Relational methods for statistical analysis and decision making in the context of non-standard data- and information structures

Kumulative Habilitationsschrift nach §12(1), Nr. 2 der Habilitationsordnung der
Ludwig-Maximilians-Universität München für die Fakultät für Mathematik, Informatik und Statistik

vorgelegt von

Dr. Georg Schollmeyer

This cumulative Habilitationsschrift is based on the scientific publications listed below in accordance with §12(1), Nr. 1 of the Habilitationsordnung as well as the publications submitted in accordance with §12(1), Nr. 2 of the Habilitationsordnung. In the following sections, the publications submitted in lieu of a Habilitationsschrift are summarized and situated in a broader scientific context in accordance with §12(3) of the Habilitationsordnung. In accordance with §12(2), Sentence 1 of the Habilitationsordnung, I describe my own contributions to the submitted works.

The published works in lieu of a Habilitationsschrift according to §12(1), Nr. 2 of the Habilitationsordnung are divided into three parts: A: *Relational data analysis for non-standard data*; B: *Decision making under weakly structured information*; and C: *Analysis of deficient data.* Section 2 lists the additional published works in accordance with §12(1), Nr. 1 of the Habilitationsordnung. All these three papers are peer-reviewed. In Section 3, I declare my own contributions to all published papers. In Section 4 to Section 6, I summarize the contributions in accordance to §12(1), Nr. 2. For every paper I start by stating the original abstracts (called *Original abstract)* of the papers to give a very quick overview. Then I describe every contribution in more detail. Finally, for every part, I give a broader embedding of the papers and indicate possibilities for further research. Despite the broader view given in this Habilitationsschrift, I would like to take the opportunity to emphasize that this habilitation is really a cumulative habilitation and therefore the Habilitationsschrift should be seen more as a profound addendum to and embedding of the original published work. Therefore, for more detail I would like to recommend to read the original papers that contribute to this cumulative habilitation.

Contents

1 Published works submitted for the Habilitation according to §12(1) Nr. 2 of the Habilitationsordnung

A: Relational data analysis for non-standard data

1. Hannah Blocher, Georg Schollmeyer, and Christoph Jansen (2022): Statistical models for partial orders based on data depth and formal concept analysis. In: Ciucci, D.; Couso, I.; Medina, J.; Slezak, D.; Petturiti, D.; Bouchon-Meunier, B.; Yager, R.R. (eds): Information Processing and Management of Uncertainty in Knowledge-Based Systems. Communications in Computer and Information Science, 1602:17-30.

2. Hannah Blocher and Georg Schollmeyer (2023): Data depth functions for non-standard data by use of formal concept analysis. Under review. Accessible under `www.foundstat.statistik.uni-muenchen.de/personen/mitglieder/blocher/blocheretal_properties23.pdf`

3. Hannah Blocher, Georg Schollmeyer, Christoph Jansen and Malte Nalenz (2023): Depth functions for partial orders with a descriptive analysis of machine learning algorithms. In: Proceedings of the Thirteenth International Symposium on Imprecise Probabilities: Theories and Applications (ISIPTA '23). Proceedings of Machine Learning Research, 215:59–71.

B: Decision Making under weakly structured information

4. Christoph Jansen, Georg Schollmeyer, and Thomas Augustin. Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences. International Journal of Approximate Reasoning, 98:112–131, 2018.

5. Christoph Jansen, Hannah Blocher, Thomas Augustin, and Georg Schollmeyer (2022): Information efficient learning of complexly structured preferences: Elicitation procedures and their application to decision making under uncertainty. International Journal of Approximate Reasoning, 144:69-91.

6. Jean Baccelli, Georg Schollmeyer, and Christoph Jansen (2022): Risk aversion over finite domains. Theory and Decision, 93:371–397.

C: Analysis of deficient data

7. Georg Schollmeyer (2021): Computing simple bounds for regression estimates for linear regression with interval-valued covariates. In Jasper de Bock, Andrés Cano, Enrique Miranda, and Seraffin Moral, editors, Proceedings of the Twelfth International Symposium on Imprecise Probabilities: Theories and Applications, Proceedings of Machine Learning Research, 147:273-279.

8. Georg Schollmeyer (2019): A short note on the equivalence of the ontic and the epistemic view on data imprecision for the case of stochastic dominance for interval-valued data. In: Jasper de Bock, Cassio de Campos, Gert de Cooman, Erik Quaeghebeur, and Gregory Wheeler, editors, Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications, Proceedings of Machine Learning Research, 103:330-337.

2 Published works submitted according to §12(1) Nr. 1 of the Habilitationsordnung

9. Christoph Jansen, Malte Nalenz, Georg Schollmeyer, and Thomas Augustin (2023): Statistical comparisons of classifiers by generalized stochastic dominance. Journal of Machine Learning Research, 24(231):1–37.

10. Christoph Jansen, Georg Schollmeyer, Hannah Blocher, Julian Rodemann, and Thomas Augustin (2023): Robust statistical comparison of random variables with locally varying scale of measurement. In: Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI 2023). Proceedings of Machine Learning Research, 216:941-952.

11. Krasymyr Tretiak, Georg Schollmeyer, and Scott Ferson (2023): Neural network model for imprecise regression with interval dependent variables. Neural Networks, 161:550-564.

3 Declaration of the contributions of the author

In this section, for every publication (termed Contribution 1 to Contribution 11 corresponding to the list above) that contributes to the habilitation in the sense of §12(1), Nr. 1 and Nr. 2, I briefly indicate my own contribution to the respective publication[1] All contributions except contribution 2 are peer-reviewed papers. Contribution 2 is currently under review (and accessible under `https://www.foundstat.statistik.uni-muenchen.de/personen/mitglieder/blocher/blocheretal_properties23.pdf`).

Contribution 1 Sections 1,2 and 7 were written jointly by all three authors. Georg Schollmeyer wrote most parts of Section 5. He developed the generalizations of the Tukeys depth, the peeling depth and the enclosing depth, as well as the proposal for the weighting within the modified Tukeys depth. The properties from Definiton 1 were developed and discussed by all authors. The problems of classical methods described in Section 2 were extensively discussed by all authors (and the solution via the concrete scaling method from Section 4 was developed by Hannah Blocher). All authors contributed to the revision of the paper.

Contribution 2 Hannah Blocher had the management and coordination responsibility and drafted most of the paper. Georg Schollmeyer wrote Section 4.4 and 5.4, Example 3 of Section 2 and the explanation to the generalized Tukeys depth in Section 3. Hannah Blocher wrote the other parts. Both authors contributed to the development of the properties (P1) - (P13) and the corresponding theorems and proofs. (Properties (P6), (P7), (P8), (P9), (P13) and (P14) and the idea of proof for Theorems 1, 5, 7, 11 and 12 are mainly due to Georg Schollmeyer. The others are mainly due to Hannah Blocher.)

Contribution 3 Hannah Blocher developed the idea of ufg depth. She wrote most of the paper. To be more precise: The introduction was written by Christoph Jansen. Hannah Blocher wrote the preliminaries and defined the (empirical) ufg depth. Furthermore, Hannah Blocher claimed and proved Lemma 1, Theorem 2, 1st and 2nd part, Corollary 3, Corollary 4, the lower bound of Theorem 5 and Theorem 6. Georg Schollmeyer made the claim and proofs of Theorem 2, 3rd part and the upper bound of Theorem 5. The claim of Theorem 7 was done by Georg Schollmeyer and Christoph Jansen. Georg Schollmeyer proved Theorem 7. Chapter 5 was written by Georg Schollmeyer. Hannah Blocher wrote Chapter 6.1 and implemented the test if a subset is an element of $\mathcal{S}$. Georg Schollmeyer contributed with the implementation of the connectedness property. Malte Nalenz provided the data set, performed the data preparation and wrote Chapter 6.2. Hannah Blocher analyzed the data set and wrote Chapter 6.3. Georg Schollmeyer, Christoph Jansen and Malte Nalenz supported the analysis with intensive discussions. Christoph Jansen provided the conclusion. Georg Schollmeyer and Christoph Jansen also helped with discussions about the definition of ufg depth and all properties. Malte Nalenz, Christoph Jansen, and Georg Schollmeyer also contributed by providing detailed proofreading and help with the general structure of the paper. All authors contributed to the revision of the paper.

Contribution 4 This contribution is based on a conference paper for the ISIPTA '17 conference (see Jansen et al. (2017)), drafted by Christoph Jansen and revised according to remarks of the other authors as well as review comments of the three anonymous ISIPTA '17 referees. Following the invitation to the conference's special issue in the International Journal of Approximate Reasoning, the conference version was significantly extended to the present version. The whole extended version was drafted and, in most parts, written by Christoph Jansen. Exceptions are the paragraph directly following Definition 6, the paragraph directly following the proof of Proposition 3, the second paragraph following Definition 8, and the first two paragraphs directly following the proof of Proposition 6. These were drafted by Georg Schollmeyer. The idea of the

[1] In cases (Contribution 3 and Contribution 4) where the specific contributions of the authors had already been declared elsewhere, the declaration was taken over mutatis mutandis (up to small additions concerning the revision of the papers).

proof of Proposition 3 was jointly developed by Christoph Jansen and Georg Schollmeyer. The concept of granularity (see Definition 3) was developed in discussions of all three authors. The name granularity is due to Thomas Augustin, who also contributed to embedding the paper into the relevant literature. Additionally, several rounds of discussing preliminary versions of the paper by all authors lead to improvements of the presentation. All authors contributed to revising the paper according to the reviewers' comments.

Contribution 5 The paper was mainly drafted by Christoph Jansen. Georg Schollmeyer worked out most parts of Example 2 (Section 6). In particular, he proposed to use the distance-based model and proposed and implemented the prediction method based on subgroup discovery to predict the next pair to elicite. (The simulation part for generating preference systems was done by Hannah Blocher.) All authors contributed to revising the paper.

Contribution 6 The paper was mainly drafted by Jean Baccelli. All authors contributed to the main result, Theorem 1. Georg Schollmeyer worked out most parts of the Examples (Example 1 till Example 3), as well as the proofs of the facts to which these examples are related. All authors contributed to revising the paper.

Contribution 7 & Contribution 8 Because both contributions are single author papers, the papers were written completely autonomously by Georg Schollmeyer. The revision process was also done completely autonomously by Georg Schollmeyer.

Contribution 9 The paper was mainly drafted by Christoph Jansen. Georg Schollmeyer supplied Proposition 2 including its proof and its usefulness for the simulation study. He also drafted parts of Section 3.3 and intensively discussed the aspect of incommensurability as well as the role of the hypotheses in the statistical test of section 4.2. All authors contributed to revising the paper.

Contribution 10 The paper was mainly drafted by Christoph Jansen. Georg Schollmeyer supplied parts of Proposition 7 and Proposition 8 including parts of the idea for their proof. He also discussed and drafted parts of the aspects related to regularization. All authors contributed to revising the paper.

Contribution 11 The paper was mainly drafted by Krasymyr Tretiak. Georg Schollmeyer wrote parts of Section 3 and supplied R-Code for an implementation of the sharp collection region.

4 A: Relational data analysis

"*It is often said that mathematics is a language. If so, group theory provides the proper vocabulary for discussing symmetry. In the same way, lattice theory provides the proper vocabulary for discussing order, and especially systems which are in any sense hierarchies. One might also say that just as group theory deals with permutations, so lattice theory deals with combinations.*" [Birkhoff, 1938, p. 793]

4.1 Introduction

The classical *numerical* approach to the analysis of data is usually underpinned by relational measurement theory where one first assumes a certain underlying empirical relative that can be captured by a measurement mapping that at least respects this relational structure. This mapping usually maps into the real numbers and one in a certain way gives only *meaning*[2] to that structure of the reals that relates to some *meaning* within the empirical relative. In a second step one quickly forgets that the obtained numbers appeared indirectly by the measurement mapping and simply operates in the reals (of course not forgetting which aspects carry *meaning*). Opposed to this, within the methodology of *relational data analysis* one thinks more directly within the empirical relation between objects or between objects and attributes. For example one would say that this rod is longer than that rod instead of firstly comparing rods with unit-rods and then saying for example this rod is 2.5 inches long whereas that rod is only 2.3 inches long. With this approach to data analysis one in a sense postpones the decision about measurement theoretic considerations as long as possible and intertwines it more directly with the actual process of data analysis. One way of concretely doing so is provided by the mathematically rigorously developed theory of formal concept analysis with its canonical demand for a thorough and substance matter reflecting conceptual scaling of all involved aspects of a data analysis that should or should not carry *meaning*. Our first three contributions combine formal concept analysis with the methodology of data depth. Since both theories are not so well known, we would like to start with a very brief introduction to both topics.

Formal Concept Analysis

Formal concept analysis (FCA) is a mathematical theory developed in the early 1980s by the Darmstadt research group around Rudolf Wille, Bernhard Ganter and Peter Burmeister. Formal concept analysis understands itself as an applied theory of order- and lattice theory. The development of formal concept analysis was intellectually influenced by the views of Hartmut von Hentig and his call for a restructuring of the sciences:

"*... dann müssen die einzelnen Wissenschaften in erster Linie ihre Disziplinarität überprüfen, und das heißt, ihre unbewußten Zwecke aufdecken, ihre bewußten Zwecke deklarieren, ihre Mittel danach auswählen und ausrichten und ihre Berechtigung, ihre Ansprüche, ihre möglichen Folgen öffentlich und verständlich darlegen und dazu ihren Erkenntnisweg und ihre Ergebnisse über die Gemeinsprache (und die von mir sogenannte 'Anschauung') zugänglich machen.*"[3][von Hentig, 1974, p. 136f]

Rudolf Wille also argued in the sense of this demand in the abstract of his essay *Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts*:

[2]Compare Suppes [1969].

[3]English translation by GS: *Then every single individual science must first and foremost examine its disciplinarity, and that means it must reveal its unconscious ends, declare its conscious ends, and it must select and align its means accordingly, and publicly and it must comprehensibly explain its justification, its claims, its possible consequences, and for this purpose it has to make its path of knowledge and its results accessible via the common language (and what I call 'Anschauung').*

"*Lattice theory today reflects the general status of current mathematics: there is a rich production of theoretical concepts, results, and developments, many of which are reached by elaborate mental gymnastics; on the other hand, the connections of the theory to its surroundings are getting weaker and weaker, with the result that the theory and even many of its parts become more isolated. Restructuring lattice theory is an attempt to reinvigorate connections with our general culture by interpreting the theory as concretely as possible, and in this way to promote better communication between lattice theorists and potential users of lattice theory.*" [Wille, 1982]

The purely mathematical foundations of formal concept analysis were already laid in the 1930s by Garrett Birkhoff in the form of general lattice theory. The starting point of formal concept analysis is the mathematical formalization of the concept *concept*, which is inspired, among others, by the writings of Charles Sanders Peirce [cf., Peirce [2017]. Nowadays, formal concept analysis finds practical application for example in data and text mining (e.g., Poelmans et al. [2012]), linguistics (e.g., Falk and Gardent [2014]), web mining (e.g., Elzinga et al. [2012]), software mining (e.g., Wermelinger et al. [2009]), bioinformatics (e.g., Keller et al. [2012], chemistry (e.g., Keller et al. [2012]), medicine (e.g., Messai et al. [2011]) or ontology engineering (e.g., De Maio et al. [2012]). For a comprehensive survey on applications of FCA, see Poelmans et al. [2013]

Concretely, in FCA[4] one starts with a so-called formal context. A formal context is a triple $\mathbb{K} := (G, M, I)$ with G a set of objects, M a set of attributes and $I \subseteq G \times M$ a binary relation with the interpretation $(g, m) \in I$ if object g has attribute m. In the context of statistical data analysis, the objects are usually the data points, for example the persons that participated in a social survey. The attributes describe the observed values of the interesting variables, for example the answer *yes* or *no* to the posed questions and then naturally $(g, m) \in I$ means that person g answered the question m with yes. If the answers to the questions in a survey are not binary, or if one wants to explicitly incorporate the answer *no* as an attribute, then one can transform the answers into binary attributes according to some coding scheme. This way to codify non-binary variables into binary variables is called *conceptual scaling* in FCA. This flexible way of incorporating the data into a FCA analysis that at the same time calls the user to very carefully decide about which measurement theoretic aspects should carry meaning and which not, should be seen as a main strength of formal concept analysis. Now, given the formal context $\mathbb{K} = (G, M, I)$, the data analysis proceeds by looking at the structure of $\mathbb{K}$ in terms of so called formal concepts: A pair (A, B) with $A \subseteq G$ a set of objects and $B \subseteq M$ a set of attributes is called a formal concept if

i) $\forall g \in A \forall m \in B : (g, m) \in I$

ii) A is maximal w.r.t. property i), i.e., there exists no further $g \in G \backslash A$ with $\forall m \in M : (g, m) \in I$

iii) B is maximal w.r.t. property i), i.e., there exists no further $m \in M \backslash B$ with $\forall g \in A : (g, m) \in I$

A formal concept (A, B) describes in a formal way a concept by specifying with A (the so-called concept extent) which objects belong to the concept and by simultaneously specifying with B (the so-called concept intent) which attributes characterize the concept. The set of all formal concepts is itself partially ordered by the subconcept/superconcept relation $\sqsubseteq$ that is naturally defined as

$$(A, B) \sqsubseteq (C, D) :\iff A \subseteq C \;\&\; B \supseteq D.$$

The analysis of this partially ordered set, which turns out to be actually a complete lattice, is the business of formal concept analysis. With the derivation operators

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

[4]For a comprehensive introduction into FCA, see Ganter and Wille [2012].

one can characterize formal concepts as

$$(A, B) \text{ is a formal concept } \iff A = \Phi(B) \,\&\, B = \Psi(A).$$

Furthermore, as can be seen from above, the extent A of a formal concept (A, B) uniquely defines its intent B as $B = \Psi(A)$ and the intent B uniquely defines the extent via $A = \Phi(B)$. Therefore, the whole structure of $\mathbb{K}$ can also be uniquely described by only looking at extents. Then, the family of all extents is a closure system (i.e., a family of subsets that contains G and that is additionally closed under arbitrary intersections). This closure system can be alternatively described by the closure operator $\gamma := \Phi \circ \Psi$: The extents are exactly the images of this closure operator. Another way for describing the structure of $\mathbb{K}$ is to use so-called formal attribute implications: A formal attribute implication is a pair (X, Y), denoted with $X \longrightarrow Y$, where $X, Y \subseteq M$ are attribute sets. A formal implication is valid in a formal context if all objects that have all attributes in X do also have all attributes in Y. Then the family of all formal concept intents can be described as the family of all attribute sets $B \subseteq M$ that respect all valid implications $X \longrightarrow Y$(, i.e., if $B \supseteq X$ then also $B \supseteq Y$ for all valid implications $X \longrightarrow Y$). Dually, by changing the roles of objects and attributes, one can work with formal implications between object sets and this is somehow building a natural bridge to more conventional approaches to data analysis: For example in a linear space, there is a natural notion of location: One can say for example that a point q lies between the points p and r, if q is a convex combination of p and r. In other words, if q lies in the *space spanned by p and r*. In terms of formal concept analysis this may be translated to q lying between p and r if q lies in the space spanned by p and r, i.e., if $q \in \gamma(\{p, r\})$, or in terms of formal object implications, if $\{p, r\} \longrightarrow \{q\}$. With this abstraction, it is in fact possible to translate certain notions like for example unimodality of a function to the setting of formal concept analysis. In our first three contributions we made much use of this idea to bring to together the theory of formal concept analysis and the methodology of data depth.

Data depth

The methodology of data depth is a non-parametric approach that goes back to Tukey (see Tukey [1975]) who used it as a graphical tool for descriptively vizualizing bivariate data. Since then, the methodology was extended to the general multivariate case of $\mathbb{R}^d$-valued data: Generally speaking, a data depth function measures how deep or outlying a given data point $z \in \mathbb{R}^d$ is located with respect to an observed data cloud or an assumed underlying distribution in $\mathbb{R}^d$. With this, data depth provides a center-outward ordering of points in $\mathbb{R}^d$. This can be used for a whole bouquet of non-parametric (and usually robust) multivariate statistical analyses. Applications range from the description of multivariate distributions (Liu et al. [1999], Serfling [2004], Wang and Serfling [2005]) over outlier detection (Serfling [2006], Zhang [2002]), depth based classification and clustering (Ruts and Rousseeuw [1996], Christmann [2002], Jörnsten [2004]), rank and sign tests (Brown and Hettmansperger [2018], Hettmansperger and Oja [2018]), multivariate density estimation (Fraiman et al. [1997]) to robust linear regression (Rousseeuw and Hubert [1999]). Starting from the first proposal of Tukey, a whole bunch of different new concrete proposals for depth functions were developed. Additionally, certain desirable properties a depth function should have, were formulated. These concern structural properties like affine equivariance (see Serfling and Zuo [2000]), statistical properties like uniform consistency (cf., e.g., Liu [1990], Arcones and Gine [1993]) and robustness properties like a high breakdown point (cf., e.g., Donoho and Gasko [1992]). Beyond the case of $\mathbb{R}^d$, in the meanwhile the methodology of data depth was also applied to functional data (Dai and Lopez-Pintado [2022]) or data in a metric space (Geenens et al. [2023], Chakraborty and Chaudhuri [2014]). Our contributions in this habilitation still prolongs the area of application of data depth by using it in the context of data that can be represented by a formal context. Due to the flexibility of the method of conceptual scaling, this in fact allows a broad application of data depth to many examples of non-standard data analysis. In our first and our third contribution, we deal with the case of data that are partially ordered sets on a given finite

basic set of items. Contribution 2 develops and analyzes properties of depth functions in the general setting of arbitrary formal contexts/data-settings.

4.2 Our contributions

Contribution 1

Blocher, H.; Schollmeyer, G.; Jansen, C. (2022): Statistical models for partial orders based on data depth and formal concept analysis. In: Ciucci, D.; Couso, I.; Medina, J.; Slezak, D.; Petturiti, D.; Bouchon-Meunier, B.; Yager, R.R. (eds): Information Processing and Management of Uncertainty in Knowledge-Based Systems. Communications in Computer and Information Science, 1602:17-30.

Original 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.

Within this paper we are dealing with the special case of data sets where each data point is a partial order (poset for short). In this situation, a relational approach appears very natural. In fact, while for the case of total orders, which build a (non-Abelian) group, also group theory can be used (and is used, cf., e.g., Diaconis [1988], Lebanon and Mao [2007], Goibert et al. [2022]), for the more general class of *partial* orders, the hierarchical approach of lattice theory is very promising[5]. As far as we are aware of, such an approach for the statistical modeling of poset-valued data is not very intensively studied and therefore we think that our contribution in this field is of high value.

Concretely, in this paper we use the language of formal concept analysis and combine this language of applied lattice theory with the concept of data depth that is usually only used for data sets with values in $\mathbb{R}^d$, functional data or data with values in a metric space, cf., e.g., Mosler [2013], Mosler and Polyakova [2012], Dai and Lopez-Pintado [2022], Geenens et al. [2023], Chakraborty and Chaudhuri [2014]. Within our contribution we see partial order data as a very interesting example of non-standard data. In particular, we take here the special viewpoint that the envisaged orders are only partially ordered not because of the non-observation of total orders, i.e., we do not assume that the partial orders arise due to a censoring process. Such a viewpoint can be found in many existing approaches like e.g., Nakamura et al. [2019], Lebanon and Mao [2007]. There, an underlying total order together with a coarsening process is assumed and modeled. In contrast to this approach that can be termed *epistemic* (cf., Couso and Dubois [2014]), here, we hold the *ontic* view that understands partial orders as partially ordered by nature. To give a concrete example: In Dittrich et al. [1998] university students were asked for their choices of foreign universities for their semester abroad. Concretely, the students preferences were collected via pairwise comparisons between (all-together 6) universities. The students could say that they

[5]Note that also for partial orders, e.g., Lebanon and Mao [2007] use group theory. However, there, partially ranked data are treated as censored data. In this way, this approach does not directly fit well to our ontic view on partial orders hold within our paper, see below.

prefer one university over another or vice versa. Additionally, they could also answer that they have no preference between two given universities. In this situation it appears very natural to assume that an observed incomparability between two universities does not mean that the asked student does not know his preference or that she forgot to answer the question about preference. Instead, one can in fact say that the student has no preference between the given universities at all.

To account for this special viewpoint, we use the method of conceptual scaling from formal concept analysis: We do not only use attributes of the form $x \leq y$ (with the interpretation that university y is preferred to university x or that one is indifferent between x and y). Instead, we also include attributes of the form $x \not\leq y$. This special way of incorporating FCA leads to a special behaviour of our method, as is analyzed in the paper (Section 4 of the contribution). The other ingredient of our proposed method is the usage of data depth to build a concise statistical model for partial order data that can be specified by defining one location and one scale parameter. With this, in fact our paper is an important contribution to the old

" *...major outstanding problem in ranking theory [at the present time]. This is simply to find some method of specifying a population of ranks in the non-null case. If we have a ranking of n which can happen in n! ways, then in order to specify such a population in general we require n! parameters. This is far too large a number for any tractable mathematics to be applied to it.*" [Kendall, 1950]

In fact, compared to the above statement that relates to total orders, for partial orders, the situation becomes far more dramatic. For example for partial orders of 8 items, compared to $8! = 80640$ total orders, there are all-together 4.431.723.379 partial orders (see, e.g., Brinkmann and McKay [2002]). Therefore (also 73 years later) it would be of high interest to be able to define parametric models for partial orders that are specifiable by only a few number of parameters and that at the same time appreciates our *ontic* view on partial orders.

Compared to our data depth approach, the most usual statistical models for (partial) orders are based (or can be based) on a notion of distance between one modal order and all other orders and a probability function that decreases with increasing distance. Therefore, in our paper we also discuss the relation of our approach to such models (Section 2 of the contribution), especially some shortcomings w.r.t. our *ontic* understanding of partial data within our paper.

Additionally, already within this paper we discuss certain properties a depth function or a probability model should have. One property that is emphasized here is the property of quasiconcavity or unimodality (cf., Definition 1 in the paper). This notion can be perfectly described in the language of formal (object) implications from FCA: If one data point y lies within the space that is spanned by other points $x_1, \ldots, x_k$ (in the sense that $\{x_1, \ldots, x_k\} \rightarrow \{y\}$), then the depth value of y should at least be not smaller then the depth value of *all* $x_1, \ldots, x_k$ (i.e., $depth(y) \geq min\{depth(x_1), \ldots, depth(x_k)\}$). This property in a sense formalizes the demand that the depth function should not have any local minimum.

In this way we in fact fruitfully combine the theory of formal concept analysis and the theory of data depth to derive a way of statistically modeling partial order data.

In Section 5 of the paper we develop concrete proposals for a depth function that - together with a probability decay function (Equation 1 in the paper) - builds our statistical model. Since we are interested in unimodal models where the notion of unimodality (cf., Definition 1) is a natural adaption of classical unimodality/quasiconcavity for classical depth functions (cf., Mosler [2013]), we have to firstly analyze the mathematical implications of unimodality. Actually it turns out that on a general formal context, as well as concretely on the formal context of all partial orders, it is impossible to define a strictly unimodal depth function (where strictly means that the $\geq$ from above is replaced by $>$) and therefore it is impossible to define a strictly unimdoal probability

model. Thus, we have to live with depth functions and statistical models that are only quasiconcave/unimodal. Furthermore, as worked out in the paper, it is generally very difficult to come up with a quasiconcave depth function that is not trivial. In our contribution, we concretely analyze three generalizations of depth functions to the case of partial order data, namely I) a generalization of Tukeys depth, II) a generalization of the convex hull peeling depth and III) a new depth function that we call enclosing depth and on which we base our final solution IV):

I) Generalized Tukeys depth The generalized Tukeys depth function for general data structures is a generalization of the classical Tukeys depth (cf., Tukey [1975]) for $\mathbb{R}^d$ to arbitrary kinds of non-standard data that can be represented by a formal context. This depth function is based on Schollmeyer [2017a,b] and was formally introduced in Blocher et al. [2022]. In Schollmeyer [2017a] it was introduced as a depth function for data that are elements in a complete lattice. Since every formal context is naturally endowed with the complete lattice of all its concept extents, this depth function can be used for arbitrary formal contexts. Note that there the depth function is not called Tukeys depth. Instead, there we speak about the level function which is an antitone transformation of what is termed Tukeys depth function in Blocher et al. [2022]. In its original form for $\mathbb{R}^d$, Tukeys depth (also called halfspace depth) defines the depth of a point $z \in \mathbb{R}^d$ w.r.t. a probability law P as the infimal probability of a halfspace that contains the point z. If one treats $\mathbb{R}^d$ in a manner of synthetic geometry, then it appears natural to look at the incidence between points and halfspaces. The connection to formal concept analysis is then given by the observation that the closure system of all (closed) convex sets is generated by all possible intersections of (closed) halfspaces. In this view, it turns out that it is technically more intuitive to think in terms of outlyingness instead of depth. The outlyingness of a point $z \in \mathbb{R}^d$ w.r.t. a probabilitiy law P can then be defined as the supremal probability of an open halfspace that does not contains the point z. This definition is equivalent to the definition of oultlyingness of a point z as the supremal probability of a convex set (i.e., in terms of FCA: a concept extent) that does not contain the point z. This straight-forward generalization of Tukeys depth gives in fact a (quasiconcave) depth function that is applicable for arbitrary data types that can be represented with a formal context (or a complete lattice).

It turns out that for our concrete statistical modeling problem for partial order data, the model that would be obtained by applying Tukeys depth in its original form leads to a quasiconcave, but in particular to a more or less trivial statistical model. Concretely, the obtained probability function has only two different values.

II) Generalized convex hull peeling depth Thus, we also analyze the possibility of generalizing the convex hull peeling depth (cf., Barnett [1976]) from the case of $\mathbb{R}^d$ to the case of general spaces that can be described by formal contexts. The idea behind the convex hull peeling depth is that one peels different layers from a data cloud from outwards to inwards to get different depth layers. One starts with the extreme points of the data cloud end removes them. Then one proceeds with the remaining data points and again peels the extreme points and so on. The reason why this construction is well defined and leads to a (to a certain extent) satisfying depth function is the fact that the complete lattice of (closed) convex sets (projected on a fixed set of finitely many observed data points) in $\mathbb{R}^d$ is a meet-distributive[6] lattice, which means that every convex set is the convex hull of its extreme points[7]. Unfortunately, the formal context of all partial orders builds a concept lattice that is not meet-distributive. Therefore, different non-uniqueness issues come into play. Within every step of peeling, there are different candidates of possible layers that one can remove. For example, in every peeling step one can remove all extreme points and all other points that otherwise would not be removed at some later step of removing extreme points. Another possi-

[6] In terms of closure operators, meet-distributivity is directly related to the anti-exchange property, cf., Edelman [1980].

[7] With convex hull we mean the usual convex hull in $\mathbb{R}^d$, but projected on the data points that are actually observed. And consequently with convex set we mean a set that is the projection of a set that is convex in the usual sense.

bility would be to remove a layer that formally implies all remaining points and that is minimal w.r.t. this property, etc. Depending on how one solves this issues, either one will get a quasiconcave depth function with many ties (i.e., many identical depth values and therefore also many identical values of the associated probability function) or a depth function that is not quasiconcave.

III) Enclosing depth A third approach for a depth function is completely newly developed in the contributing paper, namely what we now call the enclosing depth. The basic idea is very simple and in a certain way dual to the construction of the peeling depth: Compared to the classical applications of data depth, a main difference of our problem setting of defining a statistical model based on data depth is the following: Usually one uses data depth in particular to define the center or the most central point(s) of a given data cloud or probability law. Additionally, one would like to specify this center in a robust way. Therefore, especially the application of the peeling depth appears very counter-intuitive because one starts the peeling with the most outlying layers, which suggests that the obtained method is not very robust. In fact, the breakdown point[8] of the convex hull peeling depth is very low, cf., Donoho and Gasko [1992], concretely for normally distributed data in $\mathbb{R}^d$ it seems to tend to zero as n increases (cf., Donoho and Huber [1983]). However, in our very comfortable situation of specifying a statistical model, we are able to specify beforehand, where the center should be. Thus, this non-robustness aspect is not relevant at all in our situation. Additionally, opposed to the construction of the peeling depth we can start with the center. Instead of peeling layers from outside to inside, we can simply start with the center and step by step we can enclose additional layers. This is simply the basic idea behind the enclosing depth. To concretely define the enclosing depth one needs to specify an enclosing operator that encloses an additional layer to the already defined layers. Similar to the peeling depth where one needs a peeling operator that removes layers, also for the enclosing depth we have to face some serious non-uniqueness issues and additionally, we would get a depth function that is generally not quasiconcave.

IV) The solution: Weighted Tukeys depth informed by the enclosing depth Finally, to define a depth function that is both quasiconcave and non-trivial at the same time, we modify Tukeys depth (which is already quasiconcave) by weighting the maximum (Equation 2 in the contribution) that is involved in Tukeys depth. To get reasonable weights we use here a careful analysis of the enclosing depth. In this way, the enclosing depth helps in getting our final proposal. The analysis of the enclosing depth shows that for example if the modal order μ sets $x \leq_\mu y$ but another order ν sets $x \not\leq_\nu y$, then ('ceteris paribus') the depth of ν depends particularly on the number of elements between x and y (w.r.t. μ). This can be naturally reflected by accordingly weighting the maximum in Equation (2). As already stated, one main demand within our contribution is to hold an *ontic* view and to include both the aspects of comparable items, as well as incomparable ones. Therefore, one should also specify how a pair (x, y) is weighted if it is incomparable w.r.t. the mode μ but comparable w.r.t. the envisaged order ν. For this we use a general proposal for defining distances between objects in a partial order or a complete lattice, see Gäbel-Hökenschnieder et al. [2016].

All-together, with this final solution we are now able to specify with one location parameter and one scale parameter (plus maybe one parameter for specifying how comparable and incomparable items are balanced and additionally a specification of the shape of the decay function) reasonable non-null models for partial orders on a finite ground space.

Finally, in Section 6 of the contribution we present a specific algorithm for sampling from such models. The basic idea of the algorithm is to firstly sample a total order and afterwards to sample from the obtained total order a partial order by removing edges from it. To calibrate the sampling probabilities we firstly compute (analytically) the obtained sampling probabilities and secondly

[8]Roughly speaking, the breakdown point of a robust method is the maximal amount of contaminated data the method can deal with before giving completely unreasonable results.

we correct for them with an acceptance-rejection method (cf., Ganter [2011]).

Contribution 2

Hannah Blocher and Georg Schollmeyer (2023): Data depth functions for non-standard data by use of formal concept analysis. Under review. Accessible under `https://www.foundstat.statistik.uni-muenchen.de/personen/mitglieder/blocher/blocheretal_properties23.pdf`

Original Abstract

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

In this paper we extensively discuss and analyze several generalizations of structural properties of depth functions on formal contexts beyond quasiconcavity. Furthermore, we propose additional properties that become only visible and relevant within the very abstract setting of formal concept analysis. In the second part of the paper we study these properties in detail for the generalized Tukeys depth (cf., also Contribution 1). More concretely, in the first part of the paper, we formulate *representation properties*, *order preserving properties*, *empirical (sequence) properties* and *universality properties*. The representation properties and the order preserving properties were already studied (to some extent) for the case $\mathbb{R}^d$ (see Serfling [2006], Mosler [2013]) and can be more or less directly generalized to the case of depth functions for formal concept analysis. For example, regarding $\mathbb{R}^d$, the idea of the representation property of affine equivariance is motivated more or less by measurement theoretic considerations that state that the result of a data depth based analysis should not depend on the concretely used coordinate system. This idea can be translated to an underlying space that is implicitly given by a formal context. One could argue that it should not matter, which concrete attributes a data point has, it should only matter which structure - given by the lattice/closure system of all formal extents or alternatively by all valid object implications - the objects constitute. This argument can be motivated by the understanding that the implications between the objects (or the lattice/closure system of all extents) somehow characterize how the the data points/objects are located in the space that is implicitly given by the underlying formal context. This consideration constitutes our property (P1). Similarly, property (P2) demands that a depth of a data point should not depend e.g., on its name, therefore different objects with the same attributes should have the same depth.

The ordering preserving properties (P3) and (P4) are adaptions of the properties *vanishing at infinity* and *maximality at the center* known from $\mathbb{R}^d$. These properties already show that for a space of data points that is only equipped with a closure system and not with the comfortable structure of a vector space (and particularly with an operation of translation), it is more difficult to get an intuitive sense of what should naturally be the center w.r.t. a depth function and which data points should be regarded as extreme outliers. In $\mathbb{R}^d$ it is easy enough to specify notions of

symmetry, e.g., of point symmetry, and then it is intuitive to demand that for symmetric distributions with point of symmetry s should have maximal depth at point s. Also the extreme outliers could be easily and intuitively characterized as points z with very large norm $||z||$. In the context of formal concept analysis, in general, we do not have a metric and strong notions of symmetry (but compare Theorem 1 of our contribution). However, it is possible to specify a weak notion of centrality and extreme outlyingness: objects that lie in no extent (despite the trivial extent G), should have minimal depth (property (P3)) and objects that lie in every extent should have maximal depth (property (P4)).

The ordering preserving properties (P5) till (P8) are all weakenings and strengthenings of quasiconcavity with (P8) being the property of strict quasiconavity. Strict quasiconcavity is actually a very strong property. There exists formal contexts for which there exists no strictly quasiconcave depth function (cf., Theorem 3 of the contribution and also (the discussion of) Contribution 1).

The ordering preserving property (P9) (*reflecting betweenness*) is some kind of the reverse property to the strict quasiconcavity property: If one data point x has a larger depth than another point y, then this should have some reason that is rooted in the location of x compared to y and all other data points in the sense of the underlying implicational structure underlying the context. Note that this property does not rely on the underlying probability measure on the context and may therefore be judged as a little bit too strong, especially in the light of the fact that for a meet-distributive finite context without duplicates there is essentially only one depth function that is strictly quasiconcave and betweenness-reflecting. Namely, this is the peeling depth (compare Theorem 5 of the contribution and Contribution 1).

The empirical (sequence) properties (P10) and (P11) are not adaptions of existing properties for $\mathbb{R}^d$ but new developments. They are properties that make statements comparing an empirical depth function for different data sizes. Property (P10) states that adding an (arbitrary) object g to a context which is a duplicate of an already existing object from the original context will always increase the depth of this object. For property (P11) firstly define an object g_{diff} as an object that strongly differs from the other objects if this object has no attribute that any other object has. Property (P11) then states that adding a strongly differing object g_{diff} will not change the ordering of the depths of the other objects. The empirical (sequence) property (P12) is the classical property of uniform convergence of the empirical depth function to the population version depth function (almost surely) under an i.i.d. sampling.

The universality properties (P13) and (P14) are also newly developed. They try to capture the problem that there are formal contexts for which there exists no strictly quasiconcave depth function. The idea of these properties is borrowed from category theory. They try to do justice to the fact that the demand of only quasiconcavity is not enough because for example a constant depth function is always quasiconcave but not interesting. Therefore, these properties try to formulate what it means for a depth function to be *as strictly quasiconcave as possible*. We use here the notion of a free object. We call a depth function D free w.r.t. a set of properties (not necessarily quasiconcavity) if it can *imitate* every other depth function E that also has these properties by equipping D with an appropriate underlying probability law and and by composing it with an isotone function, i.e., E equipped with a given law P and D equipped with an appropriate other law Q (that may depend on E and P) are order-theoretically equivalent (except for the fact that we allow for E having more ties in its depth values than D).

Depending on if we allow for Q to be arbitrary (depending on E and P) or if we restrict Q to an arbitrary small neighborhood model of laws, we get one notion of weak freenness (P13) and one notion of strong freeness (P14). Importantly, in the contribution we show that there are strongly (and weakly) free depth functions (w.r.t. quasiconcavity), at least for certain formal contexts (cf., Theorem 7 and Theorem 11). Therefore, the newly introduced universality notions are not empty notions.

In the second part of the paper, we analyze the generalized Tukeys depth w.r.t. all properties. The generalized Tukeys depth has (partly under some assumptions) most of the properties. Exceptiions are (P8), (P9), (P11) and (P14). Particularly, the fact that there are strongly free depth functions (w.r.t. quasiconcavity) on certain contexts for which the generalized Tukeys depth is not strongly free is a motivation for searching for new depth functions, since, very intuitively speaking, the non-freenness of the generalized Tukeys depth means that one can expect to have often a small number of different depth values which is of course a clear disadvantage for any kind of data depth analysis.

Contribution 3

Hannah Blocher, Georg Schollmeyer, Christoph Jansen and Malte Nalenz (2023): Depth Functions for Partial Orders with a Descriptive Analysis of Machine Learning Algorithms. In: Proceedings of the Thirteenth International Symposium on Imprecise Probabilities: Theories and Applications (ISIPTA '23). Proceedings of Machine Learning Research, 215:59–71.

Original Abstract

We propose a framework for descriptively analyzing sets of partial orders based on the concept of depth functions. Despite intensive studies of depth functions in linear and metric spaces, there is very little discussion on depth functions for non-standard data types such as partial orders. We introduce an adaptation of the well-known simplicial depth to the set of all partial orders, the union-free generic (ufg) depth. Moreover, we utilize our ufg depth for a comparison of machine learning algorithms based on multidimensional performance measures. Concretely, we analyze the distribution of different classifier performances over a sample of standard benchmark data sets. Our results promisingly demonstrate that our approach differs substantially from existing benchmarking approaches and, therefore, adds a new perspective to the vivid debate on the comparison of classifiers.

In this contribution we applied another generalized depth function for partial orders in the context of benchmarking machine learning algorithms. The used depth function was a generalization of the simplicial depth (Liu [1990]) to the case of partial order data. Also there we used the techniques of formal concept analysis and in particular the method of conceptual scaling to define the depth function. The definition of this depth function (we call this depth function the union free generic depth function, or shortly, *ufg*-depth) is based on a special family of characterizing implications (the so-called union free generic base, which was introduced in Blocher [2020]). Such a family is a special set of formal implications from FCA that in a certain sense describes the whole closure system of all concept extents that itself characterizes the given formal context and therefore the whole data set. While the classical simplicial depth in $\mathbb{R}^d$ defines the depth of a data point z as the proportion of $d+1$-simplices in the data set that contain the point z, the *ufg*-depth of a point z is the proportion of *ufg*-premises from the *ufg*-base (potentially weighted by their cardinality) that formally imply the point z.

One aspect within the contribution is the structural analysis of the *ufg*-base and the *ufg*-depth for the special case of partial order data with the conceptual scaling that was aleardy used in Contribution 1. It shows up that for example the maximal cardinality of an *ufg*-premise can be bounded by the VC dimension of the corresponding closure system of the concept extents (cf., property (P7) in the contribution). This, together with a further property that states that the family of *ufg*-premises is in a certain sense connected (cf., property (P8) of the contribution and Schollmeyer and Blocher [2023]) allows a more efficient enumeration of all *ufg*-premises and allows therefore the computation of the *ufg*-depth in reasonable time. Properties (P3) and (P4) give insight into how the sample and particularly outliers affect the *ufg*-depth. This shows that it is

worth going also beyond the very general framework of Contribution 2 and to look at the mathematical implications of special cases of data structures together with concrete conceptual scalings.

Next, in the contribution, we use the *ufg*-depth to descriptively analyze the performance of different machine learning classification algorithms over a set of data sets and over a set of different performance measures: If for one fixed data set one defines one classifier as at least as performative as another if it is at least as performative w.r.t. all considered performance measures, then one gets a partial order over the set of all considered classifiers. Since one has not only one data set, but a whole sample of data sets, all-together one gets a sample of data points that are itself partial orders. Such a sample was descriptively analyzed in this paper. Compared to the classical benchmarking task, within this paper we were therefore able to enrich the toolbox for classifier comparison w.r.t. at least two additional aspects:

Firstly we could do an analysis w.r.t. a whole set of performance measures. As also stated in the conclusion of the paper, the methodology of using partial orders still allows far more ways of comparison: For example one could use receiver operating characteristic (ROC) curves (see, e.g., Fawcett [2006], cf., also the methodology developed in Chang [2020]) instead of a set of classical performance measures and one could say that one classifier is as least as performative as another on a given data set if its ROC curve lies always above the ROC curve of the other algorithm.[9]

Secondly, with the methodology of data depth we are able to analyze not only the obtained partial orders on every data set separately. Instead we can now analyze the whole distribution of the partial orders within our data sample: The location of the distribution can be analyzed by looking at the deepest data point(s), i.e., the deepest partial order within the sample. This partial order to a certain extent gives a 'typical' data set together with a typical ordering of classifiers. Also outliers can be detected by looking at data points with very low depth. Additionally, a qualitative impression about the dispersion of the distribution can be given by looking at all depth contours[10]. On the other hand, the dispersion can be concisely described quantitatively by looking at all partial orders, i.e., also partial orders that were not observed, and by computing the proportion of partial orders that have a depth value above a certain threshold. In this sense, in a way one "measures" the volume of the depth contours. Actually in the concretely analyzed data set the dispersion was very low given the high variability of the included data sets. This suggests that in fact the methodology proposed here seems to give insightful information about the data situation.

4.3 Related Work & Future Research

As already indicated, the generalized Tukeys depth is based on Schollmeyer [2017a,b]. For the special case of data that are total orders, this depth function was also already applied in Schollmeyer [2017b]. There, the phenomena of the *wisdom of the crowd* (cf., Surowiecki [2004]) in the special context of ranking data was analyzed. We used a ranking task from Lee et al. [2014] where the participants of a study were asked to rank the former 44 US presidents according to their time of presidency. Among other things, we analyzed the statistical model fit of different classical models for rankings (concretely the models described in Biernacki and Jacques [2013] and Lee et al. [2014]). For this, we used the generalized Tukeys depth. Another application of the generalized Tukeys depth can be found in Jansen et al. [2018a] where we used the generalized Tukeys median as an aggregate for a consensus order in the context of social choice theory.

So far, beyond some statistical tests of model fit in Schollmeyer [2017b], we did only descriptive analyses. If the analyzed data set can be assumed to be an i.i.d. sample from some underly-

[9] Alternatively, instead of looking at the whole ROC curves one could also look only at a certain range of the ROC curves where the classifiers are usually applied. This would lead to more dense partial orders. Therefore, with this one has some possibility to flexibly scale the partiality of the observed partial orders.

[10] A depth contour of level λ is the set of all data points that have a depth of at least α.

ing population, it is natural to also consider statistical inference. At the time we are preparing a paper (Blocher et al.) that deals with statistical tests as well a regression for non-standard random variables that can be treated with one of our developed depth functions. For a two sample location test, a permutation based approach can be used. Concretely, one can more or less straight-forwardly adapt the permutation test from Li and Liu [2004] which is based on depth vs. depth plots (DD plots). Both the T-based test, as well as the M-based test described in Li and Liu [2004] can be adapted. In fact, we were already able to generalize both tests to a test with one parameter $\lambda \in [0, 1]$ that controls the type of the test. For $\lambda = 1$ one gets a test that really tests for different locations (this corresponds to the M-based test). For λ very small one gets a test that more or less tests, if the distributions in both subsamples are in a specific sense separated (this corresponds to the T-based test). The idea behind this generalization of the test is to maximize the depth of a data point w.r.t. one subsample under the constraint that the depth of this point w.r.t. the other subsample is above some threshold. (This threshold is controled by the parameter λ). Opposed to the case $\mathbb{R}^d$, the question if one optimizes over all possible data points or if one optimizes e.g., only over all observed data points, becomes more crucial because for non-standard data the (notion of) dimensionality of the underlying space is to some extent unclear in the first place and could possibly be too high. Therefore, we are currently also investigating possibilities for regularizing our proposed test. For a one-sample test, the situation seems to be still more unclear, because one has to somehow specify the distribution under the null hypotheses. Another possibility would be to rely on bootstrapping. However, for example for partial order data, the situation here is still very non-standard: For example a mirroring approach (cf., e.g., [Hesterberg, 2015, Section 4.2])[11] for constructing a reverse bootstrap percentile confidence interval is not possible because we are not equipped with an invertible group operation, here. For parametric tests, the parametric models developed in Blocher et al. [2022] can be used. Also non-parametric tests for dispersion or shape seem to be more difficult to develop, but we still plan to propose some sort of such tests by working with a (possibly parametrically specified) reference distribution. For regression, our ideas are still in its first stages. Since all depth functions considered so far can be easily equipped with weighting schemes, non-parametric kernel-like techniques seem to be applicable, here.

Beyond concrete proposals and applications of depth functions, within the methodology of data depth, already for $\mathbb{R}^d$ there is usually also an emphasis of structural properties a depth function should obey (cf., also Contribution 2). This is useful firstly to understand the inner workings of a given depth function and secondly to classify and categorize the rich class of possible depth function proposals. For non-standard data, this aspect becomes at the same time i) more rich and ii) more meagre:

i) Within the approach of FCA to data depth, we are able to treat different data types within the same framework. Therefore we are able to study for example the invariants (and variances) of a depth function both within a restricted data type, as well as across different data types.

ii) Different notions known from $\mathbb{R}^d$ become more difficult to capture or can still become more or less empty concepts: For example the property *maximality at center* (cf., e.g., Serfling and Zuo [2000]) in the concrete statement that a depth function should be maximal at the center of symmetry would need the notion of a symmetric distribution. In $\mathbb{R}^d$, point symmetry can be based on the notion of a translation. However, for many non-standard data, like for example partial orders, a notion of translation is (partly) missing. A very weak notion of point symmetry that uses an involution that models a mirroring of a point on a possible center s of symmetry is used in Theorem 1 of Contribution 2. (Compare also Section 5.3 where a translation group for the case of partial order data in the context of social choice theory is shortly discussed.) However, if this notion of point symmetry can be often fulfilled in relevant cases is a question of further research. Alternatively, for $\mathbb{R}^d$ another notion of symmetry which is of course a stronger one is the notion of

[11] However, note further that such an approach is often not a good idea, see [Hesterberg, 2015, Section 4.4].

rotational symmetry: One could say that a law P is symmetric around a point z if it is invariant under all isometries with fixpoint z. If one would have a synthetic approach to the geometry of $\mathbb{R}^d$ and only looks at the incidence of e.g., points and halfspaces, then it appears natural to look at all (continuous) automorphisms that keep the combinatorial structure of incidence between points and halfspaces[12] (and in particular map halfspaces to halfspaces) and that have one fixpoint z. Then for example a stretching of one coordinate axis would also be such an automorphism and essentially there cannot be any non-trivial law P that is invariant under all such automorphisms. Therefore this notion of a symmetric law would be empty within this formalization.

As already mentioned, there are formal contexts for which there is no strictly quasiconcave depth function at all. Therefore it is of high interest to characterize, for which formal contexts there is a strictly quasiconcave depth function. Additionally, there are many interesting formal contexts (including the case of partial order data) where there exists no strictly quasiconcave depth function. Therefore, we intend to also study the existence of depth functions that are free w.r.t. quasiconcavity. Our current insights suggest that the *ufg*-depth is a promising candidate for such a (strongly) free depth function. This is only one of many reasons why we intend to further study the *ufg*-depth in another contribution (Blocher and Schollmeyer), from a structural, from a computational and also from a statistical perspective.

[12] A depth function that is invariant under such automorphisms is called a combinatorial depth in Mosler [2013]. Funnily enough, the classical notion of *affine equivariance* looks at the broader class of linear mappings instead of only isometries, therefore, the exact conceptual treatment of $\mathbb{R}^d$ in a geometric fashion seems to be a little bit unclear.

5 B: Decision Making under weakly structured information

5.1 Introduction

Measurement theoretical considerations do also play an important role in decision theory. For example the classical notion of first order stochastic dominance, which is aligned to an ordinal scale of measurement, is a basic demand in decision theory. On the other hand, if the outcome of an uncertain payoff has a cardinal scale like for example a random monetary payoff, then a reasonable rationale for decision making is to apply expected utility theory, either directly by equating money with utility or by using a specific utility function that for example accounts for decreasing returns to scale. This would be directly related to second order stochastic dominance. However, between these extreme cases there are also intermediate cases thinkable. The contributions of this part are concerned with these intermediate cases: While the first contribution introduces the basic mathematical apparatus and the general concepts, the second contribution deals with the explicit elicitation of a decision maker in such an intermediate situation that one could call *a situation of weakly structured (partially ordinal and partially cardinal) information*. The third contribution in this part deals with notions of risk aversion, which are loosely related to the concept of second order stochastic dominance. Instead of a cardinal scale of measurement that is usually assumed for second order stochastic dominance, we deal here with an only (totally) ordinal scale of measurement and show that still in this very weakly structured situation there are weak notions of relative risk aversion establishable.

5.2 Our Contributions

Contribution 4

Christoph Jansen, Georg Schollmeyer, and Thomas Augustin. Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences. International Journal of Approximate Reasoning, 98:112–131, 2018.

Original Abstract

We introduce three different approaches for decision making under uncertainty if (I) there is only partial (both cardinally and ordinally scaled) information on an agent's preferences and (II) the uncertainty about the states of nature is described by a credal set (or some other imprecise probabilistic model). Particularly, situation (I) is modeled by a pair of binary relations, one specifying the partial rank order of the alternatives and the other modeling partial information on the strength of preference. Our first approach relies on decision criteria constructing complete rankings of the available acts that are based on generalized expectation intervals. Subsequently, we introduce different concepts of global admissibility that construct partial orders between the available acts by comparing them all simultaneously. Finally, we define criteria induced by suitable binary relations on the set of acts and, therefore, can be understood as concepts of local admissibility. For certain criteria, we provide linear programming based algorithms for checking optimality/admissibility of acts. Additionally, the paper includes a discussion of a prototypical situation by means of a toy example.

In this contribution we introduce the notion of so-called *preference systems* that can be used for solving decision problems when the scale of measurement of the outcomes of the acts is neither purely ordinal nor purely cardinal but somewhere between these extreme cases. Additionally, we use these preference systems in a situation where also the information about the probabilistic uncertainty of the considered states of nature is itself under additional epistemic uncertainty, or where the uncertainty about the state of nature is not purely probabilistic, but instead described by models of imprecise probabilities under an ontological understanding (cf., e.g., Walley and Fine [1982]). Let A be a non-empty set that describes the possible outcomes within a decision problem. A preference system is then a triple $\mathcal{A} = (A, R_1, R_2)$ where $R_1 \subseteq A \times A$ is a preorder[13] on A and $R_2 \subseteq R_1 \times R_1$ is a preorder on R_1. The preference system models the preferences between certain outcomes/alternatives in the following way:

- If a pair (a, b) is in R_1, we interpret this as a is at least as desirable as b, that is a and b can be ordered by preference. If neither (a, b), nor (b, a) is in R_1, then a and b are incomparable.
- If a pair of pairs $((a, b), (c, d))$ is in R_2, then this is interpreted as exchanging alternative b by alternative a is at least as desirable as exchanging alternative d by alternative c, that is, a is more strongly preferred over b than c is preferred over d. If both $((a, b), (c, d)) \notin R_2$ and $((c, d), (a, b)) \notin R_2$, then the exchange of b by a is incomparable to the exchange of d by c.

Now, in decision theory the general problem situation is that one has to choose between a given set $\mathcal{G} \subseteq A^S$ of acts that can be seen as random variables that map from a space S of possible states of nature to the set A of outcomes, i.e., if state $s \in S$ is realized, then an act $X \in \mathcal{G}$ would lead to the outcome $X(s)$. In the most simple case the uncertainty underlying the realized state is modeled by a probability measure π on S and the outcome space A is a subset of the reals which directly measures the utility of an outcome in a cardinal manner. In this case, a classical and very natural decision criteria is to choose an act $X^* \in \mathcal{G}$ that maximizes expected utility, i.e., for which we have

$$\forall X \in \mathcal{G} : \mathbb{E}_\pi(X^*) \geq \mathbb{E}_\pi(X).$$

Now, in our more complex situation, the outcome space does not directly codify metrically scaled utilities. In fact there are no utilities present in the first place, one does only have relational statements about preferences between outcomes or between exchanges of outcomes. Therefore, one way to proceed would be (like also done in parts of the theory of stochastic dominance) to look at the family of all utility functions that in a sense respect the relational preference structure given by the preference system. Concretely, one could say for a possible utility function $u : A \longrightarrow [0, 1]$ that it respects a preference system $\mathcal{A}$ if we have

i) If $(a, b) \in R_1$, then $u(a) \geq u(b)$ with equality iff additionally $(b, a) \in R_1$

ii) If $((a, b), (c, d)) \in R_2$ then $u(a) - u(b) \geq u(c) - u(d)$ with equality iff additionally $((c, d), (a, b)) \in R_2$.

A preference system would then be called *consistent* if there is at least one utility function[14] that respects the preference system. For making decisions between acts from $\mathcal{G}$ one can then look at the whole family $\mathcal{U}$ of all utility functions that respect a given preference system $\mathcal{A}$. Due to technical reasons, in the contribution we take the subfamily $\mathcal{N}$ of all utility functions $u \in \mathcal{U}$ that are additionally normalized (i.e. $\inf\{u(a) \mid a \in A\} = 0$ and $\sup\{u(a) \mid a \in A\} = 1$).

In addition to the complexity/indeterminacy of the preferences, in our contribution we also allow for a more complex situation regarding the uncertainty about the realization of the actual state $s \in S$. Instead of a classical probability measure π over S we allow for any kind of uncertainty model that can mathematically be modelled as a set $\mathcal{M}$ of probability measures that allows an

[13] A preorder on a set A is a reflexive and transitive binary relation on A.

[14] The demand that the utility function maps to $[0, 1]$ is only a technical one, here.

interpretation as *the set of all probability measures that are in a certain sense compatible with our uncertainty assessment*[15]. Such kinds of uncertainty models can be found within a broad class of different theories that can be more or less subsumed under the umbrella term imprecise probabilities (for an introduction to imprecise probabilities, see, e.g., Augustin et al. [2014]). Examples range from robust Bayesian analysis (Insua and Ruggeri [2012]) over generalized frequentist methods (Fierens [2009]), logical probability (Weichselberger [2007]), lower previsions (Troffaes and De Cooman [2014]), interval probability (Weichselberger [2001, 2000]), non-parametric predictive inference (Coolen and Augustin [2009]) to belief functions (Shafer [1976]) and possibility theory (Dubois and Prade [1988]).

For the establishment of a concrete decision rule there are now many different ways to proceed. In the contribution we present the following concrete three proposals:

I) Criteria based on generalized expectation intervals.

II) Criteria based on global comparisons.

III) Criteria based on pairwise comparisons.

We now shortly sketch the three approaches. For a detailed description we refer to the contribution.

I) Criteria based on generalized expectation intervals Here the idea is to generalize expected utility by replacing the expected value by the whole set of expected values that is obtained by varying over all normalized utilities that respect the preference system and over all precise probabilities π which lie in the set $\mathcal{M}$ (always assuming that the respective expectations exist). Then, one can look at the upper and lower bounds of this set and define an expectation interval for an act X as

$$\mathbb{E}_{\mathcal{N},\mathcal{M}}(X) := \left[\inf_{u\in\mathcal{N},\pi\in\mathcal{M}} \mathbb{E}_\pi(u\circ X), \sup_{u\in\mathcal{N},\pi\in\mathcal{M}} \mathbb{E}_\pi(u\circ X) \right].$$

For decision making one can then choose one act X^* that maximizes i) the lower bound (maximin), ii) the upper bound (maximax) or iii) a convex combination of the lower and the upper bound. With this one will get i) a pessimistic and ii) an optimistic decision rule, as well as with iii) a whole range of decision rules between these two extremes.

II) Criteria based on global comparisons Opposed to approach I) where one first computes upper and lower bounds for each act X separately, in approach II one looks for every fixed pair of utility $u \in \mathcal{N}$ and probability $\pi \in \mathcal{M}$ globally on all acts and checks for a candidate X if $\mathbb{E}_\pi(u\circ X) \geq \mathbb{E}_\pi(Y)$ for all $Y \in \mathcal{G}$. Then, in a second step one quantifies over $u \in \mathcal{N}$ and $\pi \in \mathcal{M}$. In this sense, one looks globally on all acts in $\mathcal{G}$. Depending on how one concretely quantifies, different decision criteria appear. Concretely, in our contribution we define an act X to be

i) $\mathcal{A} \mid \mathcal{M}$-admissible if $\exists u \in \mathcal{N}\ \exists \pi \in \mathcal{M} :\ \forall Y \in \mathcal{G} :\ \mathbb{E}_\pi(u\circ X) \geq \mathbb{E}_\pi(u\circ Y)$

ii) $\mathcal{A}$-admissible if $\exists u \in \mathcal{N}\ \forall \pi \in \mathcal{M} :\ \forall Y \in \mathcal{G} :\ \mathbb{E}_\pi(u\circ X) \geq \mathbb{E}_\pi(u\circ Y)$

iii) $\mathcal{M}$-admissible if $\exists \pi \in \mathcal{M}\ \forall u \in \mathcal{N} :\ \forall Y \in \mathcal{G} :\ \mathbb{E}_\pi(u\circ X) \geq \mathbb{E}_\pi(u\circ Y)$

iv) $\mathcal{A} \mid \mathcal{M}$-dominant if $\forall u \in \mathcal{N}\ \forall \pi \in \mathcal{M} :\ \forall Y \in \mathcal{G} :\ \mathbb{E}_\pi(u\circ X) \geq \mathbb{E}_\pi(u\circ Y)$

Given one of these criteria, for decision making one would simply choose one act that fulfills this criteria (if such an act exists). Note that we used here only quantifications where a $\forall$-quantor follows an $\exists$-quantor and not vice versa. This means that at no point a certain utility u depends

[15]Some models of uncertainty, like e.g., possibility theory do not directly allow for such an interpretation but usually could also be used because despite a lack of a possible interpretation of $\mathcal{M}$ as a set of compatible probability measures, mathematically the concretely obtained results are usually reasonable and still interpretable in an appropriate decision theoretic manner.

on the concretely envisaged probability π and also a certain probability π does not depend on the concretely envisaged utility u. In this sense the above criteria are in an additional way global.

III) Criteria based on pairwise comparisons Here, we do to not look globally on all acts simultaneously. Instead we only look at pairs of acts. Then we quantify over $u \in \mathcal{N}$ and $\pi \in \mathcal{M}$ like done in approach II), but now without emphasis on a global viewpoint such that all-together six combinations of quantifying can be done. Concretely, for $X, Y \in \mathcal{G}$ we define for the placeholders $\square$ and $\triangle$ which can be replaced by the $\exists$-quantor or the $\forall$-quantor the binary relation $R_{\square\triangle}$ via

$$(X,Y) \in R_{\square\triangle} :\Longleftrightarrow \square u \in \mathcal{N}; \triangle \pi \in \mathcal{M} : \ \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y).$$

For decision making one would then choose an act from $\mathcal{G}$ that is a maximal element w.r.t. the binary relation $R_{\square\triangle}$. (Note that in general, only the relation $R_{\forall\forall}$ is transitive. Therefore, for the other relations one possibly would have to calculate the transitive hull beforehand.)

With the above decision criteria we have enlarged the toolbox for decision making, especially for cases where the scale of measurement of the outcomes is non-standard. To make an example: If $\mathcal{M} = \{\pi\}$ is a singleton and $A \subseteq \mathbb{R}$ is a bounded interval of the reals, $R_1 = \leq \cap A \times A$ where $\leq$ is the usual $\leq$-relation on the reals and $R_2 = \emptyset$. Then the relation $R_{\forall\forall}$ is exactly the relation of first order stochastic dominance.

If we replace $R_2 = \emptyset$ by

$$R_2 := \{(a,b),(c,d) \mid a \geq b; c \geq d; a - b \geq c - d\}$$

then the generalized expectation interval of approach I) collapses to a point-interval that coincides with the classical expectation.

If we replace R_2 by

$$R_2 := \{(a,b),(c,d) \mid a \geq b; c \geq d; a - b \geq c - d; a \leq d\}$$

then this would correspond to decreasing returns to scale and the relation $R_{\forall\forall}$ would essentially be second order stochastic dominance.

Now, with our notion of preference systems one is able to explicitly specify intermediate cases between the extremes from above. To give one example: If the outcome space is describing the self-described health status of a person, then usually there is one more or less clear cutting point between *good* and *bad*. Though a variable like health would usually be treated as ordinal, instead of $R_2 = \emptyset$ (which would model this situation) one may say with some reason that an exchange from a bad to a good health status (if it is not negligibly small) has more strength than an exchange within a good or within a bad health status. This could be naturally be implemented in the relation R_2.

Instead of directly varying the relation R_2 by substance matter considerations, one could also try to vary the size of the class of utility functions $\mathcal{N}$. In our contribution we explicitly do this. We introduce a kind of a granularity parameter $\delta \geq 0$ and modify the set $\mathcal{N}$ to the set

$$\mathcal{N}_\delta := \{u \in \mathcal{N} \mid \forall (a,b) \in P_{R_1} : u(a) - u(b) \geq \delta \ \& \ \forall ((a,b),(c,d)) \in P_{R_2} : a - b \geq c - d + \delta\}$$

where P_{R_1} and P_{R_2} denote the strict parts of R_1 and R_2, respectively. Then, this parameter can be seen as some kind of regularization parameter that controls the size of $\mathcal{N}_\delta$. Larger values of δ lead to a smaller set $\mathcal{N}_\delta$ and therefore for example to a stronger ordering power of the relation $R_{\forall\forall}$. The parameter δ is inspired by the notion of a *just notable difference* from psychophysics (see Luce [1956] for details): One may presuppose that if a decision maker has a preference between

outcomes or between exchanges of outcomes, then his corresponding utility differences should in a sense be large enough that he can notice a preference. Mathematically, for example in the situation from above with the relation

$$R_2 := \emptyset$$

the parameter δ would then allow to continuously switch between an implicitly ordinal versus an implicitly cardinal scale of measurement of the outcome space.

Finally, in our contribution we do not only introduce the above concepts for decision making, but we also concretely propose linear programming formulations to compute the most of the decision criteria. The basic idea is here to model the utility function by introducing a decision variable for every utility value that the utility function actually takes. The incorporation of the imprecision in the uncertainty model $\mathcal{M}$ is done as usual by formulating one decision problem for every extreme point of the imprecise uncertainty model.

In the last section of the contribution we also apply our algorithms to a small synthetic example.

Contribution 5

Christoph Jansen, Hannah Blocher, Thomas Augustin, and Georg Schollmeyer. Information efficient learning of complexly structured preferences: Elicitation procedures and their application to decision making under uncertainty. International Journal of Approximate Reasoning, 144:69–91, 2022.

Original Abstract

In this paper we propose efficient methods for elicitation of complexly structured preferences and utilize these in problems of decision making under (severe) uncertainty. Based on the general framework introduced in Jansen et al. [2018b], we now design elicitation procedures and algorithms that enable decision makers to reveal their underlying preference system (i.e. two relations, one encoding the ordinal, the other the cardinal part of the preferences) while having to answer as few as possible simple ranking questions. Here, two different approaches are followed. The first approach directly utilizes the collected ranking data for obtaining the ordinal part of the preferences, while their cardinal part is constructed implicitly by measuring the decision maker's consideration times. In contrast, the second approach explicitly elicits also the cardinal part of the decision maker's preference system, however, only an approximate version of it. This approximation is obtained by additionally collecting labels of preference strength during the elicitation procedure. For both approaches, we give conditions under which they produce the decision maker's true preference system and investigate how their efficiency can be improved. For the latter purpose, besides data-free approaches, we also discuss ways for statistically guiding the elicitation procedure if data from elicitations of previous decision makers is available. Finally, we demonstrate how the proposed elicitation methods can be utilized in problems of decision under (severe) uncertainty. Precisely, we show that under certain conditions optimal decisions can be found without fully specifying the preference system.

As described above, the relations R_1 and R_2 of a preference system give a very flexible way to specify the preferences of a decision maker. But how would one concretely elicitate these preferences? In particular, if the outcome space has n elements, then for the elicitation of R_1 one would have to ask the decision maker for $n(n-1)/2$ pair comparisons if one would do this straight-forwardly. Moreover, for R_2 it would be $n^2(n^2-1)/2$ pair comparisons. This is of course usually too much.

Therefore, in this contribution we propose information efficient elicitation procedures that try to make a decision under the demand to ask as few pair comparisons as possible. In the contribution we propose two elicitation procedures called *time elicitation* and *label elicitation*. We now shortly sketch these two methods.

Procedure 1: time elicitation The first procedure for eliciting the preference system of a decision maker uses data about consideration times that is obtained during the elicitation process of R_1. The idea behind this is the assumption that the strength of the preference between two consequences decreases in the time that ranking the two consequences takes. For a pair of alternatives $(a_i, a_j) \in P_{R_1}$, besides the ranking, we also measure the consideration time t_{ij} the decision maker needs for ranking the two consequences. This way of using additional data is inspired by the paradata-approach from survey methodology (e.g., Kreuter [2013]) that uses the fact that individual data about the surveying process itself can contain valuable information about respondents and their dispositions. The collected consideration times are then utilized for constructing the relation R_2 being a candidate for the true relation (denoted R_2^* in the sequel, analogously, the true ordinal part is denoted by R_1^*). All in all, this constructs a preference system while asking only questions on the decision maker's ordinal preferences R_1^*. The concrete procedure of time elicitation can be described as follows:

We assume a finite outcome space $A = \{a_1, \dots, a_n\}$. Then, we start with three initial relations $R_1 = \{(a, a) \mid a \in A\}$ and $R_2 = \emptyset$ and $C = \emptyset$. The relation C models the set of incomparable pairs. Then, the decision maker is iteratively asked about the preferences between certain pairs $\{a_i, a_j\}$ from the set $A_{\{2\}} := \{\{a, b\} : a, b \in A, a \neq b\}$. Note that instead of asking about all elements in $A_{\{2\}}$ one could alternatively also ask only for a subset $B \subseteq A_{\{2\}}$. Of course one can then expect to get only a subsystem (i.e., $R_1 \subseteq R_1^*$ and $R_2 \subseteq R_2^*$) of the true preference system of the decision maker. In every step of the elicitation there are four possible cases:

i) The decision maker judges a_i and a_j as incomparable. In this case R_1 and R_2 remain unchanged, but the pairs (a_i, a_j) and (a_j, a_i) are added to C . The times t_{ij} and t_{ji} are set to zero.

ii) The decision maker prefers a_i to a_j. Then we add the pair (a_i, a_j) to R_1 and measure the time $t_{ij} > 0$ and the time t_{ji} is set to zero.

iii) The decision maker prefers a_j to a_i. In this case we add (a_j, a_i) to R_1 and measure the time $t_{ji} > 0$. The time t_{ij} is set to zero.

iv) The decision maker judges a_i and a_j as equivalent[16]. We then add the pairs (a_i, a_j) and (a_j, a_i) to R_1. The times t_{ij} and t_{ji} are set to some $c_\infty \in \mathbb{R}$ with $c_\infty > \max\{t_{pq} \mid (a_p, a_q) \in P_{R_1}\}$.

The above procedure will produce a relation R_1 that is a subrelation of R_1^* and that in this sense approximates R_1^*. Regarding R_2, we proceed as follows:
First, set $t_{ii} := c_\infty$ for $i = 1, \dots, n$. Then, for constructing R_2 we use the consideration times: We successively take pairs of pairs (a_i, a_j) and $(a_k, a_l) \in R_1$ and define R_2 by $((a_i, a_j), (a_k, a_l)) \in R_2 :\iff t_{kl} \geq t_{ij} > 0$, i.e., if the decision between a_k and a_l took at least as long as the decision between a_i and a_j. All-together, this procedure produces a preference system $A = (A, R_1, R_2)$ on A. If all elements in $A_{\{2\}}$ were asked, then under the following assumption A1, the obtained preference system coincides with the true preference system of the decision maker. (If only a subset is asked, then only a subsystem is obtained.)

Assumption A1 For $(a_i, a_j), (a_k, a_l) \in R_1^*$ it holds

i) $t_{kl} > t_{ij} > 0 \iff ((a_i, a_j), (a_k, a_l)) \in P_{R_2^*}$.

[16]Two consequences are called equivalent if $(a, b) \in R_1^*$ & $(b, a) \in R_1^*$ The set of all equivalent pairs is denoted with $I_{R_1^*}$.

ii) $t_{kl} = t_{ij} > 0 \iff ((a_i, a_j), (a_k, a_l)) \in I_{R_2^*}$.

iii) The maximal consideration time c_∞ is attained exactly for equivalent consequences, i.e. $t_{ij} = t_{ji} = c_\infty \iff (a_i, a_j) \in I_{R_1^*}$.

Now the efficiency of time elicitation can be remarkably improved if we assume a decision maker with a transitive relation R_1^* that satisfies assumption $A1$ and the following two additional assumptions:

Assumption A2 For $(a_i, a_j), (a_j, a_k) \in P_{R_1^*}$ we have $1/t_{ij} + 1/t_{jk} = 1/t_{ik}$, if $(a_i, a_k) \in P_{R_1^*}$.

Assumption A3 For $(a_i, a_j) \in I_{R_1^*}$ we have $t_{ki} = t_{kj}$ whenever $(a_k, a_i), (a_k, a_j) \in P_{R_1^*}$ and $t_{ik} = t_{jk}$ whenever $(a_i, a_k), (a_j, a_k) \in P_{R_1^*}$.

These additional assumptions guarantee that the consideration times are aligned to the relation R_2^*. Now, suppose time elicitation has produced the relations R_1^k and C^k after k pairs have been presented. Transitivity of R_1^* then allows to deduce all preferences for pairs in $H^k \setminus R_1^k$ where H^k is the transitive hull of R_1^k. The pair to present in step $k+1$ can then be selected from the (usually remarkably) smaller set $A_2 \setminus \{\{a, b\} \mid (a, b) \in H^k \text{ or } (b, a) \in H^k \text{ or } (a, b) \in C^k\}$. The pairs in the set $\{\{a, b\} \mid (a, b) \in H^k \text{ or } (b, a) \in H^k \text{ or } (a, b) \in C^k\}$ and the corresponding consideration times can then be autmatically determined according to transitivity and assumption A2. Of course, the assumption A2 is a very strong assumption. An alternative way of elicitation is label elicitation.

Procedure 2: label elicitation

This elicitation produces an approximation of R_1^* by asking only questions about R_1^*. In contrast to time elicitation, this construction does not rely on the use of paradata. Instead it explicitly elicitates labels of preference strength. These labels are intended to provide ordinal information about preference strength. The idea is as follows: To every presented pair of consequences, the decision maker assigns a label from some previously fixed set of labels that are interpreted only on an ordinal scale of measurement, plus the possibility of labels that represent incomparability and indifference. In case two presented pairs are comparable, the assigned labels will be ordered and we add the corresponding pair of pairs to the relation approximating R_2^* whenever the first pair receives a strictly greater label than the latter (or both receive label 0). With this, label elicitation can be described in more detail as follows:

Again, we have a finite set $A = \{a_1, ..., a_k\}$ of outcomes. We start with two empty relations $R_1 = \emptyset$ and $R_2 = \emptyset$. We then successively ask about the preferences between some (not necessarily all) pairs $(a_i, a_j) \in A \times A$, where the decision maker assigns a label from the set $L_r := \{n, c, 0, 1, ..., r\}$ to every such pair. This labeling process can be described by a labeling function $l_r : A \times A \to L_r$. The labels from L_r have the following interpretation: The higher the label from $L_r \setminus \{0, n, c\}$ assigned to a pair $(a_i, a_j) \in A \times A$ is, the stronger is the decision maker's strict preference of a_i over a_j. If the label n is assigned to (a_i, a_j), this means that a_i and a_j are incomparable, whereas the label 0 is interpreted as indifference between a_i and a_j. If label c is assigned to (a_i, a_j), this means that a_i is strictly preferred to a_j, but no statement about intensity of preference is possible. For simplicity, we write l_r^{ij} instead of $l_r((a_i, a_j))$ in the sequel. The collected labels are utilized to successively build up a preference system: Whenever $l_r^{ij} \in L_r \setminus \{n, 0\}$, we add the pair (a_i, a_j) to R_1 . If $l_r^{ij} = 0$, we add both pairs (a_i, a_j) and (a_j, a_i) to R_1 , whereas if $l_r^{ij} = n$ the relation R_1 remains unchanged. This procedure leaves us with a (potentially non-complete) relation R_1 approximating the ordinal part R_1^* of the true preference system. Subsequently, we can utilize the labels of preference intensity that we collected during the procedure for also constructing an approximation R_2 for the cardinal part R_2^* of the decision maker's preferences. For that, we take pairs of pairs $(a_i, a_j), (a_k, a_l) \in R_1$ and add $((a_i, a_j), (a_k, a_l))$ to R_2 if and only if $l_r^{ij} > l_r^{kl}$ or if $l_r^{ij} = l_r^{kl} = 0$.

This procedure produces a preference system. Again, under assumptions that guarantee an alignment of the decision makers preference system with the labels (concretely assumptions 4, 5 and 6 in the contribution), this preference system is a subsystem of the true preference system of the decision maker (cf., proposition 4 of the contribution). Moreover, if all elements of $A_{\{2\}}$ were asked and if the label function is fine-grained enough (i.e., r is large enough), the true preference system is obtained by label elicitation.

At the end of Section 4.3 of the contribution we also discuss a further way to make label elicitation more efficient: The idea is to choose the number r of labels dynamically. Starting with an initially chosen r we start a first elicitation round. Then we make a further elicitation round with a larger r but only for pairs where we have some incomparabilities that are only due to identical labels.

Additionally, the contribution also discusses statistical ways for making the elicitation procedures more efficient (cf., Example 2 of the contribution). The idea here is that in certain situations it is possible to get some information about previous decision processes made by former decision makers. If it is reasonable to assume that these decision makers are very similar to the decision maker who is actually elicitated, then one can use their preferences to statistically guide which question to ask next in the elicitation process. Here, in principle one could use any predictive model that can handle the preferences of the former decision makers as an input. Then, for every possible pair comparison to ask next, one could make a probabilistic prediction about which preference (or indifference or incomparability) has which probability. Based on all predicted probabilities one can decide, which pair comparison should be asked next. How exactly to decide this based on the probabilities is not obvious. In principle, there are two possibilities: If one knows the exact decision criteria, then one maybe can incorporate it into the exact rule for deciding about the next pair comparison to ask. Alternatively, one can build a rule without incorporating the concretely used decision criteria.

In the Contribution, we used the second approach and simply asked in every step of the elicitation procedure that pair comparison for which a comparability is most probable. For the predictive model we used the methodology of subgroup discovery (see Atzmueller [2015]) to make probabilistic predictions. As a decision criterion, we used the $\mathcal{A} \mid \mathcal{M}$-dominance from Contribution 4. A a small simulation study showed that this can in fact increase the efficiency of the elicitation.
For the simulation of preference systems we only used the relation R_1^*. For the random generation of relations R_1^*, we used the Mallows model (see Mallows [1957]) adapted to partial orders. The relation R_2^* was not used for the simulation of the random preference systems and also not for the prediction of new pairs. We simply set $R_2^* = \emptyset$. We intent to also include the cardinal part R_2^* of the preference systems in future research, compare Section 5.3.

Contribution 6

Jean Baccelli, Georg Schollmeyer, and Christoph Jansen (2022): Risk aversion over finite domains. Theory and Decision, 93:371–397.

Original Abstract

We investigate risk attitudes when the underlying domain of payoffs is finite and the payoffs are, in general, not numerical. In such cases, the traditional notions of absolute risk attitudes, that are designed for convex domains of numerical payoffs, are not applicable. We introduce comparative notions of weak and strong risk attitudes that remain applicable. We examine how they are characterized within the rank-dependent utility model, thus including expected utility as a special case. In particular, we characterize strong comparative risk aversion under

rank-dependent utility. This is our main result. From this and other findings, we draw two novel conclusions. First, under expected utility, weak and strong comparative risk aversion are characterized by the same condition over finite domains. By contrast, such is not the case under non-expected utility. Second, under expected utility, weak (respectively: strong) comparative risk aversion is characterized by the same condition when the utility functions have finite range and when they have convex range (alternatively, when the payoffs are numerical and their domain is finite or convex, respectively). By contrast, such is not the case under non-expected utility. Thus, considering comparative risk aversion over finite domains leads to a better understanding of the divide between expected and non-expected utility, more generally, the structural properties of the main models of decision-making under risk.

This contribution establishes notions of risk aversion in a setting where the underlying scale of measurement of the outcome space of envisaged lotteries is only of (total) ordinal scale of measurement. The classical risk attitudes studied in economics usually start with lotteries (i.e., random variables with a finite support) that map in an outcome space that is assumed to be of cardinal scale of measurement. The notions of risk aversion, risk seeking, and risk neutrality are usually based on the notion of an increase in risk. A prominent notion of an increase in risk is based on the concept of a mean-preserving spread (Rothschild and Stiglitz [1970]): Informally speaking, a random variable Y is a mean-preserving spread of X if Y and X have the same expectation and Y is obtained from X by spreading out probability mass of X in the direction from the mean to the tails. In this case one could intuitively say that Y is more risky, more variable or more uncertain[17] than X. If one now thinks about ordinal random variables, then the mean is not well-suited in measurement theoretic terms, and the notion of a mean preserving spread has to be replaced by another notion. One such notion is that of a median preserving spread (Allison and Foster [2004], de la Vega [2018]): If Y is obtained from X by spreading out probability mass in the direction from the median to the tails, then it appears natural to say that Y is more risky than X. If one replaces the median by an r-quantile with $r \in (0,1)$, then one speaks about a quantile preserving spreading or an r-spread (Mendelson [1987], Bommier et al. [2012]). Additionally, if Y is an r-spread of for some $r \in (0,1)$ one may say simply that Y is a spread of X. Finally, if X is degenerate, meaning that X is a constant random variable (and therefore without any variability) one may say that Y is a basic spread of X. The notion of a spread can be easily characterized with the cumulative distribution functions: Assuming continuous distribution functions for ease of explication, for example Y is an r-spread of X if and only if the distribution functions single cross at one point z where $P(X \leq z) = P(Y \leq z) = r$ and below z the distribution function of Y lies above that from X; and above z the distribution function of Y lies below the distribution function of X, which formalizes that probability mass is spread out in the direction from z to the tails. Then, one can define Y as more risky than X if Y is a spread or a r-spread (for some fixed r) or a basic spread of X, respectively. With this one would get notions of *more risky than* of different strengths. To formulate a notion of risk aversion of a decision maker, one could say for example that a decision maker is risk averse, if she prefers X to Y whenever Y is a spread of X. But this way to proceed would lead to more or less uninteresting concepts, because these concepts would clash with core properties of decision making under risk: For instance, assuming mixture-continuity[18] - a standard property satisfied by e.g., rank dependent utility and many other models of decision theory - one can show that a decision maker who is risk averse to all spreads necessarily violates strict respect of first order stochastic dominance. Similar conclusions hold for the other kinds of spread. Therefore, in the contribution, we analyze a relative notion of *comparative risk aversion*: We say that a decision maker A is *strongly more risk averse* than a decision maker B if every spread that B accepts is also accepted by A. If we demand this only for basic spreads then

[17] In Rothschild and Stiglitz [1970] these three terms are used synonymously.

[18] A decision maker satisfies mixture continuity if for any three lotteries p, q, r the sets $S = \{\alpha \in [0,1] \mid q \text{ is weakly preferred to } \alpha p + (1-\alpha)r\}$ and $T = \{\alpha \in [0,1] \mid \alpha p + (1-\alpha)r \text{ is weakly preferred to } q\}$ are closed in $[0,1]$. Here, for example $\alpha p + (1-\alpha)r$ denotes the mixture lottery where one plays lottery p with probability α and the lottery r with probability $(1-\alpha)$.

we speak about *weak risk aversion*.

The main result of our contribution is the characterization of strong comparative risk aversion under the model of rank dependent utility (RDU, Quiggin [1982]). Therefore, let us shortly introduce RDU:

Let $A = \{a_1, \ldots, a_n\}$ be a finite set of outcomes. Assume that all considered decision makers are ordinally equivalent, which means that they all prefer a_1 over a_2 over ... over a_n. Let L be the set of all probability distributions, or lotteries, over A. Then, RDU holds if there exists a strictly increasing, continuous probability weighting function $w : [0,1] \longrightarrow [0,1]$ with $w(0) = 0$ and $w(1) = 1$, and a strictly increasing utility function $u : X \longrightarrow \mathbb{R}$, such that (defining by convention $u(a_{n+1}) = 0$) the preference of the decision maker can be represented with the function $v : L \longrightarrow \mathbb{R}$ given by

$$v(l) = \sum_{i=1}^{n} \left[w\left(\sum_{j=1}^{i} p_j \right) (u(a_i) - u(a_{i+1})) \right]$$

via *lottery l is preferred to lottery l' if and only if* $v(l) > v(l')$. Here p_j denotes the probabilities that lottery l has outcome a_j.

The characterization of comparative risk aversion is now done by analyzing the probability weighting functions and the utility functions of two decision makers, which we denote with w_1 and w_2 and u_1 and u_2 in the sequel. The basic insight in the characterization is the study of the *degree of non-convexity* of w_1 as a function of w_2 (i.e., of the function $w_1 \circ w_2^{-1}$) and the *degree of non-concavity* of u_1 as a function of u_2 (i.e., of the function $u_1 \circ u_2^{-1}$, restricted to the range of u_2). For this, we introduce an index $I_w^{1,2}$ and an index $I_u^{1,2}$ as

$$I_w^{1,2} = \inf \left\{ \left. \frac{\frac{w_1(p)-w_1(q)}{w_2(p)-w_2(q)}}{\frac{w_1(r)-w_1(s)}{w_2(r)-w_2(q)}} \right| p,q,r,s \in [0,1], p > q \geq r > s \right\}$$

$$I_u^{1,2} = \sup \left\{ \left. \frac{\frac{u_1(a)-u_1(b)}{u_2(a)-w_2(b)}}{\frac{u_1(c)-u_1(d)}{u_2(c)-u_2(d)}} \right| a,b,c,d \in A, a \succ b \succeq c \succ d \right\}.$$

Here, the symbols $\succ$ and $\succeq$ denote the strict and weak preferences between the outcomes that both decision makers commonly have. The higher these indexes are, the less convex (concave, respectively) is w_1 as a function of w_2 (u_1 as a function of u_2, respectively). More concretely, one can show that

i) u_1 is more concave than u_2 (i.e., $u_1 \circ u_2$ is concave) if and only if $I_u^{1,2} \leq 1$.

ii) w_1 is more convex than w_2 (i.e., $w_1 \circ w_2^{-1}$ is convex) if and only if $I_w^{1,2} = 1$.

The asymmetry between *i)* and *ii)* stems from the fact that the weighting functions are assumed to be continuous. This implies that $I_w^{1,2} \leq 1$ always holds. In contrast, the utility functions are not necessarily continuous, which is of special importance in our contribution because we assume that the outcome space and therefore the ranges of the utility functions are finite and particularly not intervals. With this, we can state the following

Theorem For two (ordinally equivalent) RDU decision maker with weighting functions w_1, w_2 and utility functions u_1, u_2, decision maker 1 is strongly more risk averse than decision maker 2 if and only if

$$I_w^{1,2} \geq I_u^{1,2}.$$

Intuitively speaking one can interpret this result in the following sense. The risk aversion of decision maker 2 in comparison to decision maker 1 that may be due to a certain degree of non-concavity of the utility function u_1 as a function of u_2 is compensated by a larger amount of convexity of the probability weighting function of decision maker 2 compared to decision maker 1.

There are two special cases, for which this theorem has immediate corollaries. Firstly the case of classical expected utility (EU). This is simply the case where the weighting function is the identity map on $[0, 1]$. The other special case is the case of dual expected utility (DEU,Yaari [1987]) that is obtained if A is a subset of the reals and the utility function is the identity map on A.

Corollary 1 For two EU decision makers it holds that decision maker 1 is strongly more risk averse than decision maker 2 if and only if u_1 is more concave than u_2.

Corollary 2 For two DEU decision makers it holds that decision maker 1 is strongly more risk averse than decision maker 2 if and only if w_1 is more convex than w_2.

The corollaries from above are already known for the convex case. With the convex case we mean here that either the ranges of the utility functions are convex subsets of $\mathbb{R}$ or that the outcome space A is a convex subset of $\mathbb{R}$. For the case of a finite range of the utility functions, as far as our knowledge goes the corollaries, and Theorem 1 are new. Importantly from the point of view of proving things, some techniques like playing with a certainty equivalent are not available in the finite ranged case.

Finally, concerning weak risk aversion, we were not able to give a concise characterization (but see p.383ff of our contribution for some further partial characterizations). The characterization of weak risk aversion under RDU is an open question even in the convex case. Even under the assumption that the utility functions have convex range, giving a characterization would require characterizing under RDU the *absolute* concept of weak risk aversion (i.e., aversion to mean-preserving spreads with the added condition that the less risky lottery is degenerate). But this is a longstanding open problem of the field (see Chateauneuf and Cohen [1994], Chateauneuf et al. [1997]).

5.3 Related Work & Future Research

In the meanwhile, the concept of preference systems was already successfully applied in the context of machine learning in Jansen et al. [2023a] as well as in the context of statistics in Jansen et al. [2023b].

First, in Jansen et al. [2023a] we studied the performance of different classification algorithms over a sample of data sets. We looked at different performance measures and introduced an appropriate preference system for comparing different observed performance vectors. Concretely, we looked at three classical performance measures (*accuracy*, *area under the curve* and *Brier score*) under a partial cardinal scale of measurement: We specified that one performance vector is better than (or equal to) another performance vector if and only if it is better (or equal) in every dimension. Additionally, we say that an exchange (p, q) is better than (or equal to) an exchange $(\tilde{p}, \tilde{q})$ if p is better than (or equal to) q and $\tilde{p}$ is better than (or equal to) $\tilde{q}$ and the difference $p - q$ is in every dimension better than (or equal to) the difference $\tilde{p} - \tilde{q}$.

Second, in Jansen et al. [2023b] we use preference systems to operationalize a multivariate notion of poverty/inequality: While dimensions like income can be treated as of a cardinal scale of measurement, dimensions like education or health are only of an ordinal scale of measurement. All these aspects can be handled by introducing a preference system where, compared to a purely ordinal analysis, beyond the relation R_1 also the relation R_2 is non-trivial. (Of course, one could also argue for a non-cardinal scale of measurement for dimensions like income: For a purely ordinal

assessment one would fall back to a purely multidimensional ordinal analysis, while if one only assumes decreasing returns to scale, the relation R_2 would still be non-trivial). Additionally to the need of preference systems, in both papers also the fact, that one does only have a sample of the involved random variables, has to be taken into account. We deal with this by doing a permutation based statistical test based on the basic conceptualisation given in Schollmeyer et al. [2017]. The obtained test has some potential to be a good alternative to other tests like the tests discussed in Demšar [2006].

Concerning the elicitation of preference systems we intend to develop more specialized elicitation techniques that make still more use of data from previous elicitation processes: In Jansen et al. [2022], for the prediction of the pair comparison to ask in the next step in the elicitation procedure, we used the methodology of subgroup discovery. Concretely, in every step we constrained the training set to that partial orderings that are in accordance with all already elicited pairs of the decision maker that is actually elicitated. Therefore, in every step of the procedure the training data is getting smaller. At the same time the covariates for the prediction of new pairs are getting less complex. A thorough analysis of how this effects a possible over- or underfitting of the prediction procedure would be of high interest here. Fortunately, for the case of subgroup discovery, for example an explicit analysis of the Vapnik-Chervonenkis dimension of the covariate space is possible, cf., Schollmeyer et al. [2017]. We intend to explicitly incorporate such an analysis in an appropriate dynamical regularization scheme (for which there are also already concrete proposals available, see Schollmeyer et al. [2017], Schollmeyer [2023]) for a more efficient prediction of the next pair to elicitate.

Additionally, also for the simulation study based analysis of the elicitation procedure we intend to use more elaborate statistical models for preference systems. Concretely, we intent to replace the distance based Mallows model for partially ordered data used in Jansen et al. [2022] (cf., Example 2, p. 87) by the data depth based model from Blocher et al. [2022]. Actually, one motivation for developing the ideas within Blocher et al. [2022] was a demand for more non-parametric and therefore more flexible statistical models for partial orders to be able to simulate realistic scenarios for a random version of Relation R_1. Needless to say that also concerning the relation R_2 there is still the same demand: How can one develop reasonable statistical models for the cardinal part of a random preference system? For the answer one only needs to take Birkhoff seriously: Since R_2 is simply a preorder on R_1 we have a hierarchy over a hierarchy and thus again plain and simply a hierarchy. Therefore, again the whole toolbox of formal concept analysis together with all already developed methodology for data depth in the context of formal concept analysis can be applied! At least in a conceptual sense. Computational issues of course will become still more cumbersome by moving from R_1 to R_2.

Finally, not only for the simulation of preference system, but also for statistically guiding the elicitation procedure with the methodology of subgroup discovery - a methodology that can be very naturally framed in the language of formal concept analysis (cf., e.g., Boley and Grosskreutz [2009]) - we have a FCA method at hand that can be used also for incorporating the relation R_2.

Concerning Contribution 6 and a notion of one random variable Y being more variable or more dispersed than another random variable X is also of interest for random variables taking values in very abstract spaces. Think of random variables taking values in an object set G within an abstract setting of formal concept analysis. In the setting of Contribution 6 the random variables were of ordinal scale of measurement. The notion of a median preserving spread relied on the notion of the median and the notion of being below or above the median, respectively. In more abstract spaces the notion of a median can be captured with the help of data depth as a point with maximal depth. But one generally does not have a notion of below or above. However, it is still possible to define a reasonable notion of one random variable Y being more dispersed than another random variable X in the abstract setting of formal concept analysis by using the methodology of data depth:

A very intuitive notion of a partial order of one random variable Y *being more dispersed than* another random variable X, which we will denote with $\leq_{DISP}$, could be defined using upper levels sets/contour sets [19] of depth functions as follows: For a given proportion $\alpha \in [0,1]$ define $C_{\alpha,X}$ as that set A of the deepest data points of the underlying space w.r.t. an underlying depth function D and the underlying law X such that $P(X \in A) = \alpha$. Let furthermore denote $C_{\alpha,Y}$ denote the analogous set for law Y. Then, define Y as more dispersed than X if for all $\alpha \in [0,1]$ we have

$$C_{\alpha,X} \subseteq C_{\alpha,Y}.$$

(Here, one would have to treat different objects with the same attributes as identical.) For the simple case of (continuous) random variables in $\mathbb{R}^1$ the usually used depth functions collapse to depth functions that are isotone transformations of the function $D(z,X) := \min\{F_X(z), 1 - F_X(z)\}$, where F_X is the cdf of the underlying image law P_X. For that depth functions, it is easy to show that Y is more dispersed than X if and only if Y is a median preserving spread of X. In this sense, we really have a generalization of a median preserving spread. This generalization could be used not only in cases where we have a clear ordering, like in classical income inequality analysis. In fact, we have a methodology for the *inequality analysis for non-ordinal data* (where inequality is meant here in the sense of inequality/variance within a population/random variable). One interesting example would be social choice theory data, i.e., data that are (possibly partial) rankings, which are clearly of a non-ordinal scale of measurement in the sense that there is no clear a priori *below* and *above*. Interestingly, the abstract structural properties of depth functions discussed in Contribution 2 would have a big influence on which aspects of dispersion are captured by the above depth-based dispersion order: A quasiconcave depth function is unimodal and can therefore not distinguish between dispersion that is due to *polarization*[20] and dispersion that is due to *unimodal variability*. In contrast, a depth function that is not quasiconcave/unimodal could possibly distinguish more between *unimodal dispersion* and *polarization* in the sense of only capturing unimodal dispersion.

One technical aspect is that the relation $\leq_{DISP}$ may be very weak because a necessary condition for comparability is that the median of both Y and X is the same. One way out of this would be to look not at all $\alpha \in [0,1]$ within the definition of $\leq_{DISP}$, but instead to restrict α to be larger than a threshold c. Another way to get a stronger ordering relation could be established if one has some kind of translation operation in the sense of a translation group $\mathfrak{T}$. Then one could define Y be more dispersed than X if there exists a translation $T \in \mathfrak{T}$ such that

$$C_{\alpha,T\circ X} \subseteq C_{\alpha,Y}.$$

However, it seems to be seldom the case that in very abstract settings one is equipped with an additional translation group. To give one concrete example where we in fact have a natural translation group is the case of social choice theory. There, the data are (possibly partial) orders on a set of items $\{I_1, \ldots, I_k\}$. Then, a natural translation group would be $\mathfrak{T} := \{T_\sigma \mid \sigma \text{ permutation on } \{1, \ldots, k\}\}$ where T_σ is given by $T_\sigma(p) = \{(I_{\sigma(i)}, I_{\sigma(j)}) \mid (I_i, I_j) \in p\}\}$. This translation group would then lead to a dispersion order that does only care for the dispersion of the distribution of the posets per se, but not for 'the names of the items'.

[19] An upper level set or contour set with level λ is defined as the set of all objects that have dept λ or larger.

[20] Polarization means that data points are clustered into two or more 'opposite' subgroups. A clear cut rigorous disambiguation between polarization and dispersion in the context of social choice theory is, as far as the author is aware, not yet established. For a more elaborate disambiguation between polarization and dispersion/inequality in the context of poverty measurement, see, e.g., Esteban and Ray [1994], Duclos et al. [2004].

6 C: Analysis of Deficient Data

"*...a large part of...statistics is about what you would do if you had a model; and all of us spend enormous amounts of energy finding out what would happen if the data kept pouring in. I wish we could learn to look at the data more directly, without the fictional models and priors. On the same wish-list: we stop pretending to fix bad designs and inadequate measurements by modeling*" [Freedman, 1997, p.24]

6.1 Introduction

The aim of the contributions in this part is to add some methodology to the analysis of deficient data, but not in the spirit of statistical modeling, instead in a more direct way where one tries to stay as close to the data as possible instead of imposing further assumptions that would help in making a classical statistical analysis possible at all. As an illustrative example, take *Frischs true regression* (cf., Tamer [2010], Reiersöl [1945]): Assume two random variables X and Y that are in a linear relationship, i.e. $Y = \beta_0 + \beta_1 X + \varepsilon$, but that both cannot be observed without measurement error. Instead one only observes X^* and Y^* that are noisy versions of X and Y, respectively. Then the corresponding linear model is not identified and therefore cannot be consistently estimated. In fact every slope parameter between the one obtained by regressing Y^* on X^* and the inverse of the slope parameter obtained by regressing X^* on Y^* would be a reasonable candidate for an estimate of the slope. Instead of imposing additional assumptions that would identify the model, the methodology used here is in the spirit of partial identification and simply lives with an interval or a set of estimates. Of course, the methodology of partial identification usually still comes along with some assumptions, concretely in Contribution 7 the linearity assumption is of crucial importance. In this sense we are still very far away from Freedmans wish-list.

6.2 Contributions

Contribution 7

Georg Schollmeyer (2021): Computing simple bounds for regression estimates for linear regression with interval-valued covariates. In Jasper de Bock, Andrés Cano, Enrique Mirand, and Serafin Moral, editors, Proceedings of the Twelfth International Symposium on Imprecise Probabilities: Theories and Applications, Proceedings of Machine Learning Research, 147:273-279.

Original Abstract

In this paper, we deal with linear regression where the covariates are interval-valued and the dependent variable is precise. Opposed to the case where the dependent variable is interval-valued and the covariates are precise, it is far more difficult to compute the set of all ordinary least squares (OLS) estimates as the precise values of the covariates vary over all possible values, compatible with the given intervals of the covariates. Though the exact solution is difficult to obtain, there are still some simple possibilities to compute bounds for the regression parameters. In this paper we deal with simple linear regression and present three different approaches: The first one uses a simple interval-arithmetic consideration for the equation for the slope parameter. The second approach uses reverse regression to swap the roles of the dependent and the independent variable to make the computation analytically solvable. The obtained solution for the reverse regression then gives an analytical upper bound for the slope parameter of the original regression. The third approach does not directly give bounds for

the OLS estimator. Instead, before the actual interval analysis, in a first step, we modify the OLS estimator to another linear estimator which is simply a reasonably weighted convex combination of a number of unbiased estimators, which are themselves based on only two data points of the data set, respectively. It turns out that for the degenerate case of a precise independent variable, this estimator coincides with the OLS estimator. Additionally, the third method does also work if both the independent variable, as well as the dependent variable are interval-valued. Also the case of more than one covariate is manageable. A further nice point is that because of the analytical accessibility of the third estimator, also confidence intervals for the bounds can be established. To compare all three approaches, we conduct a short simulation study.

This contribution deals with the case of a (simple) linear regression of the form $Y = X\beta + \varepsilon$ in the case that Y is precisely observed, but X can only be observed in intervals $[\underline{X}, \overline{X}]$ and we only know $P(X \in [\underline{X}, \overline{X}]) = 1$. One way of analysis that is often done, especially in the field of imprecise probabilities or partial identification, is the computation of the so-called cautious data completion (cf., Augustin et al. [2014]). The cautious data completion is obtained by varying virtual possible precise data X over the whole range within the observed bounds $[\underline{X}, \overline{X}]$, computing all parameter-estimates that would be obtained by applying a classical estimator on this virtual precise data, and then collecting all resulting estimates in a set. For the case of linear regression, precisely observed covariates X and imprecisely observed outcomes Y, this was done for example in Schollmeyer and Augustin [2015], cf., also Černỳ and Rada [2011], Beresteanu et al. [2011]. There, the obtained set-valued estimator was called *collection region*. In these papers, as a classical estimator the ordinary least sqaures (OLS) estimator was used. Because for fixed X this estimator is linear in the outcome Y, it is relatively simple to compute the resulting set. The resulting set of estimates is then mathematically a zonotope, i.e., a Minkowski sum of line segments, see Černỳ and Rada [2011].

Opposed to the simple situation of a precise X, the case of an interval-valued covariate $[\underline{X}, \overline{X}]$, the situation is far more difficult. In particular there could be precise covariates compatible with the observed bounds, for which the corresponding design matrix is singular. Thus the definition of the cautious data completion is not even well-defined. Therefore, we go a different way, within this contribution. The idea for an interval-valued estimator of the slope parameter given in the contribution is very simple: For the precise case, given a sample $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$, in principle already a pair $(x_i, y_i); (x_j, y_j)$ with $i \neq j$ identifies the slope parameter and a simple (but not very effective) estimator for the slope parameter is given by $\hat{\beta} := \frac{y_j - y_i}{x_j - x_i}$. Given this, one can construct for all pairs (i, j) a corresponding slope estimate. Then one can aggregate all these slope estimates, for example by a weighted mean. This - given the usual assumptions - will lead to an unbiased estimator of the true slope. To make the estimator efficient, one can analyse its variance and then one can choose the weights within the weighted mean in such a way that the variance is minimal. In fact, by doing so one gets exactly the least squares estimator, see Theorem 1 of the contribution. This insight is not new, see Olkin and Yitzhaki [1992]. However, as far as the knowledge of the author goes, it was never used in connection with interval-valued covariates. For interval-valued covariates, one can analytically compute upper and lower bounds for the estimate based on the pair (i, j). Then, one can compute a weighted sum of the upper bounds and a weighted sum for the lower bounds (using the same weights) to get an estimator that asymptotically covers the true unknown slope parameter. Of course, it can happen that for certain pairs one obtains trivial bounds $[-\infty, +\infty]$ because the intervals of the pair of covariates overlap. In this case one has to exclude these pairs. If there are still pairs for which the covariates do not overlap, then one still gets an unbiased estimate for (a superset of) the slope. (This is due to the fact that one excludes pairs only based on the covariates, without 'seeing' the response and therefore the random error.) In a last step one has to choose the weights. In the precise case the weights are mainly based on the differences $|x_i - x_j|$. Since these differences are not precisely

known in the situation of interval-valued covariates, one cannot directly minimize the variance of the estimator. But at least one can take for example the mid-points of the intervals and then minimize the variance analogously to the precise case. Note that this will still lead to an unbiased set-valued estimator that converges to a set that contains the true parameter. The problem of minimizing the variance can be done by solving a quadratic programming problem, see Section 2.3 of the contribution.

Beyond the approach from above, in the contribution we also present two other approaches. One applies interval-arithmetic to the classical OLS-forula (see Section 2.1 of the contribution, cf., also Tretiak et al. [2023] for a similar approach). The other approach uses reverse regression, i.e., one treats Y as the independent variable and X as the dependent variable. Then, because only the dependent variable of the reverse regression is interval-valued, the cautious data completion for the reverse regression can be simply calculated as indicated above. With this, one implicitly gets an upper bound for the slope of the original regression from the lower bound of the slope of the reverse regression via the Cauchy-Schawrz inequality as

$$|\beta_{yx}| \leq \frac{1}{|\beta_{xy}|},$$

where β_{xy} is the lower bound for the slope of the reverse regression and β_{yx} denotes the slope of the original regression.

In Section 2.5 of the contribution, we also shortly discuss inference (e.g., computing confidence intervals). For this, one would need the dispersion σ^2 of the error term, that is usually not known. In the case of interval-valued covariates, it is generally very difficult to compute tight bounds for the dispersion, cf., e.g., Ferson et al. [2002]. However, for example Hladík and Černỳ [2017] supply conservative bounds which can be used for conservative inference.

Finally, a short simulation study concludes the contribution. The general impression from the results of this simulation study is that the approach that uses the weighted mean of the estimates based on pairs is usually better than the other two approaches. For a small dispersion of the error term the results for all three approaches are very similar.

Contribution 8

Georg Schollmeyer (2019): A short note on the equivalence of the ontic and the epistemic view on data imprecision for the case of stochastic dominance for interval-valued data. In Jasper de Bock, Cassio de Campos, Gert de Cooman, Erik Quaeghebeur, and Gregory Wheeler, editors, Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications, Proceedings of Machine Learning Research, 103:330–337.

Original Abstract

In the context of the analysis of interval-valued or set-valued data it is often emphasized that one has to carefully distinguish between an epistemic and an ontic understanding of set-valued data. However, there are cases, for which an ontic and an epistemic view do still lead to exactly the same results of the corresponding data analysis. The present paper is a short note on this fact in the context of the analysis of stochastic dominance for interval-valued data.

In this contribution we are concerned with first order stochastic dominance for univariate data under the additional aspect that these data cannot be observed directly, but only in intervals. A classical example of such a situation is the question about income within a social survey that

is interested in income inequality between some populations. It is very usual that respondents within social surveys refuse to tell their income and that the probability of refusing to answer is not independent from the true income. Therefore, some social surveys like, e.g., the German General Social Survey try to decrease non-response by first asking an open question about income. Then, if the respondent refuses to answer, a categorized question about the income (e.g., *Is your income between* 400 *and* 500 €) is asked additionally. The data obtained by such a procedure are highly non-standard in the sense that the scale of measurement is highly non-standard. On the one hand, looking at two observed intervals of income, say $[a, b]$ and $[c, d]$, if $b < c$ then one could say that the interval $[a, b]$ is clearly located below the interval $[c, d]$. Therefore, at least a partially ordered scale of measurement is given. On the other hand, we do also know something about the range of the quotient (or the difference) of the true incomes. Therefore, we have also a partial interval scale of measurement. These facts are seemingly also partly due to an *epistemic* understanding of interval data in the sense of Couso and Dubois [2014]: In the epistemic viewpoint, we assume that there is some true quantity, here concretely the income of a person that participated at a social survey. Stated in one stylized sentence, the epistemic understanding of data imprecision can be summarized as

Observing a set-valued data point that represents an imprecise observation of a precise, but not directly observable data point of interest.

Opposed to this, there are other cases of interval-data that are more appropriately understood in an ontic view in the sense of Couso and Dubois [2014] (compare also the discussion in Contribution 1), which could be summarized as

Observing a set-valued data point that is understood as a precise observation of something that is set-valued by nature, but that is 'imprecise' only in the sense that we do not observe $\mathbb{R}^d$-valued data, but set-valued data. The observed set is genuinely set-valued and there are no distinguished elements in the observed set and there is actually no real imprecision at all.

In this view, the localization of the scale of measurement within, e.g., the scale hierarchy of Stevens seems to be still more unclear[21]. To give an example: If we analyze the lifetimes of, say writers, then we can see such kind of data as interval-data. One natural aspect of the underlying empirical relational structure is the notion of one writer having clearly lived before another writer. For example Franz Kafka (1883 - 1924) clearly lived before Martin Walser (1927 - 2023), and this relational fact has, among other things, at least one clear *meaning*: Franz Kafka could have had an influence on Martin Walser and his writing[22]. But Walser in fact could not have any influence on Kafka. On the other hand, the (range of) quotients or differences between time points when Kafka and Walser, respectively, were alive, seems to be more difficult to interpret or to equip with *meaning*. In this sense, it may also be important from a measurement theoretic point of view to differentiate between different situations, which seems to be also loosely related to the dichotomy of epistemic vs. ontic data imprecision. These considerations put aside[23] (and additionally appreciating the need for differentiating between the epistemic and the ontic case in particular w.r.t. the obtained concrete results of an analysis), our contribution presents a case were in fact the epistemic and the ontic view lead to the same results. Therefore, if one is confronted with epistemic data imprecision, for analyzing first order stochastic dominance, the *misuse* of the ontic view - if one is only interested in obtaining the concrete mathematical results - may be of some interest.

Concretely, in the contribution, we look at univariate random variables X and Y that can only be

[21] Admittedly, one may see the scale hierarchy of Stevens more as a stylized hierarchy than as a thoroughly worked out rigorous and extensive conceptualization that is appropriate for every thinkable situation of data analysis.

[22] And most presumably, he did, Walser wrote a whole dissertation about Kafka, see Walser [1992].

[23] Note additionally that for first order stochastic dominance, which is analyzed in this contribution, only the partial ordinal scale of measurement is used and therefore, a possible additional partial cardinal scale of measurement would not add anything.

observed in intervals $[\underline{X}, \overline{X}]$ and $[\underline{Y}, \overline{Y}]$, respectively. We are interested if X is (weakly) stochastically smaller than Y, which means that

$$\mathbb{E}(u \circ X) \leq \mathbb{E}(u \circ Y)$$

for all increasing and bounded functions $u : \mathbb{R} \longrightarrow \mathbb{R}$ (cf., also Contribution 4). This can be characterized with the concept of upsets: Given a poset $(V, \leq)$ an upset is a set $A \subseteq V$ that fulfills

$$\forall x \in V \quad \forall a \in A : x \geq a \Longrightarrow x \in A.$$

Then, assuming that all upsets are measurable, first order stochastic dominance can be characterized as

$$X \text{ is stochastically smaller than } Y \iff P(X \in A) \leq P(Y \in A) \text{ for every upset } A \subseteq V.$$

In the case that $(V, \leq)$ are the reals equipped with the usual $\leq$-relation of the reals, it is easy to see that the upsets are the sets of the form $[c, \infty]$ or $(c, \infty]$. Therefore, stochastic dominance can be easily characterized by the cumulative distribution functions (cdf) as

$$X \text{ is stochastically smaller than } Y \iff \text{the cdf of } X \text{ lies always above the cdf of } Y.$$

Under an epistemic view, concerning the cautious data completion (cf., Augustin et al. [2014]) it is easy to see what the most extreme data completion (X^*, Y^*) is w.r.t. stochastic dominance: Simply take for X^* the upper bound $\overline{X}$ of the observed interval and for Y^* the observed lower bound $\underline{Y}$. If then $\overline{X}$ is stochastically smaller than $\underline{Y}$, we can conclude that also for every other data completion (X^*, Y^*) the (virtual) random variable X^* is stochastically smaller than the (virtual) random variable Y^*. If, on the other hand, $\overline{X}$ is not stochastically smaller than $\underline{Y}$, then there exists a cautious data completion (X^*, Y^*) (namely, $\overline{X}$ and $\underline{Y}$), for which X^* is not stochastically smaller than Y^*.

This means that the computational effort of computing the cautious data completion is in fact very low.

Concerning the ontic viewpoint, one would proceed as follows: As indicated before, say an observed interval $[a, b]$ is strictly below an interval $[c, d]$, in signs: $[a, b] < [c, d]$, if and only if $b < c$. Then take the set V of all intervals in the reals, together with this strict relation $<$, add the diagonal $\Delta_V := \{(I, I) \mid I \in V\}$ to get a partial order $\leq$. Then, with the above upset characterization of first order stochastic dominance for random variables with values in a partially ordered set, check for stochastic dominance. As showed in the contribution, both the epistemic and the ontic approach will lead to the same result. Of course, checking dominance within the ontic approach is computationally more difficult, because one does not have a simple characterization via the cumulative distribution function. However, for checking stochastic dominance within drawn samples of $[\underline{X}, \overline{X}]$ and $[\underline{Y}, \overline{Y}]$ one can efficiently solve this problem by solving a classical linear programming problem, see Schollmeyer et al. [2017]. Now, what could be an advantage of solving this more complicated problem formulation compared to simply solving the problem in the epistemic formulation? One possible main advantage of the ontic approach that was only shortly discussed in the contribution is the following:

The ontic approach would boil down to computing a supremum type characteristic

$$D = \sup_{A,\ A \text{ upset}} P(X \in A) - P(Y \in A),$$

where X and Y now denote the observed random intervals (equipped with the ordering described above). If this supremum characteristic is (smaller than or equal to) zero, X would be stochastically smaller than Y. Now, an important point in statistics is that usually we do not have direct

access to X and Y. Instead, we have only i.i.d. samples of X and Y and can only compute an empirical analogue D_n of D by replacing the respective laws of X and Y by the corresponding observed empirical laws. Now, the analysis of the statistical behaviour of D_n is of high importance. Actually, the problem of the analysis of this statistic D_n is already solved to a large extent, namely in terms of Vapnik-Chervonenkis (VC) theory (Vapnik and Chervonenkis [2015]). If the so-called VC dimension of the class of events (here the class of all upsets) over which the supremum is build has a finite VC dimension, then D_n converges almost surely to D. This would make D_n a good candidate for a statistical test of stochastic dominance. (The statistic D_n is generally not distribution-free, but a permutation test like proposed in Schollmeyer et al. [2017] could be used). If the VC dimension is infinite, then convergence cannot be guaranteed. If it is finite but very high, then one would expect that usually D_n is far away from D such that for a statistical test of stochastic dominance, a high VC dimension would make the statistical problem very ill-conditioned and thus a statistical test would become very insensitive. This suggests to regularize the test statistic, for example by reducing the family of all upsets to a subfamily that is not too complex/large. It turns out that for a partially ordered set $(V, \leq)$, the VC dimension of the family of all upsets is simply the width of the partially ordered set[24], and the width can be explicitly computed, see Schollmeyer et al. [2017]. This also allows for a direct control of the VC dimension, a concrete proposal can be found in Schollmeyer et al. [2017] (cf., also Schollmeyer [2023]).

Compared to the ontic view, within the epistemic view, it seems computational intractable to make for every data completion a permutation test.[25] Maybe still more interesting, particularly w.r.t. the cautious data completion is the fact that one seemingly cannot directly handle the problem of an ill-conditioned statistical setting. More seriously, one cannot even see that one possibly has a problem of ill-conditioning: Within the cautious data completion, for every concrete completion, one treats the precise data point as it would be the true data point. But actually it is only a virtually envisaged data point. In this sense one takes this virtual data point too seriously: For the case of univariate interval-data, for every data completion the VC dimension of the class of all upsets is one and therefore the problem for every data completion is very well-posed (cf., the Glivenko-Cantelli theorem and the Dvoretzky-Kiefer-Wolfowitz inequality). So no virtual data completion can see the ill-conditioning of the whole problem. In this sense, the cautious data completion walks on the crutch of dividing a problem, namely testing for stochastic dominance (of course conceptually rightly in an epistemic understanding), into two steps of firstly thinking of every possible precise data points that is compatible with the observed intervals and then treating this possible data point as if it was the true data point. Or, to put it into Vapniks words:

"When solving a problem of interest, do not solve a more general problem as an intermediate step. Try to get the answer that you really need but not a more general one."[Vapnik, 2006, p.477]

All these considerations show that a presumable misuse of the ontic approach within situations that are clearly of epistemic nature can be a really good idea.

6.3 Related Work & Future Research

For both contributions of part C, there are interesting possibilities to relate them to the methodology used in the other parts, in particular, the methodology of formal concept analysis and data depth. Here we want to exemplary indicate two such possibilities. The linear regression under interval-valued data could also be treated descriptively with methods of formal concept analysis: First, to start, think of a classical linear regression without epistemic data imprecision. If

[24]The width of a partially ordered set is the maximal cardinality of a set of elements that are pairwise incomparable.

[25]Maybe it is possible to analytically identify that data completion, for which e.g., the p-value would be the highest, but this would be a matter of future research. Note additionally, that for the univariate case, every data completion would be linearly ordered, so we are in the one-dimensional setting for every data completion. But for this setting the corresponding test statistic would be distribution free under the assumption of continuos cdf's. Maybe, this could help here, because this means that one does not need to rely on a permutation test.

a set of data points $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ would lie perfectly on a line with intercept β_0 and slope β_1, then one could describe this in terms of formal concept analysis: Take the vectors $(x_1, y_1), \ldots, (x_n, y_n)$ as the objects of a formal context. As the attributes take statements of the form "$x = c$" and "$y = d$" with $c, d \in \mathbb{R}$ arbitrary. Then, the fact that the points lie perfectly on a line can be expressed by saying that all formal attribute implications of the form

$$\text{“}x = c\text{”} \quad \longrightarrow \quad \text{“}y = \beta_0 + \beta_1 c\text{”}$$

are valid in the context. If the points are not perfectly on a line then some of the implications are not valid in the context. In this case one can try to measure how *strongly* the implications are not valid, for example by looking at the data point(s) that violate(s) a formal implication and by taking the squared difference between y and $\beta_0 + \beta_1 x$, where x and y are the values of the corresponding data point(s) that violate(s) the implication. Then, for every implication that is violated one has a measure of the strength of violation. One can now aggregate all these measures of strength, for example by taking the average. If one then takes the regression line that minimizes this aggregate of strengths of violations, one would end up with the classical ordinary least squares (OLS) solution, but presented in the language of formal concept analysis. Within the FCA presentation, the choice of the squared distances and the use of the average for the aggregation appear very arbitrary, compared to the classical representation under the standard OLS assumptions. In fact, within the viewpoint of classical statistics, with this representation we will not improve anything under the standard assumptions and the standard criteria for a good method (remember the Gauß-Markov theorem). However, when it comes to interval-valued data, then the FCA-representation has something to add: If we have intervals $[\underline{x}, \overline{x}]$ instead of precise x, then we can modify our conceptual scaling of the context to an interordinal scaling and take "$x \leq c$","$y \leq d$" as well as "$x \geq e$","$y \geq f$" with $c, d, e, f \in \mathbb{R}$ arbitrary as attributes (with x and y being now intervals and $[a, b] \leq c \iff b \leq c$ etc.). Then one can again measure the strength of violation of a formal implication. For example (assuming $\beta_1 > 0$) for a violation of an implication

$$\text{“}x \leq c\text{”} \quad \longrightarrow \quad \text{“}y \leq \beta_0 + \beta_1 c\text{”}$$

one can take a certain distance between the intervals y and $\beta_0 + \beta_1 x$, for example, the Hausdorff distance, where x and y are now the intervals of that data point(s) that violate(s) the implication. (Note that compared to the precise case, here are still more possibilities, which all look a little bit arbitrary in the first place. Analyzing this aspect would of course be an interesting point of future research.) Now, after aggregating (where the same arbitrariness aspect from above applies) one can again find that regression lines(s) that minimize this aggregate. Compared to the cautious data completion, where one usually gets a set with more than one regression line, here one would expect that often enough there is only one regression line that minimizes the aggregate. However, in a descriptive sense this precise regression line somehow describes the data set. Additionally, one can use the modified conceptual scaling also in the case of precise data. This seems to be an interesting point to investigate. In particular, because one does not only look at the data points in isolation, but also on whole intervals of data points together, the topological structure in x (i.e., how the different residuals are arranged and not only how they are distributed) seems to play a role. Of course, again, because under the classical assumptions of linear regression, informally speaking, the distribution of the residuals is a sufficient statistic, and the arrangement of the residuals w.r.t. x does not carry further information, this aspect will presumably be only relevant for clear enough deviations from the standard assumptions.

Concerning the epistemic-ontic dichotomy of Contribution 8 it is interesting to look at the ontic understanding of partial order data hold in Contribution 1. Directly looking at the example of students preferences between universities given in Dittrich et al. [1998] one makes the following observation: Looking at the concretely collected data it appears that during the data collection process, some interviewers simply forgot to ask some of the pair comparisons for some students. This shows that, of course, also non-standard data with an ontic understanding can additionally suffer from epistemic data imprecision. Therefore, also for such data it is of interest, to analyze,

how one can handle e.g., missing or coarsened data. In the context of partial order data and the generalized Tukeys depth, we have first results concerning the computation of the cautious data completion, i.e., the set of all data depth functions that would be obtained by replacing all partly missing data with all possible precise data (under an ontic understanding) that are compatible with everything that is observed (cf., Schollmeyer et al. [2023]). Additionally, we also already have a concrete proposal for an analysis under the coarsening at random assumption (CAR, see Heitjan and Rubin [1991]) where the coarsening process can be ignored in a certain sense. In the concrete example of the university rankings this assumptions seem to be not too unreasonable, because in this setting this assumption would mean that the event that a pair comparison was forgotten to ask does not depend on the concrete answer, a student would have given.

7 Versicherung an Eides statt

Hiermit versichere ich an Eides Statt durch meine Unterschrift, dass ich bei der Anfertigung der vorliegenden Habilitationsleistung keine weiteren als die hier angegebenen Hilfsmittel benutzt habe, und dass kein wissenschaftliches Fehlverhalten im Sinne der Richtlinien der Ludwigs-Maximilians-Universität München zur Selbstkontrolle in der Wissenschaft in der Fassung vom 16. Mai 2002 (geändert durch Beschlüsse des Senats vom 22.6.2006, 11.2.2010, 30.9.2014) vorliegt.

Georg Schollmeyer

References

R. Allison and J. E. Foster. Measuring health inequality using qualitative data. *Journal of Health Economics*, 23(3):505–524, 2004.

M. A. Arcones and E. Gine. Limit theorems for u-processes. *The Annals of Probability*, 21(3): 1494–1542, 1993.

M. Atzmueller. Subgroup discovery. *WIREs Data Mining and Knowledge Discovery*, 5(1):35–49, 2015.

T. Augustin, F. P. Coolen, G. De Cooman, and M. C. Troffaes, editors. *Introduction to imprecise probabilities*. John Wiley & Sons, 2014.

V. Barnett. The ordering of multivariate data (with discussion). *Journal of the Royal Statistics Society, Series A*, 139(318-354):14, 1976.

A. Beresteanu, I. Molchanov, and F. Molinari. Sharp identification regions in models with convex moment predictions. *Econometrica*, 79(6):1785–1821, 2011.

C. Biernacki and J. Jacques. A generative model for rank data based on insertion sort algorithm. *Computational Statistics & Data Analysis*, 58:162–176, 2013. The Third Special Issue on Statistical Signal Extraction and Filtering.

G. Birkhoff. Lattices and their applications. *Bulletin of the American Mathematical Society*, 44: 793–800, 1938.

H. Blocher. A supremum type test statistic for distributional tests: An analysis in the context of VC theory, partially spatial data and regularization. *Master thesis, unpublished, LMU Munich*, 2020.

H. Blocher and G. Schollmeyer. Union-free generic depth for non-standard data. *In preparation.*

H. Blocher, G. Schollmeyer, and C. Jansen. Statistical tests and regression analysis for partial order data based on data depth and formal concept analysis. *In preparation.*

H. Blocher, G. Schollmeyer, and C. Jansen. Statistical models for partial orders based on data depth and formal concept analysis. In D. Ciucci, I. Couso, J. Medina, D. Slezak, D. Petturiti, B. Bouchon-Meunier, and R. Yager, editors, *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 17–30. Springer, 2022.

M. Boley and H. Grosskreutz. Non-redundant subgroup discovery using a closure system. In W. Buntine, M. Grobelnik, D. Mladenić, and J. Shawe-Taylor, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 179–194, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

A. Bommier, A. Chassagnon, and F. Le Grand. Comparative risk aversion: A formal approach with applications to saving behavior. *Journal of Economic Theory*, 147(4):1614–1641, 2012. Inequality and Risk.

G. Brinkmann and B. D. McKay. Posets on up to 16 points. *Order*, 19:147–179, 2002.

B. M. Brown and T. P. Hettmansperger. An affine invariant bivariate version of the sign test. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(1):117–125, 12 2018.

M. Černỳ and M. Rada. On the possibilistic approach to linear regression with rounded or interval-censored data. *Measurement Science Review*, 11(2), 2011.

A. Chakraborty and P. Chaudhuri. On data depth in infinite dimensional spaces. *Annals of the Institute of Statistical Mathematics*, 66:303–324, 2014.

L. Chang. Partial order relations for classification comparisons. *Canadian Journal of Statistics*, 48(2):152–166, 2020.

A. Chateauneuf and M. Cohen. Risk seeking with diminishing marginal utility in a non-expected utility model. *Journal of Risk and Uncertainty*, 9(1):77–91, 1994.

A. Chateauneuf, M. Cohen, and I. Meilijson. New tools to better model behavior under risk and uncertainty: An overview. *Finance*, 18(1):25–46, 1997.

A. Christmann. Classification based on the svm and on regression depth, statistical data analysis based on the L1-norm and related methods. *Statistics for industry and technology*, 2002.

F. P. A. Coolen and T. Augustin. A nonparametric predictive alternative to the imprecise Dirichlet model: the case of a known number of categories. *International Journal of Approximate Reasoning*, 50(2):217–230, 2009.

I. Couso and D. Dubois. Statistical reasoning with set-valued information: Ontic vs. epistemic views. *International Journal of Approximate Reasoning*, 55(7):1502–1518, 2014.

X. Dai and S. Lopez-Pintado. Tukey's depth for object data. *Journal of the American Statistical Association*, 0(0):1–13, 2022.

C. L. de la Vega. Social inequality: Theoretical approaches. *Handbook of Research on Economic and Social Well-Being*, pages 377–399, 2018.

C. De Maio, G. Fenza, M. Gaeta, V. Loia, F. Orciuoli, and S. Senatore. Rss-based e-learning recommendations exploiting fuzzy fca for knowledge modeling. *Applied Soft Computing*, 12(1): 113–124, 2012.

J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

P. Diaconis. Group representations in probability and statistics. *Lecture Notes-Monograph Series*, 11:i–192, 1988.

R. Dittrich, R. Hatzinger, and W. Katzenbeisser. Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(4):511–525, 1998.

D. Donoho and M. Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, 20(4):1803 – 1827, 1992.

D. Donoho and P. Huber. The notion of breakdown point. *A Festschrift for Erich L. Lehmann*, 157184, 1983.

D. Dubois and H. Prade. *Possibilty Theory*. Plenum Press, New York, 1988.

J.-Y. Duclos, J. Esteban, and D. Ray. Polarization: Concepts, measurement, estimation. *Econometrica*, 72(6):1737–1772, 2004.

P. Edelman. Meet-distributive lattices and the anti-exchange closure. *Algebra Universalis*, 10: 290–299, 1980.

P. Elzinga, K. Wolff, and J. Poelmans. Analyzing chat conversations of pedophiles with temporal relational semantic systems. In *2012 European Intelligence and Security Informatics Conference*, pages 242–249, 2012.

J.-M. Esteban and D. Ray. On the measurement of polarization. *Econometrica*, 62(4):819–851, 1994.

I. Falk and C. Gardent. Combining formal concept analysis and translation to assign frames and semantic role sets to french verbs. *Annals of Mathematics and Artificial Intelligence*, 70(1-2): 123–150, 2014.

T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. ROC Analysis in Pattern Recognition.

S. Ferson, L. Ginzburg, V. Kreinovich, L. Longpré, and M. Aviles. Computing variance for interval data is np-hard. *ACM SIGACT News*, 33(2):108–118, 2002.

P. Fierens. An extension of chaotic probability models to real-valued variables. *International Journal of Approximate Reasoning*, 50(4):627–641, 2009.

R. Fraiman, R. Liu, and J. Meloche. Multivariate density estimation by probing depth. *Lecture Notes-Monograph Series*, 31:415–430, 1997.

D. Freedman. *Some Issues in the Foundation of Statistics*, pages 19–39. Springer Netherlands, Dordrecht, 1997.

T. Gäbel-Hökenschnieder, T. Pfeiffer, M. Al Salamat, and S. E. Schmidt. Generalized metrics and their relevance for FCA and closure operators. In *CLA*, pages 175–187, 2016.

B. Ganter. Random extents and random closure systems. In *CLA*, volume 959, pages 309–318. Citeseer, 2011.

B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer, 2012.

G. Geenens, A. Nieto-Reyes, and G. Francisci. Statistical depth in abstract metric spaces. *Statistics and Computing*, 33(2):46, 2023.

M. Goibert, S. Clémençon, E. Irurozki, and P. Mozharovskyi. Statistical depth functions for ranking distributions: Definitions, statistical learning and applications. *arXiv preprint arXiv:2201.08105*, 2022.

D. F. Heitjan and D. B. Rubin. Ignorability and coarse data. *The Annals of Statistics*, 19(4): 2244–2253, 1991.

T. C. Hesterberg. What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, 69(4):371–386, 2015.

T. P. Hettmansperger and H. Oja. Affine invariant multivariate multisample sign tests. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1):235–249, 12 2018.

M. Hladík and M. Černỳ. Two optimization problems in linear regression with interval data. *Optimization*, 66(3):331–349, 2017.

D. Insua and F. Ruggeri, editors. *Robust Bayesian Analysis*. Springer, 2012.

C. Jansen, G. Schollmeyer, and T. Augustin. A probabilistic evaluation framework for preference aggregation reflecting group homogeneity. *Mathematical Social Sciences*, 96:49–62, 2018a.

C. Jansen, G. Schollmeyer, and T. Augustin. Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences. *International Journal of Approximate Reasoning*, 98:112–131, 2018b.

C. Jansen, H. Blocher, T. Augustin, and G. Schollmeyer. Information efficient learning of complexly structured preferences: Elicitation procedures and their application to decision making under uncertainty. *International Journal of Approximate Reasoning*, 144:69–91, 2022.

C. Jansen, M. Nalenz, G. Schollmeyer, and T. Augustin. Statistical comparisons of classifiers by generalized stochastic dominance. *Journal of Machine Learning Research*, 24(231):1–37, 2023a.

C. Jansen, G. Schollmeyer, H. Blocher, J. Rodemann, and T. Augustin. Robust statistical comparison of random variables with locally varying scale of measurement. In R. J. Evans and I. Shpitser, editors, *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pages 941–952. PMLR, 31 Jul–04 Aug 2023b.

R. Jörnsten. Clustering and classification based on the L1 data depth. *Journal of Multivariate Analysis*, 90(1):67–89, 2004. Special Issue on Multivariate Methods in Genomic Data Analysis.

B. J. Keller, F. Eichinger, and M. Kretzler. Formal concept analysis of disease similarity. *AMIA Summits on Translational Science Proceedings*, 2012:42, 2012.

M. Kendall. Discussion on Symposium on Ranking Methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 12(2):182–191, 12 1950.

F. Kreuter, editor. *Improving surveys with paradata: analytic uses of process information.* Wiley, 2013.

G. Lebanon and Y. Mao. Non-parametric modeling of partially ranked data. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007.

M. D. Lee, M. Steyvers, and B. Miller. A cognitive model for aggregating people's rankings. *PloS one*, 9(5):e96431, 2014.

J. Li and R. Y. Liu. New nonparametric tests of multivariate locations and scales using data depth. *Statistical Science*, 19(4):686 – 696, 2004.

R. Liu. On a notion of data depth based on random simplices. *The Annals of Statistics*, 18: 405–414, 1990.

R. Liu, J. Parelius, and K. Singh. Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by Liu and Singh). *The Annals of Statistics*, 27(3):783 – 858, 1999.

R. D. Luce. Semiorders and a theory of utility discrimination. *Econometrica*, 24(2):178–191, 1956.

C. L. Mallows. Non-null ranking models. i. *Biometrika*, 44(1/2):114–130, 1957.

H. Mendelson. Quantile-preserving spread. *Journal of Economic Theory*, 42(2):334–351, 1987.

N. Messai, J. Bouaud, M.-A. Aufaure, L. Zelek, and B. Séroussi. Using formal concept analysis to discover patterns of non-compliance with clinical practice guidelines: A case study in the management of breast cancer. In M. Peleg, N. Lavrač, and C. Combi, editors, *Artificial Intelligence in Medicine*, pages 119–128, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

K. Mosler. *Depth statistics*, pages 17–34. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

K. Mosler and Y. Polyakova. General notions of depth for functional data. *arXiv preprint arXiv:1208.1981*, 2012.

K. Nakamura, K. Yano, and F. Komaki. Learning partially ranked data based on graph regularization. *arXiv preprint arXiv:1902.10963*, 2019.

I. Olkin and S. Yitzhaki. Gini regression analysis. *International Statistical Review / Revue Internationale de Statistique*, 60(2):185–196, 1992.

C. Peirce. How to make our ideas clear. In *Chance, love, and logic*, pages 32–60. Routledge, 2017.

J. Poelmans, D. I. Ignatov, S. Viaene, G. Dedene, and S. O. Kuznetsov. Text mining scientific papers: A survey on fca-based information retrieval research. In P. Perner, editor, *Advances in Data Mining. Applications and Theoretical Aspects*, pages 273–287, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

J. Poelmans, D. I. Ignatov, S. O. Kuznetsov, and G. Dedene. Formal concept analysis in knowledge processing: A survey on applications. *Expert Systems with Applications*, 40(16):6538–6560, 2013.

J. Quiggin. A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4): 323–343, 1982.

O. Reiersöl. *Confluence analysis by means of instrumental sets of variables*. PhD thesis, Stockholm College, 1945.

M. Rothschild and J. E. Stiglitz. Increasing risk: I. A definition. *Journal of Economic Theory*, 2 (3):225–243, 1970.

P. J. Rousseeuw and M. Hubert. Regression depth. *Journal of the American Statistical Association*, 94(446):388–402, 1999.

I. Ruts and P. J. Rousseeuw. Computing depth contours of bivariate point clouds. *Computational Statistics & Data Analysis*, 23(1):153–168, 1996.

G. Schollmeyer. Lower quantiles for complete lattices. Technical Report, LMU, 2017a. URL `http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-40448-7`.

G. Schollmeyer. Application of lower quantiles for complete lattices to ranking data: Analyzing outlyingness of preference orderings. Technical Report, LMU, 2017b. URL `http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-40452-9`.

G. Schollmeyer. On the uniform control of the Vapnik-Chervonenkis dimension in subgroup discovery using formal concept analysis. *Preliminary draft*, 2023. URL `https://georgschollmeyer.weebly.com/uploads/1/3/8/4/138402857/uniform_control.pdf`.

G. Schollmeyer and T. Augustin. Statistical modeling under partial identification: Distinguishing three types of identification regions in regression analysis with interval data. *International Journal of Approximate Reasoning*, 56:224–248, 2015.

G. Schollmeyer and H. Blocher. A note on the connectedness property of union-free generic sets of partial orders. *arXiv preprint arXiv:2304.10549*, 2023.

G. Schollmeyer, C. Jansen, and T. Augustin. Detecting stochastic dominance for poset-valued random variables as an example of linear programming on closure systems, 2017. URL `http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-40416-0`.

G. Schollmeyer, H. Blocher, C. Jansen, and T. Augustin. On the analysis of epiontic data: A case study. In E. Miranda, I. Montes, E. Quaeghebeur, and B. Vantaggi, editors, *Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications*, volume 215 of *Proceedings of Machine Learning Research*. PMLR, 11–14 Jul 2023. Poster abstract.

R. Serfling. Nonparametric multivariate descriptive measures based on spatial quantiles. *Journal of Statistical Planning and Inference*, 123(2):259–278, 2004.

R. Serfling. Depth functions in nonparametric multivariate inference. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 72:1, 2006.

R. Serfling and Y. Zuo. General notions of statistical depth function. *The Annals of Statistics*, 28 (2):461 – 482, 2000.

G. Shafer. *A Mathematical theory of evidence*. Princeton University Press, 1976.

P. Suppes. *Measurement, empirical meaningfulness, and three-valued logic*, pages 65–80. Springer Netherlands, Dordrecht, 1969.

J. Surowiecki. The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. *Economies, Societies and Nations*, 296(5), 2004.

E. Tamer. Partial identification in econometrics. *Annual Review of Economics*, 2(1):167–195, 2010.

K. Tretiak, G. Schollmeyer, and S. Ferson. Neural network model for imprecise regression with interval dependent variables. *Neural Networks*, 161:550–564, 2023.

M. Troffaes and G. De Cooman. *Lower Previsions*. Wiley, 2014.

J. Tukey. Mathematics and the picturing of data. In R. James, editor, *Proceedings of the International Congress of Mathematicians Vancouver*, pages 523–531, Vancouver, 1975. Mathematics-Congresses.

V. Vapnik. *Estimation of dependences based on empirical data*. Springer, 2006. Afterword of 2006.

V. Vapnik and A. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. In V. Vovk, H. Papadopoulos, and A. Gammerman, editors, *Measures of Complexity: Festschrift for Alexey Chervonenkis*, pages 11–30. Springer, 2015.

H. von Hentig. *Magier Oder Magister?: Über Die Einheit der Wissenschaft im Verständigungsprozess*. E. Klett, 1974.

P. Walley and T. L. Fine. Towards a frequentist theory of upper and lower probability. *The Annals of Statistics*, 10(3):741 – 761, 1982.

M. Walser. *Beschreibung einer Form: Versuch über Franz Kafka*. Suhrkamp Taschenbuch, 1992.

J. Wang and R. Serfling. Nonparametric multivariate kurtosis and tailweight measures. *Journal of Nonparametric Statistics*, 17(4):441–456, 2005.

K. Weichselberger. The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2-3):149–170, 2000.

K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung. I: Intervallwahrscheinlichkeit als umfassendes Konzept*. Physica, 2001.

K. Weichselberger. The logical concept of probability: foundation and interpretation. In G. de Cooman, J. Vejnarová, and M. Zaffalon, editors, *ISIPTA '07: Proceedings of the 5th International Symposium on Imprecise Probabilities and Their Applications*, pages 455–464. SIPTA, 2007.

M. Wermelinger, Y. Yu, and M. Strohmaier. Using formal concept analysis to construct and visualise hierarchies of socio-technical relations. In *2009 31st International Conference on Software Engineering - Companion Volume*, pages 327–330, 2009.

R. Wille. Restructuring lattice theory: An approach based on hierarchies of concepts. In I. Rival, editor, *Ordered Sets*, pages 445–470, Dordrecht, 1982. Springer Netherlands.

M. Yaari. The dual theory of choice under risk. *Econometrica*, 55(1):95–115, 1987.

J. Zhang. Some extensions of Tukey's depth function. *Journal of Multivariate Analysis*, 82(1): 134–165, 2002.