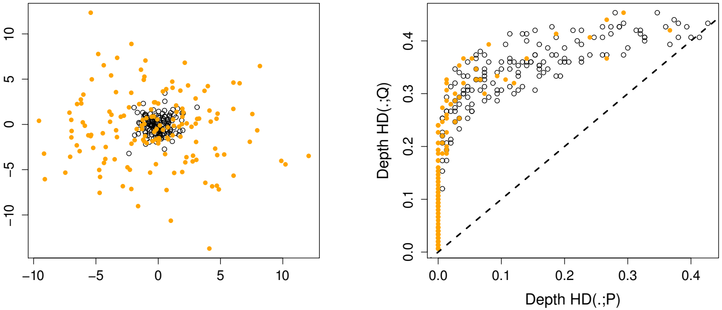
Classify a **new observation** into one of the groups (Li et al., 2012)





D-D Plots: Multivariate Q-Q Plots

D-D plots with **unequal scatters**



Computational Complexity of *hD*

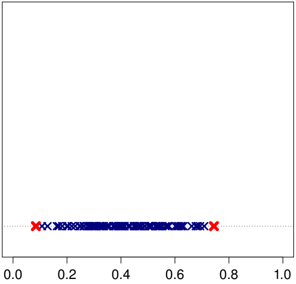
- best known exact algorithms have complexity $\mathcal{O}(\log(n)n^{d-1})$ (Rousseeuw and Struyf, 1998);
- **feasible computation** only for $n \leq 1000$ and $d \leq 5$;
- **approximations** of *hD* (Dyckerhoff, 2004)

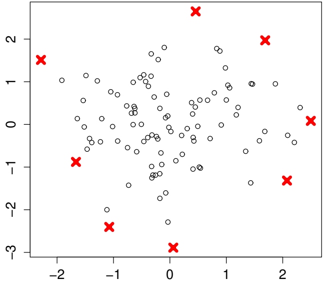
$$hD(x;P) = \inf_{u \in \mathbb{S}^{d-1}} hD(\langle x,u \rangle ; P_{\langle X,u \rangle}) \approx \min_{j=1,\dots,N} hD(\langle x,U_j \rangle ; P_{\langle X,U_j \rangle}).$$

- choice of the parameter N and the distribution of U (Nagy, 2017+).

Ties

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.

Distribution-by-Depth Characterization

Conjecture

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

Distribution-by-Depth Characterization

Conjecture

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:

- if P and Q are absolutely continuous with a compact support (Koshevoy, 2001)
- if P and Q are atomic (Koshevoy, 2002);
- if P and Q have smooth densities (Hassairi and Regaieg, 2008);
- if P and Q have smooth depth contours (Kong and Zuo, 2010).

Distribution-by-Depth Characterization

Conjecture

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:

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

Statistical Data Depth

According to Zuo and Sefling (2000), **statistical data 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 symmetric distributions;
3. **monotonicity** relative to the depth median;
4. **vanishing** at infinity.

Statistical Data Depth

According to Zuo and Sefling (2000), **statistical data 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 symmetric distributions;
3. **monotonicity** relative to the depth median;
4. **vanishing** at infinity.

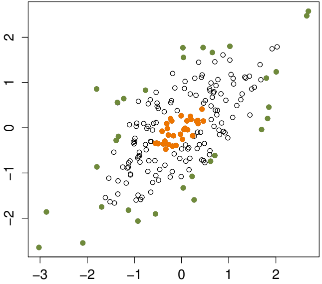
Serfling (2006) requires in addition also

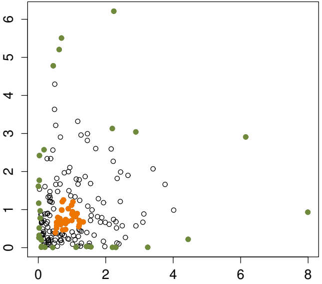
5. upper **semi-continuity** as a function of x;
6. **continuity** as a functional of P;
7. **quasi-concavity** in x.

Simplicial Depth

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

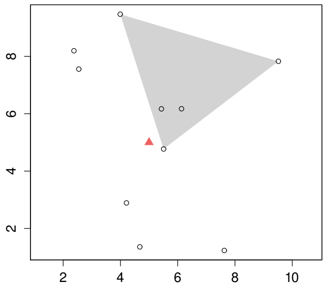
$$sD(x; P) = \mathrm{P}\left(x \in \mathbb{S}(X_1, \ldots, X_{d+1})\right).$$





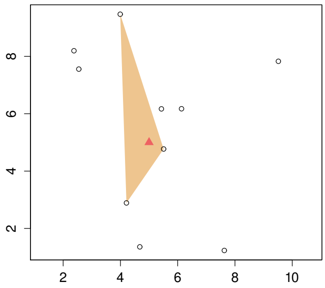
Simplicial Depth

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



Simplicial Depth

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



Simplicial Depth: Properties

Advantages:

- affine **invariant**;
- **U-statistic** (good statistical properties);
- **robust** median;
- **vanishes** at infinity.

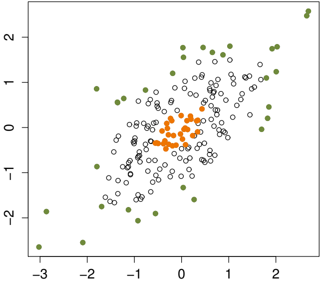
But:

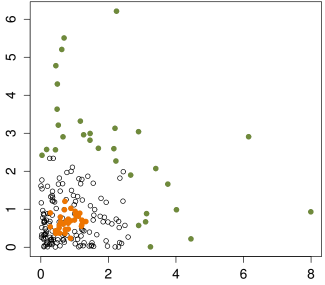
- **not quasi-concave** or monotonically decreasing;
- computationally expensive;
- population version **difficult to study** theoretically.

Simplicial Volume Depth

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

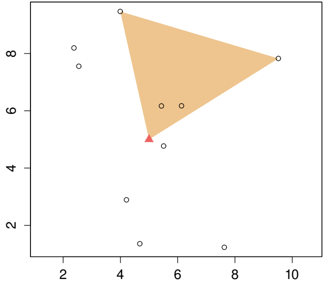
$$svD(x;P) = \left(1 + \mathrm{E}\,\lambda\left(\mathbb{S}\left(x, X_1, \ldots, X_d\right)\right)\right)^{-1}.$$





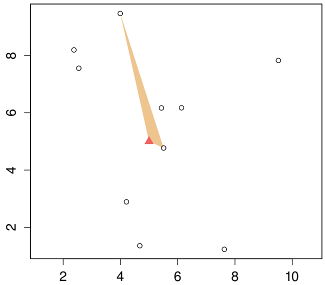
Simplicial Volume Depth (Oja's Depth)

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



Simplicial Volume Depth (Oja's Depth)

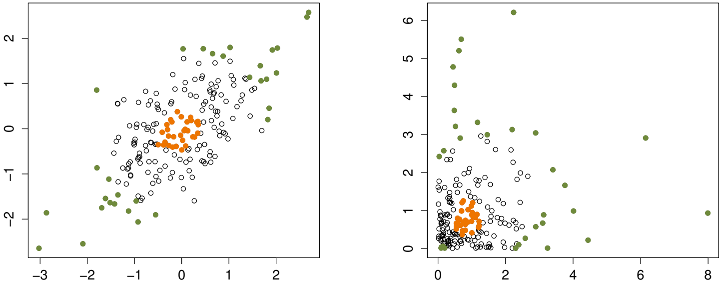
$$svD\,(x;P_n) = \left(1 + \binom{n}{d}^{-1} \sum_i \lambda(\mathbb{S}(x, X_{i_1}, \dots, X_{i_d}))\right)^{-1}$$



Spatial Depth

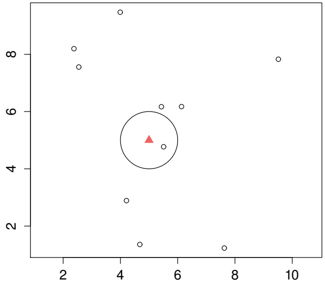
Spatial depth (Chaudhuri, 1996) of an observation in $\mathbb{R}^d$

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



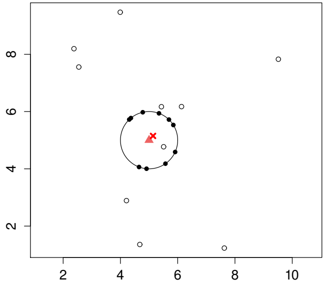
Spatial Depth

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



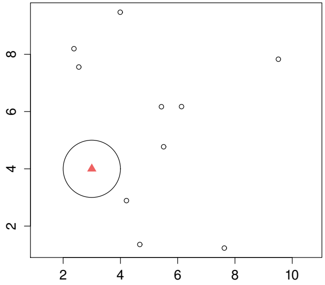
Spatial Depth

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



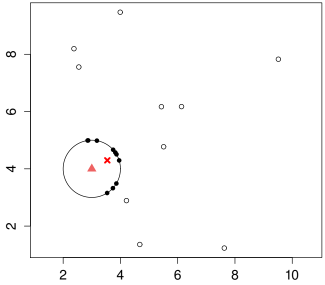
Spatial Depth

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



Spatial Depth

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



Spatial Depth: Properties

Advantages:

- rotation **invariant**;
- maximized at the **spatial median**, i.e. a point x that minimizes

$$\mathrm{E}\left\|X - x\right\|;$$

- **robust** median;
- **vanishes** at infinity;
- very **fast** computation ($\mathcal{O}(n)$);
- works also in **high-dimensional** spaces.

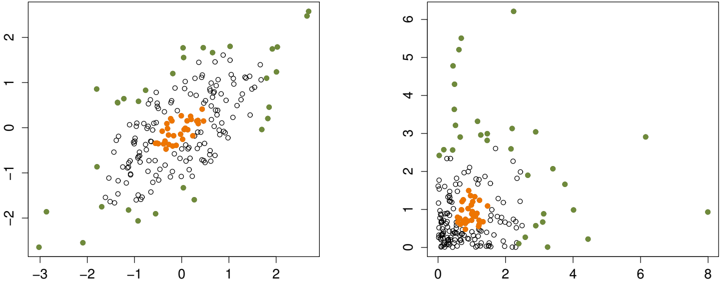
But:

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

Mahalanobis Depth

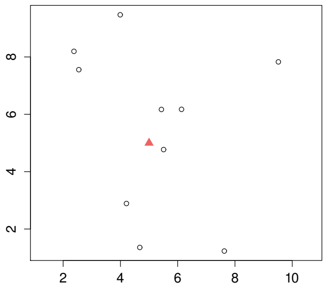
Mahalanobis depth (Mahalanobis, 1936) of an observation in $\mathbb{R}^d$

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



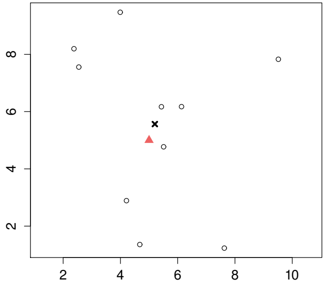
Mahalanobis Depth

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



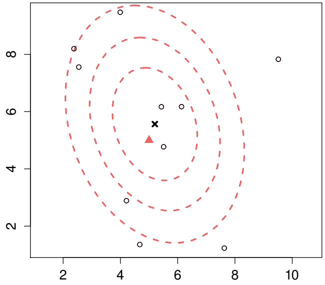
Mahalanobis Depth

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



Mahalanobis Depth

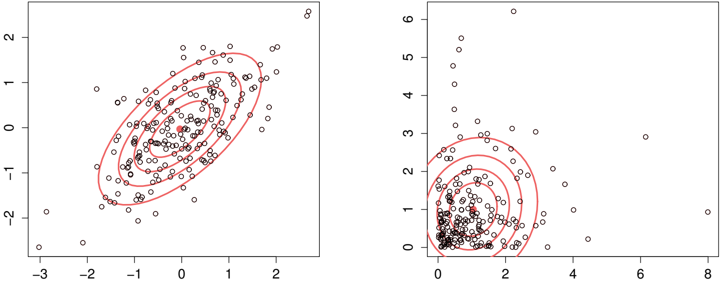
$$mD(x; P) \sim \text{Mahalanobis distance from } \mathrm{E}X$$



Mahalanobis Depth

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

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



Mahalanobis Depth: Properties

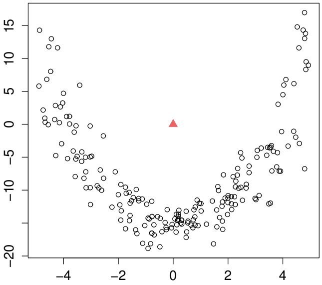
Disadvantages:

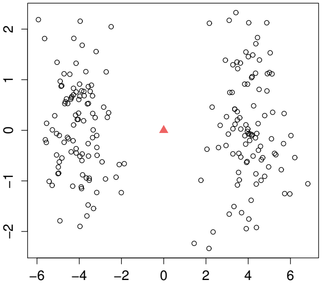
- **not always defined** (not entirely non-parametric);
- maximized at the mean ($\Longrightarrow$ not robust);
- rigid contours (concentric ellipses of the same shape).

Not really a depth.

Unimodality / Quasi-Concavity

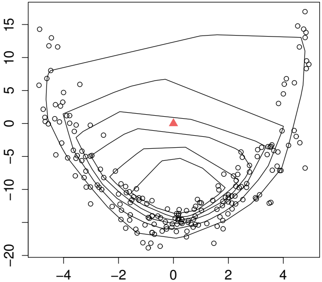
Proper depth is intended to be **unimodal** and **quasi-concave**

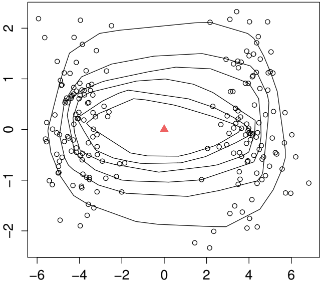




Unimodality / Quasi-Concavity

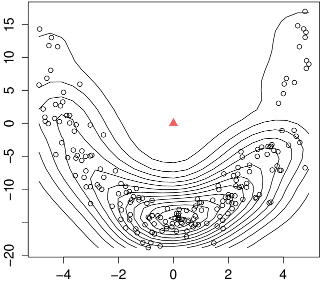
Proper depth is intended to be **unimodal** and **quasi-concave**

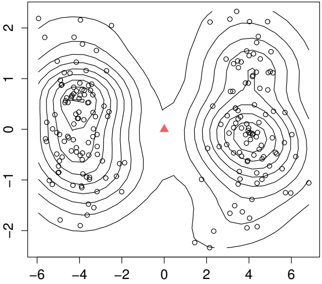




Local Depths

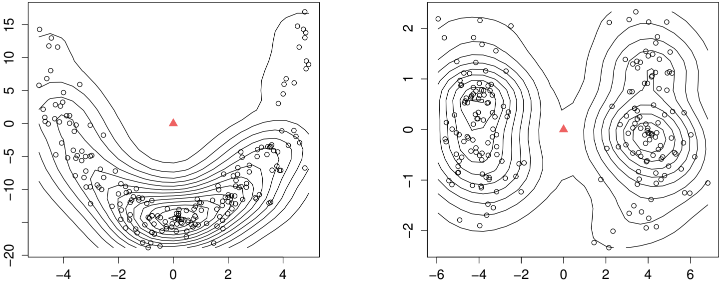
Relaxation of unimodality leads to **local depths**





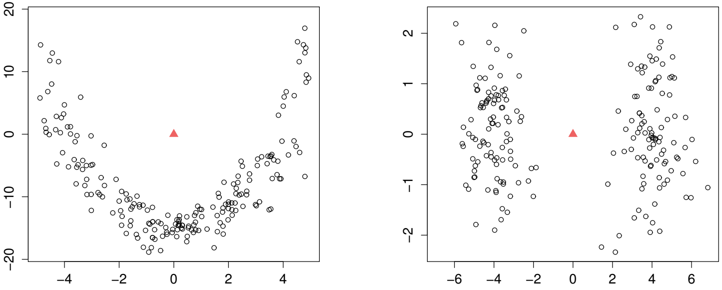
Likelihood Depth

Multivariate **density estimator** (Fraiman and Meloche, 1999)



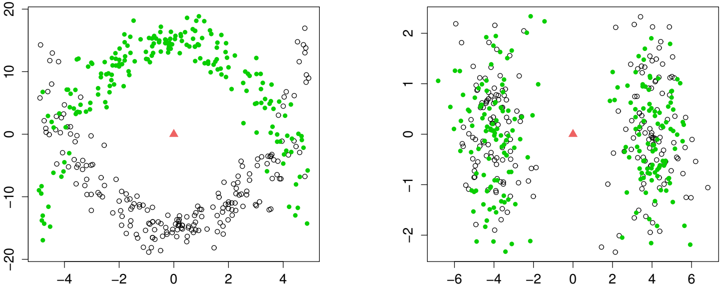
Local Halfspace Depth

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



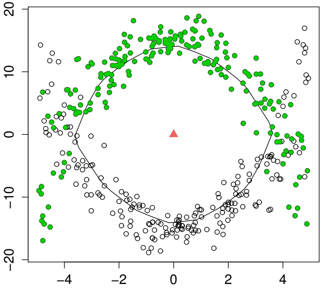
Local Halfspace Depth

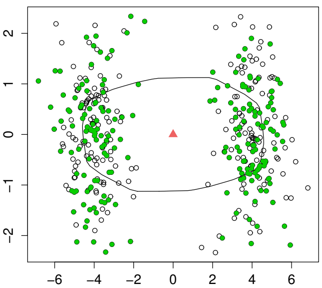
Localization of *hD* (Paindaveine and Van Bever, 2013)



Local Halfspace Depth

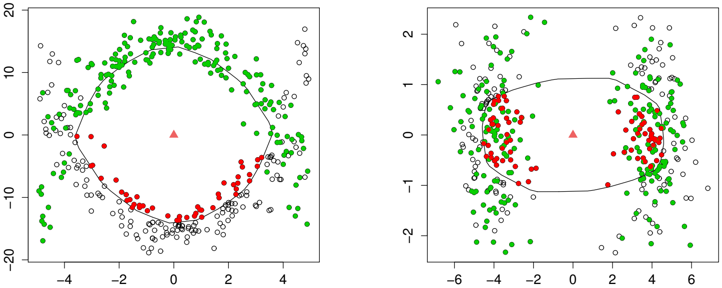
Localization of *hD* (Paindaveine and Van Bever, 2013)





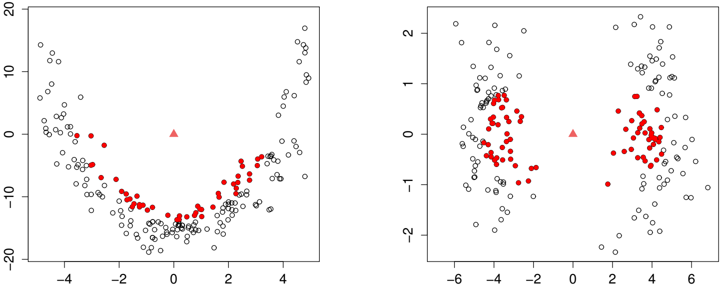
Local Halfspace Depth

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



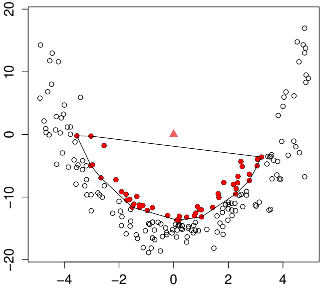
Local Halfspace Depth

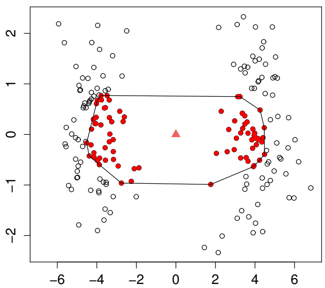
Localization of *hD* (Paindaveine and Van Bever, 2013)



Local Halfspace Depth

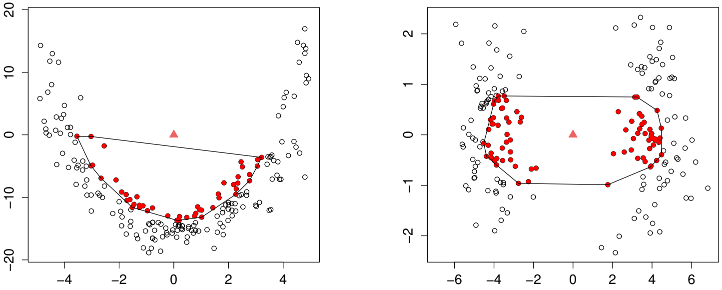
Localization of *hD* (Paindaveine and Van Bever, 2013)





Local Halfspace Depth

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 **graphs and trees** (Small, 1997);
- for **infinite-dimensional** (functional) data (Fraiman and Muniz, 2001);
- for general **metric spaces** (Carrizosa, 1996);
- in **regression** problems (Rousseeuw and Hubert, 1999);
- ...

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

Conclusions

Data depth is

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

In Parts II and III:

Connections of depth to **mathematics outside statistics**.

Selected Literature

- 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.
- Regina Y. Liu. On a notion of simplicial depth. Proc. Natl. Acad. Sci. U.S.A., 85(6):1732–1734, 1988.
- Regina Y. Liu, Jesse M. Parelius, and Kesar Singh. Multivariate analysis by data depth: descriptive statistics, graphics and inference. Ann. Statist., 27(3):783–858, 1999.
- Peter J. Rousseeuw and Ida Ruts. The depth function of a population distribution. Metrika, 49(3):213–244, 1999.
- 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.
- Yijun Zuo and Robert Serfling. General notions of statistical depth function. Ann. Statist., 28(2):461–482, 2000.
- Yijun Zuo and Robert Serfling. Structural properties and convergence results for contours of sample statistical depth functions. Ann. Statist., 28(2):483–499, 2000.