Hannah Blocher
Ludwig–Maximilians–Universität München

Depth Functions for Non-Standard Data Using Formal Concept Analysis

Kassel, December 2023

Working Group

Working Group *Foundations of Statistics and their Applications* of Prof. Dr. Thomas Augustin.
(From left to right: Dominik Kreiß (back), Hannah Blocher (front), Christoph Jansen, Thomas Augustin, Julian Rodemann, Gilbert Kiprotich, Georg Schollmeyer)

Table of Contents

Background

Depth Function

Depth Functions measure **centrality** and **outlingless** of a data point with respect to a data cloud or an underlying distribution.

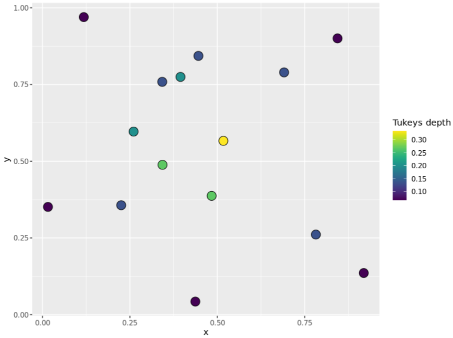


Figure: Tukey depth
15 randomly drawn points

Depth Functions measure **centrality** and **outlingless** of a data point with respect to a data cloud or an underlying distribution.

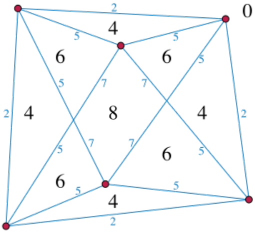


Figure: Simplicial Depth
(see `https://en.wikipedia.org/wiki/Simplicial_depth`,
visited: 20.10.23)

Depth Function: Properties

Let $\mathcal{F}$ be a set of probability measures on $\mathbb{R}^d$ with $d \in \mathbb{N}$. Let $D : \mathbb{R}^d \times \mathcal{F} \to \mathbb{R}_{\geq 0}$ be bounded, then D is[1]

1. Affine invariance: The depth function is invariant under change of the coordinate system.
2. Maximality at center: If the probability function has a unique center then the depth function has its maximum value at this center.
3. Monotonicity relative to deepest point: The depth function decreases with respect to the value with the maximal depth.
4. Vanishing at infinity: The depth function converges to zero if the norm of the point sequence converges to infinity.
5. Quasiconcavity: For every $\alpha \geq 0$ the set consisting of a depth values larger than α is a convex set.

[1] see Zou et.al. (2000) and Mosler (2013)

Depth Function: Properties

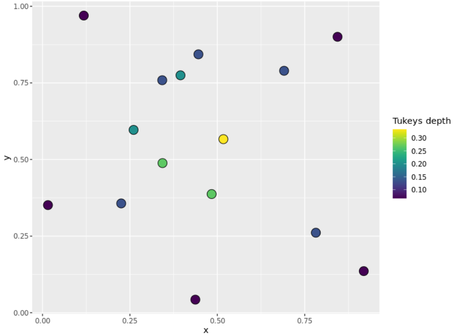


Figure: Tukey depth

Background

Non-Standard Data

Non-Standard Data

Non-Standard Data summarizes all data types that are given non standard statistical data types.
$\rightarrow$ no metric or other presupposed data structure is imposed on the observations/data.

Examples:

1. The set of partial orders (e.g. comparing ml algorithms or food)
2. Mixed (numeric $+$ nominal $+ \ldots$) data (e.g. observing spatial observations together with marks like age, education, crime, ...)

Overall Aim:

Define a Depth Function and Resulting Statistics for Non-Standard Data

$\rightarrow$ Formal Concept Analysis

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Definition (Formal Context)

A **formal context** is given by a triple $\mathbb{K} = (G, M, I)$. G corresponds to the set of **objects**, M to the set of **attributes** and I defines a binary relation between G and M.[2]

The **derivation operators**

$$\psi\colon 2^G \to 2^M, A \mapsto A' := \{m \in M \mid \forall g \in A\colon gIm\},$$
$$\varphi\colon 2^M \to 2^G, B \mapsto B' := \{g \in G \mid \forall m \in B\colon gIm\}.$$

We call the set $\varphi \circ \psi(2^G)$ the set of extents.

[2] see Ganter et.al. (2012)

This gives us

1. the **closure system** $\{A'' \mid A \subseteq G\}$ on G which describes the formal context and
2. a **family of implications** which describes the closure system completely. Let $A, B \subseteq G$. We say premise A implies conclusion B iff

   $$\psi \circ \varphi(A) \supseteq \psi \circ \varphi(B).$$

   We denote this by $A \to B$.

Summary:

$$\text{Data Set} \xrightarrow{\substack{\text{scaling}\\\text{method}}} \text{Formal Context} \xleftrightarrow{\varphi,\psi} \text{Closure System} \xleftrightarrow{\varphi,\psi} \text{Family of Implications}$$

A formal context for spatial data

The formal context $\mathbb{K}$ is given by

1. $G = \mathbb{R}^d$
2. $M = \{H \mid H \text{ is halfspace in } \mathbb{R}^d\}$
3. $I = \{(g, H) \in G \times M \mid g \in H\}$

A formal context for partial orders

Definition

Let M be a set. Then $(M, \leq)$ is a partial order if and only if for all $a, b, c \in M$

1. Reflexivity: $a \leq a$,
2. Antisymmetry: if $a \leq b$ and $b \leq a$ then a and b are the same element, and
3. Transitivity: if $a \leq b$ and $b \leq c$ then $a \leq c$

holds.

A formal context for partial orders

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Let $\mathcal{P}$ be the set of all partial orders on $\mathcal{X} = \{x_1, \dots, x_n\}$ with $n \in \mathbb{N}$.

The formal context $\mathbb{K}$ is given by

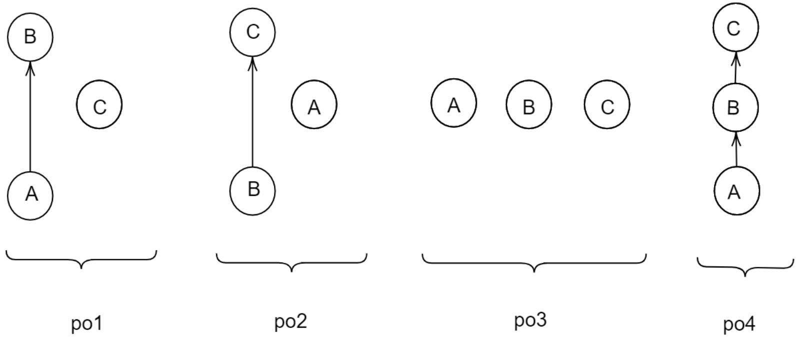
1. $G = \mathcal{P}$
2. $M = \underbrace{\{\text{“}x_i \leq x_j\text{”} \mid i, j = 1, \dots, n,\ i \neq j\}}_{=:M_{\leq}} \cup \underbrace{\{\text{“}x_i \not\leq x_j\text{”} \mid i, j = 1, \dots, n,\ i \neq j\}}_{:=M_{\not\leq}}$
3. $I = \{(g, m) \in G \times M \mid m \text{ is true for } g\}$

This corresponds to the closure operator which maps each subset $\{g_1, \dots, g_m\} \subseteq \mathcal{P} = G$, $m \in \mathbb{N}$ to

$$\{g \in \mathcal{P} \mid \cap_{i=1}^m g_i \subseteq g \subseteq \cup_{i=1}^m g_i\}.$$

A formal context for partial orders

Let $\mathcal{X} = \{A, B, C\}$. Consider the set {po1, po2} and its implications.



We obtain

1. {po1, po2} implies {po1, po2, po3}, but
2. {po1, po2} does not imply {po4}.

Adaptation to Data Represented via FCA:

What does centrality mean in this context?

Recall: Properties of the depth functions in $\mathbb{R}^d$

Let $\mathcal{F}$ be a set of probability measures on $\mathbb{R}^d$ with $d \in \mathbb{N}$. Let $D : \mathbb{R}^d \times \mathcal{F} \to \mathbb{R}_{\geq 0}$ be bounded, then D is[3]

1. Affine invariance: The depth function is invariant under change of the coordinate system.
2. Maximality at center: If the probability function has a unique center then the depth function has its maximum value at this center.
3. Monotonicity relative to deepest point: The depth function decreases with respect to the value with the maximal depth.
4. Vanishing at infinity: The depth function converges to zero if the norm of the point sequence converges to infinity.
5. Quasiconcavity: For every $\alpha \geq 0$ the set consisting of a depth values larger than α is a convex set.

[3] see Zou et.al. (2000) and Mosler (2013)

General Definition of Depth Functions on FCA

Definition

We define a depth function using formal concept analysis[4] by

$$D_G : G \times \varkappa_G \times \mathbb{P}_G \to \mathbb{R}_{\geq 0}$$

for a

1. fixed set of objects G and
2. a set of formal contexts $\varkappa_G \subseteq \{\mathbb{K} \mid G \text{ is object set of } \mathbb{K}\}$.
3. $\mathbb{P}_G$ is a set of probability measures on G defined on a σ-field which contains all extent sets of the corresponding formal contexts of $\varkappa_G$.

[4] see Blocher et.al. (2023b)

What does centrality mean in this context?

(P1) *Invariance on the extents:* Let $\mathbb{K}, \tilde{\mathbb{K}} \in \varkappa$ be two formal contexts on G and let $\Pr, \tilde{\Pr} \in \mathbb{P}$ be two probability measures on G. If there exists a bijective and bimeasureable function $i : G \to G$ such that the extents are preserved (i.e. E extent w.r.t. $\mathbb{K} \Leftrightarrow i(E)$ extent w.r.t. $\tilde{\mathbb{K}}$) and the probability is also preserved (i.e. $\Pr(E) = \tilde{\Pr}(i(E))$), then

$$D(\cdot, \mathbb{K}, \Pr) \cong D(\cdot, \tilde{\mathbb{K}}, \tilde{\Pr})$$

is true.

(P4) *Maximality:* Let $\mathbb{K} \in \varkappa, \Pr \in \mathbb{P}$. Assume there exists $g_{all} \in G$ such that for every extent E of $\mathbb{K}$ we have that $g_{all} \in E$. Then

$$D(g_{all}, \mathbb{K}, \Pr) = \max_{g \in G} D(g, \mathbb{K}, \Pr)$$

holds.

(P7ii) *Quasiconcave*: Let $\mathbb{K} \in \varkappa$ and $\Pr \in \mathbb{P}$. If for all $A \subseteq G$ and all $g \in \gamma_{\mathbb{K}}(A) \setminus A$ we have

$$D(g, \mathbb{K}, \Pr) \geq \inf_{\tilde{g} \in A} D(\tilde{g}, \mathbb{K}, \Pr),$$

we call D quasiconcave.

(P12) *Consistency*: Let $\mathbb{K} \in \varkappa$ and $\Pr \in \mathbb{P}$ be a probability measure on G. Let $\Pr^{(n)}$ be the empirical probability measure of an iid sample $g_1, \dots, g_n$ of G with $n \in \mathbb{N}$ which is drawn based on Pr. Then,

$$\sup_{g \in G} \mid D^{(n)}(g, \mathbb{K}) - D(g, \mathbb{K}, \Pr) \mid \to 0 \text{ almost surely.}$$

What does centrality mean in this context?

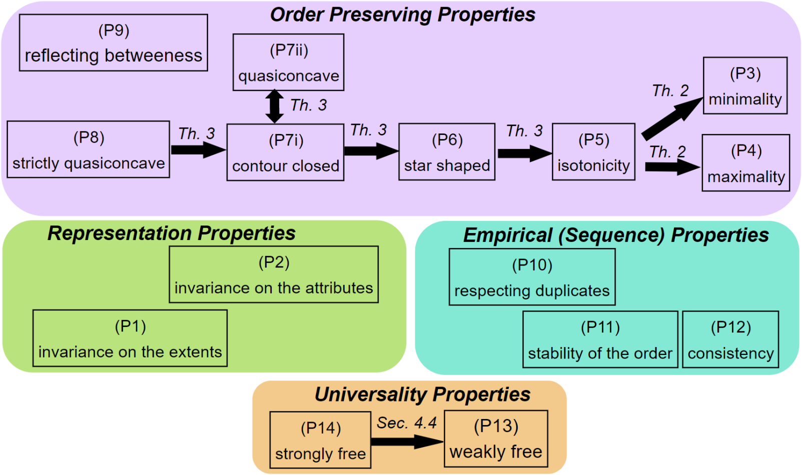


Figure: This figure can be found in Blocher et.al. (2023b).

Adaptation to Data Represented via FCA:

Concrete definition of depth functions using FCA representation?

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

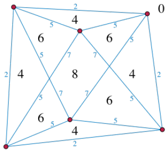
Definition

Let $(\mathbb{R}^d, \mathcal{B})$ with $d \in \mathbb{N}$ be a measurable space and let $\mathcal{F}_{\mathbb{R}^d}$ be a set of probability measures such that $(\mathbb{R}^d, \mathcal{B}, \Pr)$ defines a probability space for each $\Pr \in \mathcal{F}_{\mathbb{R}^d}$. Then the simplicial depth[5] is given by

$$D : \mathbb{R}^d \times \mathcal{F}_{\mathbb{R}^d} \to [0,1], (x, \Pr) \mapsto \Pr(x \in S^d[X_1, \ldots, X_{d+1}])$$

with

- $X_1, \ldots, X_{d+1}$ independent and identically distributed random variables from Pr, and
- $S^d[X_1, \ldots, X_{d+1}]$ being the set of points that are lie in the convex closure of $\{X_1, \ldots, X_{d+1}\}$.



[5]see Liu (1990)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Definition

The *union–free generic* family of implications, $\mathcal{U}_{\mathbb{K}}$, for a formal context $\mathbb{K}$ consists of implications $A \to B$ for which the following is true:

1. they are non–trivial (deleted implications of the form $A \to A$),
2. they have a minimal premise and a maximal conclusion (deleted implications of the form $A \to B$ if there exists $\tilde{A} \subsetneq A$ such that $\tilde{A} \to B$ or $\tilde{B} \subsetneq B$ with $A \to \tilde{B}$), and
3. cannot be constructed by union from other implications (deleted implications $A \to B$ if there is a family of implications $(A_i \to B_i)_{i \in I}$ with $A_i \subsetneq A$ for all $i \in I$ and $A = \cup_i A_i$ and $B = \cup_i B_i$ is true.[a]

[a] Compare this definition to the term proper.

↯ This family is not always sufficient to describe the corresponding closure system (e.g. $\mathcal{H} = \{A \subseteq \mathbb{N} \mid \#A < \infty\} \cup \mathbb{N}$)
↯ Is the union-free generic family of implications always unique?

Union-Free Generic Depth Function

Definition

Let

- G be a set.
- κ_G be a set of formal contexts with object set G. Moreover, for all $\mathbb{K} \in \kappa_G$ there exists a unique set of union-free generic premises $\mathcal{U}_\mathbb{K}$ that completely describes the corresponding closure operator.
- $\gamma_\mathbb{K}$ be the closure operator on G corresponding to $\mathbb{K}$.
- $\mathcal{P}_G$ gives a set of probability measures on G.

Then the union-free generic depth is defined as

$$D: \begin{array}{l} G \times \kappa_G \times \mathcal{P}_G \to \mathbb{R}_{\geq 0}, \\ (g, \mathbb{K}, P) \mapsto \sum_{j=1}^{\infty} \frac{1}{C_j} P(g \in \gamma_\mathbb{K}(X^j) \mid X^j \in \mathcal{U}_\mathbb{K}) \end{array}$$

with $X^j = \{X_1, \ldots, X_j\}$ where $X_1, \ldots, X_j \sim P$. Moreover, $C_j \in \,]0, \infty[$ for all $j \in \mathbb{N}$. We set $\sum_\emptyset = 0$ and $P(A \mid B) = 0$ for $P(B) = 0$.

Recall: Formal Context and Resulting Closure System on Partial Orders

Definition

Let $\mathcal{P}$ be the set of posets on M. We define the mapping

$$\gamma\colon \begin{array}{l} 2^{\mathcal{P}} \to 2^{\mathcal{P}} \\ P \mapsto \left\{ p \in \mathcal{P} \mid \bigcap_{\tilde{p} \in P} \tilde{p} \subseteq p \subseteq \bigcup_{\tilde{p} \in P} \tilde{p} \right\}. \end{array}$$

Example: Partial Orders

Definition

Let $\mathcal{M}$ be the set of probability measures on $\mathcal{P}$ equipped with $2^{\mathcal{P}}$ as σ-field. The *union-free generic (ufg for short) depth on posets*[6] is given by

$$D\colon \begin{array}{l} \mathcal{P} \times \mathcal{M} \to [0,1] \\ (p,\nu) \mapsto \begin{cases} 0, & \text{if for all } S \in UFG\colon \prod_{\tilde{p}\in S} \nu(\{\tilde{p}\}) = 0 \\ c \sum_{S \in UFG\colon p \in \gamma(S)} \prod_{\tilde{p}\in S} \nu\left(\{\tilde{p}\}\right), & \text{else} \end{cases} \end{array}$$

with $c = \left(\sum_{S\in UFG} \prod_{\tilde{p}\in S} \nu_n\left(\{\tilde{p}\}\right)\right)^{-1}$.

[6] see Blocher et.al. (2023a,c)

Comparison of Machine Learning Algorithms[7]

- Data Sets: 80 classification problems from OpenML.
- ML Algorithms: Random Forests (RF), Decision Tree (CART), Logistic regression (LR), L1-penalized logistic regression (Lasso) and k-nearest neighbours(KNN).
- Performance Measures: area under the curve, F-score, predictive accuracy and Brier score.

$\Rightarrow$ We obtain 80 posets

[7] see Blocher et.al. (2023a,c)

Example: Partial Orders

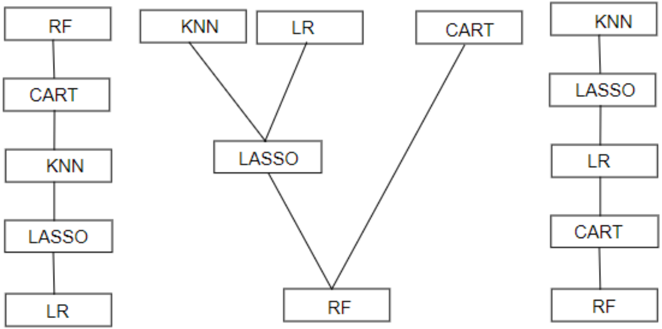


Figure: OpenML based on all four performance measures: Poset with maximal depth based on all possible posets is plotted on the left. The poset with minimal ufg depth restricted to the observed one can be seen in the middle. The poset on the right denotes the poset with minimal depth value based on all possible posets.[8]

[8] see Blocher et.al. (2023c)

Adaptation to Data Represented via FCA:

And what about inference?

Two sample test

Here, we compare the depth function evaluated based on two different (empirical) distributions: One is F and the other G.

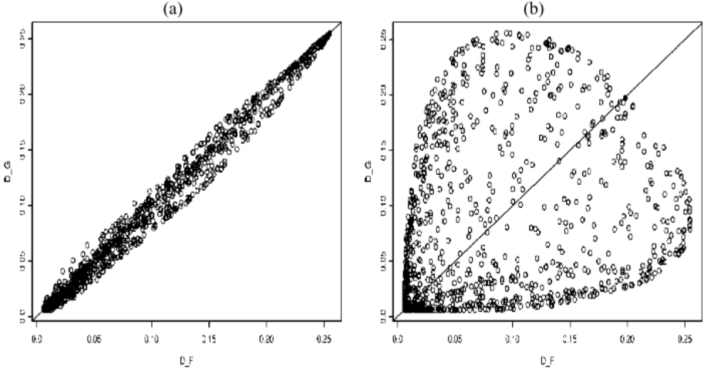


FIG. 2. *DD plots of* (a) *identical distributions and* (b) *location shift.*

Figure: This figure can be found in Li et.al. (2004).

Open Questions and Discussion

Open Questions and Discussion

- What are the conditions on a formal context such that the union-free generic family of implications is unique and sufficient to describe the corresponding closure operator?
- How to handle the difference between duplications due to sampling and duplications due to attributes of a formal context?
- How to define a one-sample test or regression?
- ...

References I

- Blocher, Hannah; Schollmeyer, Georg; Jansen, Christoph and Nalenz, Malte (2023a): Depth Functions for Partial Orders with a Descriptive Analysis of Machine Learning Algorithms. In: Proceedings of the Thirteenth International Symposium on Imprecise Probabilities: Theories and Applications (ISIPTA '23). Proceedings of Machine Learning Research, vol. 215. PMLR.
- Blocher, Hannah and Schollmeyer, Georg (2023b): Data depth functions for non-standard data by use of formal concept analysis (submitted)
- Blocher, Hannah; Schollmeyer, Georg; Nalenz, Malte and Jansen, Christoph (2023c): Comparing Machine Learning Algorithms by Union-Free Generic Depth (submitted, extenden version of Blocher et.al.(2023a))
- Ganter, Bernhard and Wille, Rudolf (2012): Formal Concept Analysis: Mathematical Foundations. Berlin, Heidelberg: Springer
- Li, Jun and Liu, Regina (2004): New Nonparametric Tests of Multivariate Locations and Scales Using Data Depth. In: Statistical Science 19, pp. 686–696.

References II

- Liu, Regina (1990): On a Notion of Data Depth Based on Random Simplices. In: The Annals of Statistics 18, pp. 405–414.
- Mosler, Karl (2013): Depth Statistics. In: Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather. Ed. by Claudia Becker, Roland Fried, and Sonja Kuhnt. Berlin, Heidelberg: Springer, pp. 17–34.
- Tukey, John (1975): Mathematics and the Picturing of Data. In: Proceedings of the International Congress of Mathematicians Vancouver. Ed. by Ralph James. Vancouver: Mathematics-Congresses, pp. 523–531.
- Zuo, Yijun and Serfling, Robert (2000): General Notions of Statistical Depth Function. In: The Annals of Statistics 28, pp. 461–482.