Statistical Depth Function

Zoltán Szabó

Machine Learning Journal Club, Gatsby Unit

June 27, 2016

Outline

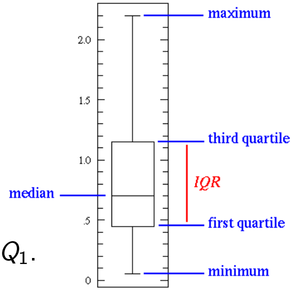
- L-statistics, order statistics, ranking.
- Instead of *moments*: median, dispersion, scale, skewness, . . .
- New visualization tools: depth contours, sunburst plot.
- Testing: symmetry.
- Depth function properties.

L-statistics, ranking

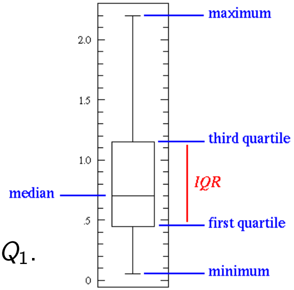
L-statistics

- Given: $x_1, \ldots, x_n \in \mathbb{R}$ samples.
- Order statistics: $x_{[1]} \leq \ldots \leq x_{[n]}$.
- Linear combination of order statistics.
- Rank: position of x_i.

- Dispersion:
  1. range: $x_{[n]} - x_{[1]}$.
  2. interquartile range: $x_{[3n/4]} - x_{[n/4]} = Q_3 - Q_1$.



L-statistics: examples



- Dispersion:
  1. range: $x_{[n]} - x_{[1]}$.
  2. interquartile range: $x_{[3n/4]} - x_{[n/4]} = Q_3 - Q_1$.
- Location:
  - median: $x_{[n/2]}$.
  - α-trimmed mean = middle of the $(1-2\alpha)$-th fraction of observations,

$$\frac{1}{(1-2\alpha)n} \sum_{i=\alpha n+1}^{(1-\alpha)n} x_{[i]}.$$

L-statistics

- Typical application: robust estimation.
- Question: Similar notions in $\mathbb{R}^d$?

L-statistics

- Typical application: robust estimation.
- Question: Similar notions in $\mathbb{R}^d$?
- Critical: $x_{[1]} \leq \ldots \leq x_{[n]}$.

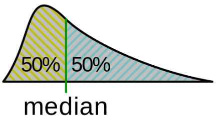
L-statistics

- Typical application: robust estimation.
- Question: Similar notions in $\mathbb{R}^d$?
- Critical: $x_{[1]} \leq \ldots \leq x_{[n]}$.
- Idea: center-outward ordering.

Median

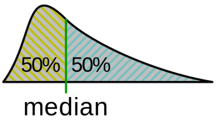
Motivation: median

- F: continuous c.d.f. on $\mathbb{R}$.
- median:
  - point splitting the probability mass to $\frac{1}{2} - \frac{1}{2}$.
  - solution of $F(m) = \frac{1}{2}$.



Motivation: median

- F: continuous c.d.f. on $\mathbb{R}$.
- median:
  - point splitting the probability mass to $\frac{1}{2} - \frac{1}{2}$.
  - solution of $F(m) = \frac{1}{2}$.

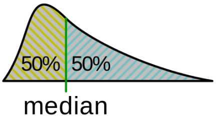


- alternative view:

$$m := \arg\max_{x \in \mathbb{R}} D(x) := 2F(x)[1 - F(x)]$$

Motivation: median

- F: continuous c.d.f. on $\mathbb{R}$.
- median:
  - point splitting the probability mass to $\frac{1}{2} - \frac{1}{2}$.
  - solution of $F(m) = \frac{1}{2}$.

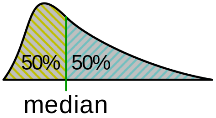


- alternative view:

$$m := \arg\max_{x \in \mathbb{R}} D(x) := 2F(x)[1 - F(x)]$$
$$= \mathbb{P}_{X_1, X_2 \sim F}(x \in [X_1, X_2]).$$

Motivation: median

- F: continuous c.d.f. on $\mathbb{R}$.
- median:
  - point splitting the probability mass to $\frac{1}{2} - \frac{1}{2}$.
  - solution of $F(m) = \frac{1}{2}$.



- alternative view:

$$m := \arg\max_{x \in \mathbb{R}} D(x) := 2F(x)[1 - F(x)]$$
$$= \mathbb{P}_{X_1, X_2 \sim F}(x \in [X_1, X_2]).$$

Moving away from m, $D(x)$ decreases monotonically to 0.

Let us extend the median to $\mathbb{R}^d$!

- Simplitical depth [Liu, 1990]:

$$SD(\mathbf{x}; F) := \mathbb{P}_F\left(\mathbf{x} \in \Delta(X_1, \ldots, X_{d+1})\right).$$

Let us extend the median to $\mathbb{R}^d$!

- Simplitical depth [Liu, 1990]:

$$SD(\mathbf{x}; F) := \mathbb{P}_F\left(\mathbf{x} \in \Delta(X_1, \ldots, X_{d+1})\right).$$

- Half-space depth [Tukey, 1975]:

$$HD(\mathbf{x}; F) := \inf_{\mathbf{x} \in H \subset \mathbb{R}^d} \mathbb{P}_F(X \in H)$$

minimum probability of any closed halfspace containing $\mathbf{x}$.

Let us extend the median to $\mathbb{R}^d$!

- Simplitical depth [Liu, 1990]:

$$SD(\mathbf{x}; F) := \mathbb{P}_F \left(\mathbf{x} \in \Delta(X_1, \ldots, X_{d+1})\right).$$

- Half-space depth [Tukey, 1975]:

$$HD(\mathbf{x}; F) := \inf_{\mathbf{x} \in H \subset \mathbb{R}^d} \mathbb{P}_F(X \in H)$$

minimum probability of any closed halfspace containing $\mathbf{x}$.

median: $m = \arg\max_{\mathbf{x} \in \mathbb{R}^d} D(\mathbf{x}; F)$.

Side-note: these quantities are also computable

- depth R-package: simplicial, half-space, Oja's depth.
- Ref. (half-space depth): [Dyckerhoff and Mozharovskyi, 2016].

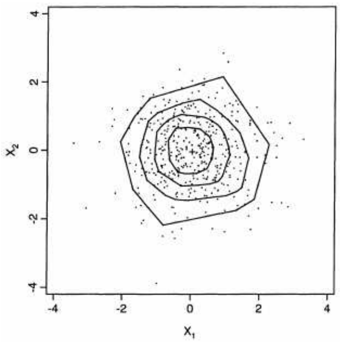
Side-note: these quantities are also computable

- depth R-package: simplicial, half-space, Oja's depth.
- Ref. (half-space depth): [Dyckerhoff and Mozharovskyi, 2016].
- Simplicial depth:

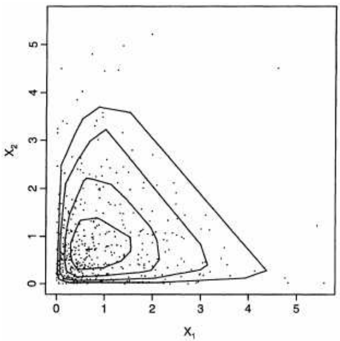
$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i, \qquad \sum_{i=1}^{n} \alpha_i = 1$$

linear equation. If $\alpha_i \geq 0$ $(\forall i) \Rightarrow$ Yes.

normal



exponential



- center-outward ordering.
- high-depth: 'center', low-depth: 'outlyingness'.
- expansion of contours: scale (dispersion), skewness, kurtosis.

Location

- Median/center: deepest point.

Location

- Median/center: deepest point.
- Other robust location statistics:
  - idea: assign smaller weights to more outlier data.
  - given: $w : [0,1] \to \mathbb{R}^{\geq 0}$ nonincreasing weight function

$$L_w = \frac{\sum_{i=1}^n \mathbf{x}_{[i]} w(i/n)}{\sum_{j=1}^n w(j/n)}.$$

Location

- Median/center: deepest point.
- Other robust location statistics:
  - idea: assign smaller weights to more outlier data.
  - given: $w : [0,1] \to \mathbb{R}^{\geq 0}$ nonincreasing weight function

$$L_w = \frac{\sum_{i=1}^{n} \mathbf{x}_{[i]} w(i/n)}{\sum_{j=1}^{n} w(j/n)}.$$

  - Specifically:
    - $w(t) = \mathbb{I}(t \leq 1/n)$: sample median.
    - $w(t) = \mathbb{I}(t \leq 1-\alpha)$: $100\alpha\%$-trimmed mean.

Location

- Median/center: deepest point.
- Other robust location statistics:
  - idea: assign smaller weights to more outlier data.
  - given: $w : [0,1] \to \mathbb{R}^{\geq 0}$ nonincreasing weight function

$$L_w = \frac{\sum_{i=1}^{n} \mathbf{x}_{[i]} w(i/n)}{\sum_{j=1}^{n} w(j/n)}.$$

  - Specifically:
    - $w(t) = \mathbb{I}(t \leq 1/n)$: sample median.
    - $w(t) = \mathbb{I}(t \leq 1 - \alpha)$: $100\alpha\%$-trimmed mean.

No expectation requirement!

Dispersion, scale

Idea-1: similar to location, the dispersion

$$\mathbf{S} = \frac{\sum_{i=1}^{n} \left(\mathbf{x}_{[i]} - \mathbf{x}_{[1]}\right) \left(\mathbf{x}_{[i]} - \mathbf{x}_{[1]}\right)^T w(i/n)}{\sum_{j=1}^{n} w(j/n)}.$$

Dispersion, scale

Idea-1: similar to location, the dispersion

$$\mathbf{S} = \frac{\sum_{i=1}^{n} \left(\mathbf{x}_{[i]} - \mathbf{x}_{[1]}\right) \left(\mathbf{x}_{[i]} - \mathbf{x}_{[1]}\right)^T w(i/n)}{\sum_{j=1}^{n} w(j/n)}.$$

- Scale: $\det(\mathbf{S})$.
- Equivariance under affine tranformations:

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{b} \Rightarrow \mathbf{S}_{\mathbf{y}} = \mathbf{A}\mathbf{S}_{\mathbf{x}}\mathbf{A}^T.$$

Dispersion, scale

Idea-1: similar to location, the dispersion

$$\mathbf{S} = \frac{\sum_{i=1}^{n} \left(\mathbf{x}_{[i]} - \mathbf{x}_{[1]}\right) \left(\mathbf{x}_{[i]} - \mathbf{x}_{[1]}\right)^T w(i/n)}{\sum_{j=1}^{n} w(j/n)}.$$

- Scale: $\det(\mathbf{S})$.
- Equivariance under affine tranformations:

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{b} \Rightarrow \mathbf{S}_\mathbf{y} = \mathbf{A}\mathbf{S}_\mathbf{x}\mathbf{A}^T.$$

No moment constraints!

Dispersion: idea-2

Idea

Dispersion := speed how the data depth decreases.

- p-th (empirical) central region:

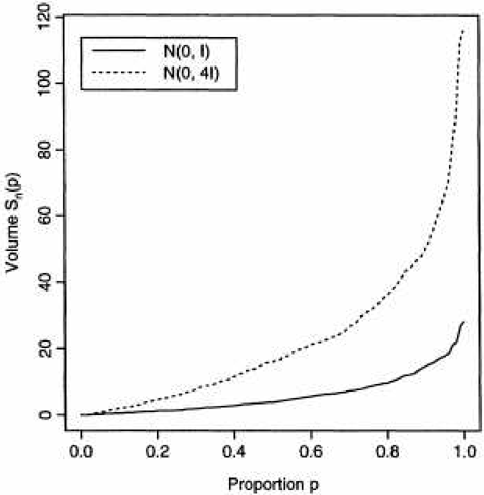
$$C_{n,p} := conv\left\{\mathbf{x}_{[1]}, \ldots, \mathbf{x}_{[np]}\right\}.$$

- scale curve:

$$S_n(p) := vol(C_{n,p}).$$

faster growing of $p \mapsto S_n(p)$ = larger scale of the distribution.

Scale curve example



Skewness: departure from spherical symmetry

- Spherical symmetry around $\mathbf{c}$:

$$X - \mathbf{c} = \mathbf{U}(X - \mathbf{c}), \ \forall \mathbf{U}: \text{ orthogonal.}$$

Skewness: departure from spherical symmetry

- Spherical symmetry around $\mathbf{c}$:

$$X - \mathbf{c} = \mathbf{U}(X - \mathbf{c}), \ \forall \mathbf{U}: \text{ orthogonal.}$$

- Strategy:
  - take the smallest sphere containing $C_{n,p}$,
  - determine the fraction of the data within this sphere,
  - plot it as function of p ($\geq p$).

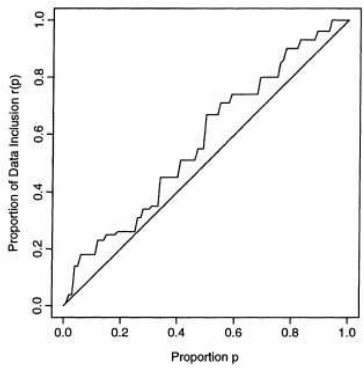
Skewness: departure from spherical symmetry

- Spherical symmetry around $\mathbf{c}$:

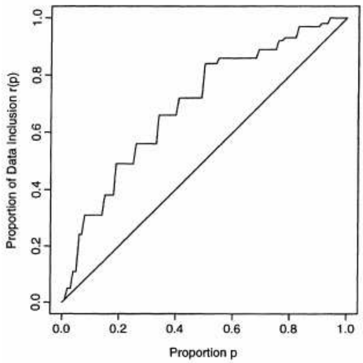
$$X - \mathbf{c} = \mathbf{U}(X - \mathbf{c}), \forall \mathbf{U}: \text{ orthogonal.}$$

- Strategy:
  - take the smallest sphere containing $C_{n,p}$,
  - determine the fraction of the data within this sphere,
  - plot it as function of p ($\geq p$).
- Sperical symmetry $\Leftrightarrow (0,0) \to (1,1)$ linear curve, i.e. area of the gap $= 0$.

Spherical (area $= 0.08$)



Asymmetric (area $= 0.2$)



Sunburst plot

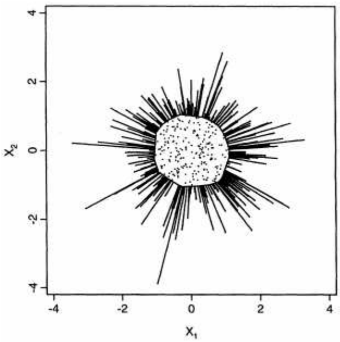
Sunburst plot = '2D boxplot'

For a given depth function:

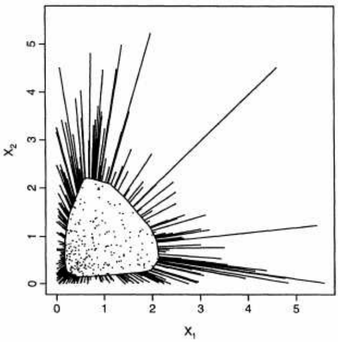
- identify the center,
- central 50% points ($\leftrightarrow$ IQR),
- ray from non-central points to center ($\leftrightarrow$ whiskers).

---

normal



exponential



Testing

Symmetries

$X \in \mathbb{R}^d$ is

- centrally symmetric around $\mathbf{c}$: $X - \mathbf{c} \stackrel{distr}{=} \mathbf{c} - X$.

Symmetries

$X \in \mathbb{R}^d$ is

- centrally symmetric around $\mathbf{c}$: $X - \mathbf{c} \overset{distr}{=} \mathbf{c} - X$.
- angularly symmetric around $\mathbf{c}$: $\frac{X-\mathbf{c}}{\|X-\mathbf{c}\|_2}$ is C-sym. around $\mathbf{0}$.

Symmetries

$X \in \mathbb{R}^d$ is

- centrally symmetric around $\mathbf{c}$: $X - \mathbf{c} \overset{distr}{=} \mathbf{c} - X$.
- angularly symmetric around $\mathbf{c}$: $\frac{X-\mathbf{c}}{\|X-\mathbf{c}\|_2}$ is C-sym. around $\mathbf{0}$.
- half-space symmetric around $\mathbf{c}$: $P(X \in H) \geq \frac{1}{2}$, for $\forall H \ni \mathbf{c}$, closed half-space.

Symmetries

$X \in \mathbb{R}^d$ is

- centrally symmetric around $\mathbf{c}$: $X - \mathbf{c} \stackrel{distr}{=} \mathbf{c} - X$.
- angularly symmetric around $\mathbf{c}$: $\frac{X-\mathbf{c}}{\|X-\mathbf{c}\|_2}$ is C-sym. around $\mathbf{0}$.
- half-space symmetric around $\mathbf{c}$: $P(X \in H) \geq \frac{1}{2}$, for $\forall H \ni \mathbf{c}$, closed half-space.

Note:

- C-symmetry $\Rightarrow$ A-symmetry $\Rightarrow$ H-symmetry.
- A-symmetry $\not\Leftarrow$ H-symmetry: in the discrete case, continuous: $\checkmark$.
- $d = 1$: all $=$ standard symmetry.

Testing: A-symmetry – idea

- SD is maximized at the center of symmetry, $= \frac{1}{2^d}$.
- F: absolutely continuous, $SD(\mathbf{c}^*; F) \leq \frac{1}{2^d}$, '$=$' $\Leftrightarrow$ F: A-sym. around $\mathbf{c}^*$.

Testing: A-symmetry – idea

- SD is maximized at the center of symmetry, $= \frac{1}{2^d}$.
- F: absolutely continuous, $SD(\mathbf{c}^*; F) \leq \frac{1}{2^d}$, '$=$' $\Leftrightarrow$ F: A-sym. around $\mathbf{c}^*$.
- $\mathbf{c}$: hypothesized center.
- Decision: if $\frac{1}{2^d} - SD(\mathbf{c}; F_n)$ is large $\Rightarrow$ reject H_0.

Testing: A-symmetry – idea

- SD is maximized at the center of symmetry, $= \frac{1}{2^d}$.
- F: absolutely continuous, $SD(\mathbf{c}^*; F) \le \frac{1}{2^d}$, '$=$' $\Leftrightarrow$ F: A-sym. around $\mathbf{c}^*$.
- $\mathbf{c}$: hypothesized center.
- Decision: if $\frac{1}{2^d} - SD(\mathbf{c}; F_n)$ is large $\Rightarrow$ reject H_0.
- Under the hood:
    - $\frac{1}{2^d} - SD(\mathbf{c}; F_n)$: degenerate U-statistic,
    - $n\left[\frac{1}{2^d} - SD(\mathbf{c}; F_n)\right] \to \infty$-sum of χ^2-s.

Depth function properties

Desirable properties

$\mathcal{F}$:=Borel probability measures on $\mathbb{R}^d$. $D(\cdot;\cdot) : \mathbb{R}^d \times \mathcal{F} \to \mathbb{R}^{\geq 0}$ bounded,

1. affine invariance:

$$D(\mathbf{Ax} + \mathbf{b}; F_{\mathbf{A}X+\mathbf{b}}) = D(\mathbf{x}; F_X) \ \forall \mathbf{A} : \text{invertible}, \mathbf{b}, X.$$

Desirable properties

$\mathcal{F}$:=Borel probability measures on $\mathbb{R}^d$. $D(\cdot;\cdot):\mathbb{R}^d \times \mathcal{F} \to \mathbb{R}^{\geq 0}$ bounded,

1. affine invariance:

$$D(\mathbf{Ax}+\mathbf{b}; F_{\mathbf{A}X+\mathbf{b}}) = D(\mathbf{x}; F_X) \ \forall \mathbf{A}: \text{invertible}, \mathbf{b}, X.$$

2. maximality at center: For F with center $\mathbf{c}$,

$$D(\mathbf{c}; F) = \sup_{\mathbf{x}\in\mathbb{R}^d} D(\mathbf{x}; F).$$

Desirable properties

$\mathcal{F}$:=Borel probability measures on $\mathbb{R}^d$. $D(\cdot;\cdot):\mathbb{R}^d\times\mathcal{F}\to\mathbb{R}^{\geq 0}$ bounded,

1. affine invariance:

$$D(\mathbf{A}\mathbf{x}+\mathbf{b};F_{\mathbf{A}X+\mathbf{b}})=D(\mathbf{x};F_X)\ \forall\mathbf{A}:\text{invertible},\mathbf{b},X.$$

2. maximality at center: For F with center $\mathbf{c}$,

$$D(\mathbf{c};F)=\sup_{\mathbf{x}\in\mathbb{R}^d}D(\mathbf{x};F).$$

3. monotonicity relative to the deepest point: For $\forall F$ with deepest point $\mathbf{c}$,

$$D(\mathbf{x};F)\leq D(\alpha\mathbf{x}+(1-\alpha)\mathbf{c};F),\forall\alpha\in[0,1].$$

Desirable properties

$\mathcal{F}$:=Borel probability measures on $\mathbb{R}^d$. $D(\cdot;\cdot) : \mathbb{R}^d \times \mathcal{F} \to \mathbb{R}^{\geq 0}$ bounded,

1. affine invariance:
$$D(\mathbf{Ax}+\mathbf{b}; F_{\mathbf{A}X+\mathbf{b}}) = D(\mathbf{x}; F_X)\ \forall \mathbf{A} : \text{invertible}, \mathbf{b}, X.$$
2. maximality at center: For F with center $\mathbf{c}$,
$$D(\mathbf{c}; F) = \sup_{\mathbf{x}\in\mathbb{R}^d} D(\mathbf{x}; F).$$
3. monotonicity relative to the deepest point: For $\forall F$ with deepest point $\mathbf{c}$,
$$D(\mathbf{x}; F) \leq D(\alpha\mathbf{x}+(1-\alpha)\mathbf{c}; F), \forall \alpha \in [0,1].$$
4. vanishing at infinity:
$$D(\mathbf{x}; F) \xrightarrow{\|\mathbf{x}\|\to\infty} 0, \forall F.$$

- Vanishing at infinity: useful for establishing

$$\sup_{\mathbf{x}} |D(\mathbf{x}; F_n) - D(\mathbf{x}; F)| \xrightarrow{n\to\infty} 0 \;\; \text{(F-a.s.)}.$$

Categorization of depth functions [Zuo and Serfling, 2000]

- Properties of HD and SD:
  - half-space (HD): 1-4 ✓ (H-sym.)
  - simplicial (SD):
    - continuous distributions: 1-4 ✓ (A-sym.)
    - discrete: 2/3 might fail.

Types

$$D_A(\mathbf{x}; F) = \mathbb{E}_F\left[h_A(\mathbf{x}; X_1; \ldots; X_r)\right] \xrightarrow{\text{example}} SD,$$

$$D_B(\mathbf{x}; F) = \frac{1}{1 + \mathbb{E}_F\left[h_B(\mathbf{x}; X_1; \ldots; X_r)\right]},$$

$$D_C(\mathbf{x}; F) = \frac{1}{1 + O(\mathbf{x}; F)},$$

$$D_D(\mathbf{x}; F) = \inf_{\mathbf{x} \in C \in \mathcal{C}} \mathbb{P}_F(X \in C) \xrightarrow{\text{example}} HD.$$

- $h_A(\mathbf{x}; \mathbf{x}_1, \ldots, \mathbf{x}_r)$: bounded, closeness of $\mathbf{x}$ to $\mathbf{x}_i$-s.
- $h_B(\mathbf{x}; \mathbf{x}_1, \ldots, \mathbf{x}_r)$: unbounded, distance of $\mathbf{x}$ from $\mathbf{x}_i$-s.
- $O(\mathbf{x}; F)$: unbounded outlyingness function.
- $\mathcal{C}$: closedness properties.

Types

$$D_A(\mathbf{x}; F) = \mathbb{E}_F\left[h_A(\mathbf{x}; X_1; \ldots; X_r)\right] \xrightarrow{\text{example}} SD,$$

$$D_B(\mathbf{x}; F) = \frac{1}{1 + \mathbb{E}_F\left[h_B(\mathbf{x}; X_1; \ldots; X_r)\right]},$$

$$D_C(\mathbf{x}; F) = \frac{1}{1 + O(\mathbf{x}; F)},$$

$$D_D(\mathbf{x}; F) = \inf_{\mathbf{x} \in C \in \mathcal{C}} \mathbb{P}_F(X \in C) \xrightarrow{\text{example}} HD.$$

- $h_A(\mathbf{x}; \mathbf{x}_1, \ldots, \mathbf{x}_r)$: bounded, closeness of $\mathbf{x}$ to $\mathbf{x}_i$-s.
- $h_B(\mathbf{x}; \mathbf{x}_1, \ldots, \mathbf{x}_r)$: unbounded, distance of $\mathbf{x}$ from $\mathbf{x}_i$-s.
- $O(\mathbf{x}; F)$: unbounded outlyingness function.
- $\mathcal{C}$: closedness properties.

Note: $D_A(\mathbf{x}; F_n)$: U/V-statistic.

Type B examples

Let $\mathbf{\Sigma} = cov(F)$. In simplicial volume depth ($D_{SVD^{\alpha}}$), $D_{\tilde{L^2}}$:

$$h(\mathbf{x}; \mathbf{x}_1, \ldots, \mathbf{x}_d) = \left(\frac{vol[\Delta(\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_d)]}{\sqrt{\det(\mathbf{\Sigma})}} \right)^{\alpha}, \alpha > 0,$$

$$h(\mathbf{x}; \mathbf{x}_1) = \|\mathbf{x} - \mathbf{x}_1\|_{\Sigma^{-1}}, \ \|\mathbf{z}\|_{\mathbf{M}} = \sqrt{\mathbf{z}^T \mathbf{M} \mathbf{z}}.$$

Properties:

- $D_{SVD^{\alpha}}$: 1-4 ✓ (C-sym.),
- $D_{\tilde{L^2}}$: 1-4 ✓ (A-sym.).

Type C example

Projection depth: worst case outlyingness w.r.t. 1d-median, for any 1d projection.

$$O(\mathbf{x};F) = \sup_{\|\mathbf{u}\|_2=1} \frac{\left|\mathbf{u}^T\mathbf{x} - med(\mathbf{u}^T X)\right|}{MAD(\mathbf{u}^T\mathbf{X})}, X \sim F,$$

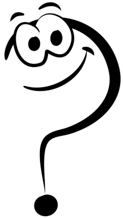
$$MAD(Y) = med(|Y - med(Y)|).$$

Properties: 1-4 ✓ (H-sym.)

Summary

- Depth functions: simplicial, half-space, ...
- They define: center-outward ordering, ranks, $\Rightarrow$
- L-statistics, symmetry test.
- location (median,...), dispersion, scale, skewness, kurtosis (no moments),
- sunburst plot.

Thank you for the attention!



Dyckerhoff, R. and Mozharovskyi, P. (2016).
Exact computation of the halfspace depth.
Technical report, University of Cologne.
http://arxiv.org/abs/1411.6927.

Liu, R. (1990).
On a notion of data depth based on random simplices.
The Annals of Statistics, 18:405–414.

Tukey, J. (1975).
Mathematics and picturing data.
In *International Congress on Mathematics*, pages 523–531.

Zuo, B. Y. and Serfling, R. (2000).
General notions of statistical depth function.
The Annals of Statistics, 28:461–482.