Data depth functions for non-standard data by use of formal concept analysis

Hannah Blocher[a,*], Georg Schollmeyer[a]

[a]*Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstr. 33, 80539 München, Germany*

Abstract

Data depth functions have been intensively studied for normed vector spaces. However, a discussion on depth functions on data where one specific data structure cannot be presupposed is lacking. In this article, we introduce a notion of depth functions for data types that are not given in statistical standard data formats and therefore we do not have one specific data structure. We call such data in general non-standard data. To achieve this, we represent the data via formal concept analysis which leads to a unified data representation. Besides introducing depth functions for non-standard data using formal concept analysis, we give a systematic basis by introducing structural properties. Furthermore, we embed the generalised Tukey depth into our concept of data depth and analyse it using the introduced structural properties. Thus, this article provides the mathematical formalisation of centrality and outlyingness for non-standard data and therefore increases the spaces centrality is currently discussed. In particular, it gives a basis to define further depth functions and statistical inference methods for non-standard data.

Keywords: conceptual scaling, depth function, formal concept analysis, generalised Tukey depth, non-standard data
2020 MSC: Primary 62H99, Secondary 62G30

1. Introduction

Data depth functions generalise the concept of *centrality* and *outlyingness* to multivariate data and provide therefore a useful concept to define *nonparametric* and *robust statistical methods*. To achieve this, depth functions denote the center and outlying areas based on an underlying distribution or a data cloud. Moreover, they classify how near other areas are to the outlying or center ones. This results in a center-outward order of the data points. Since there does not exist one unique perspective on centrality, several different depth functions on $\mathbb{R}^d$ have been developed. Some examples are simplicial depth, see [14], zonoid depth, see [9], or Tukey depth, see [5, 22]. To give a systematic basis on the notion of depth functions [23] and [15] set up properties to mathematically formalise the existing intuition about centrality and outlyingness in $\mathbb{R}^d$. In addition to $\mathbb{R}^d$, the scope of data types on which depth functions are defined increased in the last years. For example, in [11] the authors discuss properties of depth functions on functional data. In recent years, this concept has been used to construct statistical methods and to analyse different application settings. For example [13, 18] developed statistical nonparametric and robust tests using depth functions. Using the robust central-outward order [17] studied anomaly detection. In [4], data depth functions are used to visualise and describe features of hydrologic events given two real world streamflow data sets from Canada.

However, all these depth functions and analyses have in common that they build on spaces that have a strong underlying structure, like Banach spaces. Thus, to apply the concept of depth and resulting statistical methods, the data has to be embedded into statistical standard data formats, like numeric or nominal. The aim of this article is to take the next step and consider depth functions without assuming one specific data type in advance. This includes data types like, e.g., set of partial orders or mixed spatial and ordinal data. We call data that cannot be embedded into statistical standard data formats *non-standard data*. To give a notion of centrality and outlyingness for non-standard data, a unified data representation is needed. Therefore, we use the theory of *formal concept analysis* which transforms the data set into a *closure system* on the entire data set itself. This gives us a unified, flexible, and general

---

*Corresponding author. Email address: `hannah.blocher@stat.uni-muenchen.de`

applicable specification for different data structures. In particular, it does not force the user to add assumptions on the data which are not necessarily fulfilled. We use this unified representation of the data to generally define depth functions for non-standard data. To clarify the notion of centrality and outlyingness for data represented via formal concept analysis we introduce *structural properties*. Thereby, we use the structure of the closure system to transfer the properties given by [23] and [15] and develop further properties representing the connections between the data points. Finally, we embed the *generalised Tukey depth* given in [1] to our concept of depth functions for non-standard data. Moreover, we use the provided structural properties to analyse the generalised Tukey depth.

This article is intended to give a *systematic basis* on depth functions for non-standard data. By introducing structural properties we give a notion of centrality and outlyingness. Furthermore, we provide a *framework to analyse* depth functions for non-standard data and start a discussion on the notion of centrality for such kind of data.

The structure of this article is as follows: Since formal concept analysis is the basis of our depth function, we give a short introduction to the theory of formal concept analysis in Section 2. Based on this, Section 3 introduces the general definition of depth functions for non-standard data by use of formal concept analysis. There we transfer the generalised Tukey depth to our concept of depth functions. The next section states the structural properties and their restriction strength. Afterwards, we analyse the generalised Tukey depth using structural properties. In Section 6, we collect our concluding remarks.

2. Formal concept analysis

Formal concept analysis was developed by Rudolf Wille, Bernhard Ganter and Peter Burmeister to build a bridge between mathematical lattice theory[1] and applied users. It enables the analysis of relationships between the data points by representing the data in a unified and user-friendly manner. In this section, we briefly describe the aspects of formal concept analysis that are most relevant to our article. This is based on [10]. For further readings, we refer to [3].

The fundamental definition of formal concept analysis is the representation of a data set as a cross table, see [10, p. 17]

Definition 1. A *formal context* $\mathbb{K}$ is a triple (G, M, I) with G being the *object* set, M the set of *attributes* and $I \subseteq G \times M$ a binary relation between G and M.

In our case, the objects are the data points and the attributes are characteristics of these data points. The relation I then states whether an object g has an attribute m, if $(g, m) \in I$, or not, if $(g, m) \notin I$. Thus, these attributes need to be binary-valued, whether they occur or not. Naturally, there exist characteristics of the data points which are many-valued, like sex or age. To include these many-valued characteristics as well into the formal context, we use so-called *scaling methods*, see [10, p. 36ff], which transfers many-valued characteristics into a set of binary-valued attributes.

Example 1. Consider three data points p_1, p_2 and p_3 which have one numeric and one nominal observation. For example, let $p_1, p_2, p_3 \in [0, 1] \times \{a, b\}$ be the data set with $p_1 = (0, a)$, $p_2 = (0.5, b)$ and $p_3 = (0.7, a)$. Now, to obtain a formal context, there exist several different scaling methods. Here, we use the *interordinal scaling* for the numeric part and the *nominal scaling* for the nominal part, see [10, p. 42]. With this, we get the formal context given in Table 1.

	"≤ 0"	"≤ 0.5"	"≤ 0.7"	"≥ 0.7"	"≥ 0.5"	"≥ 0"	a	b	c
p_1	×	×	×			×	×		
p_2		×	×		×	×			×
p_3			×	×	×	×	×		

Table 1: Formal context: Example for interordinal and nominal scaling.

We want to point out that when referring back to this example in the following section, then $G = \{p_1, p_2, p_3\}$ is the entire underlying space and not just a sample of a larger set.

---

[1] In the context of order theory and not group theory.

2

By using scaling methods, we can represent a large variety of different data sets through a formal context. This allows us to transfer the most diverse data types into a uniform structure. Also, data sets which are not given in standard statistical data formats. We call such data *non-standard data*. An example of non-standard data is given by [1]. Here, we show how a set of partial orders can be represented by a formal context. Further examples of scaling methods can be seen in Examples 3 and 4. The scaling method can be combined with the aim to reduce data complexity. Note that this can lead to a conceptual scaling error, see, e.g., [12].

Based on this user-friendly representation of the data set by a formal context, we can now define so-called *derivation operators*, see [10, p. 18]:

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

The function Ψ maps a set of objects A onto every attribute which every object in A has. So, $A' = \Psi(A)$ is the maximal set of attributes that every object in A has. The reverse, from attribute set to object set, is provided by the function Φ. The composition of these two functions $\gamma := \Phi \circ \Psi : 2^G \to 2^G$ gives us then a family of sets which denotes the relationship between the data points. More precisely, every object set $A \subseteq G$ which is an element of the codomain of $\gamma = \Phi \circ \Psi$ is the maximal set of objects which have all the same attributes $\Psi(A)$ in common. Thus, the composition groups all those objects together that have the same attributes. By considering all possible different attribute combinations one can get the entire set $\gamma(2^G)$.

Definition 2. The set $\gamma(2^G)$ is called the set of *extents*, see [10, p. 18].

To take advantage of this slightly different representation of the data set as a family of sets, we use that γ defines a *closure operator*[2], see [10, p. 8].

Definition 3. A closure operator $\gamma : 2^G \to 2^G$ is defined as a function on a power set to itself. A closure operator needs to be extensive (i.e. for all $A \subseteq G$, $A \subseteq \gamma(A)$), isotone (i.e. if $A \subseteq B \subseteq G$, then $\gamma(A) \subseteq \gamma(B)$) and idempotent (i.e. for all $A \subseteq G$, $\gamma(A) = \gamma(\gamma(A))$).

In particular, a closure operator always induces a *closure system* $\gamma(2^G)$. A closure system $\mathcal{S} \subseteq 2^G$ is a family of sets which contains the entire space (i.e. $G \in \mathcal{S}$) and any intersection of sets in $\mathcal{S}$ is again in $\mathcal{S}$ (for all $S \subseteq \mathcal{S}$ with $S \neq \emptyset$ we have $\bigcap_{s \in S} s \in \mathcal{S}$).

Note that there exists a one-to-one correspondence between the closure system and the closure operator. Since $\gamma = \Phi \circ \Psi$ is a closure operator, the set of extents is a closures system. Thus, the closure operator γ describes the closure system and reverse. For more details on closure systems see [10, Chapter 0.3].

Example 2. Recall Example 1. We obtain $\Psi(\{p_1\}) = \{\text{"} \leq 0\text{"}, \text{"} \leq 0.5\text{"}, \text{"} \leq 0.7\text{"}, \text{"} \geq 0\text{"}, a\}$ and therefore $\gamma(\{p_1\}) = \Phi \circ \Psi(\{p_1\}) = \{p_1\}$[3]. One can show that the set of extents $\gamma(2^{\{p_1, p_2, p_3\}})$ equals the power set of $\{p_1, p_2, p_3\}$.

Let us take a closer look at how the closure operator describes the connection between the data points. This is the basic idea of how we will later use the closure operator to define structural properties for depth functions. As we pointed out above the closure operator groups data points together which have all the same attributes in common. This means if $a \in \gamma(A) \setminus A$ the object a has all attributes which every object in A has as well. Thus, one can say that *A implies a* based on the relationship structure given by the formal context which is then included in the definition of γ. Therefore using the closure system or closure operator (both describe the same since they have an one-to-one correspondence) to define the structural properties of the depth function illustrates the relationship between the data points. For further details on implications see [10, Chapter 2.3]. Note that in [10] attribute implications are discussed and we focus here on object implications. Nevertheless, the concepts can be transferred to object implications.

---

[2]In what follows, we use the term *closure* always in the context of a closure operator or a closure system. When referring to a closed set based on a topology, metric, or norm we denote this by *topological(-ly) closed/closure*.

[3]For simplicity, we denote $\Phi(m)$ instead of $\Phi(\{m\})$ for $m \in M$ in the following. The same holds for $\Psi(g)$ and $\gamma(g)$ for $g \in G$.

Example 3. Now, we introduce a further scaling method, the so-called *hierarchical nominal scaling*. This scaling method is inspired by the occupations of persons within a social survey. Usually occupations[4] are categorised within a hierarchy of different levels. On a first level occupations are split into different categories $(a_1, b_1, ...)$. For example category a_1 could be "Managers", category b_1 "Professionals", etc. Each of these categories is then split again on a more fine-grained level (Level 2) into further subcategories. In this case, category a_1 is split into a_1a_2 "Managers: Chief executives, senior officials, and legislators" and a_1b_2 "Managers: Administrative and commercial managers" and so on. Note that the Level 2 subcategories based on the first level b_1 split do not have to match those of the Level 2 splits based on the Level 1 a_1 split. Subsequently, the Level 2 categories are again subdivided into subcategories, and so on. Such data structure can be conceptually scaled naturally. For every level, we introduce attributes describing every single category based on the upper level classification, see Table 2. Here, for simplicity, we used only two levels with two categories, respectively.

	a_1	b_1	a_1a_2	a_1b_2	b_1a_2	b_1b_2
a_1a_2	x		x			
a_1b_2	x			x		
b_1a_2		x			x	
b_1b_2		x				x

Table 2: Minimal example of a hierarchical nominal scaling with two levels (1, 2) and two categories (*a*,*b*, respectively).

Now, let us take a closer look at the extents given by Table 2 which are

$$\{\emptyset, \{a_1a_2\}, \{a_1b_2\}, \{b_1a_2\}, \{b_1b_2\}, \{a_1a_2, a_1b_2\}, \{b_1a_2, b_1b_2\}, G\}.$$

Furthermore, we obtain that $a_1a_2 \in \gamma(\{a_1b_2, b_1a_2\})$.

Example 4. In the next sections, we want to transfer the idea of depth functions from $\mathbb{R}^d$ to general non-standard data which are represented by a formal context. Before looking at this, we now discuss the reverse and show how one can represent the elements in $\mathbb{R}^d$ as objects of a formal context. We consider $\mathbb{R}^d$ together with the topology induced by the Euclidean norm. The scaling method introduced here is inspired by [19, 20]. Let $G = \mathbb{R}^d$ be the object set and the attribute set $M = \{H \subseteq \mathbb{R}^d \mid H \text{ topologically closed halfspace}\}$ is the set of all topologically closed halfspaces. We say that there exists a relation $(g, H) \in I$ between an object/point $g \in G = \mathbb{R}^d$ and an attribute/halfspace $H \in M$ if and only if $g \in H$. Then $\mathbb{K}_{\mathbb{R}^d} = (G, M, I)$ gives us a formal context.

In the next step, let us consider the set $\gamma(2^G)$ with γ being induced by the formal context $\mathbb{K}_{\mathbb{R}^d}$. Let $A \subseteq G = \mathbb{R}^d$. Then $\Psi(A)$ are all halfspaces which contain every object/point in A. Further on, $\gamma(A) = \Phi \circ \Psi(A)$ are then every object/point which lies in every halfspace in $\Psi(A)$. Thus $\gamma(A)$ is the intersection of all halfspaces in $\Psi(A)$ and therefore a topologically closed convex set[5]. More generally, one can show that for every topologically closed convex set in $2^{\mathbb{R}^d}$ there exists a set $A \subseteq G = \mathbb{R}^d$ such that this convex set is given by $\gamma(A)$. Thus, γ is the convex closure operator on $\mathbb{R}^d$ and the extent set $\gamma(2^G)$ is the set of all topologically closed convex sets.

More concretely, assume that $d = 2$ and consider Figure 1. Then $d \in \gamma(\{a, b, c\})$. This means that d lies in every topologically closed halfspace which contains also a, b, and c. In other words, d shares the same attributes as a, b, c share. Thus, one can say that d is implied by a, b and c based on γ. Since $e \notin \gamma(\{a, b, c\})$, there exists a halfspace which contains a, b and c but not e. This shows how the concrete definition of the closure system enhances the connection between single objects.

---

[4]For example within the International Standard Classification of Occupations (ISCO) from 2008, see `https://www.ilo.org/public/english/bureau/stat/isco/index.htm` (accessed: Aug 10th, 2023).

[5]A set $C \subseteq \mathbb{R}^d$ is convex if and only if for all $c, g \in C$ and for every $t \in [0, 1]$, $tc + (1 - t)g \in C$ is true (i.e. the line segment between c and g is also in C).

3. Definition of depth functions for non-standard data using formal concept analysis

Our aim in this section is to give a general definition of data depth functions for non-standard data types using formal concept analysis. By representing the data points G via a formal context, with G being the object set, we obtain a unified structure that is not tailored to one specific data type. In particular, the newly provided depth function allows to analyse a large variety of different data types on centrality and oultyingness issues. With this, nonparametric methods can be developed for all these data.

Fig. 1: Convex set.

The depth function presented here only specifies the domain and codomain but not the exact mapping rule. Thus, the structural properties presented later, see Section 4, can be seen as generic properties for this kind of depth functions.
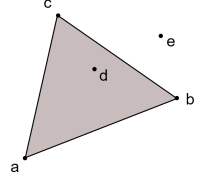
Definition 4. We define a depth function using formal concept analysis by

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

for a fixed set of objects G and a set of formal contexts $\varkappa_G \subseteq \{\mathbb{K} \mid G \text{ is object set of } \mathbb{K}\}$. $\mathscr{P}_G$ is a set of probability measures on G defined on a σ-field which contains all extent sets of the corresponding formal contexts of $\varkappa_G$[6] .

Thus, we compute the depth of an object set based on a probability measure and formal context representing the object relationships. We want to emphasise that G, $\varkappa_G$ and $\mathscr{P}_G$ mutually depend on each other. Sometimes the restriction of $\varkappa_G$ and $\mathscr{P}_G$ to a subset of all possible formal contexts or probability measures is necessary. For example, assume that $G = [0, 1]$, then for every subset of G there exists a formal context such that this subset is an extent. Thus, in this case, a restriction to a subset of contexts is necessary if we want to allow the uniform distribution to be an element of $\mathscr{P}_G$. This follows from [21, Chapter 1.1] which shows that there cannot exist a probability measure on $[0, 1]$ which formalises the intuition of volume and has the entire power set of $[0, 1]$ as input. Another aspect is that restriction can lead to a set of formal contexts fulfilling additional structural requirements. With this, it can be possible to define mapping rules which are not possible in general. (See Section 4, Property (P8) does not hold for every formal context, but only for a subset.) Another example is given in [1] and [2] where we considered one single formal context on the set of partial orders. There, we used the structure given by this concrete formal context to define the mapping rule. The same reasoning can be applied to the probability set $\mathscr{P}_G$. Thus, for a proper definition of the depth function, not only the exact mapping rule is important, but the considered formal contexts and probability measures as well.

A data set can be represented by different attribute sets and corresponding binary relations. Thus, Definition 4 can lead to many different depth functions even if the object set G, the probability $\Pr \in \mathscr{P}_G$, and a concrete mapping rule are specified. Hence, the choice of formal context and scaling method can have a huge impact on the depth values, i.e. this can also be seen in Section 4.

The empirical depth function corresponds to the depth function in Definition 4 with the empirical probability measure as input. To ensure that the empirical depth function is well defined we need to assume that every empirical probability measure $\Pr_G^{(n)}$ of every probability measure $\Pr_G \in \mathscr{P}_G$ is again an element of $\mathscr{P}_G$.

Definition 5. Let G be a set and $\varkappa_G \subseteq \{\mathbb{K} \mid G \text{ is object set of } \mathbb{K}\}$ a set of formal contexts on G. We assume that $\mathscr{P}_G$ consists of probability measures that are defined upon a σ-field containing all extents of $\varkappa_G$. Furthermore, we assume that for every sample $g_1, \ldots, g_n$ based on a probability measure $\Pr_G \in \mathscr{P}_G$, we get for the empirical probability measure $\Pr_G^{(n)}$ that $\Pr_G^{(n)} \in \mathscr{P}_G$ is true. Then the empirical depth function for a sample $g_1, \ldots, g_n$ with corresponding empirical probability measure $\Pr_G^{(n)}$ is given by

$$D_G^{(n)} : G \times \varkappa_G \to \mathbb{R}_{\geq 0}, (g, \mathbb{K}) \mapsto D_G(g, \mathbb{K}, \Pr_G^{(n)}).$$

Serving as an example, we consider the generalised Tukey depth, based on [19, 20] and introduced in [1]. The Tukey depth on $\mathbb{R}^d$ of a point $g \in \mathbb{R}^d$, compare to [5, 22], is the smallest probability of a halfspace containing g. To

---

[6]Note that since every probability measure must be compatible with every formal context, this is a slight limitation in the definition. Another formulation could use a subset of $\varkappa_G \times \mathscr{P}_G$ where each pair in the subset must satisfy the measurability condition and not every possible combination of $\varkappa_G$ and $\mathscr{P}_G$.

build the bridge to formal concept analysis, we consider the formal context $\mathbb{K}$ with the object set $G = \mathbb{R}^d$, attribute set $M = \{H \subseteq \mathbb{R}^d \mid H \text{ topologically closed halfspace}\}$ and binary relation I with $(g, H) \in I$ if and only if $g \in H$, see Example 4. Based on this the Tukey depth for a point $g \in \mathbb{R}^d$ and probability measure $\Pr_{\mathbb{R}^d}$ on $\mathbb{R}^d$ can be written as

$$\inf_{H \in \mathcal{H}(g)} \Pr_{\mathbb{R}^d}(H) = \inf_{H \in \Psi(g)} \Pr_{\mathbb{R}^d}(H) = 1 - \sup_{H \in \Psi(g)} \Pr_{\mathbb{R}^d}(G \setminus H) = 1 - \sup_{H \in M \setminus \Psi(g)} \Pr_{\mathbb{R}^d}(H) = 1 - \sup_{m \in M \setminus \Psi(g)} \Pr_{\mathbb{R}^d}(\Phi(m)) \quad (1)$$

where $\mathcal{H}(g)$ is the set of all topologically closed halfspaces containing g and Φ, Ψ correspond to the derivation operators given by the formal context $\mathbb{K}_{\mathbb{R}^d}$, see Section 2. The first equality translates the term to formal concept analysis terminology. The third equality holds because $G \setminus H$ is a topologically open halfspace and the supremum does not change when considering topologically closed halfspaces instead. The last equality is again a further translation into formal concept analysis that does not rely on the notion of a halfspace.

The right hand side of Equation (1) will now be the basis of the generalisation of Tukey depth to arbitrary formal contexts. Before we proceed, we shortly indicate, why we do not use directly the left hand side of Equation (1): Generally, the topologically closed convex sets in $\mathbb{R}^d$ correspond to the extents within our approach to use formal concept analysis to define data depth functions. On the other hand, the topologically closed halfspaces of $\mathbb{R}^d$ have no general natural equivalent in formal concept analysis[7]. On the left hand side, if we replace the family of topologically closed halfspaces with the family of all extents, we obtain a depth value of zero for all $g \in \mathbb{R}^d$ when, e.g., the probability measure is continuous w.r.t. the Lebesgue measure. This is of course unsatisfying. Moreover, note that unlike halfspaces, the complement of extents in formal concept analysis are generally not extents. This further separates the left and the right hand side of Equation (1). Therefore, we take the right hand side of Equation (1). As will be shown later, see Section 5, if we take the supremum over all halfspaces or if we take the supremum over all topologically closed convex sets in $\mathbb{R}^d$ does not change the result. This can also be applied to the case of the generalised Tukey depth.

Additionally, the supremum of the right hand side of Equation (1) has a natural interpretation as a measure of oulyingness that can be expressed in the language of formal concept analysis, see [19, Section 2] and [20, Section 5]. In these articles, the author constructs special representative extents[8]: One can say that an extent is very general if it contains at least a certain amount α of data points or probability mass. Of course, for given α there are many such extents but one can take for one α the intersection of all these extents. This intersection is again an extent which is then in a certain sense a representative extent w.r.t. a level α. The outlyingness of a point g is then given by the (empirical) probability mass of the most specific depth contour (i.e., the most specific representative extent that corresponds to the smallest possible α) that still contains g. A slightly different, but order-theoretically equivalent, definition is to take the least specific depth contour that does not contain g. This is exactly what is done by $\sup_{m \in M \setminus \Psi(g)} \Pr_{\mathbb{R}^d}(\Phi(m))$ in Equation (1). With this motivation, we get as generalised Tukey depth:

Definition 6. Let G be a set, $\varkappa_G$ a subset of formal contexts and $\mathscr{P}_G$ a subset of probability measures on G. Assume that $\varkappa_G$ and $\mathscr{P}_G$ are defined as in Definition 4. The generalised Tukey depth is given by

$$T_G : \begin{array}{l} G \times \varkappa_G \times \mathscr{P}_G \to [0, 1], \\ (g, \mathbb{K}, \Pr_G) \mapsto 1 - \sup_{m \in M \setminus \Psi(g)} \Pr_G(\Phi(m)) \end{array}$$

where $\mathbb{K}$ defines the operator Φ and Ψ. We set $\sup_{t \in \emptyset} f(t) := 0$ for every function f.

Note that $\varkappa_G$ is not restricted to any subset and, in particular, $\Pr_G$ is only restricted by $\varkappa_G$, see Definition 4. The second part of the mapping rule in Definition 6, $\sup_{m \in M \setminus \Psi(g)} \Pr_G(\Phi(m))$, corresponds to the supremum of the probabilities of the events which consists of all objects having an attribute the object of interest does not have. Thus, if $g \in G$ is an object that has all the attributes which occur, then g has a maximal depth of value one.

Now, we define the empirical version. Analogously to Definition 5, let $g_1, \ldots, g_n$ be a sample of G based on $\Pr_G \in \mathscr{P}_G$. Then the empirical Tukey depth function corresponds to T_G by inserting the corresponding empirical probability measure $\Pr_G^{(n)}$. This gives us:

---

[7]Of course, one could characterise the topologically closed halfspaces as the formal concept extents that are generated by single attributes, but a kind of generalisation that uses this idea would come into trouble w.r.t. Property (P1) (invariance on the extent) and also with Property (P7i) (contourclosedness) later on, see Section 4.

[8]In [20] these extents are treated as abstract elements in a complete lattice and there they are termed quantiles.

Definition 7. Let G, $\varkappa_G$ and $\mathscr{P}_G$ be defined as in Definition 5. Then the empirical generalised Tukey depth is

$$T_G^{(n)} : \begin{array}{l} G \times \varkappa_G \to \mathbb{R}_{\geq 0}, \\ (g, \mathbb{K}) \mapsto 1 - \sup_{m \in M \setminus \Psi(g)} \sum_{\tilde{g} \in \Phi(m)} \Pr_G^{(n)}(\tilde{g}). \end{array}$$

Example 5. Recall Example 1 and 2. Let us assume that the probability measure $\Pr_G$ on $G = \{p_1, p_2, p_3\}$ is given by $\Pr_G(p_1) = \Pr_G(p_2) = 1/4$ and $\Pr_G(p_3) = 1/2$. Then the generalised Tukey depth of p_1 and p_3 are

$$T_G^{(n)}(p_1) = 1 - \sup_{m \in M \setminus \Psi(p_1)} \Pr_G(\Phi(m)) = 1 - \sup_{m \in \{\text{``} \geq 0.7 \text{''}, \text{``} \geq 0.5 \text{''}, b, c\}} \Pr_G(\Phi(m)) = \frac{1}{4},$$

$$T_G^{(n)}(p_3) = 1 - \sup_{m \in M \setminus \Psi(p_3)} \Pr_G(\Phi(m)) = 1 - \sup_{m \in \{\text{``} \leq 0 \text{''}, \text{``} \leq 0.5 \text{''}, b, c\}} \Pr_G(\Phi(m)) = \frac{1}{2}.$$

Before presenting the structural properties, we define when two depth functions are isomorph and thus represent the same center-outward order.

Definition 8. Let G be a set and D_G and $\tilde{D}_G$ be two depth functions based on $\varkappa_G$ and $\mathscr{P}_G$, $\tilde{\varkappa}_G$ and $\tilde{\mathscr{P}}_G$ respectively, see Definition 4. Let $\mathbb{K} \in \varkappa_G$, $\tilde{\mathbb{K}} \in \tilde{\varkappa}_G$, $\Pr_G \in \mathscr{P}_G$ and $\tilde{\Pr} \in \tilde{\mathscr{P}}_G$. Then $D_G(\cdot, \mathbb{K}, \Pr_G)$ and $\tilde{D}_G(\cdot, \tilde{\mathbb{K}}, \tilde{\Pr}_G)$ on G are isomorph if and only if there exists a bijective and bimeasureable function[9] $i : G \to G$ such that

$$D_G(g, \mathbb{K}, \Pr_G) \leq D_G(\tilde{g}, \mathbb{K}, \Pr_G) \iff \tilde{D}_G(i(g), \tilde{\mathbb{K}}, \tilde{\Pr}_G) \leq \tilde{D}_G(i(\tilde{g}), \tilde{\mathbb{K}}, \tilde{\Pr}_G)$$

is true for all $g, \tilde{g} \in G$. In what follows, $\cong$ denotes the isomorphism between two depth functions.

For simplicity of notation, we write D, $D^{(n)}$, T, $T^{(n)}$, $\varkappa$, $\mathscr{P}$, $\Pr$ and $\Pr^{(n)}$ instead of D_G, $D_G^{(n)}$, T_G, $T_G^{(n)}$, $\varkappa_G$, $\mathscr{P}_G$, $\Pr_G$ and $\Pr_G^{(n)}$ in the following if the underlying object set G is clear.

4. Structural properties characterising the depth function using formal concept analysis

After this general definition of depth functions based on formal concept analysis, we want to discuss some *structural properties* a depth function can have. With this, we tackle the question of what centrality is and which object is - in some sense - *closer* to the center than another object. In particular, we provide concepts to discuss centrality and outlyingness for different data types without necessarily presupposing one specific data structure. Thus, this section gives a starting point for a discussion on centrality, outlyingness, and depth functions based on formal concept analysis. Furthermore, we define a framework upon which newly introduced depth functions can be studied and compared.

For normed vector spaces, there is already an ongoing discussion about this. The authors of [14–16, 23] are concerned about these questions for data depth functions defined on $\mathbb{R}^d$. Furthermore, [7, 11] discuss depth functions and centrality topics for functional data. All these examples have in common that they are based on a normed vector space and that the clarification of centrality and outlyingness is done by defining properties. For example, [15, 23] demand that the depth of a point $x \in \mathbb{R}^d$ should converge to zero as the norm of x, i.e. $\|x\|$, tends to infinity. This reflects the intuition about outlyingness in unbounded spaces. In contrast to outlyingness, the definition of a center point does not immediately follow. Basically, every point in $\mathbb{R}^d$ can be the center. This follows from the fact that after translation, rotation, etc. the structure of a normed vector space does not change. The center generally seems to be only naturally specified in special cases. For example [23] assumes that for a probability measure that is symmetric around a point c, see [24], this point c should then be the center. In the case of a multivariate normal distribution, the center point is the mean vector. Since a depth function should represent a center-outward order, this point c needs to have a maximal depth value. Together with the center, one needs to discuss what it means that one point is further away from the center than another. In $\mathbb{R}^d$, this is achieved by the use of line segments, see [15, 23]. A point p_1 which

---

[9] Let $G, \tilde{G}$ be two sets with σ-fields $\mathscr{A}$ and $\tilde{\mathscr{A}}$ respectfully. A bijective function $i : G \to \tilde{G}$ is called bimeasurable if and only if i and i^{-1} are measurable.
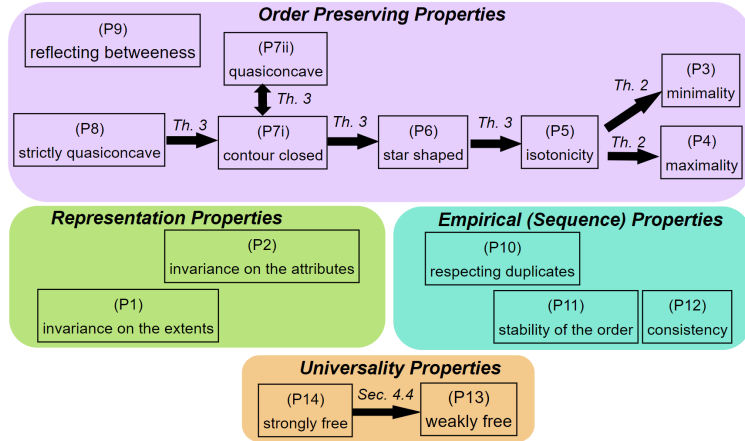
Fig. 2: Overview of the structural properties together with their mathematical connections. *Th* is Theorem and *Sec* is Section for short. This overview contains all direct implications between the properties. Further connections like, e.g., that Properties (P8) and (P9) imply one concrete center-outward order in special cases, see, e.g., Theorem 5, or under which assumptions certain properties are never or always true, see, e.g., Theorem 6, are not presented in the Figure.

lies on the line segment between the center and a further point p_2 is said to be closer to the center than p_2. In other words, p_2 is more outlying than p_1. At the first glance, here we also use the normed vector space structure.

It follows from the above that the definition of the terms centrality and outlyingness seem to highly rely on the underlying distribution. In particular, the part defining the center point which is based on a symmetric distribution, see [24]. Slightly different to [23, 24] is the approach of [15]. Here the notion of a depth function is based more directly on the structure of the normed vector space. The properties given by [15] put emphasis on the structure of the underlying spaces and not on the probability measure. Thus, in contrast to e.g., [23, 24], the definition of centrality of a point is more detached from the notion of a point of symmetry/centrality of the underlying probability measure. Note that [15] gives an even stronger definition of depth functions. In the following, we discuss the term centrality and outlyingness from the perspective of the approach in [15].

In this article, we consider a space G where the structure of the data points is given by a formal context. In the style of [23] and [15] for depth functions in $\mathbb{R}^d$, we define *structural properties* of a depth function based on formal concept analysis to clarify the notion of centrality and outlyingness. These properties express characteristics of the data set when represented via a formal context and how these characteristics are included in the depth function. Some of the structural properties build upon the existing ones in $\mathbb{R}^d$ that we transfer to our new situation. To achieve this, we use that the set of all topologically closed convex sets is a natural closure system on $\mathbb{R}^d$. Since the extent set also defines a closure system, we obtain a direct translation to formal concept analysis by representing the properties by use of the convex closure system on $\mathbb{R}^d$. For example, we transfer the idea of a line segment to our data representation via a formal context, see Property (P6). Moreover, the "quasiconcavity" property, see [15, p. 19] has a natural translation to formal concept analysis by use of the closure system, see Property (P7i and ii). Other introduced structural properties use formal concept analysis directly. An example is the maximality property (P4). Here, we say that an object which lies in every set of the closure system needs to have maximal depth value. While in $\mathbb{R}^d$ the closure system and the translated closure system are equal, this cannot be said in general for all closure systems. In particular, the concept of translation cannot be properly defined for every space G. Instead, we use that in some cases the data itself contains natural center or outward lying points, see Properties (P3) and (P4) later. All in all, one can say that the introduced structural properties are defined and discussed within two perspectives: Transferring already existing properties in $\mathbb{R}^d$ and new development of properties based on the theory of formal concept analysis.

In total 14 structural properties are presented which can be covered under four different categories: Representation properties, order preserving properties, empirical (sequence) properties, and universality properties. An overview of the structural properties together with their mathematical connections can be seen in Figure 2. In what follows, we fix the set of objects G and consider the depth function $D : G \times \varkappa \times \mathscr{P} \to \mathbb{R}_{\geq 0}$. Furthermore, we refer to Definition 4

when we say that *D* is a depth function. We use the term empirical depth function when considering a depth function in the style of Definition 5. Afterwards, in Section 5 we check these properties based on the introduced generalised Tukey depth, see Section 3.

4.1. Representation properties

Depth functions satisfying the following properties are structure preserving on *G*. This includes two parts: First, assume that we represent the data set by two formal contexts and that we consider two probability measures. Let us assume that between the corresponding extent sets and therefore the structure of the objects exists a bijective function. Furthermore, if additionally the probability values are preserved by this bijective function, then the depth function should be preserved as well. In other words, the order given by the depth function should not rely on the used scaling method unless it does substantially change the structure of the extent sets. Also, the influence of the probability measure is only based on the extent sets. This can be seen as an adaptation of the "affine invariance" property of depth functions in $\mathbb{R}^d$, see [23, p. 463]. There, the depth function equals the depth of the shifted version if the probability measure is shifted accordingly.

The second part of the representation properties considers the attribute values. Here, we say that the depth function preserves the structure if and only if two objects with the same attributes have the same depth value. From the perspective of formal concept analysis, two objects with the same attributes are duplicates and therefore they should be assigned to the same depth value.

(P1) *Invariance on the extents:* Let $\mathbb{K}, \tilde{\mathbb{K}} \in \varkappa$ be two formal contexts on *G* and let $\mathrm{Pr}, \tilde{\mathrm{Pr}} \in \mathscr{P}$ 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. $\mathrm{Pr}(E) = \tilde{\mathrm{Pr}}(i(E))$), then

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

is true.

(P2) *Invariance on the attributes:* For every $\mathbb{K} \in \varkappa$, $\mathrm{Pr} \in \mathscr{P}$ and $g_1, g_2 \in G$ with $\Psi_{\mathbb{K}}(\{g_1\}) = \Psi_{\mathbb{K}}(\{g_2\})$,

$$D(g_1, \mathbb{K}, \mathrm{Pr}) = D(g_2, \mathbb{K}, \mathrm{Pr})$$

holds.

4.2. Order preserving properties

While the representation properties ensure that similar structures on *G* lead to the same depth function, the next *order preserving properties* consider the obtained order by the depth function. Ignoring the last property in this section, these properties increase in their strength of restriction.

These properties are defined along the lines of "monotonicity relative to the deepest point" and "maximality at the center" properties, see [23, p. 463], and the "quasiconcavity" properties, see [15, p. 19], defined for $\mathbb{R}^d$. In contrast to $\mathbb{R}^d$ we neither have a norm nor a concept of translation and symmetry, but we can make use of the closure system and closure operator given by the extent set. When considering the properties defined for $\mathbb{R}^d$ in the context of the convex sets which define a closure system together with the corresponding closure operator, we can build a bridge to formal concept analysis. In particular, the "quasiconcavity" property, see [15, p. 19], has therefore a natural adaptation. Nevertheless, the convex sets are a special case of closure sets: For example, the affine invariance is reflected in the set since by shifting the set of convex sets the family of sets does not change. This does not hold in general for closure systems. For example, there exist closure systems where one point occurs more often in the sets of the closure system than another point. Another aspect is that we do not necessarily have infinitely many objects/data points which is needed for the "vanishing at infinity" property for $\mathbb{R}^d$, see [23, p. 464], where we say that the depth converges to zero for the norm of a point converging to infinity. Thus, this property cannot be transferred to our new situation. Instead, we use the concept of a formal context where two natural extreme opposite characteristics of the objects can occur. The first one: If an object has every attribute then it lies in every extent set. Conversely, if an object has no attribute at all, then the only extent containing this object is the entire set *G*. Property (P3) and (P4) now ensure that these two opposite characteristics are also reflected in the depth function.

(P3) *Minimality:* Let $\mathbb{K} \in \varkappa, \mathrm{Pr} \in \mathscr{P}$. Further, let $g_{non} \in G$ such that for every extent $E \subsetneq G$ of $\mathbb{K}$ we have that $g_{non} \notin E$, then

$$D(g_{non}, \mathbb{K}, \mathrm{Pr}) = \min_{g \in G} D(g, \mathbb{K}, \mathrm{Pr})$$

is true.

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

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

holds.

Note that a depth function which fulfils these two properties does not rely on the probability measure to set the values of the two extreme cases. On the other hand, there exist formal contexts such that the objects g_{non} or g_{all} do not exist at all. Recall Example 4 where we considered the spatial data.

Nevertheless, since we have now predefined the maximal depth value in specific cases, this has to be in line with adapting properties like "monotone on rays" or "quasiconcavity", see [15, p. 19]. In what follows we start with less restricting properties and increase their restriction strength slowly. We show that they imply Properties (P4) and (P3).

Property (P5) is inspired by the fact that in formal concept analysis, an object g_1 which lies in the closure of an other object g_2 implies that this object is more specific than g_2. In other words, g_2 has all attributes and possibly even more attributes than g_1. This is analogous to the assumption that $\gamma(\{g_1\}) \supseteq \gamma(\{g_2\})$ is true. Thus, Property (P5) says that an object g_2 must have a depth value as least as high as g_1.

(P5) *Isotonicity*: For every $\mathrm{Pr} \in \mathscr{P}$ and formal context $\mathbb{K} \in \varkappa$ with $g_1, g_2 \in G$ such that $\gamma_{\mathbb{K}}(\{g_1\}) \supseteq \gamma_{\mathbb{K}}(\{g_2\})$,

$$D(g_1, \mathbb{K}, \mathrm{Pr}) \leq D(g_2, \mathbb{K}, \mathrm{Pr})$$

is fulfilled.

There exists a natural strengthening of the isotonicity property (P5) which leads to the adaptation of the "monotonicity relative to deepest point", see [23, p. 463] property in $\mathbb{R}^d$. For start, let us assume that the depth function is bounded from above. Furthermore, we assume that the depth function has its maximum at center $c \in G$. The "monotone on rays" in $\mathbb{R}^d$ definition in [15, p. 19] states that the depth of a point that moves further away from the center c on a fixed ray should decrease. Since we do neither have a norm nor a vector space, we cannot generally define what *further away* as well as *ray* means. Thus, we translate the definition to a setting using the convex closure operator instead. In this case, if a point/object $\tilde{g}$ lies on the line segment given by a further point g and the center c, then the depth of $\tilde{g}$ must be at least as high as the depth of g. In other words, when a point $\tilde{g}$ lies in the convex closure of another point g and the center, then we have a lower bound for the depth of $\tilde{g}$. Thus, the points/objects with depth values larger or equal to a fixed value form a starshaped set[10]. This gives the name of the property. With the definition based on the convex closure system, we can easily transfer this property to formal concept analysis and obtain Property (P6).

(P6) *Starshapedness*: Let $\mathbb{K} \in \varkappa$ and $\mathrm{Pr} \in \mathscr{P}$. If there exists at least one center point $c \in G$ such that for all $g \in G$ and all $\tilde{g} \in \gamma_{\mathbb{K}}(\{c, g\})$ we have

$$D(\tilde{g}, \mathbb{K}, \mathrm{Pr}) \geq D(g, \mathbb{K}, \mathrm{Pr}),$$

then we call D starshaped.

In the above explanation, we assumed that the center point c has a maximal depth value. This assumption is not included in the definition of Property (P6) as the boundedness of the depth function, as well as the maximality at the center, follow directly. Let us fix $c \in G$ to be one center point. Since for every closure operator the isotonicity assumption (see Section 2, not to be confused with Property (P5)) holds, we get that for every object g we have $c \in \gamma(\{c, g\})$. Thus, by Property (P6) for every $g \in G$ we obtain

$$D(c, \mathbb{K}, \mathrm{Pr}) \geq D(g, \mathbb{K}, \mathrm{Pr}).$$

---

[10]A set $S \subseteq \mathbb{R}^d$ is starshaped if and only if there exists a center $c \in S$ such that for all $g \in S$ and all $t \in [0, 1]$, $tc + (1 - t)g \in S$ is true.

With this, c must have the highest depth value. Since the depth function maps to $\mathbb{R}$, this gives us also the upper bound of the depth function.

Furthermore, the starshaped property (P6) together with the invariance on the extents property (P1) has some implications on the point of maximal depth when the underlying probability measure has some symmetry property. As already indicated at the beginning of Section 4, because of the lack of a translation operation, etc., it is difficult to define symmetry in our setting. However, one can still define some notion of point symmetry, see the next theorem.

Theorem 1. *Let $\mathbb{K} \in \varkappa$ and $\Pr \in \mathscr{P}$ (both based on set G). We assume that $\Pr$ is point symmetric around $s \in G$ in the following sense: There exists a bimeasurable involutory[11] function $i : G \longrightarrow G$ such that*

1. *$\Pr(i(E)) = \Pr(E)$ for all extents $E \subseteq G$ and*
2. *for all $g \in G$ we have $s \in \gamma(\{g, i(g)\})$.*

Then for every depth function which fulfils Properties (P1) and (P6), the center of symmetry s of $\Pr$ has maximal depth.

Proof. Let c be one center point. First note that by Assumption 2. we get $s \in \gamma(\{c, i(c)\})$. Because D is starshaped, we have $D(s, \mathbb{K}, \Pr) \geq D(i(c), \mathbb{K}, \Pr)$. Now, we use that there exists a bimeasurable involutory function i and that D is invariant on the extents (P1) to get $D(i(c), \mathbb{K}, \Pr) = D(c, \mathbb{K}, \Pr)$. With this, we conclude that $D(s, \mathbb{K}, \Pr) \geq D(c, \mathbb{K}, Pr)$. This proves that s has maximal depth. □

Note that object s of Theorem 1 is not necessarily a center point in the style of Property (P6).

These properties are written down in their order of strength. More precisely, Property (P6) implies Property (P5) and Property (P5) implies Properties (P4) and (P3), see Theorem 2. Thus, if Property (P6) is satisfied and if there exists an object g which has every attribute, then c must be one of the center point c discussed in Property (P6).

Theorem 2. *Let $\mathbb{K} \in \varkappa$ and $\Pr \in \mathscr{P}$. Let D be a depth function then the following implications hold for D:*

1. *Let $c \in G$ be a center object. If D satisfies Property (P6), then Property (P5) is true for D.*
2. *If D satisfies Property (P5), then Properties (P4) and (P3) are true for D.*

Proof. Assume that $\mathbb{K} \in \varkappa$ and $\Pr \in \mathscr{P}$. Let D be a depth function.

We begin by proving that Property (P6) implies Property (P5). Therefore assume that Property (P6) is true for D and let $c \in G$ be one center point. Further let $g_1, g_2 \in G$ such that $\gamma_{\mathbb{K}}(\{g_1\}) \supseteq \gamma_{\mathbb{K}}(\{g_2\})$. Then we get that $g_2 \in \gamma(\{c, g_1\})$ and therefore we have $D(g_1, \mathbb{K}, \Pr) \leq D(g_2, \mathbb{K}, \Pr)$.

The next step is to show that Property (P5) implies Property (P4). Assume that Property (P5) is satisfied and that g_{all} lies in every extent set. Then, we get for every $g \in G$ that $\gamma_{\mathbb{K}}(\{g_{all}\}) \supseteq \gamma_{\mathbb{K}}(\{g\})$ holds. Thus, $D(g_{all}, \mathbb{K}, \Pr) \geq \max_{\{g \in G\}} D(g, \mathbb{K}, \Pr)$ is true and Property (P4) follows.

Finally, we show that Property (P3) follows from Property (P5). Let $g_{non} \in G$ be an object which lies only in the entire set and in no other extent set. Furthermore, we suppose that Property (P5) is true for D. Since $\gamma(\{g_{non}\}) = G$ for every $g \in G$ we have $\gamma(\{g\}) \subseteq \gamma(\{g_{non}\})$. Due to Property (P5) we follow that $D(g_{non}, \mathbb{K}, \Pr) \leq D(g, \mathbb{K}, \Pr)$ for every $g \in G$. This gives us Property (P3). □

Remark 1. Let us point out some consequences of Property (P6). In contrast to [23, p. 463] we do not assume that the center c is unique. For example, the depth function is allowed to have a plateau at the highest point. In particular, we allow the depth function to be constant. Especially g_{non} can only be the center point if the function is constant, due to Properties (P3), (P4) and Theorem 2. Moreover, we get that when the depth function has at least two different values, then g_{non} must have the minimal value and g_{all} the maximal value. This allows us to specify a center point and an outlying point without relying on the other observed points/objects by defining the scaling method. Note that this situation does not occur often. For example, all scaling methods described in Section 2 do not have an object g_{non} and g_{all}[12].

---

[11]A function $i : G \longrightarrow G$ is involutory if and only if $i(i(g)) = g$ for all $g \in G$

[12]An example for a formal context which contains g_{non} and g_{all} can be constructed by applying the ordinal scaling for numeric variables, see [10, p.42].

Nevertheless, this stresses the importance of a meaningful and carefully chosen scaling method. Observe that the difference between having an attribute and not having an attribute is not symmetric and cannot be switched without eventually fundamentally changing the characteristic of the corresponding closure system.

The next order preserving properties are in the style of the "quasiconcavity" property, see, e.g., [15, p. 19]. In the context of $\mathbb{R}^d$ quasiconcavity states that the set of all points with a larger (or equal) depth value than $\alpha \geq 0$ needs to be a convex set. These sets are called contour sets[13]. In this case, there is a direct transfer to formal concept analysis by the extent set as a closure system. Therefore, we first have to define the contour sets within the theory of formal concept analysis. Let $\mathbb{K} \in \varkappa$ be a formal context and $\Pr \in \mathscr{P}$ be a probability measure. For $\alpha \in \mathrm{im}(D(\cdot, \mathbb{K}, \Pr))$ the contour set $Cont_\alpha$ is defined as follows

$$Cont_\alpha := \{g \in G \mid D(g, \mathbb{K}, \Pr) \geq \alpha\}.$$

Now, we say the depth function is quasiconcave if every contour set is an extent set. This is stated in Property (P7i). Instead of considering the contour sets, an analogous statement for Property (P7i) is given in Property (P7ii). Here, we assure that the depth of an observation that lies in the closure of an input set is larger or equal to the infimum of the input. Property (P7ii) has a natural strengthening by assuming strict inequalities instead, see Property (P8).

(P7i) *Countourclosed*: For every formal context $\mathbb{K} \in \varkappa$, probability measure $\Pr \in \mathscr{P}$ and every $\alpha \in \mathrm{im}(D(\cdot, \mathbb{K}, \Pr))$ the contour set $Cont_\alpha$ is an extent of the context $\mathbb{K} \in C$.

(P7ii) *Quasiconcave*: Let $\mathbb{K} \in \varkappa$ and $\Pr \in \mathscr{P}$. 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.

(P8) *Strictly quasiconcave*: Let $\mathbb{K} \in \varkappa$ and $\Pr \in \mathscr{P}$. If for all $A \subseteq G$ and all $g \in \gamma_\mathbb{K}(A) \setminus A$ we have

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

D is strictly quasiconcave.

Again these properties have mathematical connections to each other. First of all, Property (P7i) and (P7ii) are equivalent. Secondly, Property (P8) is indeed stronger and implies Property (P7 i and ii). Thirdly, if c is a center point with maximal depth, then with this c we can show that (P7i and ii) imply the starshaped property (P6). Moreover, Property (P7i) implies Property (P5).

Theorem 3. *Let $\mathbb{K} \in \varkappa$ be a formal context and $\Pr \in \mathscr{P}$ a probability measure. Let $D(\cdot, \mathbb{K}, \Pr)$ be a depth function.*

1. *Statement (P7i) and (P7ii) are equivalent.*
2. *Property (P8) implies (P7ii).*
3. *If their exists $c \in G$ with maximal depth value, then Property (P7ii) implies (P6) with c being a center object.*
4. *Property (P7ii) implies (P5).*

Proof. Let $\mathbb{K} \in \varkappa$ and $\Pr \in \mathscr{P}$. Let $D(\cdot, \mathbb{K}, \Pr)$ a depth function. Note that the claim that Property (P8) implies (P7ii) is given immediately. Thus, we only show Part 1., 3. and 4. of Theorem 3.

1. First assume that Property (P7ii) is true and let $\alpha \in \mathrm{im}(D(\cdot, \mathbb{K}, \Pr))$ be arbitrary. We prove that Property (P7i) follows. Assume, by contradiction that $Cont_\alpha$ is not an extent set. Since γ is a closure operator and therefore idempotent, there exists $g \in \gamma(Cont_\alpha) \setminus Cont_\alpha$. Since Property (P7ii) is true, $D(g, \mathbb{K}, \Pr) \geq \inf_{\tilde{g} \in Cont_\alpha} D(\tilde{g}, \mathbb{K}, \Pr) \geq \alpha$ holds. This contradicts $g \notin Cont_\alpha$ and we get that Property (P7i) is fulfilled.
   For the reverse let (P7i) be true and let $A \subseteq G$ be arbitrary. We set $\alpha = \inf_{\tilde{g} \in A} D(\tilde{g}, \mathbb{K}, \Pr)$ and we know that $A \subseteq Cont_\alpha$. By (P7i) we know that $Cont_\alpha$ is an extent set. Since $\gamma(A)$ is the smallest extent set containing A and the set of extents is a closure system, we follow that $\gamma(A) \subseteq Cont_\alpha$. Thus the depth of every object $g \in \gamma(A)$ must be larger or equal to α which implies (P7ii).

---

[13]Note that the term used is in line with [5] and not [15] where there are called upper level sets. The contour sets correspond there to boundaries. [8] calls them alpha-trimmed regions.

3. Now, we assume that Property (P7ii) is true and we show that Property (P6) holds as well. Therefore, we assume that $c \in G$ has maximal depth value. Let $g, \tilde{g} \in G$ such that $\tilde{g} \in \gamma(\{c, g\})$ is true. Due to Property (P7ii) we get $D(\tilde{g}, \mathbb{K}, \mathrm{Pr}) \geq \min\{D(c, \mathbb{K}, \mathrm{Pr}), D(g, \mathbb{K}, \mathrm{Pr})\} \geq D(g, \mathbb{K}, \mathrm{Pr})$. The last inequality follows from the assumption that c has maximal depth value. This gives us Property (P6) with c being one center object.

4. Finally, assume that Property (P7ii) holds. Let $g_1, g_2 \in G$ with $g_1 \neq g_2$ such that $\gamma(g_2) \supseteq \gamma(g_1)$. Then the quasiconcavity property implies that $D(g_1, \mathbb{K}, \mathrm{Pr}) \geq D(g_2, \mathbb{K}, \mathrm{Pr})$. This shows Property (P5).

$\square$

Property (P8), the strictly quasiconcavity, is indeed a very strong assumption. In particular, there exist formal contexts such that Property (P8) can never be fulfilled. An example can be found in [1]. There, we discussed the special case of depth functions for partial orders and analysed Properties (P7i and ii) and (P8) in this case. We showed that the quasiconcavity cannot be fulfilled by the formal context of partial orders. Another example is given by Table 3. Here, let $G = \{g_1, g_2, g_3\}$ and let Pr be an arbitrary probability measure. In contradiction, let us assume that Property (P8) is true for a depth function D on G. Since $\gamma(g_1) = \{g_1, g_2, g_3\} = \gamma(g_2)$ we get $D(g_1, \mathbb{K}, \mathrm{Pr}) < D(g_2, \mathbb{K}, \mathrm{Pr})$ and $D(g_2, \mathbb{K}, \mathrm{Pr}) < D(g_1, \mathbb{K}, \mathrm{Pr})$ must hold at once. Since this is not possible, Property (P8) cannot hold for this context. In contrast, the quasiconcavity property (P7ii) allows equality which solves such problems. Note that in the case of the formal context given by Table 3, Property (P7ii) leads to a constant depth function.

	m_1	m_2	m_3
g_1	×		
g_2		×	
g_3			×

Table 3: Formal context: Cannot fulfil Property (P8).

More generally, one can say that the strictly quasiconcavity assumption on D can never be satisfied if one formal context $\mathbb{K} \in \varkappa$ is an element of the following set:

$$\mathscr{C}^{p8} = \left\{ (G, M, I) \text{ formal context} \;\middle|\; \begin{array}{l} \text{exist } A, \tilde{A} \subseteq G \text{ such that } \#A < \infty, \#\tilde{A} < \infty \text{ and } A \cap \tilde{A} = \emptyset \\ \text{and } A \subseteq \gamma(\tilde{A}), \; \tilde{A} \subseteq \gamma(A) \end{array} \right\}.$$

Theorem 4 proves that for every $\mathbb{K} \in \mathscr{C}^{p8}$ there exists no strictly quasiconcave depth function. Here, we get a contradiction with the strict larger assumption. This case occurs naturally when the scaling method assigns to two different objects the same attribute values.

Theorem 4. *For every $\mathbb{K} \in \mathscr{C}^{p8}$ and every $\mathrm{Pr} \in \mathscr{P}$ there exists no depth function D such that Property (P8) is fulfilled.*

Proof. Let $\mathbb{K} \in \mathscr{C}^{p8}$. We assume that there exists a depth function D and a probability measure Pr such that Property (P8) is satisfied. Since $A \subseteq \gamma(\tilde{A})$ and $\tilde{A} \subseteq \gamma(A)$ we get

$$\forall a \in A \setminus \tilde{A}: D(a, \mathbb{K}, \mathrm{Pr}) > \inf_{\tilde{a} \in \tilde{A}} D(\tilde{a}, \mathbb{K}, \mathrm{Pr}) \qquad \text{and} \qquad \forall \tilde{a} \in \tilde{A} \setminus A: \inf_{a \in A} D(a, \mathbb{K}, \mathrm{Pr}) < D(\tilde{a}, \mathbb{K}, \mathrm{Pr}).$$

Furthermore, we assumed that $\#A < \infty, \#\tilde{A} < \infty$ is true. Thus, the infimum is attained in A and $\tilde{A}$. Let $a_m \in A$ be an argument of the minimum of $D(a, \mathbb{K}, \mathrm{Pr})$, and analogously we set $\tilde{a}_m \in \tilde{A}$. Since $A \cap \tilde{A} = \emptyset$, we obtain that $\tilde{a}_m \in \tilde{A} \setminus A$ and $a_m \notin A \setminus \tilde{A}$. But with this we get

$$D(a_m, \mathbb{K}, \mathrm{Pr}) > D(\tilde{a}_m, \mathbb{K}, \mathrm{Pr}) \qquad \text{and} \qquad D(a_m, \mathbb{K}, \mathrm{Pr}) < D(\tilde{a}_m, \mathbb{K}, \mathrm{Pr}).$$

This cannot be true which contradicts the assumption of Property (P8) being true. $\square$

We want to point out that A and $\tilde{A}$ being finite is crucial since in the proof we used that the infimum is attained and not only the largest lower bound of the sets. Secondly, the intersection of A and $\tilde{A}$ being empty is also necessary since else one can set the elements in the intersection to the minimal depth value. This is then following Property (P8).

While the importance of a meaningful scaling method for Properties (P3) and (P4) is easily seen, this comes also into account for the last properties. For example Property (P7ii) and (P8) state that the depth of an object g which is implied by a set A, i.e. $g \in \gamma(A)$, must have larger (or equal) depth than the minimal depth value of the elements in A. In the context of formal concept analysis, one can say that g contains all characteristics/similarities having the objects in A in common or even more. Therefore it is at least as specific (or more) than all the elements in A together.

Thus, when defining the scaling method, not only the individual attribute values should be taken into account, but also which object combination is more specific than others.

The next property can be seen as the inverse property of Property (P8). It states that a strict inequality between depth values should be rooted in some sense in the structure of the underlying closure system which is given by the underlying formal context. More precisely, the idea behind Property (P9) is that if two objects g and $\tilde{g}$ differ in their depth, then there exists a set that separates these two objects. Assume that g has a strictly higher depth value than $\tilde{g}$. Then there exists a set A with $g \notin A$ and $g \in \gamma(A)$ such that every object in A has higher or equal depth to $\tilde{g}$. Thus, the set A, which is not necessarily an extent set, divides the two objects. Note that $\tilde{g}$ can be in A.

(P9) *Reflecting betweeness*: Let $\mathbb{K} \in \varkappa$ and $\mathrm{Pr} \in \mathscr{P}$. If for $g, \tilde{g} \in G$ with $D(g, \mathbb{K}, \mathrm{Pr}) > D(\tilde{g}, \mathbb{K}, \mathrm{Pr})$ there exists a set $A \subseteq G$ with

$$
\begin{aligned}
&i) \quad \forall a \in A\colon D(a, \mathbb{K}, \mathrm{Pr}) \geq D(\tilde{g}, \mathbb{K}, \mathrm{Pr}), \\
&ii) \quad g \in \gamma(A) \setminus A \ \text{ and} \\
&iii) \quad \tilde{g} \notin \gamma(A) \setminus A,
\end{aligned}
$$

then D fulfils the reflecting betweeness property.

Note that Condition *iii*) of Property (P9) can be easily fulfilled by adding $\tilde{g}$ to A. Property (P9) and (P8) together are very restrictive. Indeed when G is finite and the formal context satisfies the *meet distributivity assumption*, then every depth function which fulfils both properties gives the same center-outward order structure of the objects. Thus, all depth functions are isomorph to each other. In particular, this order does not rely on the probability measure Pr.

Definition 9. Let $\mathbb{K} = (G, M, I)$ be a formal context. Let E be an extent set of $\mathbb{K}$. We call an object $g \in G$ an *extreme point* of E (given by $\mathbb{K}$) if and only if

$$
g \in E \quad \text{and} \quad g \notin \gamma\left(E \setminus \{\tilde{g} \mid \Psi(\tilde{g}) = \Psi(g)\}\right).
$$

We denote the set of all extreme points of an extent E by $\mathcal{E}_E$.

A formal context is *meet distributive* if and only if every extent is the closure of its extreme points.

An example for a meet distributive formal context is the spatial formal context given in Example 4 where G is only a finite subset of $\mathbb{R}^d$. Here the extreme objects correspond to the extreme points of a convex set. Note that in the definition of the extreme points duplicates are deleted in the sense that duplication does not change the character of being an extreme point[14]. Let us assume that the formal context $\mathbb{K}$ has no duplicates. Then we immediately obtain for every extent E that there does not exist any further set $A \subseteq E$ with $A \cap \mathcal{E}_E \neq \mathcal{E}_E$ such that $\gamma(A) = E$. Thus, the extreme points are not only sufficient to imply E but also necessary.

Theorem 5. *Let G be finite and let $\mathbb{K} \in \varkappa$ be a meet distributive formal context with object set G. We assume that the context has no duplicates (i.e. $\nexists g, \tilde{g} \in G$ with $g \neq \tilde{g}$ and $\Psi(g) = \Psi(\tilde{g})$). Let D and $\tilde{D}$ be two depth functions on G and $\mathbb{K}$ which fulfil Property (P8) and (P9). Then for every probability measure $\mathrm{Pr} \in \mathscr{P}$ the orders given by D and $\tilde{D}$ are identical.*

Proof. We show this claim in two steps (see below):

- Step 1: Since Property (P7) is true, we conclude with the use of Property (P9) that all extreme points of $Cont_\alpha$ with $\alpha \in im(D(G, \mathbb{K}, \mathrm{Pr}))$ must have the same depth value for some depth function D.

- Step 2: We use Property (P8), G being finite and the extents are given by the extreme points to prove the claim.

---

[14][10, Theorem 4.4] shows that the above definition is indeed equivalent to being a meet distributive formal context.

Step 1: Let D be a depth function which fulfils Properties (P8) and (P9). For every $\alpha \in im(D(G, \mathbb{K}, \Pr))$ the extreme points $\mathcal{E}_\alpha$ of $Cont_\alpha$ have the same depth value. In contradiction assume that there exists an $\alpha \in im(D(G, \mathbb{K}, \Pr))$ and $g, \tilde{g} \in \mathcal{E}_\alpha$ such that

$$D(g, \mathbb{K}, \Pr) > D(\tilde{g}, \mathbb{K}, \Pr).$$

Since Property (P9) is true, there exists a set $A \subseteq G$ such that for all $a \in A$ we have $D(a, \mathbb{K}, \Pr) \geq D(\tilde{g}, \mathbb{K}, \Pr)$. Furthermore, we get that $g \in \gamma(A)$ and $g \notin A$. By definition of the contour set, we obtain $A \subseteq Cont_\alpha$ and, in particular, that $\gamma(A \cup (\mathcal{E}_\alpha \setminus \{g\})) = Cont_\alpha$. Since $g \notin A \cup (\mathcal{E}_\alpha \setminus \{g\})$ although g is an extreme point, this is a contradiction to the meet distributivity assumption.

Step 2: We now use Step 1 to prove the theorem. Let us consider a contour set $Cont_\alpha \subseteq G$ with $\alpha \in im(D(G, \mathbb{K}, \Pr))$. By Step 1 we get that all objects in $\mathcal{E}_\alpha$ have the same depth value. More precisely, the depth value α. Using Property (P8) we obtain that every object in $Cont_\alpha \setminus \mathcal{E}_\alpha$ must have strictly larger depth. Furthermore, since G is finite, there exists an $\tilde{\alpha} \in im(D(G, \mathbb{K}, \Pr))$ such that there exists no further $\tilde{\tilde{\alpha}} \in im(D(G, \mathbb{K}, \Pr))$ with $\tilde{\alpha} > \tilde{\tilde{\alpha}} > \alpha$. Due to Theorem 3 and Property (P7i), we know that $Cont_{\tilde{\alpha}} = Cont_\alpha \setminus \mathcal{E}_\alpha$ is again a contour set. Again the extreme points of $Cont_{\tilde{\alpha}}$ must have depth value $\tilde{\alpha}$. We can repeat this procedure with $Cont_{\tilde{\alpha}}$.

Now, let's assume that D and $\tilde{D}$ are two depth functions that both are strictly quasiconcave. We start with the extent set G. Then for both depth functions the extreme points of the extent set G have the lowest depth value $\alpha_{D,1}$ and $\alpha_{\tilde{D},1}$. Since the extreme points are unique we get

$$Cont_{\alpha_{D,1}} \setminus \{g \mid D(g, \mathbb{K}, \Pr) = \alpha_{D,1}\} = Cont_{\alpha_{\tilde{D},1}} \setminus \{g \mid \tilde{D}(g, \mathbb{K}, \Pr) = \alpha_{\tilde{D},1}\} := Cont^{(2)}.$$

We continue with $Cont^{(2)}$. Again, the extreme points of $Cont^{(2)}$ are unique and must have the same depth value in both depth functions. Furthermore, both depth functions assign the extreme points the second lowest depth value. Analogously, we delete the extreme points of $Cont^{(2)}$ and obtain the same contour set $Cont^{(3)}$ for both depth functions. Repeat this procedure until $Cont^{(i)}$ is the empty set. Since G is finite, this can be achieved in a finite number of steps. This shows that the order given by D and $\tilde{D}$ are identical. $\qquad\square$

An example which proves that G being finite in Theorem 5 is necessary is given by Figure 3. Here $G = [0, 1]$ and $\mathbb{K}$ is given by the interordinal scaling, see Section 2, with

$$G = [0, 1],$$
$$M = \{\text{``} \geq a \text{''} \mid a \in [0, 1]\} \cup \{\text{``} \leq a \text{''} \mid a \in [0, 1]\},$$
$$I = \{(g, \text{``} \geq a \text{''}) \mid g \geq a\} \cup \{(g, \text{``} \leq a \text{''}) \mid g \leq a\}.$$

For $\mathbb{K} = (G, M, I)$, the set of extents equals the topologically closed intervals in $[0, 1]$. Then both plots in Figure 3 represent a depth function based on $\mathbb{K}$ and both functions fulfil Properties (P9) and (P8), but the respective center-outward orders differ.

Finally, we want to point out that for every formal context $\mathbb{K}$ where every object has a duplicate, every quasiconcave depth function fulfils Property (P9). If every object has a duplicate, then we get that for every extent E and every $g \in E$ there exists a set $A \subseteq E \setminus \{g\}$ with $\gamma(A) = E$. Such formal contexts are summarised by the following set

Fig. 3: The functions represent two depth functions which are strictly quasiconcave and reflecting betweeness for $G = [0, 1]$.

$$\mathscr{C}^{P9} = \left\{ (G, M, I) \text{ formal context} \; \middle| \; \begin{array}{l} \text{for every extent } E \text{ of } (G, M, I) \text{ and for every } g \in E \\ \text{we have } \gamma(E \setminus \{g\}) = E \end{array} \right\}.$$

Theorem 6. *Let $\mathbb{K} \in \mathscr{C}^{P9}$. Then for every probability measure $\Pr$ on G a quasiconcave (P7ii) depth function fulfils the reflecting betweeness property (P9).*
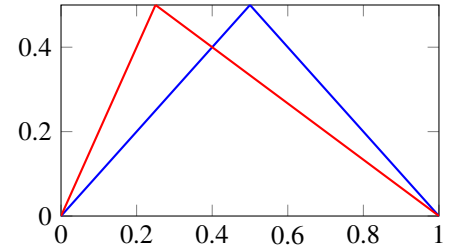
15

Proof. Let $\mathbb{K} \in \mathscr{C}^{P9}$, $\Pr \in \mathscr{P}$ and let $g_1, g_2 \in G$ with $D(g_1, \mathbb{K}, \Pr) > D(g_2, \mathbb{K}, \Pr)$. Consider the extent set $E = \gamma(\{g_1\})$. By assumption there exists a set $\tilde{A} \subseteq E$ such that $g_1 \notin \tilde{A}$ and $\gamma(\tilde{A}) = E \ni g_1$. Since D is quasiconcave, every element of $\tilde{A}$ must have at least the depth value of $D(g_1, \mathbb{K}, \Pr)$. This follows from $\tilde{A} \subseteq \gamma(\{g_1\})$. Then $A = \tilde{A} \cup \{g_2\}$ fulfils the necessary conditions in Property (P9). $\qquad\square$

All these order preserving properties seem to not rely on the probability measure $\Pr \in \mathscr{P}$. This is in the style of [15]. There the author defined a depth function as being a function that satisfies certain characteristics based on a Banach space. He showed that this implies some properties which depend on the probability measure. More precisely, he showed that these properties imply that the depth function takes its maximum at the center of a symmetric probability measure. Such considerations can be also done for the properties introduced in this section, see, e.g., Theorem 1. In the next section, we focus on the set of probability measures $\mathscr{P}$ directly.

4.3. Empirical (sequence) properties

In this section, we consider the reverse of the above section. For a fixed formal context $\mathbb{K}$ we are interested in the behaviour of the depth function when the (empirical) probability measure changes. In what follows, we regard different empirical probability measures $\Pr^{(n)}$ induced by different samples $g_1, \ldots, g_n \in G$. We assume that every sample $g_1, \ldots g_n$ is independent and identical distributed (iid) as a probability measure $\Pr \in \mathscr{P}$ with $n \in \mathbb{N}$. Moreover, let $\Pr^{(n)}$ be the corresponding empirical probability measure and assume that $\Pr^{(n)} \in \mathscr{P}$ holds. Hence, in this section, we consider the empirical depth function, recall Definition 5.

The first two properties discuss how two empirical depth functions differ if the two corresponding iid samples differ in a specific manner. The first one considers the influence of duplicates in the sample in comparison to deleting the duplication. In this case, the depth value of the duplicated object should be higher based on the sample where the duplication exists. The second property studies the impact of one single sample element on the resulting center-outward order. Here we consider two empirical probability measures. The first empirical probability measure is based on a sample that has an object which greatly differs from the other object. The second empirical probability measure is based on the same sample but without this greatly different object. An object g_{diff} differs greatly from the other objects if this object has no attribute that any other object has. Or, equivalently, g_{diff} and a further object of the sample g_s are both elements of an extent E if and only if $E = G$ is true. Thus, this object g_{diff} should have no impact on the center-outward order of the other objects.

(P10) *Respecting duplications*: Let $\mathbb{K} \in \varkappa$. Let $g_1, \ldots, g_n$ be an iid sample of G with $n \in \mathbb{N}$. Assume that there exist g_i and g_j in the sample with $i \neq j$ which have identical attribute set (so $\Psi(g_i) = \Psi(g_j)$). We set

- $Pr_G^{(n)}$ to be the empirical probability measure of $g_1, \ldots, g_n$ with $\Pr^{(n)} \in \mathscr{P}$, and

- $\Pr^{(n-1)}$ to be the empirical probability measure $g_1, \ldots, g_{i-1}, g_{i+1}, \ldots, g_{n-1}$ (without g_i) and $\Pr^{(n-1)} \in \mathscr{P}$.

Then

$$D^{(n-1)}(g_j, \mathbb{K}) < D^{(n)}(g_j, \mathbb{K}).$$

(P11) *Stability of the order*: Let $\mathbb{K} \in \varkappa$ and let $g_1, \ldots, g_n$ be an iid sample of G with $n \in \mathbb{N}$. Assume that there exists an $i \in \{1, \ldots, n\}$ such that g_i is greatly different form all other objects $g_1, \ldots, g_{i-1}, g_{i+1}, \ldots, g_n$ of the sample. This means that the only extents which contain g_i and a further object g_j with $j \in \{1, \ldots, i-1, i+1, \ldots, n\}$ is G. We set

- $Pr_G^{(n)}$ to be the empirical probability measure of $g_1, \ldots, g_n$ with $\Pr^{(n)} \in \mathscr{P}$, and

- $\Pr^{(n-1)}$ to be the empirical probability measure $g_1, \ldots, g_{i-1}, g_{i+1}, \ldots, g_{n-1}$ (without g_i) and $\Pr^{(n-1)} \in \mathscr{P}$.

Then the center-outward order of $g_1, \ldots, g_{i-1}, g_{i+1}, \ldots, g_n$ given by $D^{(n)}$ and $D^{(n-1)}$ are the same. So,

$$D^{(n)}(\cdot, \mathbb{K})\Big|_{\{g_1, \ldots, g_{i-1}, g_{i+1}, g_n\}} \cong D^{(n-1)}(\cdot, \mathbb{K})\Big|_{\{g_1, \ldots, g_{i-1}, g_{i+1}, g_n\}}.$$

Let us take a look at Definition 5 of the empirical depth function. There, the empirical probability measure is used for the definition. Thus, we start with a discussion on how duplicates or greatly different object effects the empirical probability measure of the extents. Therefore, we assume that the samples are iid. First, let us consider the case of Property (P10) where two objects g_1 and g_2 are duplicates. This means that they have the same attributes and therefore lie in the same extent set. If the invariance on the attribute Property (P2) is fulfilled, then g_1 and g_2 have the same depth. Furthermore, if we consider an iid sample and the corresponding empirical probability measure, then the probability reflects the duplication for every extent which contains g_1 (and therefore also g_2).

Let us now consider Property (P11) and let $g_1, \ldots, g_n$ be a sample with $g_i = g_{diff}$ the greatly different object w.r.t. to the rest of the sample. One can observe that there exist three distinctive cases on how the empirical probability measures differ on $E \subsetneq G$ extent:

- Case 1: $g_{diff} \in E$. Then $g_1, \ldots, g_i, g_{i+1}, \ldots, g_n \notin E$ and $\Pr^{(n)}(E) = 1/n \neq 0 = \Pr^{(n-1)}(E)$.

- Case 2: $g_\ell \in E$ for some $\ell \in \{1, \ldots, i-1, i+1, \ldots, n\}$. Then $g_{diff} \notin E$ and $\Pr^{(n)}(E) = ((n-1)/n)\Pr^{(n-1)}(E)$.

- Case 3: $g_\ell \notin E$ for all $\ell \in \{1, \ldots, n\}$. Then $\Pr^{(n)}(E) = 0 = \Pr^{(n-1)}(E)$.

Thus, Cases 2 and 3 show that the order of the extents which do not contain g_{diff} based on the probability measure $\Pr^{(n)}$ is the same as for $\Pr^{(n-1)}$. Hence, the order structure given by the probability values of extents containing $g_1, \ldots, g_i, g_{i+1}, \ldots, g_n$ is not influenced by g_{diff}.

These are observations on the empirical probability measure, but we only generally defined the (empirical) depth function. Thus, the question if a depth function fulfils these properties is concerned with how these connections between the empirical probability distributions are included in the precise definition of the mapping rule.

Properties (P10) and (P11) focus on two empirical probability measures and how their difference influences the empirical depth function. We end this subsection by discussing the behaviour of the depth function based on a sequence of empirical probability measures. Let $(\Pr^{(n)})_{n \in \mathbb{N}}$ be a sequence of empirical probability measures, with $\Pr^{(n)}$ being given by an iid sample of $\Pr \in \mathscr{P}$ with size $n \in \mathbb{N}$. By assumption we have that for all $n \in \mathbb{N}$, $\Pr^{(n)} \in \mathscr{P}$. Property (P12) discusses the consistency based on the empirical depth function towards the (population) depth function.

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

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

4.4. Universality properties

Finally, we introduce a notion of universality of depth functions. The main motivation is that in general there exists no strictly quasiconcave depth function, see Theorem 4. If one refrains from strict quasiconcavity, then a natural demand is to stick to quasiconcavity. However, the depth function that assigns every object the value zero is also quasiconcave but useless. Therefore, we develop properties that describe the richness of a depth function. Concretely, the motivation behind the following properties is the wish for a depth function to be *as strictly quasiconcave as possible*. While the motivation is based on the strictly quasiconcavity property, one can generalise this idea to all properties defined above.

Before we introduce our notion of a richness of a depth function, we first indicate that defining this notion in a more naive way does not lead to an adequate notion of richness. For now, let the formal context $\mathbb{K}$ and probability measure $\Pr$ be fixed. An intuitive way to concretising the property of a quasiconcave depth function D of being as strictly quasicconcave as possible is to demand that there exists no other quasiconcave depth function E that is more strictly quasiconcave than D. This can be formalised by saying that E is more strictly quasiconcave, then D if everywhere, where E violates strict quasiconcavity, also D violates strict quasiconcavity. By violations of the strict quasiconcavity property, we mean that only equality and not strict inequality is fulfilled. More precisely, a depth function E is more strictly quasiconcave than D if

$$quasiker(E(\cdot, \mathbb{K}, \Pr)) := \left\{(A, g) \,\middle|\, A \subseteq G, g \in G \text{ such that } g \in \gamma(A), E(g, \mathbb{K}, \Pr) = \inf_{\tilde{g} \in A} E(\tilde{g}, \mathbb{K}, \Pr)\right\}$$

$$\subsetneq quasiker(D(\cdot, \mathbb{K}, \Pr)).$$

In other words, there exists a pair $(A, g) \in quasiker(D(\cdot, \mathbb{K}, \mathrm{Pr})) \setminus quasiker(E(\cdot, \mathbb{K}, \mathrm{Pr}))$ with $g \in \gamma(A)$ such that the depth function D violates for this pair the strictly quasiconcave property but E does not.

However, this definition leads to a problem. The following situation can occur: Assume we have a quasiconcave depth function D and $(A, g) \in quasiker(D(\cdot, \mathbb{K}, \mathrm{Pr}))$ as well as $(\tilde{A}, \tilde{g}) \in quasiker(D(\cdot, \mathbb{K}, \mathrm{Pr}))$. Then it may be the case that increasing the depth value $D(g, \mathbb{K}, \mathrm{Pr})$ by some small amount ε leads to a depth function $E(\cdot, \mathbb{K}, \mathrm{Pr})$ that is still quasiconcave and that now fulfils $quasiker(E(\cdot, \mathbb{K}, \mathrm{Pr})) \subsetneq quasiker(D(\cdot, \mathbb{K}, \mathrm{Pr}))$. Now, assume that the same is true for increasing the depth of $\tilde{g}$ (without increasing the depth of g), but increasing the depth values of both g and $\tilde{g}$ at the same time is not possible without violating the quasiconcavity property. However, assume further that both the underlying probability measure Pr as well as the underlying context $\mathbb{K}$ are perfectly symmetric w.r.t. g and $\tilde{g}$. In this case, it is not reasonable to arbitrarily increase one depth value to be more strictly quasiconcave. This means that according to the above definition of being as strictly quasiconcave as possible, there does not exist a possible and reasonable depth function that is as strictly quasiconcave as possible. This underlines the importance to emphasise that a depth function always inherently codifies both the structure of the underlying space as well as the concrete underlying probability measure.

We now introduce our notion of richness of a depth function. The following two universality properties are stated in a more general way such that they do not only capture the richness of a depth function w.r.t. the property of being quasiconcave. The introduced universality properties are also applicable to other properties of depth functions. Concretely, we take inspiration from a basic notion that is folklore in category theory. We adopt here the notion of an universal or a free object: Very roughly speaking, we say that a depth function D is free w.r.t. a set of some structural properties, if for every other depth function E that obeys these properties, we can obtain $E(\cdot, \mathbb{K}, \mathrm{Pr})$ from D as a composition of $D(\cdot, \mathbb{K}, \tilde{\mathrm{Pr}})$ and an isotone function $f : \mathbb{R} \to \mathbb{R}$. Here, $\tilde{\mathrm{Pr}}$ is a suitable further probability measure. Informally, this means that we allow a meaningful depth function to have identical depth values if and only if the underlying probability measure and the structure of the space support this. In other words, a depth function is free if it is flexible enough to imitate every other depth function (with the same properties) by supplying it with a suitable probability measure $\tilde{\mathrm{Pr}}$. With this idea in mind we now state two different notions of freeness, one weak and one strong notion. Therefore, let $Q \subseteq \{P_1, \ldots, P_{11}\}$ be a set of properties from Sections 4.1, 4.2 and 4.3. A depth function D is called

(P 13) *weakly free* w.r.t. a family $\mathscr{P}$ of probability measures on the object set G and w.r.t. Q, if it satisfies every property from Q and if for every probability measure Pr and every depth function E on $\mathbb{K}$ that also satisfies all properties in Q, there exists a probability measure $\tilde{\mathrm{Pr}} \in \mathscr{P}$ and an isotone function $f : \mathbb{R} \longrightarrow \mathbb{R}$ such that

$$f \circ D(\cdot, \mathbb{K}, \tilde{\mathrm{Pr}}) \cong E(\cdot, \mathbb{K}, \mathrm{Pr}).$$

(P 14) *strongly free* w.r.t. a family $\mathscr{P}$ of probability measures on the object set G and w.r.t. Q, if it satisfies all properties from Q and if for every $\varepsilon > 0$, there exists a class $\mathscr{P}^{\varepsilon}$ of probability measures from $\mathscr{P}$ with a diameter[15] $d(\mathscr{P}^{\varepsilon})$ lower than or equal to ε such that for every other depth function E on $\mathbb{K}$ that also satisfies all properties in Q and for every $\mathrm{Pr} \in \mathscr{P}$ there exists a probability measure $\tilde{\mathrm{Pr}} \in \mathscr{P}^{\varepsilon}$ and an isotone function $f : \mathbb{R} \longrightarrow \mathbb{R}$ with

$$f \circ D(\cdot, \mathbb{K}, \tilde{\mathrm{Pr}}) = E(\cdot, \mathbb{K}, \mathrm{Pr}).$$

Remark 2. We say that a depth function D is richer than a depth function E if and only if we can modify D in such a manner that the center-outward order given by E can be reconstructed by D. This means that we can find a probability measure $\tilde{\mathrm{Pr}}$ and an isotone function f such that $f \circ D(\cdot, \mathbb{K}, \tilde{\mathrm{Pr}})$ equals E. Note that f only needs to be isotone and not bijective. Therefore, D must be at least as rich as E in distinguishing the individual objects of G. Finally, observe that $\tilde{\mathrm{Pr}}$ and f depend on $Q, \mathrm{Pr}, \mathbb{K}, D$ and E.

Within the notion weakly free we completely detach from Pr by allowing $\tilde{\mathrm{Pr}}$ to be any probability measure from $\mathscr{P}$. For the notion strongly free we restrict the allowed probability measures $\tilde{\mathrm{Pr}}$. Therefore, strongly free implies

---

[15]The diameter of a set of probability measures is simply defined as the supremal distance between two probability measures within this set. As a distance between two probability measures P and Q we take here the Kolmogorov distance $d(P, Q) := \sup_{A \in \Sigma} |P(A) - Q(A)|$ where Σ is the assumed σ-field underlying both P and Q.

weakly free. The idea behind restricting the probability measure $\tilde{\text{Pr}}$ is the requirement that a depth function should be able to detach itself to some degree from the underlying probability measure Pr. More precisely, there exists a balanced measure Pr* such that any structure within the depth values that is not due to the structure of the formal context can already be modified by an arbitrarily small modification of this balanced measure Pr*. The reason why we did not explicitly use such a balanced measure Pr* in the definition is a technical one. The balanced measure does not need to be a probability measure (e.g., there is no uniform distribution on the real numbers). Therefore we work here with the set $\mathscr{P}^\varepsilon$ with arbitrary $\varepsilon > 0$.

Theorem 7. *For the hierarchical nominal context given in Table 2 of Example 3, there exists a depth function D that is strongly free with respect to quasiconcavity (and w.r.t. the family of all probability measures on G), namely the function D given by:*

$$D(g, \mathbb{K}, \text{Pr}) = \begin{cases} 1 & \text{if } g = g^* := \underset{\tilde{g} \in G}{\arg\max} \, \text{Pr}(\{\tilde{g}\}) \\ 1/2 & \text{if on the first level, } g \text{ has the same category as } g^* \\ 0 & \text{else} \end{cases} .$$

Here we assume that the argument of the maxima is unique. Otherwise we arbitrarily choose one of the arguments of the maxima.

Proof. First observe that D is indeed quasiconcave, because the contour sets are extent sets. The reason is that the images of γ are exactly the one element sets, the whole set G and the two element sets $\{a_1 a_2, a_1 b_2\}$ and $\{b_1 a_1, b_1 a_2\}$. Now, for $\varepsilon > 0$ define the class of probability measures

$$\mathscr{P}^\varepsilon := \{\text{Pr} \mid \forall g \in G \text{ we have } \text{Pr}(\{g\}) \in [1/4 - \varepsilon/4, 1/4 + \varepsilon/4]\} .$$

This class has in fact diameter lower than or equal to ε. Now, let $\varepsilon > 0$, let Pr be an arbitrary probability measure on G, and let E be an arbitrary quasiconcave depth function. Let further g be the object with the highest depth according to E. Without loss of generality we assume that $g = a_1 a_2$. Additionally, we assume that there exists exactly one g with maximal depth (otherwise, arbitrarily choose from the objects with maximal depth). Then define $\tilde{\text{Pr}} \in \mathscr{P}^\varepsilon$ by $\tilde{\text{Pr}}(\{g\}) := 1/4 + \varepsilon$ and $\tilde{\text{Pr}}(\{\tilde{g}\}) := 1/4 - \varepsilon/12$ for $\tilde{g} \neq g$. If $\varepsilon > 3/4$, we set ε to $3/4$. Because E is assumed to be quasiconcave, and since the contour sets are always nested, the contour sets of E are exactly the sets $\{a_1 a_2\}$, $\{a_1 a_2, a_1 b_2\}$ and G. (For E it could also be the case that the set of contour sets is only a subset of these three sets). At the same time, due to construction, the contour sets of D are exactly the same which means that we can construct an isotone function $f : \mathbb{R} \to \mathbb{R}$ with $f \circ D(\cdot, \mathbb{K}, \tilde{\text{Pr}}) = E(\cdot, \mathbb{K}, \text{Pr})$. Since in particular $\varepsilon > 0$ was arbitrary, D is in fact strongly free with respect to quasiconcavity and w.r.t. the family of all probability measures on G. $\square$

Also for a finite hierarchical nominal context with more than two levels and with more than two categories in every level, one can show that there exists a depth function that is strongly free w.r.t. quasiconcavity and w.r.t. the family of all probability measures.

Remark 3. Note that while for example quasiconcavity implies starshapedness, freeness w.r.t. quasiconcavity generally does not imply freeness w.r.t. starshapedness. This is because a quasiconcave depth function D is not able to imitate a depth function E that is starshaped but not quasiconcave, if such a depth function E exists at all. In our concrete Example 3 of the hierarchical nominal context with two levels and two categories on every level, quasiconcavity is accidentally equivalent to starshapedness. For more than two levels this is not the case anymore.

5. Example: Generalised Tukey depth function

In Section 4, we generally introduced structural properties for depth functions based on formal concept analysis. One purpose of these structural properties is to give a systematic basis to analyse depth functions using formal concept analysis. Therefore, in this section, we take the opportunity to analyse the generalised Tukey depth given in Section 3.

Before we begin with discussing the structural properties, we look at some general aspects of the generalised Tukey depth. Here and in the following, we assume that G is a set, $\varkappa$ a set of formal contexts and $\mathscr{P}$ a set of probability measures on σ-fields which contain every extent set given by $\varkappa$.

In Definition 6 the generalised Tukey depth is defined by using the extent sets given by one single attribute. Thus, in the computation one apparently only takes a proper subset of all possible extents into account. The reason for this is that considering all extent sets instead of only those induced by one single attribute does not change the depth value. Compare to Section 3 where this has already been mentioned.

Theorem 8. *Let G be an object set. For every formal context $\mathbb{K}$ with object set G and every probability measure Pr on G we get for the generalised Tukey depth function*

$$T(g, \mathbb{K}, \mathrm{Pr}) = 1 - \sup_{m \in M \setminus \Psi(g)} \mathrm{Pr}(\Phi(m)) = 1 - \sup_{B \subseteq M \setminus \Psi(g)} \mathrm{Pr}(\Phi(B)).$$

Proof. We have to show that

$$\sup_{m \in M \setminus \Psi(g)} \mathrm{Pr}(\Phi(m)) = \sup_{B \subseteq M \setminus \Psi(g)} \mathrm{Pr}(\Phi(B)). \tag{2}$$

Since $B \subseteq M \setminus \Psi(g)$ is a superset of $m \in M \setminus \Psi(g)$ we immediately get $\leq$ in Equation (2). For the reverse inequality let $B \subseteq M \setminus \Psi(g)$ be arbitrary. Then for every $m \in B$ we get that $\Phi(m) \supseteq \Phi(B)$ and thus $\mathrm{Pr}(\Phi(m)) \geq \mathrm{Pr}(\Phi(B))$. Note that $m \in M \setminus \Psi(g)$. Together with the properties of a supremum, we get $\geq$ in Equation (2). $\square$

Remark 4. Analogously to Theorem 8 one can show that

$$T(g, \mathbb{K}, \mathrm{Pr}) = 1 - \sup_{m \in M \setminus \Psi(g)} \mathrm{Pr}(\Phi(m)) = 1 - \sup_{\substack{A \subseteq G \setminus \{g\} \\ A \text{ extent}}} \mathrm{Pr}(A)$$

is true for $g \in G$, $\mathbb{K}$ a formal context and probability measure Pr. Again, we obtain $\leq$ immediately. For the reverse, assume for simplicity that the supremum is attained at extent $A \subseteq G \setminus \{g\}$. Since $g \notin A$ holds, there exists $m \in \Psi(A)$ such that $g \notin \Phi(m)$. In particular, we get for this m that $\mathrm{Pr}(\Phi(m)) \geq \mathrm{Pr}(A)$ is true which gives indeed equality above.

Nevertheless, Theorem 8 shows that the generalised Tukey depth uses only the marginal distribution over the attributes. Furthermore, the generalised Tukey depth takes only those attributes with large marginal distribution into account. Small marginal distribution values have a lower or sometimes even no impact. Thus, it does not necessarily include the dependency structure of the objects given by the attributes. This means that some objects $g_1, \ldots, g_m$ can change in a manner such that only small marginal distributions over the attributes change. This does change the dependency structures of the objects, but this does then not influence the depth function values. For example look at the two formal contexts in Table 4. Here, we have (with uniform probability measure on $\{g_1, g_2, g_3\}$) $T(g_1, \mathbb{K}_1, \mathrm{Pr}) = 1/3 = T(g_1, \mathbb{K}_2, \mathrm{Pr})$, $T(g_2, \mathbb{K}_1, \mathrm{Pr}) = 1/3 = T(g_2, \mathbb{K}_2, \mathrm{Pr})$ and $T(g_3, \mathbb{K}_1, \mathrm{Pr}) = 2/3 = T(g_3, \mathbb{K}_2, \mathrm{Pr})$. Although in the second formal context, we have that g_2 implies g_3 as it is more specific than g_2 which is not given by the first formal context.

	m_1	m_2	m_3	m_4
g_1	×			×
g_2		×	×	
g_3			×	×

	m_1	m_2	m_3	m_4
g_1	×			×
g_2		×	×	
g_3		×	×	×

Table 4: Two formal contexts illustrating the limitation of Tukey depth. (left: $\mathbb{K}_1$ and right: $\mathbb{K}_2$)

After this general discussion of the generalised Tukey depth, we now analyse the depth by the use of our structural properties introduced in Section 4.

5.1. Representation properties

The first two structural properties, the representation properties, aim to ensure that the structure in the extent set is reflected in the depth function. The generalised Tukey depth function is based on the set of extents, see Theorem 8 and Remark 4. More precisely, for each g the depth is based on the probabilities of those extents which do not contain g. Since the function i in Property (P1) is bijective and preserves extents and probabilities, this implies that the depth values are preserved as well. This shows that Property (P1) holds. Furthermore, since two objects which equal in their attributes, always lie in the same extent sets, they have to have the same depth values. Thus, Property (P2) is true for the generalised Tukey depth.

5.2. Order preserving properties

In Section 4, the order preserving properties are ordered in their strength. Thus, we prove that the contourclosed property (P7i) holds for the generalised Tukey depth and with this further order preserving properties follow by Theorem 3 and 2.

Theorem 9. *The generalised Tukey depth fulfils Property (P7i).*

Proof. We prove Property (P7i) by contradiction. Assume that there exists a formal context $\mathbb{K}$ on an object set G and a probability measure Pr together with an $\alpha \in im(T(\cdot, \mathbb{K}, \mathrm{Pr}))$ such that $Cont_\alpha$ is not an extent set. This means that

$$\gamma(Cont_\alpha) \supsetneq Cont_\alpha$$

since γ is a closure operator. Thus, there exists $g \in \gamma(Cont_\alpha) \setminus Cont_\alpha$. The generalised Tukey depth, see Definition 6, is based on the extent sets induced by a single attribute that the object g does not have. Since $g \in \gamma(Cont_\alpha)$ we know that for every attribute which g does not have there exists at least one $\tilde{g} \in Cont_\alpha$ which does not have this attribute either. Hence, we have

$$\forall m \in M \setminus \Psi(g) \ \exists \tilde{g} \in Cont_\alpha : 1 - \mathrm{Pr}(\Phi(m)) \geq 1 - \sup_{\tilde{m} \in M \setminus \Psi(\tilde{g})} \mathrm{Pr}(\Phi(m)).$$

By definition of the contour set α, we get

$$T(g, \mathbb{K}, \mathrm{Pr}) = 1 - \sup_{m \in M \setminus \Psi(g)} \mathrm{Pr}(\Phi(m)) \geq \inf_{\tilde{g} \in Cont_\alpha} \left(1 - \sup_{\tilde{m} \in M \setminus \Psi(\tilde{g})} \mathrm{Pr}(\Phi(\tilde{m}))\right) = \inf_{\tilde{g} \in Cont_\alpha} T(\tilde{g}, \mathbb{K}, \mathrm{Pr}).$$

This is contradicting $g \notin Cont_\alpha$ and we can conclude that Property (P7i) holds for every formal context $\mathbb{K}$ and probability measure Pr. □

Remark 5. Now we apply Theorem 2 and 3, see overview in Figure 2, and obtain that since Property (P7i) is true for the generalise Tukey depth, the following properties hold as well: minimality property (P3), maximality property (P4), isotonicity property (P5), starshapedness property (P6) and the quasiconcavity property (P7ii).

In Section 4 we showed that there exist formal contexts such that the strictly quasiconcavity property (P8) does not hold for any depth function and any probability measure. Hence, here we are interested in formal contexts such that the generalised Tukey depth is strictly quasiconcave. A small example of a formal context such that the quasiconcavity is fulfilled is given in Table 5.

To check this claim, we take a look at Property (P8) applied on the generalised Tukey depth. Thus, Property (P8) is satisfied if and only if for all $A \subseteq G$ we have

$$\forall \tilde{g} \in \gamma(A) \setminus A : 1 - \sup_{m \in M \setminus \Psi(\tilde{g})} \mathrm{Pr}(\Phi(m)) > \inf_{g \in A} \left(1 - \sup_{m \in M \setminus \Psi(g)} \mathrm{Pr}(\Phi(m))\right)$$

$$\Leftrightarrow \quad \forall \tilde{g} \in \gamma(A) \setminus A : \sup_{m \in M \setminus \Psi(\tilde{g})} \mathrm{Pr}(\Phi(m)) < \sup_{g \in A} \sup_{m \in M \setminus \Psi(g)} \mathrm{Pr}(\Phi(m)). \tag{3}$$

	m_1	m_2	m_3	m_4
g_1	×	×	×	
g_2	×	×		×
g_3	×	×	×	×

Table 5: Formal context where the generalised Tukey depth fulfiles Property (P8).

Note that $\leq$ in Inequality (3) follows by Property (P7ii). To check whether the generalised Tukey depth based on the formal context given in Table 5 is strictly

quasiconcave, we go through every subset $A \subseteq G$ with $\gamma(A) \setminus A \neq \emptyset$. This is only true for $A = \{g_1, g_2\}$ with $\gamma(A) = \{g_1, g_2, g_3\}$. The attributes of g_3 are a superset of the union of the attributes of g_1 and g_2 (i.e. $\Psi(\{g_3\}) \supseteq \Psi(\{g_1\}) \cup \Psi(\{g_2\})$). Observe that the attributes of g_1 and g_2 which maximise the supremum of the corresponding generalised Tukey depth function are attributes of g_3. Thus, when ensuring that there exists an attribute $m \in \Psi(g_3) \setminus \Psi(g_2) \cap \Psi(g_1)$ such that $\Pr(\Phi(m))$ is strictly larger than $\Pr(\Phi(\tilde{m}))$ based on those attributes $\tilde{m}$ which are not an element of $\Psi(g_3)$, the generalised Tukey depth is strictly quasiconcave. Note that it is sufficient that the depth of either g_1 or g_2 must be below the depth of g_3. Furthermore, we want to point out that we also have to assume that there does not exist a further object g_4 which is a duplicate of g_3, because then $\Psi(g_3) \setminus \Psi(g_2) \cap \Psi(g_1) = \emptyset$. With this, we get the strict part in Inequality (3). Based on the idea above, we can say that the generalised Tukey depth satisfies Property (P8) for every formal context which is an element of the following set:

$$\mathscr{C}^{P8} = \left\{ \begin{array}{l|l} (G, M, I) \text{ formal context without} & \text{for every subset } A \text{ and } g \in \gamma(A) \setminus A \text{ we have } \Psi(g) \supseteq \cup_{a \in A} \Psi(a) \text{ and} \\ \text{duplicates according to attributes} & \exists m \in \Psi(g) \setminus \cap_{a \in A} \Psi(a) \colon \Pr(\Phi(m)) > \sup_{m \in M \setminus \Psi(g)} \Pr(\Phi(m)) \end{array} \right\}.$$

Finally, let us take a closer look at the reflecting betweenes property (P9). An example of a formal context $\mathbb{K}$ where the generalised Tukey depth does not fulfil Property (P9) is given in Table 6. Consider $G = \{g_1, g_2, g_3\}$ together with the following probability measure $\Pr$: $\Pr(g_1) = 1/5$, $\Pr(g_2) = 3/5$ and $\Pr(g_3) = 1/5$.

The problem is that $T(g_1, \mathbb{K}, \Pr) > T(g_2, \mathbb{K}, \Pr)$ is true, but there exists no set A with $g_1 \notin A$ such that $g_1 \in \gamma(A)$. Thus, such a set A in Property (P9) does not exist. A way around this problem is given by Theorem 6.

	m_1	m_2	m_3
g_1	$\times$		
g_2		$\times$	
g_3		$\times$	$\times$

Table 6: Formal context where the generalised Tukey depth does not fulfil Property (P9) and (P11).

5.3. Empirical (sequence) properties

The discussion about empirical (sequence) properties profits from the strong impact of the (empirical) probability measure on the definition of the generalised Tukey depth, see Definition 6 and 7. With this we can immediately follow that Property (P10) is fulfilled by the generalised Tukey depth. The stability of the order property does not hold. Consider again the formal context given by Table 6. Then g_1 is an outlier w.r.t. $\{g_2, g_3\}$. Let $\{g_1, g_2, g_3\}$ be the iid sample. For the empirical generalised Tukey depth based on $\Pr^{(2)}$ for $\{g_2, g_3\}$ we get $T^{(2)}(g_3, \mathbb{K}) = 1 > 1/2 = T^{(2)}(g_2, \mathbb{K})$. When adding object g_1 to the probability measure $\Pr^{(3)}$ we get that $T^{(3)}(g_3, \mathbb{K}) = 2/3 = T^{(3)}(g_2, \mathbb{K})$. The order of g_2 and g_3 is not stable. Since this can always occur for small $n \in \mathbb{N}$, we cannot restrict the set of probability measures or formal contexts.

It remains to discuss the consistency property (P12). One can show that if the set of extents induced by single attributes

$$\mathcal{ES} := \{\Phi(m) \mid m \in M\}$$

is a Glivenko-Cantelli class w.r.t. the underlying probability measure $\Pr$, see [6], then the generalised Tukey depth is indeed consistent. This, for example, is given when $\mathcal{ES}$ has a finite VC dimension. The proof is similar to the proof given by [5, p. 1816f].

Theorem 10. *Let $\mathbb{K}$ be a formal context with G as object set. Let $\Pr_G^{(n)}$ be the empirical probability measure given by an iid sample of sizes $n \in \mathbb{N}$ and based on probability measure* $\Pr$. *When $\mathcal{ES}$ is a Glivenko-Cantelli class with respect to a probability measure* $\Pr$, *then*

$$\sup_{g \in G} \left| T^{(n)}(g, \mathbb{K}) - T(g, \mathbb{K}, \Pr) \right| \overset{n \to \infty}{\longrightarrow} 0 \text{ almost surely.}$$

Thus, $T^{(n)}$ is consistent and Property (P12) is fulfilled.

Proof. Since we assumed that $\mathcal{ES}$ is a Glivenko-Cantelli class w.r.t. probability measure $\Pr$, we get

$$\sup_{E \in \mathcal{ES}} \left| \Pr^{(n)}(E) - \Pr(E) \right| \overset{n \to \infty}{\longrightarrow} 0 \text{ almost surely.}$$

22

With this, we now obtain

$$\sup_{g\in G}\left|1-\sup_{m\in M\setminus\Psi(g)}\Pr{}^{(n)}(\Phi(m))-\left(1-\sup_{m\in M\setminus\Psi(g)}\Pr(\Phi(m))\right)\right|$$

$$=\sup_{g\in G}\left|\sup_{m\in M\setminus\Psi(g)}\Pr{}^{(n)}(\Phi(m))-\sup_{m\in M\setminus\Psi(g)}\Pr(\Phi(m))\right|$$

$$\leq\sup_{E\in\mathcal{ES}}\left|\Pr{}^{(n)}(E)-\Pr(E)\right|\overset{n\to\infty}{\longrightarrow}0\text{ almost surely,}$$

which proves the claim. $\qquad\square$

Remark 6. We know that the generalised Tukey depth fulfils Property (P1). Let us assume that for a formal context $\tilde{\mathbb{K}}$ the set of all extents induced by one single attribute is not a Glivenko-Cantelli class w.r.t. some probability measure $\tilde{\Pr}$. Then, in some cases, one can define a second formal context $\mathbb{K}$ and probability measure $\Pr$ such that there exists a bijective and bimeasurable function i which preserves the extents and the probability measure. If the extents based on a single attribute given by $\mathbb{K}$ is a Glivenko-Cantelli class w.r.t. probability measure $\Pr$, then we can transfer the consistency based on $\mathbb{K}$ onto $\tilde{\mathbb{K}}$. An example for this is given by the following two formal contexts: Let $G=\mathbb{R}^d$ and $\mathbb{K}$ be the formal context defined in Example 4. For $\tilde{\mathbb{K}}$ let $\tilde{M}$ be the set of all topologically closed convex sets and $\tilde{I}$ the binary relation with $(g,\tilde{m})\in I$ if and only if g is an element of the corresponding convex set $\tilde{m}$. The extents equal again all topologically closed convex sets. Thus by applying Property (P1) we can replace $\tilde{\mathbb{K}}$ with $\mathbb{K}$ and can prove that also for $\tilde{\mathbb{K}}$ the generalised Tukey depth is consistent.

5.4. Universality properties

Let us end this discussion on the generalised Tukey depth by considering the univerality properties. The next two theorems prove that the generalised Tukey depth is weakly free but not strongly free w.r.t the quasiconcavity property.

Theorem 11. *Let G be finite. Then the generalised Tukey depth is weakly free w.r.t. the family of all probability measures on G and w.r.t. the quasiconcavity property (P7ii).*

Proof. Let G be finite, let $\mathbb{K}$ be an arbitrary context with object set G and let $\Pr$ be an arbitrary probability measure on G. Moreover, let E be an arbitrary quasiconcave depth function. We set $E_1,\ldots,E_K$ to be the unique increasingly ordered depth values of $E(\cdot,\mathbb{K},\Pr)$ and let k_i be the number of objects with depth value E_i. Furthermore, let G_i denote the boundaries of the contour sets, i.e., the set of objects with depth E_i. We now have to construct a probability measure $\tilde{\Pr}$ and an isotone function f such that $f\circ T(\cdot,\mathbb{K},\tilde{\Pr})=E(\cdot,\mathbb{K},\Pr)$.

To achieve this, we first consider some properties this probability function needs to fulfil. Therefore, let $\tilde{\Pr}\in\mathscr{P}$ be a possible candidate. We set $p_i:=\tilde{\Pr}(G_i)$ and $p_i^{min}:=\min_{g\in G_i}\tilde{\Pr}(\{g\})$. Since E is quasiconcave, we have $g\notin\gamma(G_{\ell+1}\cup\ldots\cup G_K)$ for all $g\in G_\ell$ with $\ell\in\{1,\ldots,K-1\}$. This implies that for every $\ell\in\{1,\ldots,K-1\}$ and every $g\in G_\ell$ there exists an attribute $m\in M$ with $(g,m)\notin I$, but $(h,m)\in I$ for all $h\in G_{\ell+1}\cup\ldots\cup G_K$. Therefore, we get for $g\in G_\ell$

$$p_\ell^{min}\leq T(g,\mathbb{K},\tilde{\Pr})\leq\sum_{i=1}^{\ell}p_i. \tag{4}$$

Now, we construct probability measure $\tilde{\Pr}$ needed for Property (P13) in three steps: First set $\tilde{\Pr}(g)=1$ for $g\in G_1$. Then recursively set $\tilde{\Pr}(g)=\sum_{i=1}^{\ell}\tilde{\Pr}(G_i)+1$ for $g\in G_{\ell+1}$. Finally, we normalise $\tilde{\Pr}$ by $\tilde{\Pr}(g):=\tilde{\Pr}(g)/\sum_{g\in G}\tilde{\Pr}(g)$ to get a probability measure. Due to construction and Inequality (4), it is ensured that $T(g,\mathbb{K},\tilde{\Pr})<T(h,\mathbb{K},\tilde{\Pr})$ for $g\in G_\ell$ and $h\in G_{\ell+1}$. Thus, one can define function f by setting $f(T(g,\mathbb{K},\tilde{\Pr}))=E_\ell$ for $g\in G_\ell$ and isotonically extending it to function $f:\mathbb{R}\to\mathbb{R}$. With this isotone function, we get $f\circ T(\cdot,\mathbb{K},\tilde{\Pr})=E(\cdot,\mathbb{K},\Pr)$ which shows that T is weakly free. $\qquad\square$

Theorem 12. *For the context describing hierarchical nominal data given in Table 2 the generalised Tukey depth is not strongly free w.r.t quasiconcavity and w.r.t. the family of all probability measures on G.*

Proof. First note that the depth function D from Theorem 7 is quasiconcave. Furthermore, D is flexible enough to assign every arbitrary object g the highest depth 1. This implies that the object h with the same category on Level 1 as g has the second highest depth $1/2$ and all other objects have depth 0. Now, the generalised Tukey depth $T(\cdot, \mathbb{K}, \text{Pr})$ needs to be as flexible as D to be strongly free w.r.t. quasiconcavity. Looking at Table 2, first observe that for object $g = a_1 a_2$ to have the highest generalised Tukey depth, it is necessary that $\text{Pr}(\{a_1 a_2, a_1, b_2\}) \geq 0.5$. This is because if we have $\text{Pr}(\{a_1 a_2, a_1, b_2\}) < 0.5$, we can conclude that $T(a_1 a_2, \mathbb{K}, \text{Pr}) \leq 1 - \text{Pr}(\{b_1 a_2, b_1 b_2\}) = \text{Pr}(\{a_1 a_2, a_1, b_2\}) < 0.5$. Since either $\text{Pr}(\{b_1 a_2\})$ or $\text{Pr}(\{b_1 b_2\})$ is smaller or equal to 0.5, we get that one of the objects $b_1 a_2$ or $b_1 b_2$ has a generalised Tukey depth of larger or equal to 0.5. Now, for object $h = a_1 b_2$ to have the second highest depth, it is necessary that the corresponding supremum in Definition 6 is attained for an attribute that differs from $g = a_1 a_2$. Because both g and h do not have attribute b_1, it is clear that attribute b_1 is relevant for the supremum within the generalised Tukey depth both for objects g and h. For h to have the second highest depth, it is therefore necessary that for h the supremum is attained for attribute $a_1 a_2$. But for this it is necessary that

$$\text{Pr}(\{a_1 b_2\}) \geq \text{Pr}(\{b_1 a_2, b_1 b_2\}). \tag{5}$$

Now, let $\varepsilon = 0.1$ and assume that T is strongly free w.r.t. quasiconcavity. Thus, we can find a family $\mathscr{P}^\varepsilon$ of diameter $\varepsilon = 0.1$ with the corresponding properties. Now take $\text{Pr}^* \in \mathscr{P}^\varepsilon$ such that the generalised Tukey depth has the highest depth at $g = a_1 a_2$ and for $h = a_1 b_2$ the second highest depth. With the above considerations we can conclude first that $\tilde{Pr}(\{a_1 a_2, a_1 b_2\}) \geq 0.5$ and, thus, $\text{Pr}^*(\{a_1 a_2, a_1 b_2\}) \geq 0.5 - \varepsilon = 0.4$. Additionally, because also object $b_1 a_2$ could be the object with highest depth w.r.t. D and some probability measure Pr, an analogous argumentation implies that also $\text{Pr}^*(\{b_1 a_2, b_1 b_2\}) \geq 0.5 - \varepsilon = 0.4$. But this, together with Inequality (5) implies that $\text{Pr}^*(a_1 b_2) \geq 0.4$. With analogous argumentations, one can also show $\text{Pr}^*(\{g\}) \geq 0.4$ for all other objects. But this is a contradiction to Pr^* being a probability measure because $\text{Pr}^*(G) = 0.4 * 4 = 1.6$. Therefore, the generalised Tukey depth is not strongly free w.r.t. quasiconcavity.

$\square$

6. Conclusions

This article introduces a general notion of depth functions for non-standard data based on formal concept analysis. This covers the analysis of centrality and outlyingness for a large variety of different data types. In particular, this approach does not involve a presupposed data structure and thus closes the gap of discussing centrality only on one specific data type at once. To enable a discussion about these different data types, especially also those which are not given in statistical standard data format, we use formal concept analysis. To mathematically formalise the notion of depth functions using formal concept analysis, we generally define depth functions based on a formal context and introduce structural properties. With this, we start the discussion on centrality on data represented via a formal context. Moreover, we provide a framework for analysing depth functions based on formal concept analysis. Finally, we use this framework to discuss and analyse the generalised Tukey depth. Building on this article, there are several promising further research areas, which are (but not limited to):

Further concrete mapping rules for depth functions: In Theorem 7 we defined a depth function D which is strongly free with respect to Property (P8) and the formal context defined in Table 2. One can show that this depth function D only uses the probabilities of Level 1 to obtain the object g with the highest depth value. In further research, it is of interest to define mapping rules for depth functions which take more levels into account. We already introduced the ufg depth function for the special case of G being the set of partial orders, see [2]. As shortly denoted there this can be generalised to general formal contexts.

Larger scale application: In this article the focus lies on the general analysis of depth functions and the generalised Tukey depth. Hence, the examples discussed here aim to support and clarify claims of theoretical/structural properties of depth functions. In further research, a discussion based on the perspective of a larger scale application is of interest. For example, in [2] we applied the ufg depth function, which can be embedded into our concept of depth upon partial orders, which represent the performance of machine learning algorithms.

Statistical inference: Building on this not only a descriptive analysis, but also statistical inference methods can be developed. Since depth functions describe in a robust and nonparametric manner the order of the data this can be used to define statistical tests and models. Analogously to the approach here, where we, among others, transferred

already existing properties to our concept, this can be done for statistical inference methods. A starting point can be [13].

Analysis of generalised Tukey depth based on one specific scaling method: Here, we generally analysed the generalised Tukey depth function without focus on one specific scaling method. In [1] we discussed the special case of G being the set of partial orders and one scaling method. Since one fixed scaling method on a set G gives us more structure on the corresponding closure system an analysis can lead to more structural properties of the depth function.

Acknowledgments

References

[1] H. Blocher, G. Schollmeyer, C. Jansen, Statistical models for partial orders based on data depth and formal concept analysis, in: D. Ciucci, I. Couso, J. Medina, D. Ślęzak, D. Petturiti, B. Bouchon-Meunier, R. Yager (Eds.), Information Processing and Management of Uncertainty in Knowledge-Based Systems, volume 1602, Springer, Cham, 2022, pp. 17–30.

[2] H. Blocher, G. Schollmeyer, C. Jansen, M. Nalenz, Depth functions for partial orders with a descriptive analysis of machine learning algorithms, in: E. Miranda, I. Montes, E. Quaeghebeur, B. Vantaggi (Eds.), Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications, volume 215, PMLR, Oviedo, 2023, pp. 59–71.

[3] C. Carpineto, G. Romano, Concept data analysis: Theory and applications, Wiley, Chichester and Weinheim, 2004.

[4] F. Chebana, T. B. M. J. Ouarda, Depth-based multivariate descriptive statistics with hydrological applications, Journal of Geophysical Research 116 (2011).

[5] D. L. Donoho, M. Gasko, Breakdown properties of location estimates based on halfspace depth and projected outlyingness, The Annals of Statistics (1992) 1803–1827.

[6] M. Dudley, E. Gin, J. Zinn, Uniform and universal Glivenko-Cantelli classes, Journal of Theoretical Probability 4 (1991) 485–510.

[7] S. Durocher, S. Szabados, Curve stabbing depth: Data depth for plane curves, in: Y. Bahoo, K. Georgiou (Eds.), 34th Canadian Conference on Computational Geometry, pp. 121–128.

[8] R. Dyckerhoff, Convergence of depths and depth-trimmed regions, ArXiv, 2017. https://arxiv.org/abs/1611.08721 Accessed: Aug 11th, 2023.

[9] R. Dyckerhoff, K. Mosler, G. Koshevoy, Zonoid data depth: Theory and computation, in: COMPSTAT, Physica-Verlag HD, 1996, pp. 235–240.

[10] B. Ganter, R. Wille, Formal concept analysis: Mathematical foundations, Springer, Berlin, Heidelberg, 2012.

[11] I. Gijbels, S. Nagy, On a general definition of depth for functional data, Statistical Science 32 (2017) 630 – 639.

[12] T. Hanika, J. Hirth, Quantifying the conceptual error in dimensionality reduction, in: T. Braun, M. Gehrke, T. Hanika, N. Hernandez (Eds.), Graph-Based Representation and Reasoning - 26th International Conference on Conceptual Structures, volume 12879, Springer, Cham, Cham, 2021, pp. 105–118.

[13] J. Li, R. Y. Liu, New nonparametric tests of multivariate locations and scales using data depth, Statistical Science 19 (2004) 686–696.

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

[15] K. Mosler, Depth statistics, in: C. Becker, R. Fried, S. Kuhnt (Eds.), Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather, Springer, Berlin, Heidelberg, 2013, pp. 17–34.

[16] K. Mosler, P. Mozharovskyi, Choosing among notions of multivariate depth statistics, Statistical Science 37 (2022) 348–368.

[17] P. Mozharovskyi, Anomaly detection using data depth: Multivariate case, ArXiv, 2022. https://arxiv.org/abs/2210.02851 Accessed: Aug 11th, 2023.

[18] S. Pawar, D. Shirke, Data depth-based nonparametric tests for multivariate scales, Journal of Statistical Theory and Practice 16 (2022) 1–21.

[19] G. Schollmeyer, Application of lower quantiles for complete lattices to ranking data: Analyzing outlyingness of preference orderings, Technical Report, LMU, 2017. http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-40452-9 Accessed: Aug 11th, 2023.

[20] G. Schollmeyer, Lower quantiles for complete lattices, Technical Report, LMU, 2017. http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-40448-7 Accessed: Aug 11th, 2023.

[21] T. Tao, An introduction to measure theory, volume 126 of Graduate studies in mathematics, American Mathematical Society, Providence, RI, 2011.

[22] J. Tukey, Mathematics and the picturing of data, in: R. James (Ed.), Proceedings of the International Congress of Mathematicians Vancouver, Mathematics-Congresses, Vancouver, 1975, pp. 523–531.

[23] Y. Zuo, R. Serfling, General notions of statistical depth function, The Annals of Statistics 28 (2000) 461–482.

[24] Y. Zuo, R. Serfling, On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry., Journal of Statistical Planning and Inference 84 (2000) 55–79.