Data depth

Pavlo Mozharovskyi*

*LTCI, Telecom Paris, Institut Polytechnique de Paris

Tail events analysis:
Robustness, outliers and models for extreme values

Palaiseau, February 18, 2020
Outline of the course

**Format:** 6 × 3.5 hours + exam

- **Class 1:** Introduction to robust statistics
- **Class 2:** Lab session I
- **Class 3:** Data depth
- **Weeks 4:** Extreme value statistics
- **Week 5:** Multi-dimensional setting
- **Week 6:** Lab session II

**Programming language:** R

**Grading:** Exam
Today

The concept of data depth

Some examples
  - Mahalanobis depth
  - Spatial depth
  - Projection depth
  - Zonoid depth

Depth-based outlier detection

Halfspace depth
  - Properties
  - Trimmed regions

Classes of depth functions
  - Depths satisfying the projection property

A concept of depth for the functional framework
  - Functional outlier detection

References
Contents

The concept of data depth

Some examples
  Mahalanobis depth
  Spatial depth
  Projection depth
  Zonoid depth

Depth-based outlier detection

Halfspace depth
  Properties
  Trimmed regions

Classes of depth functions
  Depths satisfying the projection property

A concept of depth for the functional framework
  Functional outlier detection

References
Statistical data depth (Zuo & Serfling, 2000)

A data depth measures how “close” a given point is located to the “center” of a distribution. For \( x \in \mathbb{R}^d \) and a \( d \)-variate random vector \( X \) distributed as \( P \in \mathcal{P} \), a data depth is a function

\[
D : \mathbb{R}^d \times \mathcal{P} \rightarrow \mathbb{R}^1, (x, P) \mapsto D(x|P)
\]

that is:

(i) affine invariant: \( D(Ax + b|AX + b) = D(x|X) \) for any \( d \times d \) non-singular matrix \( A \) and any \( b \in \mathbb{R}^d \);

(ii) maximal at symmetry center: \( D(x^*|X) = \sup_{x \in \mathbb{R}^d} D(x|X) \) holds for any \( P \in \mathcal{P} \) having center \( x^* \);

(iii) monotone on rays: for any \( x^* \) with \( D(x^*|X) = \sup_{x \in \mathbb{R}^d} D(x|X) \), any \( x \in \mathbb{R}^d \), and any \( 0 \leq \alpha \leq 1 \) it holds:

\[
D(x|X) \leq D(x^* + \alpha(x - x^*)|X);
\]

(iv) vanishing at infinity: \( \lim_{||x|| \rightarrow \infty} D(x|X) = 0. \)

A distribution is called centrally symmetric about \( x^* \) if distributions of \( X - x^* \) and \( x^* - X \) coincide.
A **data depth** measures how “close” a given point is located to the “center” of a distribution. For \( x \in \mathbb{R}^d \) and a \( d \)-variate random vector \( X \) distributed as \( P \in \mathcal{P} \), a data depth is a function

\[
D : \mathbb{R}^d \times \mathcal{P} \rightarrow [0, 1], (x, P) \mapsto D(x|P)
\]

that is:

D1 **translation invariant:** \( D(x + b|X + b) = D(x|X) \) for any \( b \in \mathbb{R}^d \);

D2 **linear invariant:** \( D(Ax|AX) = D(x|X) \) for any \( d \times d \) non-singular matrix \( A \);

D3 **vanishing at infinity:** \( \lim_{||x|| \to \infty} D(x|X) = 0 \);

D4 **monotone on rays:** for any \( x^* \in \arg \max_{x \in \mathbb{R}^d} D(x|X) \), any \( x \in \mathbb{R}^d \), and any \( 0 \leq \alpha \leq 1 \) it holds: \( D(x|X) \leq D(x^* + \alpha(x - x^*)|X) \);

D5 **upper semicontinuous in \( x \):** the upper-level sets

\[
D_\alpha(X) = \{ x \in \mathbb{R}^d : D(x|X) \geq \alpha \}
\]

are closed for all \( \alpha \).
Statistical data depth (Mosler, 2013)

Some remarks:

- **D1 – D5** implies (ii) due to Dyckerhoff (2004):
  If $X$ is centrally symmetric about $x^*$, then any depth function satisfying D1 – D5 $D(\cdot|X)$ is maximal at $x^*$.

- **D4** implies star-shaped upper-level sets of $D$.
  One can strengthen to:
  - **D4con**: $D(\cdot|X)$ is a quasiconcave function, i.e. the upper-level sets $D_\alpha(X)$ are convex for all $\alpha$.

- **D1** and **D2** define affine invariante depth.
  One can also weaken to:
  - **D2iso**: $D(Ax|AX) = D(x|X)$ for every isometric linear $A$ to define orthogonal invariant depth;
  - **D2sca**: $D(\lambda x|\lambda X) = D(x|X)$ for any $\lambda > 0$ to define scale invariant depth.
Central regions

- For given $P$ and $\alpha \in [0, 1]$, the level sets $D_\alpha(P)$ form a family of depth-trimmed of central regions.

- The innermost region arises at some depth $\alpha_{\text{max}} \leq 1$, which depends on the depth notion $D$ and distribution $P$. Then $D_\alpha(X)$ is the set of deepest points.

- Central regions describe distribution w.r.t. location, dispersion, and shape.

- Properties of central regions, for any $\alpha$:
  - Due to D1 and D2 $D_\alpha(X)$ is affine equivariant: $D_\alpha(AX + b) = AD_\alpha(X) + b$ for any $d \times d$ non-singular matrix $A$ and any $b \in \mathbb{R}^d$.
  - Due to D3 $D_\alpha(X)$ is bounded;
  - Due to D4 $D_\alpha(X)$-s are nested: if $\alpha \geq \beta$, then $D_\alpha(X) \subseteq D_\beta(X)$, and star-shaped; due to D4con $D_\alpha(X)$ is in addition convex;
  - Due to D5 $D_\alpha(X)$ is closed.
Depth lifts and orderings

- Assume $\alpha_{\text{max}} = 1$ and add a real dimension to $D_\alpha, \alpha \in [0, 1]$, then construct the depth lift:

$$\hat{D}(P) = \{(\alpha, y) \in [0, 1] \times \mathbb{R}^d : y = \alpha x, x \in D_\alpha(P), \alpha \in [0, 1]\}.$$

- This gives rise to an ordering of probability distributions in $\mathcal{P}$:

$$P \prec_D Q \quad \text{if} \quad \hat{D}(P) \subset \hat{D}(Q).$$

If (and only if) the family of central regions induced by $D$ completely characterizes the underlying distribution, then $\prec_D$ is an order, otherwise it is a preorder.

- $D$ introduces a probability semi-metric on $\mathcal{P}$:

$$\delta_D(P, Q) = \delta_H(\hat{D}(P), \hat{D}(Q)),$$

where $\delta_H(X, Y)$ is the Hausdorff distance between two non-empty sets in $\mathbb{R}^d$ with metric $\delta$:

$$\delta_H(X, Y) = \max\{\sup_{x \in X} \inf_{y \in Y} \delta(x, y), \sup_{y \in Y} \inf_{x \in X} \delta(x, y)\}.$$
Applications

- **Multivariate data analysis** (Liu, Parelius, Singh '99);
- **Statistical quality control** (Liu, Singh '93);
- **Cluster analysis and classification** (Mosler, Hoberg '06; Li, Cuesta-Albertos, Liu '12; Lange, Mosler, Mozharovskyi '14);
- **Tests for multivariate location, scale, symmetry** (Liu '92; Dyckerhoff '02; Dyckerhoff, Ley, Paindaveine '15);
- **Outlier detection** (Hubert, Rousseeuw, Segaert '15);
- **Multivariate risk measurement** (Cascos, Mochalov '07);
- **Robust linear programming** (Bazovkin, Mosler '15);
- **Missing data imputation** (Mozharovskyi, Josse, Husson '17);
- etc.

R-package **ddalpha** (Pokotylo, Mozharovskyi, Dyckerhoff, Nagy): calculates a number of depths; performs depth-based classification of multivariate and functional data; contains 50 multivariate and 5 functional data sets.
Contents

The concept of data depth

Some examples
  Mahalanobis depth
  Spatial depth
  Projection depth
  Zonoid depth

Depth-based outlier detection

Halfspace depth
  Properties
  Trimmed regions

Classes of depth functions
  Depths satisfying the projection property

A concept of depth for the functional framework
  Functional outlier detection

References
Mahalanobis depth (Mahalanobis, 1936)

- Mahalanobis depth is defined as:
  \[ D^{Mah}(x|X) = \frac{1}{1 + (d^{Mah})^2(x|X)} , \]
  
  based on Mahalanobis distance:
  \[ (d^{Mah})^2(x|X) = (x - \mu_X)^T \Sigma_X^{-1} (x - \mu_X) . \]

- In the empirical version, \( \mu_X \) and \( \Sigma_X \) are substituted by suitable estimates:
  - moment estimates;
  - robust estimates such as minimum volume ellipsoid or minimum covariance determinant (MCD).

- Properties:
  - satisfies D1 – D5 and D4con, is continuous;
  - being defined by \( d(d + 1) \) parameters, can be seen as a parametric depth;
  - by a single elliptical contour characterizes a multivariate normal distribution or one within an affine family of non-degenerate elliptical distributions.
Mahalanobis depth (Mahalanobis, 1936)
Mahalanobis depth (Mahalanobis, 1936)
Spatial depth (Vardi & Zhang, 2000; Serfling 2002)

Exploiting the idea of spatial quantiles of Chaudhuri (1996) and Koltchinskii (1997), Vardi & Zhang (2000) and Serfling (2002) formulate the spatial depth (also $L_1$-depth) as:

$$D^{spt}(x|X) = 1 - \left\| \mathbb{E} \left[ \nu \left( \Sigma^{-\frac{1}{2}} (x - X) \right) \right] \right\|,$$

$$\nu(y) = \begin{cases} \frac{y}{\|y\|} & \text{if } y \neq 0, \\ 0 & \text{if } y = 0. \end{cases}$$

Properties:

- satisfies D1 – D5, but not D4con, is continuous;
- if $\Sigma$ is orthogonal, satisfies D2iso only;
- with D2iso its maximum (say $x^*$) is referred to as spatial median, a multivariate location estimator having asymptotic breakdown point of 0.5;
- spatial sign covariance matrix can be defined as:

$$S(X) = \mathbb{E} \left[ \nu(X - x^*) \nu(X - x^*)^T \right].$$
Spatial depth (Vardi & Zhang, 2000; Serfling 2002)
Projection depth (Zuo & Serfling, 2000)

According to Zuo & Serfling (2000), projection depth is defined as:

\[
D^{prj}(x|X) = \frac{1}{1 + O^{prj}(x|X)},
\]

where

\[
O^{prj}(x|X) = \sup_{r \in S^{d-1}} \frac{|x^T r - \text{med}(X^T r)|}{\text{MAD}(X^T r)}
\]

is the projected outlyingness (Stahel, 1981; Donoho, 1982), med(Y) and MAD(Y) = \(\text{med}(|Y - \text{med}(Y)|)\) are the univariate median and median absolute deviation from the median, respectively.

Properties:

- Satisfies D1 – D5 and D4con, is continuous;
- its median has asymptotic breakdown point of 0.5.
Projection depth (Zuo & Serfling, 2000)
Contents

The concept of data depth

Some examples
  - Mahalanobis depth
  - Spatial depth
  - Projection depth
  - Zonoid depth

Depth-based outlier detection

Halfspace depth
  - Properties
  - Trimmed regions

Classes of depth functions
  - Depths satisfying the projection property

A concept of depth for the functional framework
  - Functional outlier detection

References
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

Koshevoy & Mosler (1997) and Mosler (2002) define the zonoid trimmed region as:

\[
D_{\alpha}^{\text{zon}}(P) = \left\{ \int_{\mathbb{R}^d} xg(x)dP(x) : g : \mathbb{R}^d \mapsto [0, \frac{1}{\alpha}] \text{ measurable and } \int_{\mathbb{R}^d} g(x)dP(x) = 1 \right\}
\]

for \( \alpha \in (0, 1] \), and

\[
D_{0}^{\text{zon}}(P) = \text{cl}\left( \bigcup_{\alpha \in (0,1]} D_{\alpha}(P) \right)
\]

for \( \alpha = 0 \), where “cl” denotes closure.

Zonoid depth is then defined as:

\[
D^{\text{zon}}(x|X) = \begin{cases} 
\sup\{\alpha : x \in D_{\alpha}^{\text{zon}}(X)\} & \text{if } x \in \text{conv}(\text{supp}(X)) \\
0 & \text{otherwise} 
\end{cases}
\]
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For \( \mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_n\} \), \( \alpha \in \left[\frac{k}{n}, \frac{k+1}{n}\right] \), \( k = 1, ..., n - 1 \), \( N = \{1, ..., n\} \)

\[
D_{z\alpha}^{\text{zon}}(\mathbf{X}) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \left(1 - \frac{k}{\alpha n}\right) \mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \right\}:
\]
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For \( \mathbf{X} = \{ \mathbf{x}_1, ..., \mathbf{x}_n \} \), \( \alpha \in \left[ \frac{k}{n}, \frac{k+1}{n} \right] \), \( k = 1, ..., n - 1 \), \( N = \{1, ..., n\} \)

\[
D_{\alpha}^{zôn}(\mathbf{X}) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \left(1 - \frac{k}{\alpha n}\right) \mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \right\}
\]
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $X = \{x_1, \ldots, x_n\}$, $\alpha \in \left[\frac{k}{n}, \frac{k+1}{n}\right]$, $k = 1, \ldots, n-1$, $N = \{1, \ldots, n\}$

$$D_{\alpha}^{\text{zon}}(X) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} x_{i_j} + \left(1 - \frac{k}{\alpha n}\right) x_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N \right\}$$
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $X = \{x_1, \ldots, x_n\}$, $\alpha \in \left[\frac{k}{n}, \frac{k+1}{n}\right]$, $k = 1, \ldots, n - 1$, $N = \{1, \ldots, n\}$

$$D_{\alpha}^{\text{zon}}(X) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} x_{ij} + \left(1 - \frac{k}{\alpha n}\right) x_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N \right\}:
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $X = \{x_1, ..., x_n\}$, $\alpha \in \left[\frac{k}{n}, \frac{k+1}{n}\right]$, $k = 1, ..., n - 1$, $N = \{1, ..., n\}$

$$D_{\alpha}^{\text{zon}}(X) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} x_{i_j} + \left(1 - \frac{k}{\alpha n}\right)x_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \right\} :$$
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $X = \{x_1, \ldots, x_n\}$, $\alpha \in \left[\frac{k}{n}, \frac{k+1}{n}\right]$, $k = 1, \ldots, n-1$, $N = \{1, \ldots, n\}$

$$D_{\alpha}^{z\text{on}}(X) = \text{conv}\left\{\frac{1}{\alpha n} \sum_{j=1}^{k} x_{i_j} + \left(1 - \frac{k}{\alpha n}\right)x_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N\right\} :$$
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $X = \{x_1, \ldots, x_n\}$, $\alpha \in \left[ \frac{k}{n}, \frac{k+1}{n} \right]$, $k = 1, \ldots, n - 1$, $N = \{1, \ldots, n\}$

$$D_{\alpha}^{zon}(X) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} x_{i_j} + \left(1 - \frac{k}{\alpha n}\right)x_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N \right\} :$$
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $X = \{x_1, \ldots, x_n\}$, $\alpha \in \left[\frac{k}{n}, \frac{k+1}{n}\right]$, $k = 1, \ldots, n - 1$, $N = \{1, \ldots, n\}$

$$D_{\alpha}^{z\text{on}}(X) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} x_{i_j} + \left(1 - \frac{k}{\alpha n}\right)x_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N \right\} :$$
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, $\alpha \in \left[\frac{k}{n}, \frac{k+1}{n}\right]$, $k = 1, \ldots, n - 1$, $N = \{1, \ldots, n\}$

$$D_{\alpha}^{\text{zon}}(\mathbf{X}) = \text{conv}\left\{\frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \left(1 - \frac{k}{\alpha n}\right) \mathbf{x}_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N\right\} :$$
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, $\alpha \in \left[\frac{k}{n}, \frac{k+1}{n}\right]$, $k = 1, \ldots, n-1$, $N = \{1, \ldots, n\}$

$$D_{\alpha}^{\text{zon}}(\mathbf{X}) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{ij} + \left(1 - \frac{k}{\alpha n}\right) \mathbf{x}_{ik+1} : \{i_1, \ldots, i_{k+1}\} \subset N \right\}.$$
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, $\alpha \in \left[\frac{k}{n}, \frac{k+1}{n}\right]$, $k = 1, \ldots, n-1$, $N = \{1, \ldots, n\}$

$$D^\text{zon}_\alpha(\mathbf{X}) = \text{conv}\left\{\frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \left(1 - \frac{k}{\alpha n}\right)\mathbf{x}_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N\right\} :$$
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For \( \mathbf{X} = \{x_1, \ldots, x_n\} \), \( \alpha \in \left[\frac{k}{n}, \frac{k+1}{n}\right] \), \( k = 1, \ldots, n-1 \), \( \mathbb{N} = \{1, \ldots, n\} \)

\[
D_{\alpha}^{\text{zon}}(\mathbf{X}) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} x_{i_j} + \left(1 - \frac{k}{\alpha n}\right)x_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subseteq \mathbb{N} \right\}:
\]
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, $\alpha \in \left[\frac{k}{n}, \frac{k+1}{n}\right]$, $k = 1, \ldots, n - 1$, $N = \{1, \ldots, n\}$

$$D_{\alpha}^{\text{zon}}(\mathbf{X}) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \left(1 - \frac{k}{\alpha n}\right) \mathbf{x}_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N \right\}$$
Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)
Zonoid depth (Mosler, 2002), properties:

- satisfies \( D1 - D5 \) and \( D4\text{con} \);
- \( D_{\alpha}^{\text{zon}}(X) \) is equivariant to any linear transformation, \textit{i.e.} any marginal projection of \( D_{\alpha}^{\text{zon}} \) is \( D_{\alpha}^{\text{zon}} \) of the marginal;
- \( D_{\alpha}^{\text{zon}}(x|X) \) is continuous on \( \text{conv}(\text{supp}(X)) \) in \( x \) and in \( X \);
- \( D_{\alpha}^{\text{zon}}(x|X) \) is monotone on dilation, \textit{i.e.} for two random vectors \( X \) and \( Y \) in \( \mathbb{R}^d \):
  \[
  D_{\alpha}^{\text{zon}}(x|X) \leq D_{\alpha}^{\text{zon}}(x|Y) \quad \text{if} \quad Y \overset{d}{=} \lambda X \text{ with } \lambda > 1 ;
  \]
- \( D_{\alpha}^{\text{zon}}(x|X) = 0 \) for all \( x \notin \text{conv}(\text{supp}(X)) \);
- \( D_{\alpha}^{\text{zon}}(\cdot|X) \) characterizes the distribution;
- \( D_{\alpha}^{\text{zon}}(X) \)-s are subadditive, \textit{i.e.} for \( X \) and \( Y \) in \( \mathbb{R}^d \):
  \[
  D_{\alpha}^{\text{zon}}(X + Y) \subset D_{\alpha}^{\text{zon}}(X) \oplus D_{\alpha}^{\text{zon}}(Y)
  \]
  where \( \oplus \) signifies the Minkowski sum of sets;
- If \( P \) has finite first moment and \( P(\partial H) = 0 \) for every halfspace \( H \) with \( P(H) = 1 \), then (Cascos, Lopez-Diaz, 2016):
  \[
  \sup_{x \in \mathbb{R}^d} \left| D_{\alpha}^{\text{zon}}(x|P) - D_{\alpha}^{\text{zon}}(x|P_n) \right| \overset{a.s.}{\to} 0.
  \]
Zonoid depth (Dyckerhoff, Koshevoy, Mosler, 1996):

- Zonoid depth can be defined as:

\[
D^{\text{zon}}(x|X) = \sup \{ \alpha \in [0, 1] : x \in D^{\text{zon}}_{\alpha}(X) \},
\]

where \(\sup\) of \(\emptyset\) is defined to be 0, with:

\[
D^{\text{zon}}_{\alpha}(X) = \left\{ \sum_{i=1}^{n} \lambda_i x_i : \sum_{i=1}^{n} \lambda_i = 1, 0 \leq \lambda_i, \alpha \lambda_i \leq \frac{1}{n} \text{ for all } i \right\}.
\]

- \(D^{\text{zon}}(x|X) = \frac{1}{n\gamma^*}\) with \(\gamma^*\) being an optimizer of:

\[
\min \gamma \quad \text{s. t.} \quad X^T \lambda = x, \quad \lambda^T 1_n = 1, \quad \gamma 1_n - \lambda \geq 0_n, \quad \lambda \geq 0_n.
\]

(Notation: \(X = (x_1, \ldots, x_n)^T, \lambda = (\lambda_1, \ldots, \lambda_n)^T\).)
Contents

The concept of data depth

Some examples
  Mahalanobis depth
  Spatial depth
  Projection depth
  Zonoid depth

Depth-based outlier detection

Halfspace depth
  Properties
  Trimmed regions

Classes of depth functions
  Depths satisfying the projection property

A concept of depth for the functional framework
  Functional outlier detection

References
Detection of multivariate outliers: Mahalanobis depth

Regard again the two measurements during a test:

- Label $x$ as outlier $D_{Mah}^x(x|X) < \min(D_{Mah})$. 
Detection of multivariate outliers: Mahalanobis depth

Regard again the two measurements during a test:

- Label \( \mathbf{x} \) as outlier if \( D^{Mah}(\mathbf{x} \mid \mathbf{X}) < \min(D^{Mah}) \).
- With less available data, e.g., at the beginning of the production process, when abnormal behavior is in addition more likely.
Detection of multivariate outliers: Mahalanobis depth

Regard again the two measurements during a test:

- Label $x$ as outlier $D^{Mah}(x|X) < \min(D^{Mah})$.
- With less available data, e.g. at the beginning of the production process, when abnormal behavior is in addition more likely.
- Mahalanobis depth upper-level set is not robust.
Detection of multivariate outliers: projection depth

Regard again the two measurements during a test:

- Projection depth of \( \mathbf{x} \) w.r.t. \( \mathbf{X} = \{ \mathbf{x}_i \}_{i=1}^n \):

\[
D^{\text{proj}}(\mathbf{x}|\mathbf{X}) = \left( 1 + \max_{\mathbf{u} \in S^{d-1}} \frac{||\mathbf{x}^\top \mathbf{u} - \text{med}(\mathbf{X}\mathbf{u})||}{\text{MAD}(\mathbf{X}\mathbf{u})} \right)^{-1}.
\]
Detection of multivariate outliers: projection depth

Regard again the two measurements during a test:

\[ D^{prj}(x|X) = \left(1 + \max_{u \in S^{d-1}} \frac{|x^T u - \text{med}(X u)|}{\text{MAD}(X u)} \right)^{-1}. \]

- Projection depth of \( x \) w.r.t. \( X = \{x_i\}_{i=1}^n \):

- Label \( x \) as outlier \( D^{prj}(x|X) < \min(D^{prj}) \).
Detection of multivariate outliers: projection depth

Regard again the two measurements during a test:

![Graph showing projection depth]

- **Projection depth** of $\mathbf{x}$ w.r.t. $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{n}$:

  $$D^{prj}(\mathbf{x}|\mathbf{X}) = \left(1 + \max_{u \in S^{d-1}} \frac{|\mathbf{x}^T u - \text{med}(\mathbf{X} u)|}{\text{MAD}(\mathbf{X} u)} \right)^{-1}.$$

- Label $\mathbf{x}$ as **outlier** $D^{prj}(\mathbf{x}|\mathbf{X}) < \min(D^{prj})$.
- Stable rule for more observations.
Tukey (=halfspace, location) data depth

Babies with low birth weight

Weight, in grams

Age, in weeks

800 1000 1200 1400

20 25 30 35
Tukey (=halfspace, location) data depth

Babies with low birth weight

Age, in weeks

Weight, in grams
Tukey (halfspace, location) data depth

Babies with low birth weight

Weight, in grams

Age, in weeks

120 / 161
Tukey (=halfspace, location) data depth

Babies with low birth weight

Weight, in grams

Age, in weeks

112 / 161
Tukey (=halfspace, location) data depth

Babies with low birth weight

Weight, in grams
Age, in weeks

47 / 161
Tukey (=halfspace, location) data depth

Babies with low birth weight

Weight, in grams

Age, in weeks

26 / 161
Tukey (halfspace, location) data depth

Babies with low birth weight

Age, in weeks

Weight, in grams

41 / 161
Tukey (=halfspace, location) data depth

Babies with low birth weight

Weight, in grams

Age, in weeks

49 / 161
Tukey (halfspace, location) data depth

Babies with low birth weight

Weight, in grams

Age, in weeks

114 / 161
Tukey (=halfspace, location) data depth

Babies with low birth weight

Weight, in grams

Age, in weeks

135 / 161
Tukey (=halfspace, location) data depth

Babies with low birth weight

Weight, in grams

Age, in weeks

800 1000 1200 1400

20 25 30 35

Babies with low birth weight

Weight, in grams
Tukey (halfspace, location) data depth

Babies with low birth weight

Age, in weeks

Weight, in grams
Tukey (=halfspace, location) data depth

Babies with low birth weight

157 / 161

Weight, in grams

Age, in weeks
Tukey (=halfspace, location) data depth

Babies with low birth weight

152 / 161
Tukey (=halfspace, location) data depth

Babies with low birth weight

14 / 161

Weight, in grams

Age, in weeks
Tukey (=halfspace, location) data depth

Babies with low birth weight

Weight, in grams
Tukey (halfspace, location) data depth
Tukey (=halfspace, location) data depth

Babies with low birth weight

Weight, in grams

Age, in weeks

9 / 161
Tukey (=halfspace, location) data depth

Babies with low birth weight

Age, in weeks

Weight, in grams

147 / 161
Tukey (=halfspace, location) data depth

Babies with low birth weight

Weight, in grams

Age, in weeks

3 / 161
Tukey (=halfspace, location) data depth
Tukey (halfspace, location) depth

Tukey (1975) — “Mathematics and the picturing of data”

Tukey depth of $x \in \mathbb{R}^d$ w.r.t. a $d$-variate random vector $X$ distributed as $P$ is defined as the smallest probability mass of a closed halfspace containing $x$:

$$D_{Tuk}(x|X) = \inf \{ P(H) : H \text{ is a closed halfspace, } x \in H \},$$

and w.r.t. a sample $X = \{x_1, \ldots, x_n\} \in \mathbb{R}^d$:

$$D_{Tuk(n)}(x|X) = \frac{1}{n} \min_{u \in S^{d-1}} \# \{i : u^T x_i \geq u^T x \}.$$

Tukey depth defines a family of (depth-)trimmed (central) regions $D_{Tuk}^\alpha(X)$, the upper-level sets of the depth function:

$$D_{Tuk}^\alpha(X) = \{x \in \mathbb{R}^d : D_{Tuk}(x|X) \geq \alpha\}.$$
Halfspace depth: Properties

- satisfies $D1 – D5$ and $D4\text{con}$;
- $D^{Tuk}(x|X) = 0$ for all $x \notin \text{conv}(\text{supp}(X))$;
- $D^{Tuk}(\cdot|X)$ attains its supremum (Rousseeuw & Ruts, 1999);
- at least one sample point has depth $\geq \frac{1}{d+1}$ (Mizera, 2002);
- if distribution smoothness condition is satisfied, i.e.
  \[ P(\partial H) = 0 \text{ for all } H \in \mathcal{H}, \mathcal{H} \text{ class of closed halfspaces in } \mathbb{R}^d, \]
  then the function $x \mapsto D^{Tuk}(x|\cdot)$ is continuous (Donoho & Gasko, 1992; Masse, 2004);
- determines the empirical distribution uniquely (Struyf & Rousseeuw, 1999; Koshevoy, 2002);
- is strongly uniformly consistent (Donoho & Gasko, 1992):
  \[ \sup_{x \in \mathbb{R}^d} \left| D^{Tuk}(x|P) - D^{Tuk}(x|P_n) \right| \overset{a.s.}{\to} 0. \]
Let $m(n)$ be the maximum number of subsets formed by intersecting finite sets of $n$ points with halfspaces in $\mathbb{R}^d$. One has $m(n) \leq \frac{3n^{d+1}}{2(d+1)!}$.

Let $\mathcal{H}$ be the set of all halfspaces in $\mathbb{R}^d$ and suppose $\epsilon > 0$. Then, for $n$ sufficiently large,

$$P\left( \sup_{H \in \mathcal{H}} \left| \frac{1}{n} \# \{X \cap H \} - P(X \in H) \right| \geq \epsilon \right) \leq 4m(n^2) e^{4\epsilon + 4\epsilon^2} e^{-2n\epsilon^2}.$$ 

$X$ decays exponentially if $P(\|X\| > R) = O\left(e^{-\lambda \frac{R^2}{2}}\right)$.

$X$ is Lipschitz continuous in projection if $F_{X^T r}$ is Lipschitz continuous $\forall \ r \in S^{d-1}$.

$X$ is radially Lipschitz continuous if $F_{X^T r}(t)$ is a Lipschitz continuous function of $r$ for any fixed $t$.

Suppose that $X$ decays exponentially, is Lipschitz continuous in projection and radially. Fix $\epsilon$, then there exists a constant $C$ such that for $n$ sufficiently large,

$$P\left( \sup_{x \in \mathbb{R}^d} \left| D_{Tuk}^{X}(x|X) - D_{Tuk}^{X}(x|X) \right| \geq \epsilon \right) \leq Cn^{\frac{3}{2}(d-1)} e^{-2n\epsilon^2}.$$
Halfspace depth: Smoothness

For a $P$ on $\mathbb{R}^d$ and $x \in \mathbb{R}^d$, the halfspace function of $P$ at $x$ is the function

$$G(r; x, X) = P(X^T r \geq x^T r) \text{ for } r \in S^{d-1}.$$ 

Let $X$ satisfy distribution smoothness condition, let

$$\arg\min_{r \in S^{d-1}} G(r; x, X)$$

exist and be unique at some $x \in \mathbb{R}$, and let

$$D_{\text{Tuk}}(x|X) = \alpha > 0,$$

then (Masse, 2004)

$$\sqrt{n}(D_{\text{Tuk}}(x|X) - D_{\text{Tuk}}(x|X)) \xrightarrow{L} \mathcal{N}(0, \alpha(1 - \alpha)).$$

$P$ has contiguous support if there is no intersection of any two halfspaces with parallel boundaries that has nonempty interior but zero probability and divides the support of $P$ into two parts.

Halfspace depth characterizes $P$, if $P$ is contiguous and all depth contours are smooth (Kong & Zuo, 2010).
Halfspace depth: Smoothness

Let $P$ be a contiguous distribution on $\mathbb{R}^d$, let $x \in \mathbb{R}^d$ such that $D^{Tuk}(x|X) = \alpha > 0$, and let $D^{Tuk}_{\alpha}(X)$ have non-empty interior. Then $G(r; x, X)$ attains more than one different global minima at $x$, i.e. for some $r_1 \neq r_2$:

$$\liminf_{r \to r_1} G(r; x, X) = \liminf_{r \to r_2} G(r; x, X) = D^{Tuk}(x|X),$$

if and only if the halfspace depth $\alpha$-contour is not smooth at $x$ (Gijbels & Nagy, 2016).

Regard the condition

“Halfspace depth contours of $P$ are smooth for all $x \in \mathbb{R}^d$”.

This condition (Gijbels & Nagy, 2016):

- holds for elliptically symmetric densities;
- does not hold for normal mixtures;
- does not hold for $L_p$-symmetric distributions;
- does not hold for centrally symmetric distribution with smooth quasi-concave density;
- does not hold for strictly quasi-concave distributions.
Connection to the densest hemisphere problem

- For $x$, $X$, denote the set of directions attaining $D^{Tuk}(x|X)$

$$R(x, X) = \arg\min_{r \in S^{d-1}} \#: \{i : x_i^T r \geq x^T r\}.$$  

- The problem of computing $D^{Tuk}(x|X)$ can be seen as invariant in more than just affine way.

- For $X$ with $x_i \neq x$ for $i = 1, \ldots, n$ define:

$$Y = \{y_i \mid y_i = \frac{x_i - x}{\|x_i - x\|}\}.$$  

- One can easily show that

$$D^{Tuk}(x|X) = D^{Tuk}(0|Y).$$

- Calculating $D^{Tuk}(0|Y)$ and picking $r \in R(0|Y)$ corresponds to finding the open densest hemisphere, i.e. an open hemisphere containing the highest portion of observations. It is of non-polynomial complexity (Johnson & Preparata, 1987).
Connection to the supervised classification

Let $X_1 = \{x_1, \ldots, x_{n_1}\}$ and $X_2 = \{x_{n_1+1}, \ldots, x_{n_1+n_2}\}$ be two training classes in $\mathbb{R}^d$.

Regarding the univariate case, if the two classes are well separated, one would expect that most of the observed differences $x_i - x_j$, $i = 1, \ldots, n_1$, $j = n_1 + 1, \ldots, n_1 + n_2$ will have the same sign (positive or negative).

Generalizing this idea to multivariate case, we define the separating hyperplane by $r \in S^{d-1}$ and search $r$ by maximizing (Ghosh & Chaudhuri, 2005):

$$U_{(n_1,n_2)}(r) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n_1+n_2} I(r^T(x_i - x_j) > 0).$$

Letting $Y = \{x_i - x_j : i = 1, \ldots, n_1, j = n_1 + 1, \ldots, n_1 + n_2\}$ this is equivalent to calculating

$$D^{Tuk}(0|Y)$$

and picking $r \in R(0|Y)$. 

The concept of data depth

Some examples
  - Mahalanobis depth
  - Spatial depth
  - Projection depth
  - Zonoid depth

Depth-based outlier detection

Halfspace depth
  - Properties
    - Trimmed regions

Classes of depth functions
  - Depths satisfying the projection property

A concept of depth for the functional framework
  - Functional outlier detection

References
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Age, in weeks

Weight, in grams
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Weight, in grams

Age, in weeks

![Graph showing babies with low birth weight]
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Age, in weeks

Weight, in grams
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Age, in weeks

Weight, in grams
Tukey (halfspace, location) depth-trimmed regions

Babies with low birth weight

Age, in weeks

Weight, in grams
Tukey (halfspace, location) depth-trimmed regions

Babies with low birth weight

![Graph showing the relationship between age and weight for babies with low birth weight. The graph includes a scatter plot with points plotted on a grid with age on the y-axis and weight on the x-axis. There are two lines indicating possible trends or models for the data.]
Tukey (halfspace, location) depth-trimmed regions

Babies with low birth weight

Weight, in grams

Age, in weeks
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Age, in weeks

Weight, in grams

800 1000 1200 1400
20 25 30 35

Babies with low birth weight
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Age, in weeks

Weight, in grams

800 1000 1200 1400
20 25 30 35

Babies with low birth weight

Weight, in grams

Age, in weeks
Tukey (=halfspace, location) depth-trimmed regions
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Weight, in grams
Age, in weeks
Tukey (halfspace, location) depth-trimmed regions
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Age, in weeks

Weight, in grams
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Weight, in grams

Age, in weeks
Tukey (halfspace, location) depth-trimmed regions

Babies with low birth weight

Weight, in grams

Age, in weeks

800 1000 1200 1400
20 25 30 35

Babies with low birth weight

800 1000 1200 1400
20 25 30 35
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Weight, in grams
Age, in weeks
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Weight, in grams

Age, in weeks
Tukey (halfspace, location) depth-trimmed regions

Babies with low birth weight

Age, in weeks

Weight, in grams
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Weight, in grams

Age, in weeks

800 1000 1200 1400

20 25 30 35
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Weight, in grams

Age, in weeks
Tukey (=halfspace, location) depth-trimmed regions
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Weight, in grams

Age, in weeks
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Age, in weeks

Weight, in grams

800 1000 1200 1400

20 25 30 35
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Weight, in grams
Age, in weeks
Tukey (=halfspace, location) depth-trimmed regions

Babies with low birth weight

Age, in weeks

Weight, in grams
Tukey (=halfspace, location) data depth
Directional quantiles

Kong, Mizera (2008)

For a random vector $X \in \mathbb{R}^d$ define the $u$-directional $\alpha$-th quantile:

$$Q(\alpha, u, X) = \inf\{x : P(u^T X \leq x) \geq \alpha\}.$$  

Directional quantile envelope:

$$R_\alpha(X) = \bigcap_{u \in S^{d-1}} H(u, Q(\alpha, u, X)),$$

where $H(u, q) = \{x : u^T x \geq q\}$ is the supporting halfspace determined by $u \in S^{d-1}$ and $q \in \mathbb{R}$.

Then for every $p \in (0; \frac{1}{2}]$ directional quantile envelopes coincide with the Tukey regions:

$$R_\alpha(X) = D^{Tuk}_\alpha(X) \text{ for every } \alpha \in (0; \frac{1}{2}].$$
Directional quantiles (illustration)
Directional quantiles (illustration)
Directional quantiles (illustration)
Directional quantiles (illustration)
Directional quantiles (illustration)
Directional quantiles (illustration)
Multiple-output regression quantiles

Hallin, Paindaveine, Šiman (2010); Paindaveine, Šiman (2011)

Regress \( Y \in \mathbb{R}^m \) on \( X = (1, W^T)^T \in \mathbb{R}^p \).

For a sample \( (X, Y) = \{(x_i, y_i) \in \mathbb{R}^p \times \mathbb{R}^m; i = 1, \ldots, n\} \), for \( \alpha \in (0, 1) \), and for \( u \in S^{m-1} \) a \((\alpha u)\)-quantile positive halfspace is any

\[
H_{\alpha u}^{(n)+} := \{(w^T, y^T)^T \in \mathbb{R}^{p-1} \times \mathbb{R}^m : \hat{b}_{\alpha u}^T y - \hat{a}_{\alpha u}^T (1, w^T)^T \geq 0\}
\]

with

\[
(\hat{a}_{\text{HPS};\alpha u}^T, \hat{b}_{\text{HPS};\alpha u}^T)^T = \arg \min \sum_{i=1}^{n} \rho_{\alpha}(b_{\text{HPS};\alpha u}^T y_i - a_{\text{HPS};\alpha u}^T x_i) \text{ subject to } u^T b = 1,
\]

or

\[
(\hat{a}_{\text{proj};\alpha u}^T, \hat{b}_{\text{proj};\alpha u}^T)^T = \arg \min \sum_{i=1}^{n} \rho_{\alpha}(b_{\text{proj};\alpha u}^T y_i - a_{\text{proj};\alpha u}^T x_i) \text{ subject to } u = b,
\]

where \( \rho_{\alpha}(x) = x(\alpha - I(x < 0)) \) is the \( \alpha \)-quantile check function.
Multiple-output regression quantiles (location)

Hallin, Paindaveine, Šiman (2010); Paindaveine, Šiman (2011)

Regress $X \in \mathbb{R}^d$ on $1 \in \mathbb{R}$.

For a sample $X = \{x_i \in \mathbb{R}^d; i = 1, ..., n\}$, for $\alpha \in (0, 1)$, and for $u \in S^{d-1}$ a $(\alpha u)$-quantile positive halfspace is any

$$H_{\alpha u}^{(n)+} := \{x^T \in \mathbb{R}^d : \hat{b}_{\alpha u}^T x - \hat{a}_{\alpha u} \geq 0\}$$

with

$$(\hat{a}_{\text{HPŠ}}; \alpha u, \hat{b}_{\text{HPŠ}}; \alpha u)^T = \arg\min \sum_{i=1}^{n} \rho_{\alpha}(b^T x_i - a) \text{ subject to } u^T b = 1,$$

or

$$(\hat{a}_{\text{proj}}; \alpha u, \hat{b}_{\text{proj}}; \alpha u)^T = \arg\min \sum_{i=1}^{n} \rho_{\alpha}(b^T x_i - a) \text{ subject to } u = b,$$

where $\rho_{\alpha}(x) = x(\alpha - I(x < 0))$ is the $\alpha$-quantile check function.

Then $D_{\alpha}^{Tuk}(X) = R_{\alpha}(X) = \bigcap_{u \in S^{d-1}} \{H_{\alpha u}^{(n)+}\}$. 
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Multiple-output regression quantiles (illustration)
Tukey (halfspace, location) depth region
Tukey (=halfspace, location) depth region: $\alpha = 2/161$
Tukey (=halfspace, location) depth region: $\alpha = \frac{5}{161}$
Tukey (halfspace, location) depth region: $\alpha = 9/161$
Tukey (=halfspace, location) depth region: $\alpha = \frac{13}{161}$
Tukey (halfspace, location) depth region: $\alpha = \frac{17}{161}$
Tukey (=halfspace, location) depth region: $\alpha = \frac{25}{161}$
Tukey (=halfspace, location) depth region: $\alpha = \frac{33}{161}$
Tukey (=halfspace, location) depth region: $\alpha = \frac{41}{161}$
Tukey (=halfspace, location) depth region: $\alpha = \frac{49}{161}$
Tukey (=halfspace, location) depth region: $\alpha = \frac{57}{161}$
Tukey (=halfspace, location) depth region: $\alpha = \frac{65}{161}$
Tukey (=halfspace, location) depth region: $\alpha = 68/161$
Contents

The concept of data depth

Some examples
  - Mahalanobis depth
  - Spatial depth
  - Projection depth
  - Zonoid depth

Depth-based outlier detection

Halfspace depth
  - Properties
  - Trimmed regions

Classes of depth functions
  - Depths satisfying the projection property

A concept of depth for the functional framework
  - Functional outlier detection

References
Projection property (Dyckerhoff, 2004)

- Let \( D \) be a depth. \( D \) satisfies the (\emph{weak}) projection property, if for each point \( x \in \mathbb{R}^d \) and each random vector \( X \) it holds:

\[
D(x|X) = \inf_{r \in S^{d-1}} D(r^T x|r^T X). 
\]

- If a depth satisfies the weak projection property, then for each direction \( r \in S^{d-1} \) we have:

\[
D(x|X) \leq D(r^T x|r^T X). 
\]

- Let \( D \) be a depth which satisfies the weak projection property. If in addition for every \( z \in \mathbb{R} \) and \( r \in S^{d-1} \) there exists a \( x \in \mathbb{R}^d \) such that \( r^T x = z \) and

\[
D(x|X) = D(r^T x|r^T X),
\]

then we say that \( D \) satisfies the \emph{strong} projection property.
Projection property (Dyckerhoff, 2004)

Let $D$ be a depth, then

1. $D$ satisfies the weak projection property if and only if for every random vector $X$ and $\alpha \geq 0$ holds
   \[
   D_{\alpha}(X) = \bigcap_{r \in S^{d-1}} r^{-1}(D_{\alpha}(r^T X)) ;
   \]

2. $D$ satisfies the strong projection property if and only if for every random vector $X$, $\alpha \geq 0$, and $r \in S^{d-1}$ holds the above equation and in addition
   \[
   r^T D_{\alpha}(X) = D_{\alpha}(r^T X) .
   \]

If a depth $D$ is quasiconcave (i.e. satisfies $D4con$), we say that $D$ is a convex depth.

Let $D$ be a depth that satisfies the weak projection property. Then $D$ is a convex depth.
Projection property (Dyckerhoff, 2004)

- For a compact convex subset $K \in \mathbb{R}^d$ the support function $h_K: \mathbb{R}^d \rightarrow \mathbb{R}$ is defined by
  $$h_K(r) = \max \{ r^T x : x \in K \}.$$

- Let $D$ be a depth. $D$ satisfies the strong projection property if and only if for every random vector $X$ and direction $r \in S^{d-1}$
  $$h_{D_\alpha}(X)(r) = h_{D_\alpha}(r^T X)(1).$$

- Let $D^1$ be a univariate depth. If $D$ is defined by
  $$D(x|X) = \inf_{r \in S^{d-1}} D^1(r^T x| r^T X),$$
  then $D$ is a multivariate convex depth that satisfies the weak projection property.

- Satisfaction of the property by depths:
  - Mahalanobis depth satisfies the strong projection property.
  - Halfspace depth satisfies the weak projection property.
  - Zonoid depth satisfies the strong projection property.
  - Projection depth satisfies the weak projection property.
Let $X$ be a stochastic process with continuous paths defined on $[0, 1]$, and $x$ its realization. Then:

\[
D(x|X) = \int_0^1 D(x(v)|X(v)) \, dv.
\]

see Fraiman, Muniz (2001); also López-Pintado, Romo (2011).
Let $X$ be a stochastic process with continuous paths defined on $[0, 1]$, and $x$ its realization. Then:

$$D(x|X) = \int_0^1 \min\{F_{X(v)}(x(v)), 1 - F_{X(v)}(x(v)^-)\} dv.$$ 

see Fraiman, Muniz (2001); also López-Pintado, Romo (2011).
Detection of functional outliers

- Functional (projection) depth of $f$ w.r.t. $\mathcal{F} = \{f_i\}_{i=1}^n$:

$$D^{prj}(f|\mathcal{F}) = \int_{t_{min}}^{t_{max}} D^{prj}(f(t)|\mathcal{F}(t)) \, dt,$$

where $D_{prj}(\cdot|\cdot)$ is the projection depth, as above.
Detection of functional outliers

- Functional (projection) depth of $f$ w.r.t. $\mathcal{F} = \{f_i\}_{i=1}^n$:

$$D^{prj}(f|\mathcal{F}) = \int_{t_{\min}}^{t_{\max}} D^{prj}(f(t)|\{f_1(t), \ldots, f_n(t)\}) \, dt,$$

where $D^{prj}(\cdot|\cdot)$ is the projection depth, as above.
Detection of functional outliers

- Functional (projection) depth of $f$ w.r.t. $\mathcal{F} = \{f_i\}_{i=1}^n$:

$$D_{proj}(f|\mathcal{F}) = \int_{t_{\text{min}}}^{t_{\text{max}}} D_{proj}(f(t)|\{f_1(t), \ldots, f_n(t)\}) \, dt,$$

where $D_{proj}(\cdot|\cdot)$ is the projection depth, as above.
Detection of functional outliers

Functional (projection) depth of $f$ w.r.t. $\mathcal{F} = \{f_i\}_{i=1}^n$:

$$D_{prj}(f|\mathcal{F}) = \int_{t_{min}}^{t_{max}} D_{prj}(f(t)|\{f_1(t), \ldots, f_n(t)\}) \, dt,$$

where $D_{prj}(\cdot|\cdot)$ is the projection depth, as above.
Detection of functional outliers

▶ Functional (projection) depth of \( f \) w.r.t. \( \mathcal{F} = \{ f_i \}_{i=1}^n \):

\[
D^{\text{prj}}(f | \mathcal{F}) = \int_{t_{\text{min}}}^{t_{\text{max}}} D^{\text{prj}}(f(t) | \{ f_1(t), \ldots, f_n(t) \}) \, dt,
\]

where \( D^{\text{prj}}(\cdot | \cdot) \) is the projection depth, as above.

▶ Label \( f \) as outlier if \( D^{\text{prj}}(f; \mathcal{F}) < \min(D^{\text{prj}}) \).
Contents

The concept of data depth

Some examples
  Mahalanobis depth
  Spatial depth
  Projection depth
  Zonoid depth

Depth-based outlier detection

Halfspace depth
  Properties
  Trimmed regions

Classes of depth functions
  Depths satisfying the projection property

A concept of depth for the functional framework
  Functional outlier detection

References
Thank you for your attention! Questions?

Thank you for your attention! Questions?

Thank you for your attention! Questions?

Thank you for your attention! Questions?