Statistical Models for Partial Orders Based on Data Depth and Formal Concept Analysis

Hannah Blocher, Georg Schollmeyer, and Christoph Jansen

Department of Statistics, Ludwig-Maximilians-Universität Munich, Germany
{hannah.blocher,georg.schollmeyer,christoph.jansen}@stat.uni-muenchen.de

Abstract. In this paper, we develop statistical models for partial orders where the partially ordered character cannot be interpreted as stemming from the non-observation of data. After discussing some shortcomings of distance based models in this context, we introduce statistical models for partial orders based on the notion of data depth. Here we use the rich vocabulary of formal concept analysis to utilize the notion of data depth for the case of partial orders data. After giving a concise definition of unimodal distributions and unimodal statistical models of partial orders, we present an algorithm for efficiently sampling from unimodal models as well as from arbitrary models based on data depth.

Keywords: partial orders · partial rankings · data depth · formal concept analysis · unimodality · quasiconcavity.

1 Introduction and Motivation

Partial orders play a role in a broad range of scientific disciplines. In many of these disciplines like revealed preference theory, social choice theory, decision making under uncertainty, social-economics (Human Development Index, costumer preference rankings etc.) or statistics and machine learning, studying *partial* orders has attracted more and more researchers (see [27], [5], [15, 16], [9, 22, 17] and [11] for recent works in the respective discipline). Consequently, there are many approaches that can deal with partial orders. However, in most approaches known to the authors, the incompleteness of the involved orders is interpreted as stemming from missing data, see, e.g., [21, 24]. In other words, an explicit missing mechanism is modeled or at least assumed. In contrast, in this paper we explicitly assume that the incompleteness of the order is not due to missing of data. Instead, we understand an observed incomparability between two items as a precise observation of a factual incomparability that actually exists.

Many of the existing models are based on a distance measure on the set of partial orders. They obtain a center-outward ordering of all partial orders with respect to a predefined partial order that represents the center. However, often the distance measures are based on the linear extensions of the partial order and thus do not take into account the incomparable character, but imitate an underlying true linear order. Therefore (and for other reasons given below), in what follows, we use a (statistical) depth function instead which allows an ordering of the partial orders w.r.t all other (observed) orders. Depth functions are commonly used in robust and nonparametric statistics. Based on these depth functions, we present unimodal statistical models for partial orders. Therefore, we first need to define unimodality in the context of partial orders, where we make use of the theory of formal concept analysis, see Section 2 and Section 3. In Section 4, we propose some concrete depth functions, and in Section 5 we introduce an algorithm for sampling from the proposed statistical models. Finally, we give a brief conclusion in Section 6.

To illustrate how the currently used distance measures implicitly mimic the missing mechanism and other counter-intuitive structures, let us start by discussing the current approaches that use distance measures for (partial) orders. There are several proposals for adequately defining a meaningful distance concept between (partial) orders in the literature (cf, e.g., [6, 10]) which can be used to establish distance based statistical models for partial orders. Throughout the paper let $\mathcal{X}$ be a finite ground space with $n \geq 3$ elements and let $\mathscr{P}$ denote the set of all partial orders[1] (i.e., all reflexive, transitive and anti-symmetric binary relations) on $\mathcal{X}$. Two prominent distance measures for partial orders are discussed for example in [6]: The *nearest neighbour* and the *Hausdorff distance*. Both of these distances rely on the idea of first computing the set of all linear extensions of the considered partial orders and then, each in its own manner, generalizing the well-known *Kendall's* τ-distance (see [18]) for linear orders (i.e. counting pairs that are ranked oppositely by the considered orders). However, such an approach has the following counter-intuitive property: The nearest neighbour distance systematically assigns lower distance values if sparse partial orders are involved. The nearest neighbour distance is defined as $d_{NN}(P_1, P_2) := \min_{L_1 \in lext(P_1)} \min_{L_2 \in lext(P_2)} \tau(L_1, L_2)$ for two orders P_1, P_2 where $lext(P)$ denotes the set of all linear extensions of a partial order P and τ denotes the Kendall's τ-distance for linear orders mentioned before. Then it is immediate from the definition that $d_{NN}(\tilde{P}_1, P_2) \leq d_{NN}(P_1, P_2)$ for arbitrary partial orders $\tilde{P}_1 \subseteq P_1$, since this implies $lext(P_1) \subseteq lext(\tilde{P}_1)$ and therefore the minimum is taken over a super-set of the original one. Most extremely, the minimal distance is attained whenever one of the considered partial orders is the trivial one consisting solely of the diagonal $D_{\mathcal{X}} := \{(x, x) : x \in \mathcal{X}\}$, whereas two partial orders differing only in few pairs receive non-minimal distance value. This seems to be a very counter-intuitive property of this generalized distance measure. An analogous line of argumentation applies when the nearest neighbour distance is replaced by the directed Hausdorff hemi-metric $m_H(P_1, P_2) := \max\limits_{L_1 \in lext(P_1)} \min\limits_{L_2 \in lext(P_2)} \tau(L_1, L_2)$. Then, in a dual manner, $D_{\mathcal{X}}$ (if seen as the first argument in the Hausdorff hemi-metric,) has always the maximal distance to other orders whereas a linear order L has always a smaller distance to other orders compared to any other partial order $P \subseteq L$. Similar arguments can be given for the usual symmetrized non-directed Hausdorff dis-

[1] In the sequel, we will also shortly say order instead of partial order.

tance defined by $d_H(P_1, P_2) := max\{m_H(P_1, P_2), m_H(P_2, P_1)\}$. Alternatively, one could directly generalize Kendall's τ to partial orders without looking at linear extensions. This would result for example in one of the two expressions $\tau_s(P_1, P_2) := |\Delta(P_1, P_2)| = |(P_1 \cup P_2) \setminus (P_1 \cap P_2)|$ or $\tau_a(P_1, P_2) := |\{(x, y) \mid x \neq y, (x, y) \in P_1, (y, x) \in P_2\}|$, both, in a way, generalizing the idea of counting pairs that are ranked oppositely by the considered *partial* orders. However, whereas τ_a has the same problem like the nearest neighbour distance, the expression τ_s would lead, as will be shown in Section 2, Example 1, to statistical models that are not completely quasiconcave, which means that it seems to be impossible to build a simple unimodal model with such a distance (cf., Definition 1). Furthermore, τ_s treats pairs which are in the relation and pairs being not in the relation in exact the same way, and one can ask if this is natural. As we will see later, our approach that uses a depth function treats pairs being in the relation or not seemingly differently. (Note that a partial order is transitive but not necessarily negatively transitive, so there is in fact some asymmetry between a pair being in the relation or not.) With these problems in mind, we propose statistical modelling of partial orders based on a depth function. The model idea is analogous to a distance based version of the form $P(X = x) = C_\lambda \cdot \Gamma(\lambda \cdot d(\mu, x))$, where, C_λ is a normalizing constant, $d : \mathscr{P} \times \mathscr{P} \longrightarrow \mathbb{R}_{\geq 0}$ is a distance, $\Gamma : \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{\geq 0}$ is a (weakly decreasing) decay function, $\mu \in \mathscr{P}$ is a location parameter and $\lambda \in \mathbb{R}_{>0}$ is a scale parameter. Now, instead of a distance function, in this paper we work with a depth function and a corresponding statistical model given by

$$P(X = x) = C_\lambda \cdot \Gamma(\lambda \cdot (1 - D^\mu(x))) \tag{1}$$

where now D^μ is a depth function that is maximal at partial order μ. Since depth functions are usually only used for data in $\mathbb{R}^d$ we have to adapt the notion of data depth to partial order data, for which we use formal concept analysis.

2 Formal Concept Analysis, Data Depth and Unimodality

In this section we only touch a few aspects about the theory of formal concept analysis and we refer the reader to [14] for more details. The basis of formal concept analysis is the definition of a **formal context** $\mathbb{K} = (G, M, I)$ which is a generalization and formalization of a cross table. Here, G is a set of objects, M a set of binary attributes and $I \subseteq G \times M$ a relation. We say that an object g has an attribute m if $(g, m) \in I$ is true. For example cross table 1 describes a formal context with $G = \{\mu, g, h, i\}$, $M = \{m_1, \ldots, m_6\}$ and the relation I is given by the crosses. By the use of the following **derivation operators**, we obtain a description of the relation between the object and attribute set:

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

Here $\Psi(A)$ contains all the attributes that each object in A has, and $\Phi \circ \Psi(A) \subseteq G$ are all objects that have all attributes in $\Psi(A)$. The tuple $(\Phi \circ \Psi(A), \Psi(A))$ for

$A \subseteq G$ is called a **formal concept**, $\Phi \circ \Psi(A)$ its **extent**, and $\Psi(A)$ its **intent**. The construction of the two derivation operators allows to determine the relation I when the set of all concepts is known. Note that $\Psi(A) = \Psi \circ \Phi \circ \Psi(A)$ holds, and thus each concept is uniquely described by its extent or intent. Moreover, the set of extents and the set of intents yield a closure system with $\Phi \circ \Psi$ and $\Psi \circ \Phi$, respectively, the corresponding closure operator. Note that if $A \subseteq G$ lies in an extent E, then the closure operator $\Phi \circ \Psi$ ensures that every object having all attributes of $\Psi(A)$ is also an element of E. Thus, $A \subseteq E$ implies that $\Phi \circ \Psi(A) \subseteq E$. With this, we say that the pair $A, B \subseteq G$ is an **(object) implication** (we denote this by $A \to B$) if $\Phi \circ \Psi(A) \supseteq \Phi \circ \Psi(B)$ holds. Moreover, one can show that the set of all implications that follow from the extent set completely describe the extent set itself. Within this paper, we use formal implications between objects to model a notion of betweenness. For example $\{g, h\} \longrightarrow \{i\}$ can be interpreted as "object i lies between object g and object h" (or "object i lies in the space that is spanned by the objects g and h"), because object i has all attributes that are shared by both g and h. (Note that we do not restrict the premise of a formal implication to have exactly two objects.) For further discussion of a family of implications, see [2] and [14]. If non-binary attributes are considered, then they are converted into a set of binary attributes by using a so-called conceptual scaling method (see Section 3).

Our approach is to represent the set of partial orderings by a formal context and, using the properties of a formal context, to define the notion of unimodality and depth function. By using the following properties that a function $f\colon G \to \mathbb{R}$ can satisfy on a formal context $\mathbb{K}$, we define the notion of unimodality.

Definition 1. *Let $\mathbb{K} = (G, M, I)$ be a formal context and let $f : H \longrightarrow \mathbb{R}$ with $H \subseteq G$ be a function. Then f is called*

i) ***isotone*** *if for all $g, h \in H$ we have $\{g\} \longrightarrow \{h\} \implies f(g) \leq f(h)$;*
ii) 2-***quasiconcave*** *if for arbitrary objects $g, h, i \in H$ we have $\{g, i\} \longrightarrow \{h\} \implies f(h) \geq \min\{f(g), f(i)\}$;*
iii) ***completely quasiconcave*** *if for every finite set of objects $\{g_1, \ldots, g_n\} \subseteq H$ we have $\{g_1, \ldots, g_{n-1}\} \longrightarrow \{g_n\} \implies f(g_n) \geq \min\{f(g_1), \ldots f(g_{n-1})\}$;*
iv) ***strongly quasiconcave*** *if for every finite set $\{g_1, \ldots, g_n\} \subseteq H$ of size $n \geq 2$ we have $\{g_1, \ldots, g_{n-1}\} \longrightarrow \{g_n\} \implies f(g_n) > \min\{f(g_1), \ldots f(g_{n-1})\}$;*
v) ***star-shaped*** *if there exists a center $c \in H$ such that for all $g \in H$ we have $\{c, g\} \longrightarrow \{h\} \Longrightarrow f(h) \geq \min(f(c), f(g))$.*

Additionally, a probability measure P on a finite G is called ***unimodal (strictly unimodal)*** *if its probability function, restricted to its support $\{g \in G \mid P(\{g\}) > 0\}$, is completely quasiconcave (strongly quasiconcave).*

In general, depth functions measure outlyingness and centrality of an observation w.r.t. a data cloud or an underlying probability measure. We apply the concept of data depth to partial order data represented by a formal context and we denote it by $D\colon G \to \mathbb{R}_{\geq 0}$. Note that it depends on the formal context. Moreover, if we ensure that the depth function is completely quasiconcave (strongly quasiconcave), then the statistical model given in (1) is unimodal (strictly unimodal).

Our notion of quasiconcavity is an adaption of classical quasiconcavity which was already used (e.g., in [23]) for classical data depth for $\mathbb{R}^d$. In particular, here we emphasize (complete) quasiconcavity because it most adequately renders the idea of an unimodal distribution of partial orders that would be induced by a statistical model that uses a quasiconcave depth function: Quasiconcavity would ensure that we have no point that is a local minimum of the probability function w.r.t. the notion of betweenness that is appropriate for a formal concept analysis view on partial orders. Another nice feature of complete quasiconcavity is the fact that this property is equivalent to the property that the upper level sets $D_\alpha := \{g \in G \mid D(g) \geq \alpha\}$ of the depth function D are extents. Thus, every upper level set can be nicely described by a formal concept which makes them descriptively accessible, especially the fact that they cannot only be exactly described by objects, but also by attributes, is very convincing.

Example 1. Let $\mathbb{K} = (G, M, I)$ be given by cross table 1. Then, the depth function D^μ with mode μ given by $D^\mu(g) := |\Psi(\mu) \cap \Psi(g)|$, together with the conceptual scaling of Section 3 can be shown to be exactly the depth-based formulation of a distance based approach with τ_s. It is 2-quasiconcave but in general not completely quasiconcave and therefore is not appropriate to define a unimodal distribution. Note that for arbitrary contexts, D^μ is generally not 2-quasiconcave. Note further that D^μ is at least star-shaped for arbitrary contexts. Furthermore, a generalization of Tukey's depth $\mathcal{T}$ (cf., [25]) and a localized version of Tukey's depth $\mathcal{T}^\mu$ with mode μ can be defined via

$$\mathcal{T}(g) := 1 - \max_{m \in M \setminus \Psi(\{g\})} \frac{|\Phi(\{m\})|}{|G|}; \quad \mathcal{T}^\mu(g) := 1 - \frac{\max\limits_{\substack{m \in M \setminus \Psi(\{g\}),\\ \mu I m}} |\Phi(\{m\})|}{|G|}, \quad (2)$$

respectively. (Here the empty maximum is defined as 0.) Both $\mathcal{T}$ and $\mathcal{T}^\mu$ are completely quasiconcave functions.

	m_1	m_2	m_3	m_4	m_5	m_6
μ		x	x	x	x	x
g		x				
h	x				x	x
i	x		x	x		

Table 1. Illustration of the difference between complete and 2-quasiconcavity.

3 Formal Context Defined by All Partial Orders

In our case the set G is exactly the set $\mathscr{P}$ of all partial orders on $\mathcal{X}$. Note that we regard a partial order not necessarily as a linear order together with a missing mechanism. Therefore, as attributes we also include the property of being incomparable pairs and get

$$M := \underbrace{\{\text{``}x_i \leq x_j\text{''} \mid i,j = 1,\dots,n,\, i \neq j\}}_{=:M_{\leq}} \cup \underbrace{\{\text{``}x_i \not\leq x_j\text{''} \mid i,j = 1,\dots,n,\, i \neq j\}}_{:=M_{\not\leq}}.$$

Since we consider only reflexive relations the attributes "$x_i \leq x_i$" and "$x_i \not\leq x_i$" are redundant and therefore not included here. Note that each order g has $n(n-1)$ many attributes $B = \Psi(\{g\})$ which can be divided into the set $B_{\leq} \subseteq M_{\leq}$ and $B_{\not\leq} \subseteq M_{\not\leq}$. In particular, we have that either (x_i, x_j) lies in g or not and thus we can conclude $(g, \text{``}x_i \leq x_j\text{''}) \in I \Leftrightarrow (g, \text{``}x_i \not\leq x_j\text{''}) \notin I \;\&\; (g, \text{``}x_j \not\leq x_i\text{''}) \in I$. This means if a pair (x_i, x_j) exists then the attribute "$x_i \not\leq x_j$" cannot hold, but "$x_j \not\leq x_i$" must be true. The same is true for the reverse. Indeed, ensuring that a pair $(x_i, x_j), i \neq j$ is in an order g or not has a different strength of restriction, i.e., if we assume that $(x_i, x_j) \in g$, then $g^{-1} := \{(x_j, x_i) \mid (x_i, x_j) \in g\}$ satisfies the condition $(x_i, x_j) \notin g^{-1}$. Thus, the number of orders $\tilde{g}$ fulfilling the condition $(x_i, x_j) \notin \tilde{g}$ is larger than the number of orders g fulfilling $(x_i, x_j) \in g$. Additionally, because of symmetry these numbers are independent of the concrete pair (x_i, x_j).

First let us go one step back and consider the formal context given only by the attribute set $M_{\leq}$. In this case, for an isotone depth function D and two orders g, h with $g \subseteq h$ we have $D(g) \leq D(h)$. Thus, we would obtain again a depth concentration on linear orders. Furthermore, if the depth function is additional 2-quasiconcave and we consider the space of all partial orders, then at least half of all partial orders must have equal depth. More precisely, the depth must be minimal. To see this, let g be an order and let g^{-1} be the inverse order. We obtain that $\{g, g^{-1}\} \longrightarrow G$ and therefore either the depth of g or the depth of g^{-1} is minimal. Thus, because the map $g \mapsto g^{-1}$ is a bijection, half of all orders have minimal depth. Note that the stronger property of strong quasiconcavity cannot be fulfilled by any depth function: Assume we have four orders $g_1, \dots, g_4$ where all pairs of orders have no attribute $m \in M_{\leq}$ in common. Then $\{g_1, g_2\} \longrightarrow G$ and therefore $\min\{D(g_1), D(g_2)\} < D(g_i), i = 3, 4$. But since $\{g_3, g_4\} \longrightarrow G$ this is a contradiction to $\min\{D(g_3), D(g_4)\} < D(g_i), i = 1, 2$. Note that for $|\mathcal{X}| \geq 3$ there exist four linear orders fulfilling this property. Let us now return to the formal context given by the entire attribute set M.

Then, the same argument for the non-existence of a strongly quasiconcave depth function from above would still apply for the extended attribute set M. Beyond this, now the context defined here contains no two different orders g and h such that $\Psi(g) \subseteq \Psi(h)$. Thus, for an isotone depth, isotonicity alone does not imply that the depth value of one order is constrained by the depth value of any other order. In contrast, 2-quasiconcavity would still lead to some restriction on the depth function: Let g be an order such that the complement order (i.e. $g^c := \mathcal{X} \times \mathcal{X} \setminus g$) is also an order. Then one of the orders must have minimal depth, since $\{g, g^c\} \longrightarrow G$. If we take G as the set of all partial orders, then examples of such orders are exactly the linear orders. If, in contrast, one had chosen G as the set of all quasiorders, then exactly all negatively transitive orders g would have the property that also g^c is in G and therefore one of g or g^c would have minimal depth.

4 Specifying Unimodal Distributions of Partial Orders

In this section, we discuss methods for generating unimodal distributions of partial orders based on three concrete depth functions. Firstly, we will discuss Tukey's depth defined by equation (2). Secondly, we define a generalization of the convex hull peeling depth (see [4]), which we will call peeling depth, here. It is sometimes said that the convex hull peeling depth has the disadvantage that it can only order the data points from outwards to inwards. In contrast, in our situation, we are able to directly specify a mode of the distribution and therefore we know beforehand, where 'the inwards', i.e., the mode, is exactly located. With this, we can in fact order the data points from inwards to outwards by starting from the mode and successively enclosing further layers. Thus, thirdly, we can define a new depth function that we call enclosing depth, here. The generalization of Tukey's depth for data values or probability distributions on arbitrary complete lattices or formal contexts was introduced in [25] and applied to the case of ranking data in [26]. The definition is given in equation (2). Before discussing all three data depths, we firstly define the remaining two:

Definition 2. *Define the* ***peeling depth*** $\mathcal{P}$ *by* $\mathcal{P}(extr(G)) := \frac{1}{|G|}$ *and*

$$\mathcal{P}\left(extr\left(G\backslash\mathcal{P}^{-1}\left(\left[0, \frac{i}{|G|}\right]\right)\right)\right) = \frac{i+1}{|G|}, \quad i = 1, 2, \ldots$$

Additionally, define the ***localized peeling depth*** $\mathcal{P}^\mu$ *w.r.t. mode* $\mu \in G$ *simply by adding a high enough amount of objects which have exactly the same attributes as* μ *to the original context* G*. The operator extr is here the extreme point operator which maps a set* A *to the set of all its extreme points.*[2] *Note that this definition is only well defined if the underlying context is meet-distributive.*[3] *Furthermore, let us define the* ***enclosing depth*** $\mathcal{E}^\mu$ *w.r.t. mode* μ *by* $\mathcal{E}^\mu\left((\Phi \circ \Psi)(\{\mu\})\right) = 1$ *and*

$$\mathcal{E}^\mu\left(encl\left((\mathcal{E}^\mu)^{-1}\left(\left[\frac{i}{|G|}, 1\right]\right)\right)\right) = \frac{i-1}{|G|}; \quad i = |G|, |G|-1, \ldots$$

[2] A point $g \in A$ is an extreme point of A if $A\backslash\{h \in G \mid \Psi(\{h\}) = \Psi(\{g\})\} \nrightarrow \{g\}$.

[3] A context is called meet-distributive, if every extent is generated by all extreme points of the extent. In our situation, the underlying context is not meet-distributive, but it is possible to replace the extreme point operator by another appropriate operator that maps a set A to a set $B \subseteq A$ that implies A and is minimal w.r.t. this property. Note that for such an operator the obtained depth function is generally not quasiconcave anymore. Another possibility would be the operator $\widetilde{extr}(A) := extr(A) \cup A\backslash(\Phi \circ \Psi)(extr(A))$. This operator would lead to a completely quasiconcave depth function. Note further that this operator is generally not minimal which means that the number of depth layers is usually lower compared to a minimal operator.

Here, encl denotes an operator which we would like to call an enclosing operator. Concretely, we have in mind an operator encl $: H \longrightarrow 2^G$ *with* $H \subseteq 2^G$ *that for all* $A \in H$ *satisfies the three properties i): encl*$(A) \cap A = \emptyset$*, ii): encl*$(A) \longrightarrow A$ *and iii):* $(\Phi \circ \Psi)(encl\,(A))$ *is minimal w.r.t. properties i) and ii).*

Now we discuss, how one can specify with the above depth functions a unimodal distribution of orders with a given mode and one scale parameter. The simplest distribution, which can be always defined in a finite setting, is the uniform distribution. To specify a distribution that is in some certain sense distributed around a given mode, one simple approach would be to first generate every partial order exactly one time (this would correspond to a uniform distribution) and then to simply add a big amount of partial orders that are identical to the mode. Then, based on the corresponding data depth that is obtained for this data set, one can define a distribution according to equation (1). (Note that generally the obtained distribution is different from a mixture of a uniform distribution and a distribution that equals the mode with probability one.) However, for Tukey's depth, due to reasons of symmetry one can show that the obtained distribution of orders would assign the mode one probability p and every other order that differs from the mode exactly one of two probability values q or r. More concretely, the localized Tukey's depth could then be written as $\mathcal{T}^{\mu}(g) = 1 - \max\left\{\max_{(p,q)\in\mu\setminus g} \alpha_{p,q}, \max_{(p,q)\in g\setminus\mu} \beta_{p,q}\right\}$ with[4] $\alpha_{p,q}, \beta_{p,q} \in [0,1]$, where actually $\alpha_{p,q}$ and $\beta_{p,q}$ do not depend on p or q. This seems to be somehow unsatisfying. Of course, one can use Tukey's depth based on another (empirically or analytically) given distribution, but then, in the first place one is back at the "... major outstanding problem in ranking theory ..." and has to specify a "... suitable population of ranks in non-null cases..." ([19]). Alternatively, one can replace $\alpha_{p,q}$ and $\beta_{p,q}$ by other weights that depend on the pairs (p,q), actually fortunately without losing the quasiconcavity. For this weighted Tukey-type depth function one would have to specify only n^2 values instead of $2^{n^2/4}$ or more values (cf., [20]), which would be needed for a completely nonparametric approach. Because this can still be very demanding, we will later use an analysis of the enclosing depth to get a rough guidance for specifying the weights. For the peeling depth there seems to be not so much ties compared to Tukey's depth. However, it seems a little bit counter-intuitive to specify a distribution of orders that are distributed around a mode by not locally looking at a neighbourhood of the mode but instead by globally ordering the data points from outwards to inwards. Compared to other applications of data depth where one does not know the location beforehand but where the problem is actually the estimation of the mode of the distribution, here we are in the comfortable situation that we can simply specify the mode. Therefore, in the sequel, we will focus on the enclosing depth (applied for the case $G := \mathscr{P}$). Also here, because of the high amount of symmetries there are many ways of defining an enclosing operator and corresponding depth layers. One way out of this would be to compute in a first step

[4] This also shows an asymmetry between pairs that are in relation and pairs that are not in relation.

for every partial order the expected depth value under a stochastic choice of the layer that is built in every step. This is of course possible and also a simulation from such a model can be exactly done. However, the obtained depth function is not completely quasiconcave. Therefore, one can in a second step build the closure of every depth contour to obtain a completely quasiconcave depth function. For this, one has to analyze in detail, how the expected depth values of the first step exactly look like, which seems to be a very difficult problem. Therefore, we only analyze the situation for total orders and a totally ordered mode μ under a conceptual scaling of the partial orders that uses only $M_{\leq}$. With this analysis we are able to roughly oversee the situation for the enclosing depth and we will use the results to guide the specification of the weights within the modified Tukey's depth (see above) under a conceptual scaling that uses both $M_{\leq}$ and $M_{\not\leq}$: Let $(p, q) \in \mu$. Define $\Delta_\mu(p, q)$ simply as the "distance" between p and q w.r.t. the mode μ measured by the number of pairs between p and q w.r.t. the covering relation of μ. Furthermore, for $x \in \mathscr{P}$ define $s_\mu(x) := \max_{(p,q)\in\mu\setminus x} \Delta_\mu(p, q)$. Then one can show that total orders x with a higher $s_\mu(x)$ have a lower depth value w.r.t. the enclosing depth $\mathscr{E}^\mu$. Thus, for a weighted version of Tukey's depth function one can weight pairs (p, q) with a higher $\Delta_\mu(p, q)$ correspondingly with a higher weight, e.g., via $\alpha_{p,q} \propto \Delta_\mu(p, q)$. (For pairs with the same value it would be natural to choose the same weight). Now, the problem is to specify the corresponding weights for pairs $(p, q) \in x\setminus\mu$. Because we would like to think from the direction of the mode μ and not from the perspective of x, we do not want to simply change the roles of μ and x. The problem here is that it seems to be somehow difficult to order pairs (p, q) w.r.t. the mode μ that are not in relation w.r.t. μ. However, there are some possibilities to rank such pairs. The following definition is somehow inspired by the work in [12]: For $(p, q) \notin \mu$ one could define[5] $\Delta_\mu(p, q) := |\{r \in \mathcal{X} \mid p \wedge_\mu q \leq_\mu r \leq_\mu p \vee_\mu q\}| - 1$. This definition extends the original definition of Δ_μ and it can be used to specify the weights (e.g., via $\alpha_{p,q} \propto \Delta_\mu(p, q); \beta_{p,q} \propto \Delta_\mu(p, q)$) for the modified Tukey's depth that uses the whole attribute set M for the conceptual scaling.

5 Simulation

We derive an algorithm to sample from statistical models on the set of partial orders on $\mathcal{X}$.The algorithms is based on the acceptance-rejection method and the idea of the algorithm is given in [13]. For a small number of elements, we can directly compute all reflexive, transitive, and anti-symmetric orders. Thus, we can easily draw a sample from one of the above distributions. Since the runtime of the computation of all partial orderings grows with the number of elements faster or equal to $2^{n^2/4}$ (see [20]), the direct computation is not feasible for larger n. Therefore, we provide an algorithm based on the following structure: First, we systematically draw a partial order and calculate the number of possible paths to

[5] If the considered partial order does not build a complete lattice one could simply compute the Dedekind-MacNeille completion beforehand.

obtain this partial order. Finally, we compute the acceptance probability such that we sample with probability of interest f. The algorithm uses that each partial order is a subset of at least one linear order. A linear order has $\frac{1}{2}(n-1)n$ many pairs of the form (x_i, x_j) with $i \neq j$ and, in particular, if we randomly delete some of these pairs, then, by computing the transitive hull, we obtain a partial order. To obtain step 1, we first take a uniform sample of a linear order and then randomly delete some pairs by a uniform variable on all subsets. By computing all linear extensions, we can compute the probability that this partial order was sampled. Finally, we adjust the acceptance probability so that the sample ends up consisting of the probability function f we are interested in. More precisely, the probability that a given order g is computed in step 1 is:

$$P_{algo\_select}(g) = |lext(g)| \cdot 2^{|g|-|reduc(g)|} \cdot \left(n! 2^{n(n-1)/2}\right)^{-1}$$

where $reduc(g)$ is the transitive reduction[6] of g and $n! \cdot 2^{n(n-1)/2}$ is the number of all paths to obtain a partial order by the procedure above. Since the number of pairs of each linear order is the same, the probability that the partial order g is sampled is identical for each linear order from the linear extension of g. Let f be the probability function from which we want to draw a sample, then the acceptance function is given by

$$acc(g) = f(g) \cdot \left(P_{algo\_select}(g) \cdot n! 2^{n(n-1)/2}\right)^{-1}. \tag{3}$$

Algorithm 1: Sampling a partial order based on linear orders

Input: n: number of items;
f: probability function with the set of all partial orders as domain;
Result: partial order sampled according to the probability given by f.
repeat
  # sampling the order
  LIN_ORDER ← sample uniformly a linear order;
  DEL_PAIRS ← sample uniformly a subset of $\{1, \ldots, (n-1)n/2\}$;
  PARTIAL_ORDER ← uniformly delete DEL_PAIRS many pairs and compute the transitive closure;
  # compute the acceptance probability (thereby we have to compute the transitive reduction)
  ACCEPT_PROB ← computation of (3);
until *random]0,1]* ≤ ACCEPT_PROB;

Lemma 1. *Algorithm 1 samples a partial order with probability function f on all partial orders of G.*

[6] The transitive closure of a relation is the smallest transitive relation containing it, and the transitive reduction is a minimal relation having the same transitive closure.

The proof is analogously to the one given in [13]. Note we could use also a modified version of the acceptance function: $\tilde{acc} = c \cdot acc$ with constant $c \geq \max_g f(g)/P_{algo\_select}(g)$. This modified version must assure that for all partial orders g, $f(g) \leq c \cdot P_{algo\_select}(g)$ is true. Unfortunately, the computation of all linear extensions is # P-complete (see [7]). Note that for some subsets of all partial orders the running time of the computation of the linear extension is smaller, i.e., if we consider only the set of trees (see [3]). Additionally, to improve the runtime of the algorithm we generally could also use an approximation for the number of all linear extensions $|lext(g)|$, for which e.g. [8] gives approximation approaches.

6 Conclusion

In this paper, we developed statistical models for partial orders based on data depth and formal concept analysis. We think that with this approach, opposed to statistical models based on distances, we are in fact able to appropriately incorporate the notion of unimodality of a statistical model for partial orders. In particular, we think that a notion of unimodality based on concepts of lattice theory is more appropriate compared to notions based on metrics or based on the embedding of partial orders into a linear space. What is left open for further research is the question how to exactly specify the decay function and the weights within the approach that uses Tukey's depth. A further analysis of the newly developed enclosing depth, especially w.r.t. the question if this depth function can be also applied if one does not know the mode beforehand, is also of high interest. Additionally, the application of our approach to concrete data situations is another important line of further research.

Acknowledgments Hannah Blocher and Georg Schollmeyer gratefully acknowledge the financial and general support of the LMU Mentoring program.

References

1. Alvo, M., Cabilio, P.: Rank correlation methods for missing data. *Canadian Journal of Statistics*, **23**(4), 345–358 (1995)
2. Armstrong, W.: Dependency structures of data base relationships. *International Federation for Information Processing congress*, **74**, 580–583 (1974)
3. Atkinson, M.: On computing the number of linear extensions of a tree. *Order,* **7**, 23–25 (1990)
4. Barnett, V.: The ordering of multivariate data. *Journal of the Royal Statistical Society Series A*, **139**(3), 318–352. With discussion (1976)
5. Boutilier, C., Rosenschein, J.: Incomplete information and communication in voting. In: Moulin, H., Brandt, F., Conitzer, V., Endriss, U., Lang, Procaccia, A. (eds.), *Handbook of Computational Social Choice.* 223-258, Cambridge University Press (2016)
6. Brandenburg, F., Gleißner, A., Hofmeier, A.: Comparing and aggregating partial orders with Kendall Tau distances. In: LNCS. Rahman, S., Nakano, S. (eds.), *WALCOM: Algorithms and Computation 2012,* 88–99, Springer (2012)

7. Brightwell G., Winkler P.: Counting linear extensions is # P-complete. *Proceedings 23rd ACM Symposium on the Theory of Computing*, 175—181 (1991)
8. Bubley R., Dyer M.; Faster random generation of linear extensions. *Discrete Mathematics*, **201**(1-3), 81–88 (1999)
9. Comim, F.: Beyond the HDI? Assessing alternative measures of human development from a capability perspective. In: *Background paper of the Human Development Report.* UNDP Human Development Report (2016)
10. Critchlow, D.: Metric methods for analyzing partially ranked data. *Lecture Notes in Statistics*, **34**, Springer (1985)
11. Fahandar, M., Hüllermeier, E., Couso, I.: Statistical inference for incomplete ranking data: The case of rank-dependent coarsening. *Proceedings of the 34th International Conference on Machine Learning*, **70**, 1078–1087 (2017)
12. Gäbel-Hökenschnieder, T., Schmidt, S.: Generalized metrics and their relevance for FCA and closure operators. *Concept Lattices and their Applications*, 175–186 (2016)
13. Ganter, B.: Random extents and random closure systems. *Concept Lattices and their Applications*, 309–318 (2011)
14. Ganter, B., Wille, R.: *Formal concept analysis: Mathematical foundations.* Springer Science & Business Media (2012)
15. Jansen, C., Schollmeyer, G., Augustin, T.: Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences. *International Journal of Approximate Reasoning*, **98**, 112–131 (2018)
16. Jansen, C., Blocher H., Augustin, T., Schollmeyer, G.: Information efficient learning of complexly structured preferences: Elicitation procedures and their application to decision making under uncertainty. Revision under review for *International Journal of Approximate Reasoning* (minor revisions)
17. Jena, S., Lodi, A., Palmer, H., Sole, C.: A partially ranked choice model for large-scale data-driven assortment optimization. *Informs Journal on Optimization,* **2**(4), 297–319. (2020)
18. Kendall, M.: A new measure of rank correlation. *Biometrika* **30** (1/2), 81—93 (1938)
19. Kendall, M.: Discussion on symposium on ranking methods. *Journal of the Royal Statistical Society: Series B,* **12**, 153–162 (1950)
20. Kleitman, D., Rothschild, B.: The number of finite topologies. *Proceedings of the American Mathematical Society*, **25**(2), 276-–282 (1970)
21. Lebanon, G., Mao, Y.: Non-parametric modeling of partially ranked data. *Journal of Machine Learning Research*, **9**(10), 2401–2429 (2008)
22. Mangaraj, B., Aparajita, U.: Constructing a generalized model of the human development index. *Socio-Economic Planning Sciences*, **70** (2020)
23. Mosler, K.: Depth statistics. *Robustness and complex data structures.* Becker, C., Fried, R., Kuhnt, S. (eds.), 17–34, Springer (2013)
24. Nakamura, K., Yano, K., Fumiyasu, K.: Learning partially ranked data based on graph regularization. arXiv:1902.10963 (2019)
25. Schollmeyer, G.: Lower quantiles for complete lattices. *Technical Report 207.* Department of Statistics. LMU Munich (2017)
26. Schollmeyer, G.: Application of lower quantiles for complete lattices to ranking data: Analyzing outlyingness of preference orderings. *Technical Report 208.* Department of Statistics. LMU Munich (2017)
27. Stewart, R.: Weak pseudo-rationalizability. *Mathematical Social Sciences,* **104**, 23-28 (2020)