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

Hannah Blocher *, Georg Schollmeyer

Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstr. 33, 80539 München, Germany

ARTICLE INFO

ABSTRACT

In this article we introduce a notion of depth functions for data types that are not given in standard statistical data formats. We focus on data that cannot be represented by one specific data structure, such as normed vector spaces. This covers a wide range of different data types, which we refer to as non-standard data. Depth functions have been studied intensively for normed vector spaces. However, a discussion of depth functions for non-standard data is lacking. In this article, we address this gap by using formal concept analysis to obtain a unified data representation. Building on this representation, we then define depth functions for non-standard data. Furthermore, we provide a systematic basis by introducing structural properties using the data representation provided by formal concept analysis. Finally, we embed the generalised Tukey depth into our concept of data depth and analyse it using the introduced structural properties. Thus, this article presents the mathematical formalisation of centrality and outlyingness for non-standard data and increases the number of spaces in which centrality can be discussed. In particular, we provide a basis for defining further depth functions and statistical inference methods for non-standard data.

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. 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 [1], zonoid depth, see [2], or Tukey depth, see [3,4]. To give a systematic basis on the notion of depth functions [5,6] 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 [7] the authors discuss properties of depth functions on functional data. [8] developed a depth function for total orders. [9] went the next step and defined depth functions for general metric spaces. In recent years, this concept has been used to construct statistical methods and to analyse different application settings. For example, [10,11] developed statistical nonparametric and robust tests using depth functions. Using the robust central-outward order [12] studied anomaly detection. In [13], 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 spatial. The aim of this article is to 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, flexible and general applicable 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. In particular, it does not force the user to make assumptions about the data that are not necessarily true, simply for the sake of mathematical feasibility. 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 [5,6] and develop further properties representing the connections between the data points. In this context, we investigate that, unlike $\mathbb{R}^d$, non-standard data can have data elements that are naturally outlying or central based on the structure of the underlying space. Finally, we embed the *generalised Tukey depth* given in [14] into our concept of depth functions for non-standard data. Moreover, we use the provided structural properties to analyse the generalised Tukey depth. Note that, analogous to the generalised Tukey depth, the concrete outlier measures developed in [15] can also be analysed using the structural properties introduced here. These outlier measures are also based on formal concept analysis. Therefore, only a slight adaptation to the concept of depth introduced here is necessary.

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: We begin with concrete motivation examples. Since formal concept analysis is the basis of our general notion of a depth function, we give a short introduction to the theory of formal concept analysis in Section 3. Based on this, Section 4 introduces the general definition of depth functions for non-standard data by use of formal concept analysis. There we define the Tukey depth within 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 7, we collect our concluding remarks.

2. Motivation

The purpose of this section is to present motivating examples to show that depth functions based on formal concept analysis cover a wider range of data types than the currently discussed depth functions known to the authors. Instead of using observations directly to define a depth function, we first apply formal concept analysis. This gives us a definition of depth functions that can be applied to a wide variety of data types simultaneously. Most importantly, it is simple for the practitioner to use, as it only relys on the definition of a cross-table, as will be seen in this section.

One of the goals of formal concept analysis is to discover relational connections between observations. Therefore, the observations are grouped according to the observed values. For example, consider the Titanic data set, where each passenger's age, sex and the passenger's class of travel is recorded. One group might be all passengers aged between 20 and 40 who are female. Another smaller group are all female passengers between the ages of 20 and 40 travelling in first class. In this case, the smaller group is more specific, e.g. in the smaller group the passengers also fulfil the condition that they all travel in first class. Formal concept analysis formalises the arrangement into different groups. The first step is to define a formal context that represents the data set. In a nutshell, a formal context is a generalisation of a cross-table.

To provide an intuitive introduction to the main concepts of formal concept analysis and to illustrate the motivation for defining depth functions based on formal concept analysis, we give now two examples. These examples also demonstrate the variety of different data types that can be represented by a formal context.

Example 1. For the Titanic data example provided by the Titanic machine learning competition, see [16], we get the formal context in Table 1 (this is just a snippet). Here each row represents an observation/passenger (i.e. called *objects* in formal concept analysis). The columns are binary *attributes* that can be either true or false. Transforming the non-binary observations (such as age, spatial values, etc.) into binary attributes is called *conceptual scaling*, see [17, p. 36ff]. Since the number of attributes can be infinite, this is a flexible, but also accessible approach to data representation. For classical data types there are already standard scaling methods, see [17, p. 36ff]. As an example, consider Table 1. The first two columns represent the sex of the passenger (f: female and m: male). The next three columns are the class level of the passenger (I, II and III). These two nominal observations are included via *nominal scaling*, see [17, p. 42]. The remaining columns represent the age of the passenger, e.g. g_1 is 34.5 years old. The corresponding attributes are the statements "$\leq x$" and "$\geq x$" for all $x \in \mathbb{R}$. This scaling method is called *interordinal scaling*, see [17, p. 42]. Note that this scaling method implies that there exists no object that has every attribute. The crosses in Table 1 indicate that the attribute applies to the passenger/observation. How different observation values are represented as a set of binary attributes depends on the scaling method used, see [17, p. 36ff].

We obtain the groups that represent the relationship between the observations by considering all possible combinations of attributes and summarising all the observations that agree on all these attributes. Note that these groups may overlap. They also have a certain mathematical structure, i.e. they show the implications between the passengers/observations. For example, in Table 1 we see that if g_3 and g_4 are in a group, we can directly imply that g_2 is also in the group, since there is no group with g_3 and g_4 that does not also contain g_2. By collecting all these implications, we can fully determine the arrangement of the groups given by the formal context. For further details see Section 3.

Table 1

This formal context contains five observations/passengers of the Titanic. Each row represents a passenger. The data is from the Titanic machine learning competition, see [16], with PassengerIds 892, 893, 894, 904 and 908. Note that the age 34.5 is due to the encoding notation. With "...", we denote that the attribute set is indeed infinite and not all attributes are denoted within the table. The same is true for the objects as this is only a snippet of the titanic data set.

	f	m	I	II	III	...	'≤23'	...	'≤24'	...	'≤34.5'	...	'≤35'	...	'≤47'	...	'≤67'	...	...	'≥67'	...	'≥47'	...	'≥35'	...	'≥34.5'	...	'≥24'	...
g_1		x			x						x		x		x	x	x	x								x	x	x	x
g_2	x				x										x	x	x	x				x	x	x	x	x	x	x	x
g_3	x			x													x	x		x	x	x	x	x	x	x	x	x	x
g_4	x		x				x		x		x		x		x	x	x	x											
g_5		x		x									x		x	x	x	x						x	x	x	x	x	x
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

We emphasise that the choice of binary attributes, as well as the decision which observation has an attribute and which does not, results in a concrete, easily accessible representation of the data. In particular, it exposes many implicit assumptions about the data. Moreover, the data are represented in such a way that we do not impose assumptions that are not present in the data, e.g. that the data can be embedded in a normed vector space. However, we want to point out that the choice of attributes can have a strong influence on the arrangement of the groups and therefore on the depth values. Instead of using interordinal scaling for the age component in the Titanic data set, another approach called ordinal scaling, which includes only the upper bound '≤x', can be used, see [17, p. 42]. With this smaller formal context, age can only influence the groups through its upper bound. Thus, instead of looking at all passengers between the ages of, e.g., 20 and 40, with this smaller formal context, only all passengers with an age smaller than or equal to 40 are grouped without a lower bound.

We now use the structure given by the different groups and the resulting implications to give a notion of depth to the observations. In the Titanic example, we want to rank the five passengers according to how central or outlying they are compared to the other passengers. In the spirit of [18], in this paper we introduce properties that represent the notion of centrality and outlyingness. For example, Property (P1) states that if we use different scaling methods of the observations (i.e., use different binary attributes in the formal context) that lead to the same groups, the depth function should stay the same. As another example, consider the relationship between g_2, g_3, and g_4. Since any group that includes g_3 and g_4 also includes g_2, we have that g_2 has all the properties that g_3 and g_4 share. So we can say that g_2 is at least as specific as g_3 and g_4 combined. Following the idea of the quasiconcavity property in $\mathbb{R}^d$, see [6], Property (P7i and ii) states that the depth of g_2 must be at least as high as the minimum depth of g_3 or g_4. Other properties discuss that non-standard data may inherit a natural center-outward order, which is not present in $\mathbb{R}^d$. For example, Property (P4) states that an object lying in every group must be central, regardless of the probability measure.

Example 2. Another example that illustrates the wide range of data types that can be discussed using formal concept analysis is the set of partial orders. A data example for partial orders is given in [19], where the objects are partial orders of classifiers. In that article, we considered a set of classifiers applied to a set of data sets. These classifiers were evaluated for each individual data set by several performance measures. Then an algorithm i outperforms another algorithm j on a given data set if and only if there exists one performance measure that states that algorithm i is better than algorithm j and all other performance measures agree that algorithm i is not worse than algorithm j. This gives us a partial order for each single data set that represents the performance structure of the classifiers for that data set. Thus, a benchmark suite of 80 data sets results in 80 observed partial orders. Now, to represent this kind of data, in [14] we introduced a formal context for partial orders in general. Here, each row corresponds to a partial order. The columns of the formal context are the binary attributes needed to describe the partial order. In [14], the first $n \cdot (n-1)$, where n is the number of classifiers/items being compared, denotes all possible pairs where we have stated that one classifier dominates the other. The next $n \cdot (n-1)$ rows indicate the reverse, i.e. that one classifier does not dominate another. Note that including the second part as binary attributes is by no means straightforward. But as in [14,19] we decided to state that non-existence of dominance is indeed a precise observation and we do not want a push towards the linear extensions of the partial orders (which exists when deleting the second $n \cdot (n-1)$ columns), we have therefore included this part.

As can be seen in the two examples above, and also in Examples 4 and 5 below, defining a formal context gives us a flexible tool for evaluating data ranging from mixed nominal and ordinal observations to spatial data or partial orders. These examples demonstrate that formal concept analysis clearly reveals the relationships between data points, providing a simple and unified view of the underlying data structure. Finally, we want to highlight differences and connections between our approach and other existing approaches. In contrast to [8,9], we have completely turned away from a metric approach in our definition of depth functions. [8] discusses total orders and therefore a translation to our approach is straightforward, just restrict the space of all partial orders to the total orders, see Example 2. This is not so clear for [9] as the representation via a formal context depends on the concrete structure of the space. Example 5 gives an approach for $\mathbb{R}^d$ equipped with the Euclidean distance.

3. Formal concept analysis

This section briefly describes the parts of formal concept analysis (FCA) required for this article and illustrates them using the two examples above. It is based on [17]. For further readings, we refer to [20]. FCA was developed by Rudolf Wille, Bernhard Ganter and Peter Burmeister to build a bridge between mathematical lattice theory and applied users. It enables the analysis of relationships between the data points by representing the data in a unified and user-friendly manner.

The fundamental definition of FCA is the representation of a data set as a cross-table, see [17, 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 *conceptual scaling methods*, see [17, p. 36ff] and Section 2, which transfers many-valued characteristics into a set of binary-valued attributes. 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*. Examples are given in Examples 1, 2, 4 and 5. The scaling method can also be used to reduce data complexity. This can lead to a conceptual scaling error, see for example [21].

Based on this user-friendly representation of the data set by a formal context, we can now define so-called *derivation operators*, see [17, p. 18]:

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

The function Ψ maps a set of objects A onto every attribute which every object in A has. So, $A' = \Psi(A) = \cap\{m \in M \mid (g, m) \in I\}$ 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. We set $\Phi(\emptyset) = G$ and $\Psi(\emptyset) = M$. 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. We have $\gamma(A) = \Phi \circ \Psi(A) = \cap_{m \in \Psi(A)}\{g \in G \mid (g, m) \in I\}$. In other words, every object set $E \subseteq G$ which is an element of the codomain of $\gamma = \Phi \circ \Psi$ is the maximal set of objects which have all the same attributes $\Psi(E)$ in common. Thus, the composition groups all those objects together that have the same attributes. With slight abuse of notation, for any $m \in M$ we write $\Phi(m)$ for $\Phi(\{m\})$. The same convention holds for $\Psi(g)$ and $\gamma(g)$ with $g \in G$.

Definition 2. The set $\gamma(2^G)$ is called the set of *extents*, see [17, p. 18]. The set $\gamma(A)$ is mentioned as *extent set*, shortly *extent*, of $A \subseteq G$.

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*, see [17, p. 8]. 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*.

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 $S \subseteq 2^G$ is a family of sets which contains the entire space (i.e. $G \in S$) and any intersection of sets in S is again in S (for all $S \subseteq S$ with $S \neq \emptyset$ we have $\bigcap_{s \in S} s \in S$).

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 closure system. Thus, the closure operator γ describes the closure system and vice versa. For more details on closure systems see [17, Chapter 0.3].

Example 3. Recall the Titanic Example 1. We obtain $\Psi(g_2) = \{$f, III, '$\leq$47', ..., '$\leq$67', ..., '$\geq$47', ..., '$\geq$35'$\}$. Therefore, we get $\gamma(g_2) = \{g_2\}$. One can show that the set of extents $\gamma\left(2^{\{g_1, g_2, g_3, g_4, g_5\}}\right)$ equals

$$\{\emptyset, \{g_1\}, \{g_2\}, \{g_3\}, \{g_4\}, \{g_5\}, \{g_1, g_4\}, \{g_1, g_5\}, \{g_2, g_5\}, \{g_2, g_3\}$$
$$\{g_1, g_4, g_5\}, \{g_1, g_2, g_5\}, \{g_2, g_3, g_5\}, \{g_1, g_2, g_4, g_5\}, \{g_1, g_2, g_3, g_5\}, \{g_1, g_2, g_3, g_4, g_5\}\}.$$

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 certain attributes in common. This means if $a \in \gamma(A) \backslash A$ 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 a 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 [17, Chapter 2.3]. Note that in [17] attribute implications are discussed and we focus here on object implications. Nevertheless, the concepts can be transferred to object implications.

Table 2

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

	a_1	b_1	$a_1 a_2$	$a_1 b_2$	$b_1 a_2$	$b_1 b_2$
$a_1 a_2$	x		x			
$a_1 b_2$	x			x		
$b_1 a_2$		x			x	
$b_1 b_2$		x				x

Example 4. 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 are categorised within a hierarchy of different levels, for example within the International Standard Classification of Occupations (ISCO) of 2008, see https://www.ilo.org/publications/international-standard-classification-occupations-2008-isco-08-structure (accessed: 24.09.2024). On a first level occupations are split into different categories ($a_1, b_1, \ldots$). 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 $a_1 a_2$ "Managers: Chief executives, senior officials, and legislators" and $a_1 b_2$ "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 in a natural way. For every level, we introduce attributes describing every single category based on the higher level classification, see Table 2. Here, for simplicity, we used only two levels with two categories, respectively.

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

$$\{\emptyset, \{a_1 a_2\}, \{a_1 b_2\}, \{b_1 a_2\}, \{b_1 b_2\}, \{a_1 a_2, a_1 b_2\}, \{b_1 a_2, b_1 b_2\}, G\}.$$

Furthermore, we obtain that $a_1 a_2 \in \gamma(\{a_1 b_2, b_1 a_2\}) = \Phi \circ \Psi(\{a_1 b_2, b_1 a_2\}) = \Phi(\emptyset) = G$.

Example 5. 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 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 [22,23]. Let $G = \mathbb{R}^d$ be the object set and the attribute set $M = \{H \subseteq \mathbb{R}^d \mid H \text{ topologically closed halfspace}\}$. Define the relation I between M and G by $(g, H) \in I \Leftrightarrow g \in H$, and let $\mathbb{K} = (G, M, I)$ denote the corresponding formal context.

With this definition, let us consider $\gamma(2^G)$ induced by $\mathbb{K}$. Let $A \subseteq G = \mathbb{R}^d$, then $\Psi(A)$ are all halfspaces containing 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. 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 equals $\gamma(A)$. 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 $f \in \gamma(\{a, b, c\})$. This means that f lies in every topologically closed halfspace which contains also a, b and c. In other words, f shares the same attributes as a, b, c share. Thus, one can say that f is implied by a, b and c based on γ. If $e \notin \gamma(\{a, b, c\})$, there exists a halfspace which contains a, b and c but not e. To summarise, an object g lies in $\gamma(\{a, b, c\})$ if and only if g lies within the triangle given by the vertices a, b and c. This shows how the concrete definition of the closure system enhances the connection between single objects.

Finally, as we have now introduced several different data situations with different scaling methods, see Example 1 (nominal, numeric data), Example 2 (partial orders), Example 4 (hierarchical nominal data) and Example 5 (spatial data), we want to point out the importance of the scaling method used. First of all, deciding which observations (age, spatial, etc.) to include in the analysis is a general issue in statistics. At first sight, FCA may seem to add a further difficulty by requiring the definition of binary attributes. However, FCA provides a clear understanding of the relationships between observations. Thus, to check whether the scaling method represents the objective, one can look at all the extents as well as all the implications. If the groups do not represent an important fact about the data (e.g. the non-dominance part of the partial orders or the lower bound of the age value), then more attributes are needed. Conversely, if the extents are close to the power set of G, one should examine the data set to see if there really are no implications between the data points. If so, some sets of attributes need to be replaced or even deleted. While nominal, ordinal and interordinal scaling are commonly used scaling methods, it is important to discuss carefully for each data set how fine/coarse the arrangement of the groups should be and to consider several different scaling methods.

4. 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 using FCA. 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 outlyingness issues. With this, nonparametric methods can be developed for all these data.

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 5, can be seen as generic properties for this kind of depth functions.

Definition 4. We define a depth function using FCA 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 formal contexts in $\varkappa_G$.

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 $\mathscr{P}_G$ depend on each other. Note that Definition 4 allows the restrictions of $\varkappa_G$ and $\mathscr{P}_G$ to a subset of all possible formal contexts or probability measures, which is sometimes 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 [24, 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 5: Property (P8) does not hold for every formal context, but only for a subset.) Another example is given in [14,19] where we considered one single formal context on the set of partial orders, see also Example 2. 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 also 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 5.

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 on a σ-field containing all extents of $\varkappa_G$. Furthermore, for every sample $g_1, \dots, g_n$ based on a probability measure $\Pr_G \in \mathscr{P}_G$, we have $\Pr_G^{(n)} \in \mathscr{P}_G$, where $\Pr_G^{(n)}$ is the empirical probability measure based on the sample. Then the empirical depth function for a sample $g_1, \dots, g_n$ 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 [22,23] and introduced in [14]. The Tukey depth on $\mathbb{R}^d$ of a point $g \in \mathbb{R}^d$, c. f. [3,4], is the smallest probability of a halfspace containing g. To build the bridge to FCA, 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 5. 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

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

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 $\mathbb{K}$ defined in Example 5. The first and the last equality are translations into FCA language, where importantly the final right hand side does not involve the notion of a halfspace, which allows the generalisation of the Tukey depth to any arbitrary FCA setting. The third equality holds because $G \setminus H$ is a topologically open halfspace and the supremum does not change when considering topologically open halfspaces instead. The right hand side of Eq. (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 Eq. (1): Generally, the topologically closed convex sets in $\mathbb{R}^d$ correspond to the extents within our approach to use FCA to define data depth functions. On the other hand, the topologically closed halfspaces of $\mathbb{R}^d$ have no general natural equivalent in FCA. 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 FCA are generally not extents. This further separates the left and the right hand side of Eq. (1). Therefore, we take the right hand side of Eq. (1). As will be shown later in Theorem 6 of Section 6, taking the supremum over all halfspaces or taking it over all topologically closed convex sets in $\mathbb{R}^d$ gives the same result, which also translates to the generalised Tukey depth where the supremum over all extents and the supremum over all extents generated by one single attribute coincide.

Additionally, the supremum of the right hand side of Eq. (1) has a natural interpretation as a measure of outlyingness that can be expressed in the language of FCA, see [23, Section 2] and [22, Section 5]. In these articles, the author constructs special representative extents: For given $\alpha \in [0, 1]$ we say that an extent is α-extensive if it contains at least a proportion of $\alpha \cdot 100\%$ 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 not containing g. This is exactly what is expressed by $\sup_{m \in M \setminus \Psi(g)} \Pr_{\mathbb{R}^d}(\Phi(m))$ in Eq. (1). With this motivation (and with the insight of Theorem 6) 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

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

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

Note that here, we take the supremum only over all extents generated by one single attribute, as this gives the same result as taking the supremum over all extents (c.f., Theorem 6) and is at the same time easier to compute. Additionally, from this definition it becomes clear that the (generalised) Tukey depth does not dependent on the dependence structure between the attributes, but only on the marginal distribution of the attributes. Observe that the term marginal is meant here as the probability of lying in a certain halfspace and should not be confused with the marginal distribution of points in $\mathbb{R}^d$ in the sense of the distribution of one coordinate. In the sequel, we will always refer to the marginal distribution as the distribution of the attributes.

Furthermore, 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, $\sup_{m \in M \setminus \Psi(g)} \mathrm{Pr}_G(\Phi(m))$, 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. Note, however, that there are usually no objects that have all attributes. For example, consider the spatial context described in Example 5, where the generalised Tukey depth corresponds to the well-known Tukey depth on $\mathbb{R}^d$. Thus, in this situation, the (generalised) Tukey depth is bounded by 0.5 and, in particular, strictly below 1. Another example is the interordinal scaling described in Section 2.

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

Definition 7. Let G, $\varkappa_G$ and $\mathscr{P}_G$ be defined as in Definition 5. Let $(g_1, \ldots, g_n)$ be a sample from G according to $\mathrm{Pr}_G \in \mathscr{P}_G$ with associated empirical measure $\mathrm{Pr}_G^{(n)}$. Then the empirical generalised Tukey depth is given by

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

As derived above, the generalised Tukey depth applied to the spatial context defined in Example 5 gives us the well-known Tukey depth on $\mathbb{R}^d$, see [3]. In particular, on $\mathbb{R}^1$ we get the ranking given by the classical quantiles, with the median being the central one and the depth values decreasing outwards.

Example 6. Recall Examples 1 and 3. Let us assume that the probability measure Pr_G in $G = \{g_1, g_2, g_3, g_4, g_5\}$ is uniform and the formal context $\mathbb{K}$ is given by Table 1. Then the generalised Tukey depths are

$$T_G(g_2, \mathbb{K}, \mathrm{Pr}_G) = 1 - \sup_{m \in M \setminus \Psi(g_2)} \mathrm{Pr}_G(\Phi(m)) = 1 - \frac{3}{5} = \frac{2}{5} = T_G(g_1, \mathbb{K}, \mathrm{Pr}_G) = T_G(g_5, \mathbb{K}, \mathrm{Pr}_G),$$

$$T_G(g_4, \mathbb{K}, \mathrm{Pr}_G) = 1 - \sup_{m \in M \setminus \Psi(g_4)} \mathrm{Pr}_G(\Phi(m)) = 1 - \frac{4}{5} = \frac{1}{5} = T_G(g_3, \mathbb{K}, \mathrm{Pr}_G).$$

Example 7. In Example 2 we briefly described the formal context in which partial orders are represented. See also [14,19]. Recall that the attributes represent the existence or non-existence of a dominance structure between two different algorithms i and j. Thus, each pair of algorithms i and j gives rise to four binary attributes: First, whether i dominates j or vice versa, and second, whether i does not dominate j and vice versa. As can be seen from the definition, the generalised Tukey depth only considers the marginal distribution of the attributes. Thus, based on the formal context describing partial orders, the generalised Tukey depth relies only on the probability of how often dominance between two classifiers occurs and how often it does not. More specifically, if we consider a sample, then the marginal distribution is given by the proportion of observed dominance between two classifiers, or the proportion of observed non-dominances. In the following, we omit the case where all partial orders have exactly the same probability, see [14] for details on this. This assumption implies that the partial order(s) with the minimum depth value are those that do not fulfil a (non-)dominance structure that has the highest proportion. There is no such simple answer for the maximum depth partial order. This is due to the transitivity assumption on the dominance structure. A partial order with highest depth value is one that summarises the most dominance and non-dominance structures that are most often observed.

Example 8. Recall the nominal hierarchical data structure discussed in Example 4. Note that the groups are already disjoint after the first partition and are then successively partitioned into smaller subsets. In the following we assume that the probability mass is not strongly focused on a small subset of possible observations, see the proof of Theorem 10 for details. Since the generalised Tukey depth includes only the maximum marginal probability of the attributes that are not true for the object of interest, we can observe that only the first partition is used for the generalised Tukey depth. Therefore, dropping all other attributes (representing the finer splits) and thus considering a smaller and less informative formal context does not change the generalised Tukey depth based on this nominal hierarchical formal context. In other words, the generalised Tukey depth ignores the further information of the data.
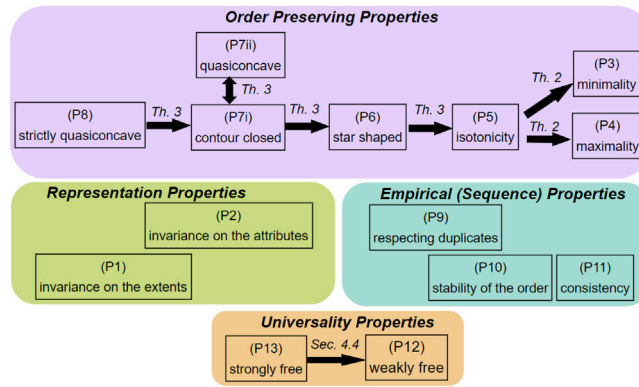
Fig. 1. Overview of the structural properties together with their mathematical connections. *Th* is Theorem and *Sec* is Section for short. This overview contains all direct implications between the properties. Further results like, e.g., limitations of the properties, see, e.g., Theorem 4, are not presented in the Figure.

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 $\mathbb{K} \in \varkappa_G$, $\tilde{\mathbb{K}} \in \tilde{\varkappa}_G$, $\mathrm{Pr}_G \in \mathscr{P}_G$ and $\tilde{\mathrm{Pr}}_G \in \tilde{\mathscr{P}}_G$. Then $D_G(\cdot, \mathbb{K}, \mathrm{Pr}_G)$ and $\tilde{D}_G(\cdot, \tilde{\mathbb{K}}, \tilde{\mathrm{Pr}}_G)$ on G are *isomorph* if and only if there exists a bijective and bimeasureable function $i : G \to G$ such that

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

is true for all $g, \tilde{g} \in G$. We call a bijective function *bimeasurable* if and only if i and i^{-1} are measurable w.r.t. the corresponding σ-fields. 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 $\mathrm{Pr}^{(n)}$ instead of D_G, $D_G^{(n)}$, T_G, $T_G^{(n)}$, $\varkappa_G$, $\mathscr{P}_G$, Pr_G and $\mathrm{Pr}_G^{(n)}$ in the following if the underlying object set G is clear.

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

After this general definition of depth functions based on FCA, 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 FCA. 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 [1,5,6,25] are concerned about these questions for data depth functions defined on $\mathbb{R}^d$. Furthermore, [7,26] 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, [5,6] 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 [5] assumes that for a probability measure that is symmetric around a point c, see [27], 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 [5,6]. A point p_1 which 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 [27]. Slightly different to [5,27] is the approach of [6]. The properties given by [6] put emphasis on the structure of the underlying spaces and not on the probability measure. Thus, in contrast to e.g., [5,27], the definition of centrality of a point is more detached from the notion of symmetry/centrality of the underlying probability measure. Note that [6] 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 [6].

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 [5] and [6] for depth functions in $\mathbb{R}^d$, we define *structural properties* of a depth function based on FCA 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 FCA 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 [6, p. 19] has a natural translation to FCA by use of the closure system, see Property (P7i and ii). Other structural properties introduced discuss the possibility that the data itself may have a center-outward order. This inherited order is then represented by the formal context. For example, consider the extreme case that an object lies in every set of the closure system. This means that it has every attribute that is true for any element of the data set. Thus, this object is implied by every other object and is therefore more specific than any other object. So it should have maximum depth for any possible probability measure, see Property (P4). Of course, the other extreme case, where an object lies only in the extent corresponding to the entire set and therefore shares no attributes with the other data elements, should have minimal depth by default, see Property (P3). Note that these two extremes do not necessarily occur. For example, there is no such element in $\mathbb{R}^d$ together with the convex sets. All in all, one can say that the introduced structural properties are defined and discussed within two perspectives: The transfer of already existing properties in $\mathbb{R}^d$ and the new development of properties based on the theory of FCA, reflecting the inherited center-outward structure of the data.

In total 13 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 Fig. 1. 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 6 we check these properties based on the introduced generalised Tukey depth, see Section 4.

5.1. 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. In Section 2 we motivated this property by illustrating that the observation values are represented within the arrangement of the different groups of the passengers/observations. Thus, if neither the groups nor the probabilities of these groups change, then the depth value should also remain the same. This can be seen as an adaptation of the "affine invariance" property of depth functions in $\mathbb{R}^d$, see [5, 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 FCA, two objects with the same attributes are duplicates. Therefore they should be assigned to the same depth value. In the formal context of the Titanic example, see Example 1, this means that two passengers of exactly the same age, sex and class of travel must have the same depth values.

(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 $\Psi(\{g_1\}) = \Psi(\{g_2\})$,

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

holds.

Remark 1. Property (P1) is the only property that directly discusses the connection between two different formal contexts. This is due to the fact that even just adding more attributes can strongly change the implications and the whole data structure representation, even though this procedure "only" adds more extent sets. Therefore, in general, it is not clear how the depth function should behave when more attributes are added. However, for example for interordinal scaling, which is "finer" than ordinal scaling, see Example 1, it can be shown that the following Property (P7i and ii) together with Property (P1) in many cases already cover the desired changes in the order of the data elements. More precisely, for ordinal scaling, Property (P7i and ii) ensures that the smallest

value has maximum depth, while for interordinal scaling it gives more flexibility and, depending on the underlying probability measure, any element can have the highest depth value. In particular, in many cases (e.g. uniform distribution) Properties (P1) and (P7i and ii) imply that the classical median has the highest depth. For the generalised Tukey depth, we always obtain the classical median as the highest depth for interordinal scaling.

5.2. 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. 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 [5, p. 463], and the "quasiconcavity" properties, see [6, 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, we can build a bridge to FCA. In particular, the "quasiconcavity" property, see [6, 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. 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 $\mathbb{K} \in \varkappa, \Pr \in \mathscr{P}$. 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

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

is true.

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

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

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 5 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 [6, 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 FCA, an object g_2 which lies in the closure of an other object g_1 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 $\Pr \in \mathscr{P}$ and formal context $\mathbb{K} \in \varkappa$ with $g_1, g_2 \in G$ such that $\gamma_{\mathbb{K}}(\{g_1\}) \supseteq \gamma_{\mathbb{K}}(\{g_2\})$,

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

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 [5, 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 [6, 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. This gives the name of the property. With the definition based on the convex closure system, we can easily transfer this property to FCA and obtain Property (P6).

(P6) *Starshapedness:* Let $\mathbb{K} \in \varkappa$ and $\Pr \in \mathscr{P}$. 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

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

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 3, 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

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

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. Note that a center c is not always naturally given. Besides the spatial data, see Example 5, where one can use the probability measure to get a center, in the framework of FCA one can say that g_{all} in Property (P4) could be such a center. In contrast, in the example of the Titanic data, see Table 1, there is no passenger who is naturally a center. Each passenger has at least one component where their observation is not in the center (median for age and modus for sex and class of travel).

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 5, 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 $\mathrm{Pr} \in \mathscr{P}$. We assume that Pr is point symmetric around $s \in G$ in the following sense: There exists a bimeasurable involutory function $i : G \longrightarrow G$ (i.e. $i(i(g)) = g$ for all $g \in G$) such that*

1. *$\mathrm{Pr}(i(E)) = \mathrm{Pr}(E)$ 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 the second assumption of this theorem we get $s \in \gamma(\{c, i(c)\})$. Because D is starshaped, we have $D(s, \mathbb{K}, \mathrm{Pr}) \geq D(i(c), \mathbb{K}, \mathrm{Pr})$. Now, we use that there exists a bimeasurable involutory function i and that D is invariant on the extents (P1) to get $D(i(c), \mathbb{K}, \mathrm{Pr}) = D(c, \mathbb{K}, \mathrm{Pr})$. With this, we conclude that $D(s, \mathbb{K}, \mathrm{Pr}) \geq D(c, \mathbb{K}, Pr)$. This proves that s has maximal depth. $\square$

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 $\mathrm{Pr} \in \mathscr{P}$. 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 $\mathrm{Pr} \in \mathscr{P}$. 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 $D(g_1, \mathbb{K}, \mathrm{Pr}) \leq D(g_2, \mathbb{K}, \mathrm{Pr})$.

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, $D(g_{all}, \mathbb{K}, \mathrm{Pr}) \geq \max_{\{g \in G\}} D(g, \mathbb{K}, \mathrm{Pr})$ 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 $D(g_{non}, \mathbb{K}, \mathrm{Pr}) \leq D(g, \mathbb{K}, \mathrm{Pr})$ for every $g \in G$. This gives us Property (P3). $\square$

Remark 2. Let us point out some consequences of Property (P6). In contrast to [5, 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 3 do not have an object g_{non} and g_{all}.

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., [6, 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. In this case, there is a direct transfer to FCA by the extent set as a closure system. Therefore, we first have to define the contour sets within the theory of FCA. Let $\mathbb{K} \in \varkappa$ be a formal context and $\Pr \in \mathscr{P}$ be a probability measure. For $\alpha \in \text{im}(D(\cdot, \mathbb{K}, \Pr))$ 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 countourclosed 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 $\alpha \in \text{im}(D(\cdot, \mathbb{K}, \Pr))$ the contour set $Cont_\alpha$ is an extent of the formal context $\mathbb{K}$.

(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.

Recall Example 1. We showed that a group containing g_3 and g_4 must also contain g_2. Property (P7i and ii) now formalises the statement in Example 1 that the depth of g_2 is at least as high as the minimum depth of g_3 and g_4. In Example 6 we saw that this is true for the generalised Tukey depth. In the next section we will prove that Property (P7i and ii) holds in general for the generalised Tukey depth, but that this is not true for Property (P8).

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 (P7i 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 (P7ii) 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 $\alpha \in \text{im}(D(\cdot, \mathbb{K}, \Pr))$ 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).

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 $D(\tilde{g}, \mathbb{K}, \Pr) \geq \min\{D(c, \mathbb{K}, \Pr), D(g, \mathbb{K}, \Pr)\} \geq D(g, \mathbb{K}, \Pr)$. 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 $D(g_1, \mathbb{K}, \Pr) \geq D(g_2, \mathbb{K}, \Pr)$. This shows Property (P5). □

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 [14] and Example 2. 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 (left). 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. We have $\gamma(\{g_i, g_j\}) = \{g_1, g_2, g_3\}$ for all $i, j \in \{1, 2, 3\}$ with $i \neq j$. With this, we obtain that $\min\{D(g_1, \mathbb{K}, \Pr), D(g_2, \mathbb{K}, \Pr)\} < D(g_3, \mathbb{K}, \Pr)$. W.l.o.g. let $D(g_1, \mathbb{K}, \Pr) < D(g_3, \mathbb{K}, \Pr)$. Together with $\min\{D(g_2, \mathbb{K}, \Pr), D(g_3, \mathbb{K}, \Pr)\} < D(g_1, \mathbb{K}, \Pr)$ due to Property (P8), we have $D(g_2, \mathbb{K}, \Pr) < D(g_1, \mathbb{K}, \Pr) < D(g_3, \mathbb{K}, \Pr)$. However, this is a contradiction to $\min\{D(g_1, \mathbb{K}, \Pr), D(g_3, \mathbb{K}, \Pr)\} < D(g_2, \mathbb{K}, \Pr)$. Hence,

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 (left), Property (P7ii) leads to a constant depth function.

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:

$$\mathscr{C}^{\mathcal{P}8} = \left\{ (G, M, I) \text{ formal context} \,\middle|\, \begin{matrix} \text{exist } A, \tilde{A} \subseteq G \text{ such that } \#A < \infty, \#\tilde{A} < \infty \text{ and } A \cap \tilde{A} = \emptyset \\ \text{and } A \subseteq \gamma(\tilde{A}), \, \tilde{A} \subseteq \gamma(A) \end{matrix} \right\}.$$

Theorem 4 proves that for every $\mathbb{K} \in \mathscr{C}^{\mathcal{P}8}$ 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 $\mathbb{K} \in \mathscr{C}^{\mathcal{P}8}$ and every $\mathrm{Pr} \in \mathscr{P}$ there exists no depth function D such that Property (P8) is fulfilled.*

Proof. Let $\mathbb{K} \in \mathscr{C}^{\mathcal{P}8}$. 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 $\forall a \in A \setminus \tilde{A} : D(a, \mathbb{K}, \mathrm{Pr}) > \inf_{\tilde{a} \in \tilde{A}} D(\tilde{a}, \mathbb{K}, \mathrm{Pr})$ and $\forall \tilde{a} \in \tilde{A} \setminus A : \inf_{a \in A} D(a, \mathbb{K}, \mathrm{Pr}) < D(\tilde{a}, \mathbb{K}, \mathrm{Pr})$. 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 $D(a_m, \mathbb{K}, \mathrm{Pr}) > D(\tilde{a}_m, \mathbb{K}, \mathrm{Pr})$ and $D(a_m, \mathbb{K}, \mathrm{Pr}) < D(\tilde{a}_m, \mathbb{K}, \mathrm{Pr})$. 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 FCA, 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.

All these order preserving properties aim at representing the underlying data structure in the depth function. This includes structures that are also valid for $\mathbb{R}^d$, e.g. quasiconcavity, but also the extreme cases where a natural center or outlying point is given by the data structure itself, see maximality and minimality property. This emphasises that non-standard data can be very different from $\mathbb{R}^d$, and these properties aim to preserve the different structures. In the next section, we focus on the set of probability measures $\mathscr{P}$ directly.

5.3. 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 $\mathrm{Pr}^{(n)}$ induced by different samples $g_1, \dots, g_n \in G$ with $n \in \mathbb{N}$. We assume that $\mathrm{Pr}^{(n)} \in \mathscr{P}$ 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 samples differ in a specific manner. The first one considers the influence of duplicates in the sample in comparison to deleting the duplicates. In this case, the depth value of the duplicated object should be higher based on the sample where the duplicates 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. Note that there are scaling methods that exclude the existence of an object g_{diff}. In the Titanic example, see Table 1, we used the interordinal scaling method to include the numerical observation. This tells us that for two objects/passengers there must be at least one attribute that both objects have in common. Conversely, if ordinal scaling is used instead in the Titanic example, object g_{diff} may occur.

(P9) *Respecting duplicates*: Let $\mathbb{K} \in \varkappa$. Let $g_1, \dots, g_n$ be a 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

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

Then

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

(P10) *Stability of the order*: Let $\mathbb{K} \in \varkappa$ and let $g_1, \ldots, g_n$ be a 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 from 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. Then the center-outward order of $g_1, \ldots, g_{i-1}, g_{i+1}, \ldots, g_n$ given by $D^{(n)}$ and $D^{(n,-i)}$ are the same. Recalling Definition 8 and the notation in Property (P9), we get

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

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 effect the empirical probability measure of the extents. First, let us consider the case of Property (P9) 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, the probability reflects the duplicates for every extent which contains g_1 (and therefore also g_2).

Let us now consider Property (P10) and let $g_1, \ldots, g_n$ be a sample with $g_i = g_{diff}$ the greatly different object w.r.t. 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 $\Pr^{(n)}(E) = 1/n \neq 0 = \Pr^{(n,-i)}(E)$.
- Case 2: $g_\ell \in E$ for some $\ell \in \{1, \ldots, i-1, i+1, \ldots, n\}$. Then $g_{diff} \notin E$ and $\Pr^{(n)}(E) = ((n-1)/n)\Pr^{(n,-i)}(E)$.
- Case 3: $g_\ell \notin E$ for all $\ell \in \{1, \ldots, n\}$. Then $\Pr^{(n)}(E) = 0 = \Pr^{(n,-i)}(E)$.

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 $\Pr^{(n,-i)}$. 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 (P9) and (P10) 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 independent and identical distributed (i.i.d.) 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 (P11) discusses the consistency based on the empirical depth function towards the (population) depth function.

(P11) *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 i.i.d. sample $g_1, \ldots, g_n$ of G with $n \in \mathbb{N}$ which is drawn based on $\Pr$. Then,

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

5.4. 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. The idea behind these properties is now the wish for a depth function to be *as strictly quasiconcave as possible*. While the following 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 quasiconcave 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 than 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

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

In other words, there exists a pair $(A, g) \in quasiker(D(\cdot, \mathbb{K}, \Pr)) \setminus quasiker(E(\cdot, \mathbb{K}, \Pr))$ 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 $(A, g) \in quasiker(D(\cdot, \mathbb{K}, \Pr))$ as well as $(\tilde{A}, \tilde{g}) \in quasiker(D(\cdot, \mathbb{K}, \Pr))$. Then it may be the case that increasing the depth value $D(g, \mathbb{K}, \Pr)$ by some small amount ε leads to a depth function $E(\cdot, \mathbb{K}, \Pr)$ that is still quasiconcave and that now fulfils $quasiker(E(\cdot, \mathbb{K}, \Pr)) \subsetneq quasiker(D(\cdot, \mathbb{K}, \Pr))$. 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 $E(\cdot, \mathbb{K}, \Pr)$ from D as a composition of $D(\cdot, \mathbb{K}, \tilde{\Pr})$ and an isotone function $f : \mathbb{R} \to \mathbb{R}$. Here, $\tilde{\Pr}$ 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 $\tilde{\Pr}$. With this idea in mind we now state two different notions of freeness, one weak and one strong notion. Therefore, let $Q \subseteq \{(P1), \dots, (P10)\}$ be a set of properties from Sections 5.1, 5.2 and 5.3 and recall Definition 8. A depth function D is called

(P12) *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 $\tilde{\Pr} \in \mathscr{P}$ and an isotone function $f : \mathbb{R} \longrightarrow \mathbb{R}$ such that

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

(P13) *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 $d(\mathscr{P}^\varepsilon)$ 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 $\Pr \in \mathscr{P}$ there exists a probability measure $\tilde{\Pr} \in \mathscr{P}^\varepsilon$ and an isotone function $f : \mathbb{R} \longrightarrow \mathbb{R}$ with

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

Here, the diameter of a set $\mathscr{P}$ of probability measures is defined as $d(\mathscr{P}) := \sup_{P,Q \in \mathscr{P}} \sup_{A \in \Sigma} |P(A) - Q(A)|$ where Σ is the assumed underlying σ-field.

Remark 3. 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 $\tilde{\Pr}$ and an isotone function f such that $f \circ D(\cdot, \mathbb{K}, \tilde{\Pr})$ 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 $\tilde{\Pr}$ and f depend on $Q, \Pr, \mathbb{K}, D$ and E.

Within the notion weakly free we completely detach from $\Pr$ by allowing $\tilde{\Pr}$ to be any probability measure from $\mathscr{P}$. For the notion strongly free we restrict the allowed probability measures $\tilde{\Pr}$. Therefore, strongly free implies weakly free. The idea behind restricting the probability measure $\tilde{\Pr}$ 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 5. *For the hierarchical nominal context given in Table 2 of Example 4, there exists a depth function D that is strongly free w.r.t. quasiconcavity (and w.r.t. the family of all probability measures on G), namely the function D given by:*

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

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

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 $\mathscr{P}^\varepsilon := \{\Pr \mid \forall g \in G \text{ we have } \Pr(\{g\}) \in [1/4 - \varepsilon/4, 1/4 + \varepsilon/4]\}$. 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 $\tilde{\Pr} \in \mathscr{P}^\varepsilon$ by $\tilde{\Pr}(\{g\}) := 1/4 + \varepsilon$ and $\tilde{\Pr}(\{\tilde{g}\}) := 1/4 - \varepsilon/3$ 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 $f \circ D(\cdot, \mathbb{K}, \tilde{\mathrm{Pr}}) = E(\cdot, \mathbb{K}, \mathrm{Pr})$. Since in particular $\epsilon > 0$ was arbitrary, D is in fact strongly free w.r.t. 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 4. 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 4 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.

6. Example: Generalised Tukey depth function

In Section 5, we generally introduced structural properties for depth functions based on FCA. One purpose of these structural properties is to give a systematic basis to analyse depth functions using FCA. Therefore, in this section, we take the opportunity to analyse the generalised Tukey depth given in Section 4. 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 of single attributes. Thus, in the computation one apparently only takes a proper subset of all possible extents into account, see, e.g., Examples 6 and 7. The reason for this is that considering all extent sets instead of only those induced by a single attribute does not change the depth value. Compare to Section 4 where this has already been mentioned.

Theorem 6. *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*

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

Proof. We have to show that

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

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

Remark 5. Analogously to Theorem 6 one can show that

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

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 $g \notin \varPhi(m)$. In particular, we get for this m that $\mathrm{Pr}(\varPhi(m)) \geq \mathrm{Pr}(A)$ is true which gives indeed equality above.

Remark 6. Theorem 6 shows that the generalised Tukey depth uses only the marginal distribution over the attributes. Thus, it is possible to change the incidence relation I and thereby change the dependency structures of the objects, but the generalised Tukey depth remains the same. Obviously, this statement is particularly true if the marginal distribution remains the same. However, since only the supremum of the extents sets of single attributes is used, even the marginal distribution can change and still the generalised Tukey depth remains the same. An example is given in Table 3 (middle and right). Here, we have $T(g, \mathbb{K}_1, \mathrm{Pr}) = T(g, \mathbb{K}_2, \mathrm{Pr})$ for all $g \in \{g_1, g_2, g_3\}$ (with uniform probability measure on $\{g_1, g_2, g_3\}$), although in the second formal context, we have that g_2 implies g_3.

6.1. 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 6 and Remark 5. 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.

Table 3
For the formal context on the left Property (P8) is never true (see Section 5.2). The middle ($\mathbb{K}_1$) and right ($\mathbb{K}_2$) formal contexts show the effect of the supremum in the definition of the generalised Tukey depth.

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

Table 4
For the formal context on the left holds Property (P8) for the generalised Tukey depth. The formal context does not fulfil Property (P10) for the generalised Tukey depth.

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

6.2. Order preserving properties

In Section 5, the order preserving properties are ordered in their strength. Thus, we prove that the contourclosed property (P7i) holds and with this further order preserving properties follow by Theorems 2 and 3.

Theorem 7. *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 $\alpha \in im(T(\cdot, \mathbb{K}, \text{Pr}))$ 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 for all $m \in M \setminus \Psi(g)$ there exists $\tilde{g} \in Cont_\alpha$ such that $1 - \text{Pr}(\Phi(m)) \geq 1 - \sup_{\tilde{m} \in M \setminus \Psi(\tilde{g})} \text{Pr}(\Phi(m))$. By definition of the contour set α, we get $T(g, \mathbb{K}, \text{Pr}) = 1 - \sup_{m \in M \setminus \Psi(g)} \text{Pr}(\Phi(m)) \geq \inf_{\tilde{g} \in Cont_\alpha} \left(1 - \sup_{\tilde{m} \in M \setminus \Psi(\tilde{g})} \text{Pr}(\Phi(\tilde{m}))\right) = \inf_{\tilde{g} \in Cont_\alpha} T(\tilde{g}, \mathbb{K}, \text{Pr})$. 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. $\square$

Remark 7. Now we apply Theorems 2 and 3, see overview in Fig. 1, 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 5 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 4 (left).

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 for all $\tilde{g} \in \gamma(A) \setminus A$

$$1 - \sup_{m \in M \setminus \Psi(\tilde{g})} \text{Pr}(\Phi(m)) > \inf_{g \in A} \left(1 - \sup_{m \in M \setminus \Psi(g)} \text{Pr}(\Phi(m))\right) \Leftrightarrow \sup_{m \in M \setminus \Psi(\tilde{g})} \text{Pr}(\Phi(m)) < \sup_{g \in A} \sup_{m \in M \setminus \Psi(g)} \text{Pr}(\Phi(m)). \tag{3}$$

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 4 (left) 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 $\text{Pr}(\Phi(m))$ is strictly larger than $\text{Pr}(\Phi(\tilde{m}))$ 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:

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

6.3. 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 Definitions 6 and 7. With this we can immediately follow that Property (P9) is fulfilled by the generalised Tukey depth. The stability of the order property does not hold. Consider again the formal context given by Table 4 (right). Then g_1 is an outlier w.r.t. $\{g_2, g_3\}$. 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 (P11). 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 [28], then the generalised Tukey depth is indeed consistent. This, for example, is given when $\mathcal{ES}$ has a finite VC dimension.

Theorem 8. *Let $\mathbb{K}$ be a formal context with G as object set. Let $Pr^{(n)}$ be the empirical probability measure given by an i.i.d. sample of sizes $n \in \mathbb{N}$ and based on probability measure Pr. When $\mathcal{ES}$ is a Glivenko–Cantelli class w.r.t. Pr, then $T^{(n)}$ is consistent and Property (P11) is fulfilled with*

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

Proof. The proof is similar to the proof given by [4, p. 1816f]. Since we assume that $\mathcal{ES}$ is a Glivenko–Cantelli class w.r.t. Pr, we get $\sup_{E \in \mathcal{ES}} \left| Pr^{(n)}(E) - Pr(E) \right| \xrightarrow{n \to \infty} 0$ almost surely. With this, the claim follows from

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

$$\leq \sup_{E \in \mathcal{ES}} \left| \mathrm{Pr}^{(n)}(E) - \mathrm{Pr}(E) \right| \xrightarrow{n \to \infty} 0 \text{ almost surely.} \quad \square$$

Remark 8. If $\mathcal{ES}$ has a finite VC dimension, then the convergence stated in Theorem 8 is additionally uniform over all possible underlying probability measures. A finite VC dimension of $\mathcal{ES}$ is given, for example, if M is finite.

Remark 9. 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 $\tilde{\mathrm{Pr}}$. 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 5. 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.

6.4. Universality properties

Let us end this discussion on the generalised Tukey depth by considering the universality properties. The next two theorems prove that the generalised Tukey depth is weakly free but not strongly free w.r.t. the quasiconcavity property. Therefore, recall the hierarchical nominal example, see Examples 4 and 8.

Theorem 9. *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 $E(\cdot, \mathbb{K}, \mathrm{Pr})$ 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 $\tilde{\mathrm{Pr}}$ and an isotone function f such that $f \circ T(\cdot, \mathbb{K}, \tilde{\mathrm{Pr}}) = E(\cdot, \mathbb{K}, \mathrm{Pr})$.

To achieve this, we first consider properties this probability function needs to fulfil. Let $\tilde{\mathrm{Pr}} \in \mathscr{P}$ be a possible candidate. We set $p_i := \tilde{\mathrm{Pr}}(G_i)$ and $p_i^{min} := \min_{g \in G_i} \tilde{\mathrm{Pr}}(\{g\})$. Since E is quasiconcave, we have $g \notin \gamma(G_{\ell+1} \cup \cdots \cup G_K)$ 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 $h \in G_{\ell+1} \cup \cdots \cup G_K$. Therefore, we get for $g \in G_\ell$

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

Now, we construct probability measure $\tilde{\mathrm{Pr}}$ needed for Property (P12) in three steps: First set $\tilde{\mathrm{Pr}}(g) = 1$ for $g \in G_1$. Then recursively set $\tilde{\mathrm{Pr}}(g) = \sum_{i=1}^{\ell} \tilde{\mathrm{Pr}}(G_i) + 1$ for $g \in G_{\ell+1}$. Finally, we normalise $\tilde{\mathrm{Pr}}$ by $\tilde{\mathrm{Pr}}(g) := \tilde{\mathrm{Pr}}(g) / \sum_{g \in G} \tilde{\mathrm{Pr}}(g)$ to get a probability measure. Due to construction and Inequality (4), it is ensured that $T(g, \mathbb{K}, \tilde{\mathrm{Pr}}) < T(h, \mathbb{K}, \tilde{\mathrm{Pr}})$ for $g \in G_\ell$ and $h \in G_{\ell+1}$. Thus, one can define function f by setting $f(T(g, \mathbb{K}, \tilde{\mathrm{Pr}})) = E_\ell$ for $g \in G_\ell$ and isotonically extending it to function $f : \mathbb{R} \to \mathbb{R}$. With this, we get $f \circ T(\cdot, \mathbb{K}, \tilde{\mathrm{Pr}}) = E(\cdot, \mathbb{K}, \mathrm{Pr})$ which shows that claim. $\square$

Theorem 10. *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 5 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 $T(\cdot, \mathbb{K}, \mathrm{Pr})$ 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 $\mathrm{Pr}(\{a_1 a_2, a_1 b_2\}) \geq 0.5$. This is because if we have $\mathrm{Pr}(\{a_1 a_2, a_1 b_2\}) < 0.5$, we can conclude that $T(a_1 a_2, \mathbb{K}, \mathrm{Pr}) \leq 1 - \mathrm{Pr}(\{b_1 a_2, b_1 b_2\}) = \mathrm{Pr}(\{a_1 a_2, a_1 b_2\}) < 0.5$. Since either $\mathrm{Pr}(\{b_1 a_2\})$ or $\mathrm{Pr}(\{b_1 b_2\})$ 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

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

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 $\mathrm{Pr}^* \in \mathscr{P}^\varepsilon$ 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 $\tilde{\mathrm{Pr}}(\{a_1 a_2, a_1 b_2\}) \geq 0.5$ and, thus, $\mathrm{Pr}^*(\{a_1 a_2, a_1 b_2\}) \geq 0.5 - \varepsilon = 0.4$. 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 $\mathrm{Pr}^*(\{b_1 a_2, b_1 b_2\}) \geq 0.5 - \varepsilon = 0.4$. But this, together with Inequality (5) implies that $\mathrm{Pr}^*(a_1 b_2) \geq 0.4$. With analogous argumentations, one can also show $\mathrm{Pr}^*(\{g\}) \geq 0.4$ for all other objects. But this is a contradiction to Pr^* being a probability measure because $\mathrm{Pr}^*(G) = 0.4 * 4 = 1.6$. This provides the claim. $\square$

7. Conclusions

This article introduced a general notion of depth functions for non-standard data based on FCA. This covers the analysis of centrality and outlyingness for a wide variety of different data types. In order to enable a discussion, especially about these non-standard data, which are not given in a standard statistical data format, we used FCA and introduced structural properties. In addition to adopting the properties discussed in $\mathbb{R}^d$, we also addressed the issue that non-standard data may inherit a central-outward order that should also be reflected in the depth function. We also provided a framework for analysing depth functions based on FCA. Finally, we used this framework to discuss and analyse the generalised Tukey depth. Building on this, there are several promising areas for further research, including (but not limited to)

Further concrete mapping rules for depth functions: In Theorem 5 we defined a depth function D which is strongly free w.r.t. 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 union-free generic depth function for the special case of G being the set of partial orders, see [19]. This is an adaptation of the simplicial depth in $\mathbb{R}^d$, see [1], to the set of all partial orders. As shortly denoted there this can be generalised to general formal contexts. [15] give another examples for a depth function using FCA. The authors there propose concrete outlier measures based on formal concept analysis. Since there is a strong relationship between outlier measures and depth functions, the properties here can also be used to analyse this measure.

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 [19] we applied the union-free generic depth function, which can be embedded into our concept of depth upon partial orders that 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, see [10] as starting point.

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 [14] 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.

CRediT authorship contribution statement

Hannah Blocher: Writing – original draft, Visualization, Supervision, Project administration, Methodology, Formal analysis, Conceptualization, Writing – review & editing. **Georg Schollmeyer:** Writing – original draft, Supervision, Methodology, Formal analysis, Conceptualization, Writing – review & editing.

Acknowledgements

References

[1] R. Liu, On a notion of data depth based on random simplices, Ann. Statist. 18 (1990) 405–414.

[2] R. Dyckerhoff, K. Mosler, G. Koshevoy, Zonoid data depth: Theory and computation, in: A. Prat (Ed.), COMPSTAT, Physica-Verlag HD, Heidelberg, 1996, pp. 235–240.

[3] 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.

[4] D.L. Donoho, M. Gasko, Breakdown properties of location estimates based on halfspace depth and projected outlyingness, Ann. Statist. 20 (1992) 1803–1827.

[5] Y. Zuo, R. Serfling, General notions of statistical depth function, Ann. Statist. 28 (2000) 461–482.

[6] 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.

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

[8] M. Goibert, S. Clémençn, E. Irurozki, P. Mozharovskyi, Statistical depth functions for ranking distributions: Definitions, statistical learning and applications, in: G. Camps-Valls, F. Ruiz, I. Valera (Eds.), Proceedings of the 25th International Conference on Artificial Intelligence and Statistics AISTATS 2022, in: Proceedings of Machine Learning Research, Vol. 151, PMLR, 2022, pp. 10376–10406.

[9] G. Geenens, A. Nieto-Reyes, G. Francisci, Statistical depth in abstract metric spaces, Stat. Comput. 33 (2023) 46.

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

[11] S. Pawar, D. Shirke, Data depth-based nonparametric tests for multivariate scales, J. Stat. Theory Pract. 16 (2022) 1–21.

[12] P. Mozharovskyi, Anomaly detection using data depth: Multivariate case, 2022, arXiv https://arxiv.org/abs/2210.02851 (Accessed 24 September 2024).

[13] F. Chebana, T.B.M.J. Ouarda, Depth-based multivariate descriptive statistics with hydrological applications, J. Geophys. Res. 116 (2011) 1–19.

[14] 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, Vol. 1602, Springer, Cham, 2022, pp. 17–30.

[15] Q. Hu, Z. Yuan, K. Qin, J. Zhang, A novel outlier detection approach based on formal concept analysis, Knowl.-Based Syst. 268 (2023) 110486.

[16] W. Cukierski, Titanic - machine learning from disaster, 2012, https://kaggle.com/competitions/titanic (Accesses 24 September 2024).

[17] B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer, Berlin, Heidelberg, 2012.

[18] R. Serfling, Y. Zuo, Structural properties and convergence results for contours of sample statistical depth functions, Ann. Statist. 28 (2000).

[19] 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, Vol. 215, PMLR, Oviedo, 2023, pp. 59–71.

[20] C. Carpineto, G. Romano, Concept Data Analysis: Theory and Applications, Wiley, Chichester and Weinheim, 2004.

[21] 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, Vol. 12879, Springer, Cham, Cham, 2021, pp. 105–118.

[22] 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 24 September 2024).

[23] 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 24 September 2024).

[24] T. Tao, An Introduction to Measure Theory, in: Graduate Studies in Mathematics, vol. 126, American Mathematical Society, Providence, RI, 2011.

[25] K. Mosler, P. Mozharovskyi, Choosing among notions of multivariate depth statistics, Statist. Sci. 37 (2022) 348–368.

[26] 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.

[27] Y. Zuo, R. Serfling, On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry, J. Statist. Plann. Inference 84 (2000) 55–79.

[28] M. Dudley, E. Giné, J. Zinn, Uniform and universal Glivenko–Cantelli classes, J. Theoret. Probab. 4 (1991) 485–510.