Reliable statistical modeling of weakly structured information: Contributions to partial identification, stochastic partial ordering and imprecise probabilities

Georg Schollmeyer

München 2017

Reliable statistical modeling of weakly structured information: Contributions to partial identification, stochastic partial ordering and imprecise probabilities

Georg Schollmeyer

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Georg Schollmeyer
aus Pirna

München, den 31.08.2017

Erstgutachter: Prof. Dr. Thomas Augustin
Zweitgutachter: Prof. Frank P. A. Coolen
Drittgutachter: Prof. Lev V. Utkin
eingereicht am 31.08.2017
Tag der Disputation: 06.11.2017

Contents

Summary

The statistical analysis of "real-world" data is often confronted with the fact that most standard statistical methods were developed under some kind of idealization of the data that is often not adequate in practical situations. This concerns among others i) the potentially deficient quality of the data that can arise for example due to measurement error, non-response in surveys or data processing errors and ii) the scale quality of the data, that is idealized as "the data have some clear scale of measurement that can be uniquely located within the scale hierarchy of Stevens (or that of Narens and Luce or Orth)".

Modern statistical methods like, e.g., correction techniques for measurement error or robust methods cope with issue i). In the context of missing or coarsened data, imputation techniques and methods that explicitly model the missing/coarsening process are nowadays well-established tools of refined data analysis. Concerning ii) the typical statistical viewpoint is a more pragmatical one, in case of doubt one simply presumes the strongest scale of measurement that is clearly "justified". In more complex situations, like for example in the context of the analysis of ranking data, statisticians often simply do not worry about purely measurement theoretic reservations too much, but instead embed the data structure in an appropriate, easy to handle space, like e.g. a metric space and then use all statistical tools available for this space.

Against this background, the present cumulative dissertation tries to contribute from different perspectives to the appropriate handling of data that challenge the above-mentioned idealizations. A focus here is on the one hand on analysis of interval-valued and set-valued data within the methodology of partial identification, and on the other hand on the analysis of data with values in a partially ordered set (poset-valued data). Further tools of statistical modeling treated in the dissertation are necessity measures in the context of possibility theory and concepts of stochastic dominance for poset-valued data.

The present dissertation consists of 8 contributions, which will be detailedly discussed in the following sections:

Contribution 1 analyzes different identification regions for partially identified linear models under interval-valued responses and develops a further kind of identification region (as well as a corresponding estimator). Estimates for the identification regions are compared to each other and also to classical statistical approaches for a data set on wine quality.
Contribution 2 deals with logistic regression under coarsened responses, analyzes point-identifying assumptions and develops likelihood-based estimators for the identified set. The methods are illustrated with data of a wave of the panel study "Labor Market and Social Security" (PASS).
Contribution 3 analyzes the combinatorial structure of the extreme points and the edges of a polytope (called credal set or core in the literature) that plays a crucial role in imprecise probability theory. Furthermore, an efficient algorithm for enumerating all extreme points is given and compared to existing standard methods.
Contribution 4 develops a quantile concept for data or random variables with values in a complete lattice, which is applied in Contribution 5 to the case of ranking data in the context of a data set on the wisdom of the crowd phenomena.
In Contribution 6 a framework for evaluating the quality of different aggregation functions of Social Choice Theory is developed, which enables analysis of quality in dependence of group

specific homogeneity. In a simulation study, selected aggregation functions, including an aggregation function based on the concepts of Contribution 4 and Contribution 5, are analyzed. Contribution 7 supplies a linear program that allows for detecting stochastic dominance for poset-valued random variables, gives proposals for inference and regularization, and generalizes the approach to the general task of optimizing a linear function on a closure system. The generality of the developed methods is illustrated with data examples in the context of multivariate inequality analysis, item impact and differential item functioning in the context of item response theory, analyzing distributional differences in spatial statistics and guided regularization in the context of cognitive diagnosis models.
Contribution 8 uses concepts of stochastic dominance to establish a descriptive approach for a relational analysis of person ability and item difficulty in the context of multidimensional item response theory. All developed methods have been implemented in the language[1] R ([R Development Core Team, 2014]) and are available from the author upon request.

The application examples corroborate the usefulness of weak types of statistical modeling examined in this thesis, which, beyond their flexibility to deal with many kinds of data deficiency, can still lead to informative substance matter conclusions that are then more reliable due to the weak modeling.

[1]Some computational expensive parts have been outsourced to Fortran 95, for solving the linear and mixed integer programs, special software (Gurobi: Gu, Rothberg, and Bixby [2012]), for which an R-interface exists, was used.

Zusammenfassung

Die statistische Analyse real erhobener Daten sieht sich oft mit der Tatsache konfrontiert, dass übliche statistische Standardmethoden unter einer starken Idealisierung der Datensituation entwickelt wurden, die in der Praxis jedoch oft nicht angemessen ist. Dies betrifft i) die möglicherweise defizitäre Qualität der Daten, die beispielsweise durch Vorhandensein von Messfehlern, durch systematischen Antwortausfall im Kontext sozialwissenschaftlicher Erhebungen oder auch durch Fehler während der Datenverarbeitung bedingt ist und ii) die Skalenqualität der Daten an sich: Viele Datensituationen lassen sich nicht in die einfachen Skalenhierarchien von Stevens (oder die von Narens und Luce oder Orth) einordnen.

Modernere statistische Verfahren wie beispielsweise Messfehlerkorrekturverfahren oder robuste Methoden versuchen, der Idealisierung der Datenqualität im Nachhinein Rechnung zu tragen. Im Zusammenhang mit fehlenden bzw. intervallzensierten Daten haben sich Imputationsverfahren zur Vervollständigung fehlender Werte bzw. Verfahren, die den Entstehungprozess der vergröberten Daten explizit modellieren, durchgesetzt. In Bezug auf die Skalenqualität geht die Statistik meist eher pragmatisch vor, im Zweifelsfall wird das niedrigste Skalenniveau gewählt, das klar gerechtfertigt ist. In komplexeren multivariaten Situationen, wie beispielsweise der Analyse von Ranking-Daten, die kaum noch in das Stevensche "Korsett" gezwungen werden können, bedient man sich oft der einfachen Idee der Einbettung der Daten in einen geeigneten metrischen Raum, um dann anschließend alle Werkzeuge metrischer Modellierung nutzen zu können.

Vor diesem Hintergrund hat die hier vorgelegte kumulative Dissertation deshalb zum Ziel, aus verschiedenen Blickwinkeln Beiträge zum adäquaten Umgang mit Daten, die jene Idealisierungen herausfordern, zu leisten. Dabei steht hier vor allem die Analyse intervallwertiger bzw. mengenwertiger Daten mittels partieller Identifikation auf der Seite defizitärer Datenqualität im Vordergrund, während bezüglich Skalenqualität der Fall von verbandswertigen Daten behandelt wird. Als weitere Werkzeuge statistischer Modellierung werden hier insbesondere Necessity-Maße im Rahmen der Imprecise Probabilities und Konzepte stochastischer Dominanz für Zufallsvariablen mit Werten in einer partiell geordneten Menge betrachtet.

Die vorliegende Dissertation umfasst 8 Beiträge, die in den folgenden Kapiteln näher diskutiert werden:

Beitrag 1 analysiert verschiedene Identifikationsregionen für partiell identifizierte lineare Modelle unter intervallwertig beobachteter Responsevariable und schlägt eine neue Identifikationsregion (inklusive Schätzer) vor. Für einen Datensatz, der die Qualität von verschiedenen Rotweinen, gegeben durch ExpertInnenurteile, in Abhängigkeit von verschiedenen physikochemischen Eigenschaften beschreibt, werden Schätzungen für die Identifikationsregionen analysiert. Die Ergebnisse werden ebenfalls mit den Ergebissen klassischer Methoden für Intervalldaten verglichen.
Beitrag 2 behandelt logistische Regression unter vergröberter Responsevariable, analysiert punktidentifizierende Annahmen und entwickelt likelihoodbasierte Schätzer für die entsprechenden Identifikationsregionen. Die Methode wird mit Daten einer Welle der Panelstudie „Arbeitsmarkt und Soziale Sicherung“ (PASS) illustriert.
Beitrag 3 analysiert die kombinatorische Struktur der Extrempunkte und der Kanten eines Polytops (sogenannte Struktur bzw. Kern einer Intervallwahrscheinlichkeit bzw. einer nicht-

additiven Mengenfunktion), das von wesentlicher Bedeutung in vielen Gebieten der Imprecise Probability Theory ist. Ein effizienter Algorithmus zur Enumeration aller Extrempunkte wird ebenfalls gegeben und mit existierenden Standardenumerationsmethoden verglichen.
In Beitrag 4 wird ein Quantilkonzept für verbandswertige Daten bzw. Zufallsvariablen vorgestellt. Dieses Quantilkonzept wird in Beitrag 5 auf Ranking-Daten im Zusammenhang mit einem Datensatz, der das „Weisheit der Vielen"-Phänomen untersucht, angewendet.
Beitrag 6 entwickelt eine Methode zur probabilistischen Analyse der „Qualität" verschiedener Aggregationsfunktionen der Social Choice Theory. Die Analyse wird hier in Abhängigkeit der Homogenität der betrachteten Gruppen durchgeführt. In einer simulationsbasierten Studie werden exemplarisch verschiedene klassische Aggregationsfunktionen, sowie eine neue Aggregationsfunktion basierend auf den Beiträgen 4 und 5, verglichen.
Beitrag 7 stellt einen Ansatz vor, um das Vorliegen stochastischer Dominanz zwischen zwei Zufallsvariablen zu überprüfen. Der Anstaz nutzt Techniken linearer Programmierung. Weiterhin werden Vorschläge für statistische Inferenz und Regularisierung gemacht. Die Methode wird anschließend auch auf den allgemeineren Fall des Optimierens einer linearen Funktion auf einem Hüllensystem ausgeweitet. Die flexible Anwendbarkeit wird durch verschiedene Anwendungsbeispiele illustriert.
Beitrag 8 nutzt Ideen stochastischer Dominanz, um Datensätze der multidimensionalen Item Response Theory relational zu analysieren, indem Paare von sich gegenseitig empirisch stützenden Fähigkeitsrelationen der Personen und Schwierigkeitsrelationen der Aufgaben entwickelt werden.

Alle entwickelten Methoden wurden in R ([R Development Core Team, 2014]) implementiert[2]. Die Anwendungsbeispiele zeigen die Flexibilität der hier betrachteten Methoden relationaler bzw. „schwacher" Modellierung insbesondere zur Behandlung defizitärer Daten und unterstreichen die Tatsache, dass auch mit Methoden schwacher Modellierung oft immer noch nichttriviale substanzwissenschaftliche Rückschlüsse möglich sind, die aufgrund der inhaltlich vorsichtigeren Modellierung dann auch sehr viel stärker belastbar sind.

[2] Computational aufwendige Teile wurden teilweise in Fortran 95 ausgelagert, für die Berechnung der linearen und der gemischt ganzzahligen Programme wurde spezielle Software (Gurobi: Gu, Rothberg, and Bixby [2012]), für die auch eine R-Schnittstelle existiert, genutzt.

Contributions of the thesis

The 8 contributions of the thesis constitute 4 parts:

A Challenges for statistical modeling under partial identification in the context of regression analysis

1) G. Schollmeyer, 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.
2) J. Plass, T. Augustin, M. Cattaneo, G. Schollmeyer, Statistical modelling under epistemic data imprecision: Some results on estimating multinomial distributions and logistic regression for coarse categorical data. In: T. Augustin, S. Doria, E. Miranda, E. Quaeghebeur, editors, *ISIPTA '15, Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*, pages 247-256. Aracne, 2015.

B Computational challenges in the theory of imprecise probabilities

3) G. Schollmeyer. On the number and characterization of the extreme points of the core of necessity measures on finite spaces. In: T. Augustin, S. Doria, E. Miranda, E. Quaeghebeur, editors, *ISIPTA '15, Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications* pages 277-286. Aracne, 2015.

C Analysis of complex data structures

4) G. Schollmeyer. Lower quantiles for complete lattices. Technical Report 207, Department of Statistics, LMU Munich, 2017a.
URL `http://www.statistik.uni-muenchen.de/forschung/technical_reports/index.html`.
5) G. Schollmeyer. Application of lower quantiles for complete lattices to ranking data: Analyzing outlyingness of preference orderings. Technical Report 208, Department of Statistics, LMU Munich, 2017b. URL `http://www.statistik.uni-muenchen.de/forschung/technical_reports/index.html`.
6) C. Jansen, G. Schollmeyer and T. Augustin. A probabilistic evaluation framework for preference aggregation reflecting group homogeneity. *Revised version under review in Mathematical Social Sciences*, original version available as a technical report under `https://epub.ub.uni-muenchen.de/29269/`, the revised version can be found under `http://gschollmeyer.userweb.mwn.de/prob_evaluation_revision.pdf`.

D Relational data analysis: Stochastic dominance for partially ordered sets

7) G. Schollmeyer, C. Jansen, and T. Augustin. Detecting stochastic dominance for poset-valued random variables as an example of linear programming on closure systems. Technical Report 209, Department of Statistics, LMU Munich, 2017a.
URL `http://www.statistik.uni-muenchen.de/forschung/technical_reports/index.html`.

8) G. Schollmeyer, C. Jansen, and T. Augustin. A simple descriptive method for multidimensional item response theory based on stochastic dominance. Technical Report 210, Department of Statistics, LMU Munich, 2017b. URL `http://www.statistik.uni-muenchen.de/forschung/technical_reports/index.html`.

Declaration of the specific contributions of the author

The following list indicates for every contributing paper the own contribution of the author:

Contribution 1: The analysis of existing approaches as well as the newly developed concepts (the set-domained loss function and the set-loss region) are mainly due to the author. Thomas Augustin helped in writing and structuring the paper, wrote the introductory part and suggested to compare the partial identification approach also to classical methods. Both authors contributed to revising the paper.

Contribution 2: Big parts of the manuscript were drafted by Julia Plass. The author contributed to the development of the observation model: The idea of using an observation-model and to rely on the analysis of the mapping Φ was jointly developed by all authors at a workshop (WPMSIP 2014, Ghent, see `http://users.ugent.be/~slopatat/wpmsiip2014/`). All authors contributed to revising the paper.

Contributions 3, 4 and 5 were autonomously written by the author. Thomas Augustin proof read Contribution 3. With the members of the working group the author had some stimulating discussions during the research seminar about Contribution 4 and 5. The revision of Contribution 3 was autonomously done by the author.

Contribution 6 was mainly drafted by Christoph Jansen. The author supplied the aggregation function “commonality sharing” (based on Contributions 4 and 5) and drafted some parts of the paper, concretely, the overview about homogeneity measures in related work and the analysis on the homogeneity measure W of Kendall and Smith and distance based measures. (The author showed that Kendall’s and Smith’s W does not satisfy the conditions for a “reasonable” homogeneity measure developed in contribution 6 and that distance-based measures satisfy the proposed conditions). All authors contributed to revising the paper.

Contribution 7: The main ideas (especially the characterization of stochastic dominance by a linear program, the embedding of the stochastic dominance problem into the problem of optimizing a linear function on a closure system, as well as the statistical analysis and the taming ideas based on Vapnik-Chervonenkis theory) were developed by the author. Christoph Jansen worked out the part about duality for the linear program for detecting stochastic dominance. Thomas Augustin helped in writing and structuring the paper and discussed the involved concepts and especially the application example of stochastic dominance and differential item functioning.

Contribution 8: The basic idea of pairs of person ability and item difficulty relations is due to the author and arose after a seminar of the working group, where methods for stochastic dominance were discussed (cf., Contribution 7). With Christoph Jansen and Thomas Augustin, the author discussed about the reasonability of the developed relational concepts. Christoph Jansen and Thomas Augustin helped in writing and structuring the paper.

Preface

Statistical data analysis in the vein of statistical modeling always also involves a step of translating a substance matter problem into a more or less formal framework of statistical modeling. During this step, usually some idealizations and schematizations, as well as some value judgments about what is the adequate translation of the actual substance matter question into a mathematical/statistical question and what could be accepted as an answer, have to be made. The present thesis is concerned with different situations of statistical data analysis, where such idealizations involve a pronounced risk of weakening the adequacy of the data analysis for solving the envisaged substance matter problem. This comprises for example

Imperfectly observed data, especially interval data in the context of the methodology of partial identification.

Precisely observed data with a non-standard measurement of scale, for example ranking data.

Linear models, where the status of the model (as a descriptive or a structural model, to put it in Freedmans terms ([Freedman, 1987])) is unclear and at the same time of crucial importance.

Psychometric models of item response theory, where one can ask for the empirical status, the scale of measurement and the metaphysic character of the assumed latent traits.

Probabilistic models that are only/could only be partially specified or that are genuinely imprecise (a.k.a. imprecise probability theory).

Notions of inequality in the context of poverty/inequality analysis that can only be concretized in a very non-decisive manner because of the need of avoiding a too restrictive value judgment.

From the diversity of these points it becomes apparent, that the present cumulative thesis is not devoted to one single specific topic of statistics, it is more or less bundled around different topics, also showing that by working on different problems one comes across during some academic undertaking like a PhD, only a small amount of envisaged problems will come to an acceptable solution, if at all. This cumulative dissertation contributes to a variety of topics that are related to each other through the presence of some sort of non-ideal data or modeling situation.

One first line of research at the beginning of this PhD was the mathematical analysis of models of imprecise probabilities, namely the class of necessity measures that are special models from possibility theory ([Dubois and Prade, 1988]), which mathematically could be also seen as special cases of imprecise probability models (cf., e.g., [De Cooman and Aeyels, 1996]). The interest in such models came from the former interest in the *Dempster-Shafer theory of evidence* ([Shafer, 1976]) that was also partly a topic of my diploma theses written at the institute of algebra at TU Dresden. The mathematical background of the author becomes apparent in Contribution 3 describing the combinatorial structure of a polytope (called the credal set or core in the literature) that is a useful tool for the mathematical description of imprecise probability models.

A second line of research is also a typical research focus of the working group *Methodological Foundations of Statistics and its Applications* I am working in, namely the methodology of partial identification that is also loosely connected to the theory of imprecise probabilities. The methodology of partial identification is typically applied if one has to deal with imprecisely observed data that, without imposing further additional assumptions, lead to only partially identified statistical models. Attempts of generalizing statistical methods (here concretely linear regression) from precise data situations to the case of interval-valued observations, often come along with additional subtle conceptual questions that are genuinely addressing the understanding of statistical modeling: Questions like "What is the empirical status of the statistical model, is it the truly underlying structure (a structural model in Freedmans terms) or more a descriptive tool (a descriptive model in Freedmans terms) guiding the statistical analysis?" have to be thoroughly answered for a second time (the first time one hopefully envisaged the question for the point-identified case), because they lead to still more pronounced differences in the partially identified case, compared to the point-identified case. Of course, such conceptual questions seemingly usually lead not to new methods, but often only to a relativization of existing methods that could also be seen as a form of non-constructive carping. As the working group always gave some space for also discussing such conceptual questions, I always felt comfortable being interested in such questions and did not have the feeling of wasting worthy time. This said, I have to say some words about the general working atmosphere in the working group and at the department of statistics, because it surely had a big influence on the way the present thesis evolved:

One very valuable point was that beyond pure research, also teaching activities were always highly appreciated and both my supervisor Thomas Augustin and the whole professorship had much credit of trust. This concerns especially firstly the possibility to play a part in interdisciplinary seminars: Seminars related to the philosophy of science and especially to the philosophy of probability together with Thomas Augustin, Marco Cattaneo, Christina Schneider and Rudolf Seising were always exciting and a seminar with Clemens Draxler about psychometric modeling initiated my interest in conceptual aspects of psychometric modeling that also inspired Contribution 8. A second point was the possibility of doing self-reliant teaching. The opportunity to give a lecture (supported by Lehre@LMU[3]) on mathematical aspects of order- and lattice theory in the context of statistics led also to some ideas that substantially influenced the evolution of Contributions 7 and 8. Especially concerning Contribution 7, more or less by accident I became aware of some nice interrelations between two seemingly unrelated fields, namely, firstly lattice theory ([Birkhoff, 1940]) in the concretion of formal concept analysis ([Ganter and Wille, 2012]) as an applied mathematical theory for "*conceptual data analysis and knowledge processing*" [Ganter and Wille, 2012, p. VII] and the theory of statistical learning in the guise of Vapnik-Chervonenkis theory that can be seen as the natural "*continuation of the Kolmogorov-Glivenco-Cantelli line of theoretical statistics*" [Vapnik, 2006, p. 482]. Also Contribution 6 and the general interest in Social Choice theory, mainly triggered by discussions with Christoph Jansen, has its multiple roots in both this lecture and in a seminar about decision theory together with Christoph Jansen and Christina Schneider. Contributions 4 and 5 arose from a general interest in dealing with distribution

[3]The program Lehre@LMU (`http://www.uni-muenchen.de/studium/lehre_at_lmu/index.html`) aims at strengthening both research-oriented and practice-oriented teaching, see also `http://www.mathematik-informatik-statistik.uni-muenchen.de/studium_lehre/lehre_lmu/index.html` for the activities of the faculty of mathematics, informatics and statistics.

functions and my background in lattice and order theory.

Of course, most teaching experience has no directly visible trace in research, however, it subtly shapes the own understanding of statistics. This is surely also true for some lose joint activities with the Munich Center for Mathematical Philosophy (MCMP). A reading group about the philosophy of statistics conducted by Barbara Osimani, and regular joint informal meetings (called “Formal Informal”) organized by Marco Cattaneo from the statistics department and Roland Poellinger from the MCMP gave still other types of insights, also especially into the different “cultural background” of philosophers and statisticians.

Finally, the research seminar of the working group was the actual place for discussing the own research and the research of the other working group members. The possibility to present current research in a very informal way and the generous willingness of the participants to spend much time in discussing topics that are, due to the broad field of research interests of the working group, often only of direct interest for a subset of the participants, was actually very helpful.

As a department of statistics and especially a working group and its surrounding is not only an institution of research and teaching, but also an organic system build by its members, I want to take the opportunity to express my sincere thanks:

> to Thomas Augustin, for always taking time for me, especially if he had no time, for the big ostinato of confidence and trust, for his openness for contemplating peripheral research topics that are not directly related to the already very broad area of research of the working group, for the leap of faith of giving me the possibility of self-reliant teaching, for generous support and advice in all kind of administrative duties;
>
> to Frank Coolen and Lev Utkin for their willingness to review the present thesis;
>
> to Christian Heumann and Helmut Küchenhoff for their willingness to be a part of the thesis committee;
>
> to Thomas Augustin, Johann Brandt, Marco Cattaneo, Gero Walter and Andrea Wiencierz for welcoming me in the working group with much milk of human kindness;
>
> to Brigitte Maxa and Elke Höfner for the smooth cooperation in all kinds of administrative and other stuff;
>
> to Thomas Augustin, Clemens Draxler, Paul Fink, Christoph Jansen, Julia Plass, Rudolf Seising, Christina Schneider and Andrea Wiencierz, for inspiring joint seminars;
>
> to the whole working group (Thomas Augustin, Johanna Brandt, Marco Cattaneo, Eva Endres, Paul Fink, Christoph Jansen, Julia Plass, Gero Walter and Andrea Wiencierz) and all other participants of the research seminars for stimulating discussions triggering joint research;
>
> to Uli Pötter for being always a critical and sharp-witted contemporary, especially in discussions at the research seminar;
>
> again to the whole working group for a good collaboration in stemming big administrative work;

to the rest of the department of statistics for an always very pleasant working atmosphere;

to the colleagues of the MCMP, especially to Jean Bacelli, Samuel Fletcher, Jürgen Landes, Arthur Merin, Barbara Osimani, Roland Pöllinger for a lose but constant exchange of ideas;

to my former professors at TU Dresden, especially Stefan E. Schmidt and Bernhard Ganter for setting a solid ground not only in mathematical, but also in cultural terms;

to my parents and my family for a constant backing.

Contributing material: Overview, general aspects, remarks and perspectives

We now give an overview over all contributing papers. For a quick overview we begin every contribution with a very brief summary, followed by a detailed description of the contribution. Parts consisting of more than one contribution are firstly introduced by giving a short general overview. For every of the four parts of the thesis, we additionally discuss interrelations between contributions, possible further research that can be done and general aspects in separate sections called *Remarks and perspectives*.

Part A: Challenges for statistical modeling under partial identification in the context of regression analysis

In many applied situations of data analysis, the variables of actual interest cannot be directly observed in the precision actually needed from a substance matter point of view. Data are often rounded or heaped or are missing such that one generally only knows that the true unobserved variables of actual interest lie in some observed intervals. Beyond classical correction methods for rounding or heaping (cf., e.g., [Schneeweiß et al., 2010]), advanced statistical methods for interval-censored data deal with this problem in a constructive way by either explicitly modeling the coarsening process, or by identifying conditions like *Missing At Random* (MAR) or *Coarsening At Random* (CAR) under which the coarsening process can be ignored in some sense (see [Heitjan and Rubin, 1991]). Often cited works are Torelli and Trivellato [1993] for heaping or Lindsey and Ryan [1998] for interval censoring. Recent applications illustrating the flexibility of statistical modeling include for example Arulampalam et al. [2017], Allen et al. [2017], Eid et al. [2016] dealing with heaped or rounded data, McDonough and Millimet [2017] which is concerned with missing data in endogenous variables or Schomaker et al. [2015] who simultaneously treat missing data and measurement error. However, for such classical statistical approaches, the underlying statistical model has to be **point-identified**[4] to allow for an analysis. This is a very strong demand that often forces the researcher to impose further substantial assumptions or to use a very narrow model class for modeling the coarsening process to point-identify the model. These ways of point-identifying a statistical model are thus necessarily often not only guided by substance matter considerations but are simply driven by the demand of a point-identified model, no matter if the presumed point-identifying assumptions are doubtful or reasonable from a substance matter point of view. Beyond this, point-identifying assumptions like MAR or CAR are typically not statistically testable in many situations, see Jaeger [2006] (cf., also Plass, Cattaneo, Schollmeyer, and Augustin [2017b] for a somehow testable point-identifying assumption).
The methodology of **partial identification**, starting with early, mostly econometric works like e.g., Frisch [1934], Reiersl [1941], Marschak and Andrews [1944], tries to go one step beyond the paradigm of point-identification and treats the question of identification as a substance matter question and not as a rhetoric question of convenience. If one refrains from the demand of a point-identified model, then generally one cannot consistently estimate the true underlying parameter(s) of the model. Thus, one does not estimate the true parameter(s), but instead one estimates the so-called **identification region**. The identification region is the set of all parameters that are empirically indistinguishable from the true parameter(s), i.e., the set of all parameters that lead to the same distribution of all observable random variables like under

[4] A parametric statistical model is called point-identified, if for every two different parameters, the distributions of the corresponding observable random variables under these parameters also differ. Informally speaking, this means that with an infinite amount of data one can exactly identify the true underlying parameter.

the true parameter(s). Note that the identification region is not a statistical estimate, but the unknown true set-valued estimand that could alternatively be understood as that set of possible true parameters one could statistically infer with an infinite amount of data. Typical for partial identification and maybe a little bit unusual in classical statistics is the explicit decomposition of the statistical problem into the problem of identification and the problem of statistical estimation that accounts for sampling uncertainty. In applications, the estimates of the identification regions are then often still informative enough to allow for non-trivial substance matter conclusions that are then somehow more reliable because of the avoidance of imposing further assumptions.

However, the early works on partial identification were largely ignored in both econometrics and statistics before the 1990's. The works of Manski ([Manski, 1989, 1990]), starting in the late 1980s, led to some revival of the partial identification view within empirical economics. Works in the flavor of partial identification can nowadays be found in many areas (mainly in econometrics) and in different contexts like for example interval data (e.g., [Manski and Tamer, 2002],) missing data (e.g., [Manski, 2005]), treatment effects (e.g., [Manski, 1990, Stoye, 2009], measurement error/misclassification (e.g., [Molinari, 2008, Watson, 1964, Küchenhoff et al., 2012]), entry games (e.g., [Gentry and Li, 2014, Ciliberto and Tamer, 2009]) or generally in economic models with inequalities (e.g., [Romano and Shaikh, 2010, Chernozhukov et al., 2013]). A somehow parallel development has taken place in biometrics and medical statistics under the keyword **systematic sensitivity analysis** (e.g., [Gilbert et al., 2003, Vansteelandt et al., 2006, Chiba and VanderWeele, 2011]): For a statistical model that is not point-identified one specifies some reasonable nuisance parameter(s) with the property that if one knows this nuisance parameter(s), then the model is identified and can be estimated. Then, one does a statistical analysis for a whole range of the actually unknown nuisance parameter(s) to get an impression about how big the influence of the sensitivity parameter on the result of the analysis is. Apart from the field of partial identification and classical statistics, the analysis of interval-valued data/imprecise data is also a research topic in imprecise probabilities (e.g., [Coolen and Yan, 2004, Utkin and Augustin, 2007, Couso and Sánchez, 2011]), in fuzzy set theory (e.g., [Denœux et al., 2005, Spadoni and Stefanini, 2011]) or machine learning (e.g., [Do and Poulet, 2009, Carrizosa et al., 2007, Utkin and Coolen, 2011, Utkin and Chekh, 2015, Steyer, 2017]). Also likelihood approaches for dealing with interval data have been established, see, e.g., Zhang [2010], Cattaneo and Wiencierz [2012], Couso and Dubois [2017].

Contribution 1) G. Schollmeyer, 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.

> **Short summary:** In Contribution 1 we treat partially identified linear models where the outcome variable is observed in intervals. In the framework of partial identification we analyze different identification regions known from literature, introduce and analyze a new kind of identification region and furthermore provide estimation techniques for this region based on quadratic programming. In an application example on wine quality assessed by experts in dependence on 11 physicochemical covariates, we compare estimates of the different identification regions with each other and to results from classical statistical approaches of interval data analysis.

Contribution **1** deals with linear regression of the form $Y = X\beta + \varepsilon$ where X is precisely observed, but Y can only be observed in intervals $[\underline{\mathrm{Y}}, \overline{\mathrm{Y}}]$, so all we know about Y is that $Y \in [\underline{\mathrm{Y}}, \overline{\mathrm{Y}}]$. A typical example of interval-valued observed Y is the answer to the question about income in social surveys: To reduce the usually relatively high non-response rate to this sensitive question, one often does not only directly ask the participants about their income. Instead, after a direct question about income, one adds further categorized questions about the income (like, e.g.: "Is your income between 200 and 299 Euro?") to get further imprecise information about participants that did not answer the direct question but are willing to supply some imprecise information about their income. In such situations, one can expect that the propensity to give only an imprecise answer or to completely refuse to answer is dependent on the actual income. Thus, from a substance matter point of view, assumptions like CAR seem to be not appropriate, here.
The contribution analyzes different variants of identification regions (and their estimation) that arise from different choices of the exact underlying statistical model: Compared to the point-identified case, the question of possible misspecification of the linear model plays a crucial role: For the classical, point-identified linear model, a misspecification of the linear relationship between X and $\mathbb{E}(Y \mid X)$ would make the notion of the "true parameter" meaningless, but one can replace this notion by the notion of the parameter of the best linear predictor loss (BLP). Actually, the question if one assumes an underlying true regression parameter or if one is only interested in the best linear predictor plays no role w.r.t. the estimation: If one assumes normally distributed homoscedastic noise, then both the true parameter, as well as the BLP-parameter under squared loss are reasonably estimated by the least squares estimator. This coincidence disappears in the partially identified situation. The question if one assumes a true underlying parameter or if one anticipates the possibility of misspecification and only looks at the BLP leads to two very different types of identification regions:

1. If one assumes a truly underlying linear relationship between $\mathbb{E}(Y \mid X)$ and X and understands the model as a **structural model** in the sense of Freedman (cf., [Freedman, 1987]) one obtains a region that is called the **Marrow Region** in the contribution: Given the interval-valued response $[\underline{\mathrm{Y}}, \overline{\mathrm{Y}}]$ one takes every possible random variable $Y \in [\underline{\mathrm{Y}}, \overline{\mathrm{Y}}]$ that can be obtained through a linear relationship $Y = X\beta + \varepsilon$ and takes as the identification region the set of all associated parameter vectors β. This set can in fact be also empty in some situations.

2. If one understands the model only as a **descriptive model** (in the sense of Freedman) that takes the assumed linear relationship between $\mathbb{E}(Y \mid X)$ and X only as a crude approximation guiding the data analysis, one presumably would go a different way: One would look at every arbitrary random variable $Y \in [\underline{\mathrm{Y}}, \overline{\mathrm{Y}}]$ (also if it does not fit to a linear relationship with X) and would compute the BLP. Finally, one would collect every such obtained BLP in a set. This type of identification region is called **Collection Region** in the contribution. Due to its construction, the Collection Region is always a superset of the Marrow Region.

These two identification regions have already been considered in the literature: The Marrow region was studied for example in Chernozhukov et al. [2004, 2007], where confidence regions for the Marrow region were developed. The Collection Region was used for example in Beresteanu et al. [2011], Beresteanu and Molinari [2008], Černý and Rada [2011], Rohwer and Pötter [2001]. A first analysis of the differences between the two regions can be found in Ponomareva and Tamer [2011]. Contribution 1 also analyzes the two identification regions and furthermore contrasts them with another identification region developed in Section 4 of the contribution. It turns out that the Marrow Region is not continuously dependent on the distribution of the observable random variables. (Actually it is not even continuous w.r.t. the weak-*-topology applied to the conditional expectations and the support function of the identification region.) This also means that in the case of misspecification, the Marrow Region can be very small (it can actually also be empty). But this does not mean that one has much information about the true parameter, it could also only mean that the model is simply misspecified and the Marrow Region only pretends much information about the (non-existing) true parameter. In contrast to this, the continuity properties and also the asymptotic properties of the Collection Region are satisfactory, but in some situations the Collection Region seems to be unnecessarily too uninformative.

Beyond the analysis of the two identification regions the contribution also develops a new kind of identification region (called **Set-loss Region**, see Section 4 of the contribution) that could be understood as a compromise between the two other regions. The idea is here to look not at single parameter vectors β, but at whole (convex and compact) sets Γ of parameter vectors. (Technically, this is done by describing compact and convex sets by their support functions.) For a given interval-valued response $[\underline{\mathrm{Y}}, \overline{\mathrm{Y}}]$ one searches for that set Γ of parameter vectors, whose predictions $\hat{\underline{\mathrm{Y}}} := \inf_{\beta \in \Gamma} X\beta$ and $\hat{\overline{\mathrm{Y}}} := \sup_{\beta \in \Gamma} X\beta$ fits best to the observed intervals $\underline{\mathrm{Y}}$ and $\overline{\mathrm{Y}}$ w.r.t. a given metric. This typical statistical approach of minimizing a metric (or a loss function), which in our situation turns out to be a projection onto a convex cone of a Hilbert space, has then nice continuity properties. The Set-loss Region can be estimated by relying on sample analogues (cf., Section 4.3 of the contribution) that share the same nice continuity properties. To compute the Set-loss Region one has to solve a quadratic program, see Section 4.2 of the contribution. From a statistical point of view, the estimation problem is very similar to the problem of nonparametrically estimating a convex function under square loss. Thus, nice asymptotic properties like asymptotic normality cannot be expected, here (cf., Groeneboom et al. [2001]).

Finally, the identification regions are also investigated by means of a data example where the quality of red wine, measured by the median of the ratings of 3 experts, in dependence of different physicochemical properties of the red wine is analyzed. Beside the analysis with the methodology of partial identification, we also analyzed the data set with classical methods, firstly by replacing the interval-valued response with the interval-midpoints and secondly by the application of an interval censoring model.

Contribution 2) J. Plass, T. Augustin, M. Cattaneo, G. Schollmeyer, Statistical modelling under epistemic data imprecision: Some results on estimating multinomial distributions and logistic regression for coarse categorical data. In: T. Augustin, S. Doria, E. Miranda, E. Quaeghebeur, editors, *ISIPTA '15, Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*, pages 247-256. Aracne, 2015.

Short summary: Contribution 2 deals with logistic regression under partial identification that is due to coarsened responses. On the one hand we study point-identifying assumptions (*Coarsening at Random* and *Subgroup Independence*) and on the other hand we provide a likelihood-based method for estimating the identified set based on an observation model. Finally, we also address the possibility of incorporating further substance matter assumptions on the coarsening process for the response variable. We illustrate the method with data from the German panel study "Labor Market and Social Security" (PASS).

Contribution **2** is concerned with partial identification in regression analysis under coarsened responses variables, too. Here, we deal with logistic regression under categorical covariates. Since we focus on the case of categorical covariates, the difference between the point-identified and the partially identified situation w.r.t. the question of misspecification seems to bo less pronounced. (Note that e.g., for the simplest case of one categorical covariate, the logistic link essentially imposes no additional structural assumption.) Concretely, we consider a binary response $Y \in \{0,1\}$ that can only be observed in a coarsened form as $\mathcal{Y}$ with $\mathcal{Y} \in 2^{\{0,1\}}$, and one only knows that $Y \in \mathcal{Y}$. We are interested in the relationship between Y and the covariate(s) X. We use here an observation model $\mathcal{Q}$ that explicitly models the conditional distribution of $Y \mid x$ and $\mathcal{Y} \mid x$ for every possible value x of X. The main difficulty is then the estimation of the coarsening probabilities in the observation model. Since Y cannot be observed, the coarsening probabilities are generally only partially identified. For estimation, we use here a maximum likelihood approach. Instead of directly maximizing the likelihood, in a first step we maximize the likelihood $L(\theta) = P_\theta(\mathcal{Y}_1 = \mathfrak{y}_1, \ldots, \mathcal{Y}_n = \mathfrak{y}_n \mid X_1 = x_1, \ldots, X_n = x_n)$ of the observable events w.r.t. the parameters $\theta_{\mathfrak{y}x} := P_\theta(\mathcal{Y} = \mathfrak{y} \mid x)$. From the estimated parameters $\hat{\theta}_{\mathfrak{y}x} = \hat{P}(\mathcal{Y} = \mathfrak{y} \mid x)$, in a second step, with the help of the observation model, we deductively obtain the set of all coarsening probabilities and the set of all values $P(Y = y \mid x)$ of actual interest, that are compatible with the estimates $\hat{\theta}_{\mathfrak{y}x} = \hat{P}(\mathcal{Y} = \mathfrak{y} \mid x)$ of the first step.

The explicit modeling with an observation model also allows for imposing further substance matter assumptions, if such assumptions are reasonable. For example, in specific situations it can be reasonable to assume that we have $P(\mathcal{Y} = \{0,1\} \mid X = x, Y = 0) = P(\mathcal{Y} = \{0,1\} \mid X = x, Y = 1)$ (This is exactly the Coarsening At Random assumption.[5]) or $P(\mathcal{Y} = \{0,1\} \mid X = x, Y = 0) \leq P(\mathcal{Y} = \{0,1\} \mid X = x, Y = 1)$, or $P(\mathcal{Y} = \{0,1\} \mid X = x, Y = y) = P(\mathcal{Y} = \{0,1\} \mid X = x', Y = y)$ for arbitrary x, x'. The last assumption is called **Subgroup Independence** in the contribution and can be understood as an assumption that

[5] The *Coarsening At Random* assumption is usually not directly statistically testable, however, in principle one could test the assumption with further validation studies, where Y can be observed.

is dual to the Coaresning At Random assumption, because, informally speaking, one assumes that the coarsening is independent of the values of the covariates, where for CAR one assumes independence from the values of the response. Except from degenerate cases, the Subgroup Independence assumption is point-identifying, but at the same time, opposed to CAR, the Subgroup Independence assumption is also statistically testable to some extent, see Plass, Cattaneo, Schollmeyer, and Augustin [2017b], where also an appropriate likelihood-ratio test is developed.

Remarks and perspectives Part A

Contribution 1 is an extended version for a special issue on the Eighth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2013) that is based on an ISIPTA conference paper ([Schollmeyer and Augustin, 2013]). In the ISIPTA conference paper one can find further more algebraic/lattice theoretic and geometric analyses of the Marrow Region and the Collection Region that were not included in Contribution 1 because of a more application oriented focus of the contribution.

Based on Contribution 2, a follow-up paper (Plass, Cattaneo, Augustin, Schollmeyer, and Heumann [2017a]) further develops the conceptual ideas more into the direction of a conceptualization in the spirit the Marrow region by relying on further parametric assumptions. There, we develop estimates as well as confidence regions based on the profile likelihood function. An application of the methods of Contribution 2 in the context of small area estimation can be found in Plass, Omar, and Augustin [2017].

Ill-posedness: Contribution 1 made clear that beyond the question of identification, within the methodology of partial identification, the model-understanding plays a crucial role and one should analyze the derived methods also w.r.t. their continuity properties to make sure that one does not hopelessly try to solve ill-posed problems. So, we have to distinguish the question about identification from the question of ill-posedness/ill-conditioning. The question of ill-posedness seems to be not explicitly addressed in the literature about partial identification very often. In Contribution 2, ill-posedness is also not explicitly addressed, but the results of Contribution 2 are all given in closed form such that one can easily analyze possible ill-conditioning of the estimation functions. The difference between identification and ill-conditioning becomes especially clear in the point-identifying assumption of Subgroup Independence in Contribution 2: If one has at least one categorical covariate with at least two possible values, then the parameters of interest are usually point-identified, but as one can see from the equations (18) at page 254, if for both values x of the covariate the conditional distribution of $\mathcal{Y}$ given x is very similar or identical (this would be exactly the homogeneous case were X has no influence on $\mathcal{Y}$), then the estimation functions (18) can be arbitrary ill-conditioned or still undefined (because one would get a fraction $0/0$).

Nonparametric alternatives: For the case of linear regression, if one does not believe in the linear relationship, one can also think about nonparametric approaches. Actually, some nonparametric approaches like kernel regression methods (see [Nadaraya, 1964, Watson, 1964]) can easily deal with imprecisely observed responses, since, at least for a fixed band-width, the nonparametric regression result of a kernel regression is monotonically increasing in the response vector[6]. However, especially in the social sciences, in the context of action planning, one is mainly interested in one single number for every covariate that describes the global effect of increasing the covariate about one unit: ‘‘One wants to know how much achievement a given amount of integration will bring, how many dollars will buy a given amount of health, or what increased chance of conviction results from pretrial detention... . The [regression] coefficients do not describe the behavior in the sample, but instead delineate the process at work.” [Achen, 1982, p. 69]. This demand is understandable but should obviously not lead to a Panglossian attitude that ignores the possibility that the true underlying relationship is very non-linear, which could possibly discredit the usefulness of the schematic linear modeling. Furthermore, in the context of partial identification, we have the additional problem that the linear relationship cannot be rigorously tested[7]. A further point is also getting more problematic in the partially identified situation: Since one cannot straight forwardly assess the fit of the model, one has no direct way of doing model selection. Furthermore, if one takes a descriptive model perspective, then already in the point-identified case one has to face the problem, that generally, the BLP-parameter for one covariate depends upon which other covariates were selected. Thus, thought through the end, the true BLP-parameter for a selected covariate is not a fixed parameter, but a random parameter.[8]

Inference: Concerning inference, in the partially identified situation one has to firstly decide, which kind of coverage property one is actually interested in. One can demand that a coverage region covers either the true unknown parameter or the whole true unknown identification region with probability greater than or equal to a prespecified level γ. Both in Chernozhukov et al. [2007, 2004], as well as in Beresteanu and Molinari [2008] the second covering property is used. In Beresteanu and Molinari [2008], one utilizes random set theory and the confidence regions are obtained by using the metric structure of the parameter-space $\mathrm{co}\,(\mathbb{R}^p)$ of all compact and convex subsets of $\mathbb{R}^p$, equipped with the Hausdorff distance

$$d_H(A,B) = \max\left\{\sup_{a\in A}\inf_{b\in B} d(a,b) \quad , \quad \sup_{b\in B}\inf_{a\in A} d(b,a)\right\},$$

where d is the Euclidean distance in R^p and p is the number of covariates including the intercept. One issue with this approach is that it is not affine equivariant: If for example one multiplies one covariate with a constant, then this would possibly result in a very different

[6]For an application of a two-step semi-parametric procedure, see [Manski and Tamer, 2002]. There, one firstly estimates the conditional expectations $\mathbb{E}(\underline{\mathrm{Y}} \mid x)$ and $\mathbb{E}(\overline{\mathrm{Y}} \mid x)$ nonparametrically via kernel regression. Then one can “deduce” from the estimated conditional expectations an estimate for the identification region of the parameters with a modified minimum-distance approach. Another interesting application of kernel techniques in the context of local linear regression ([Fan, 1993]) under interval-valued outcomes can be found in the third contribution of [Peng, 2017].

[7]Actually, it is maybe a little bit too schematic to distinguish only between assumptions that are testable and assumptions that are not, because in the sense of the Duhem-Quine problem of under determination, every assumption is not testable in isolation, it is only testable via the interplay with other assumptions.

[8]For the point-identified case there is currently much research about model selection and especially inference after model selection in the framework of a descriptive model understanding, where one firstly has to think about the meaning of the model parameters. For example in Berk et al. [2013] one directly deals with inference concerning the true BLP parameters obtained by projecting the conditional expectation of the response onto the finally selected covariates. Another nice approach is the “Leave one covariate out” approach in Lei et al. [2017]. There, similarly to variable importance measures used in random forests ([Breiman, 2001], cf., also [Strobl et al., 2008] who discuss problems of bias towards correlated predictor variables and propose a conditional variable importance measure that reflects the true impact of each predictor variable more reliably.) one measures the importance of one covariate by the difference in prediction power if that covariate is excluded from prediction.

confidence region for the transformed covariates. One way out of this invariance issue would be to standardize every covariate before doing the actual analysis. However, it could still appear that the obtained confidence regions are very broad in one parameter direction where actually the statistical uncertainty is very low compared to other directions. (This seems to also apply to the confidence procedure in Chernozhukov et al. [2007, 2004]). The reason seems to be the following: To construct the confidence region, Beresteanu and Molinari [2008] use as a test statistic the Hausdorff distance between the true identification region and the estimated regions and then they simply invert this unidimensional statistic. Because of this unidimensionality, one cannot flexibly account for different "shapes" of the distribution of the set-valued estimates in this approach. Similarly, in Chernozhukov et al. [2007, 2004] the authors use a characterization of the identified set as the set of minimizers of a real-valued criterion function $Q : \mathbb{R}^p \longrightarrow \mathbb{R}$. As a confidence region the authors then take an α-cut $CI = \{\theta \in \mathbb{R}^p \mid Q_n(\theta) \leq \alpha\}$ of an empirical analogue Q_n of the criterion function Q, where α is an appropriately chosen real-valued threshold. Thus, also here one seems to have no flexibility for accounting for different shapes.

A simple idea to account for the above issue could be to do not look at the metric structure of $\mathrm{co}(\mathbb{R}^p)$, but to utilize the structure of $\mathrm{co}(\mathbb{R}^p)$ as a partially ordered set that is additionally equipped with an addition, namely Minkowski addition[9]: To do inference, one can draw bootstrap samples and would obtain a sample of convex sets. Then, one can try to define as a "descriptive" confidence region a convex set that is a superset of at least $\gamma \cdot 100\%$ of the bootstrapped sets and that is minimal w.r.t. this property. This would of course only give a descriptive confidence region that does not necessarily (also not asymptotically) have the right covering probability. One reason for this is best explained if we think of convex compact sets as represented by its support functions: Every convex and compact set $A \subseteq \mathbb{R}^p$ can uniquely be described by its support function h_A via

$$h_A : \mathbb{R}^p \longrightarrow \mathbb{R} : u \mapsto \sup_{a \in A} \langle u, a \rangle.$$

Then, for two sets $A, B \in \mathrm{co}(\mathbb{R}^p)$ we have $A \subseteq B \iff \forall u \in \mathbb{R}^p : \; h_A(u) \leq h_B(u)$ and the Minkowski addition can be described in terms of support functions by pointwise addition as $h_{A \oplus B}(u) = h_A(u) + h_B(u)$. The bootstrap samples can thus be understood as bootstrap samples of special functions for which one could try to get some sort of a joint confidence band. If the distribution of the support functions is not symmetric, which one can expect, the bootstrap will generally be biased. One can try to deal with this issue by applying some kind of symmetrization procedure, for example by appropriately mirroring[10] the confidence band like it also done in the univariate reversed percentile method (called basic bootstrap confidence interval in [Davison and Hinkley, 1997]). Actually, this will only work well if one has a situation that is close to a pure translation family, otherwise, the mirroring can make the coverage still worse, see the explanations given in Hesterberg [2015] for the univariate case. Alternatively, one can try to additionally account for scale heterogeneity by applying a further transformation to construct a t-type statistic that is approximately pivotal. However, since by dealing with the space of convex compact sets or with the space of support functions, respectively, one has an infinite-dimensional space and the question, if and when the bootstrap

[9]The Minkowski sum of two sets A and B is simply the set $A \oplus B := \{a + b \mid a \in A, b \in B\}$.

[10]By mirroring the confidence band one possibly would obtain an associated boundary of the mirrored confidence band that is actually no support function of some convex compact set, but this is no problem if one already thinks not in the space of support functions but in the linear span of the space of the support functions.

will actually work, seems to be far from clear (at least for the author[11]). Thus, this idea was not directly followed by the author in subsequent studies, but it actually inspired the development of a simple descriptive notion of a quantile for samples/random variables with values in a complete lattice described in Contribution 4. Note here that the set of all convex compact sets of $\mathbb{R}^p$, together with the set inclusion builds a complete lattice.

Part B: Computational challenges in the theory of imprecise probabilities

Contribution 3) G. Schollmeyer. On the number and characterization of the extreme points of the core of necessity measures on finite spaces. In: T. Augustin, S. Doria, E. Miranda, E. Quaeghebeur, editors, *ISIPTA '15, Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications* pages 277-286. Aracne, 2015.

Short summary: In Contribution 3 we analyze the combinatorial structure of the extreme points and the edges of a convex polytope (called credal set or core in the literature) that plays an important role in the analysis of non-additive set functions used in theories of imprecise probabilities. We supply an effective procedure for enumerating all extreme points that could be used for efficient computations in imprecise probabilities. In a small study, we also compare the enumeration procedure with standard enumeration procedures like the reverse search method and the double description method.

The theory of imprecise probabilities (for an introduction, see, e.g., Augustin et al. [2014]) comprises a bunch of different theories[12]that attempt at generalizing classical probability theory to allow for an appropriate handling of different facets of the pleomorphic meaning of the broad terms "uncertainty" and "ambiguity" that are traditionally modeled only within one single mathematical paradigm, namely the concept of precise probabilities. Mathematically, much of these attempts boil down to use non-additive set-functions instead of finitely additive or σ-additive ones. (For a general mathematical introduction to measure- and integration theory for such non-additive measures, see [Denneberg, 1994].) For the mathematical analysis of non-additive set functions $\nu : \mathcal{A} \longrightarrow [0, 1]$, where $\mathcal{A}$ is some σ-algebra on some space Ω, the so-called credal set is an important tool. The **credal set** of a non-additive measure ν is the set

$$\mathcal{M}(\nu) := \{p \mid p \text{ probability measure on } (\Omega, \mathcal{A}),\ \forall A \in \mathcal{A} :\ p(A) \geq \nu(A)\}$$

of all classical (finitely additive or σ-additive, depending on the theory) probability measures that are in a sense compatible with the non-additive measure ν. For finite Ω the credal set is a convex polytope with finitely many extreme points. Characterizing the polytope and especially efficiently computing the extreme points is of interest in different areas of application,

[11] Some informal simulations showed different results, for the collection region, where the asymptotics of the estimated confidence regions are very regular, cf. [Beresteanu and Molinari, 2008], the results were acceptable, but for the Set-loss Region the results were not so promising.

[12] This bundle ranges 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]), nonparametric predictive inference ([Coolen and Augustin, 2009] to belief functions ([Shafer, 1976]) and possibility theory ([Dubois and Prade, 1988]).

for example in decision making (cf., e.g., [Utkin and Augustin, 2005], [Jansen, Schollmeyer, and Augustin, 2017b, Jansen, Augustin, and Schollmeyer, 2017a]) or in statistical hypothesis testing under imprecise probabilistic models (see [Augustin, 1998]). More generally, the aspect of computation in the context of imprecise probabilities is a very prominent topic in this field, because much of the computations that are very simple for the case of precise probabilities become easily very hard for imprecise generalizations. Works that try to make computations practically manageable include for example Bronevich and Augustin [2009], Destercke [2011] where one tries to approximate complex models by simpler models, Antonucci et al. [2015] where linear programming techniques are used or Utkin and Augustin [2005] which additionally utilizes duality theory of linear programming. Beyond imprecise probability theory, the credal set is also an object of interest in cooperative game theory[13] (cf., e.g., [Shapley, 1971, Gillies, 1959, Bondareva, 1963, Derks et al., 2000, Studený and Kroupa, 2016, Yokote, 2017, Abe and Funaki, 2017]), see [Studený and Kratochvíl, 2017] for a recent discussion of the relations between concepts used in imprecise probabilities and notions used in game theory.
Contribution **3** is concerned with a special class of non-additive measures, namely necessity measures that play a role in possibility theory ([Dubois and Prade, 1988]). A necessity measure is a non-additive set function ν satisfying[14]

$$\forall A, B, \in \mathcal{A} : \nu(A \cap B) = \min\{\nu(A), \nu(B)\}.$$

One can show that necessity measures are mathematically special kinds of so-called coherent lower probabilities[15] (see [De Cooman and Aeyels, 1996]). The contribution analyzes the exact combinatorial structure of the credal set of necessity measures. In Wallner [2007] it is shown that for the general case of lower coherent probabilities on a finite space Ω, the credal set has at most $|\Omega|!$ extreme points and that this bound is sharp[16]. For the special case of necessity measures, in Miranda et al. [2003] it is shown that the number of extreme points is bounded by the sharp bound $2^{|\Omega|-1}$ and in Kroupa [2008] further bounds are given in the language of possibility measures.[17] There, it is shown that the associated credal set is a simple polytope and general bounds for simple polytopes are used to get bounds for the number of extreme points.

The contribution sharpens these known results and gives an exact combinatorial description of the extreme points providing also exact formulae for the number of extreme points and furthermore supplies an efficient enumeration method for the extreme points. The contribution makes substantial use of the representation of necessity measures as special kinds of belief functions which can be efficiently described by their Moebius-inverses[18]. The enumeration method is efficient in the sense of not enumerating any extreme point twice, the time

[13] There, it is called *core* instead of *credal set*.

[14] Here, we only consider the case of a finite space Ω.

[15] A lower coherent probability is a non-additive measure satisfying $\nu(A) = \inf\limits_{p \in \mathcal{M}(\nu)} (p(A))$ for all sets A of the domain of ν. Note that in the context of lower probabilities one does not demand the domain of ν to be a σ-algebra and furthermore one takes all finitely additive classical probability measures (called linear previsions in this context) in the construction of the credal set.

[16] Actually, this result was already achieved earlier by [Derks and Kuipers, 2002] in the context of cooperative game theory.

[17] A possibility measure is the dual measure associated to a necessity measure: If ν is a necessity measure, then the associated possibility measure μ is defined via $\mu(A) = 1 - \nu(A^c)$.

[18] The Moebius inverse of a set-function $\nu : 2^\Omega \longrightarrow \mathbb{R}$ on a finite set Ω is that uniquely defined set function $m : 2^\Omega \longrightarrow \mathbb{R}$ satisfying $\forall A \subseteq \Omega : \nu(A) = \sum_{B \subseteq A} m(B)$. A belief function is a normalized set function (i.e.: $\nu(\emptyset) = 0$; $\nu(\Omega) = 1$), whose Moebius inverse is non-negative (i.e.: $\forall A \in A : m(A) \geq 0$).

complexity is $\mathcal{O}(m \cdot |\mathcal{F}(\nu)|)$, where m is the number of extreme points and $\mathcal{F}(\nu)$ is the set of the so-called focal elements of ν, that is the set of all sets $A \in \mathcal{F}$ where the Moebius inverse is positive. (The cardinality $|\mathcal{F}(\nu)|$ is bounded by $|\Omega|$ for the case of necessity measures.) The enumeration procedure is also compared to standard enumeration procedures for polytopes like the reverse search method ([Avis and Fukuda, 1996, Avis, 2000]) or the double description method ([Motzkin et al., 1953, Fukuda and Prodon, 1996]). Because the combinatorial description of the extreme points obtained in the contribution is very explicitly in nature, the associated enumeration procedure is more efficient than the above mentioned standard methods, furthermore it allows for a closed form for the number of extreme points as

$$m = |A_1| \cdot \prod_{i=2}^{k} (|A_i \backslash A_{i-1}| + 1),$$

where $A_1 \subset A_2 \subset \ldots \subset A_k$ is the increasingly ordered set of the focal elements. (For the case of necessity measures, the focal elements are always linearly ordered w.r.t. set inclusion.)

Beyond the extreme points, also the edge structure is analyzed in the contribution: A sharp characterization of adjacent extreme points and a closed form of the number of edges of the polytope as

$$\frac{1}{2}|A_1| \cdot \prod_{i=2}^{k} (|A_i \backslash A_{i-1}| + 1) \cdot \left(|A_1| - 1 + \sum_{i=2}^{k} |A_i \backslash A_{i-1}| \right)$$

is given. In a final step, a general enumeration procedure for arbitrary belief functions is also proposed.

Remarks and perspectives Part B

Contribution 3 gave the exact number of the extreme points and an efficient enumeration procedure for computing all extreme points. However, since a necessity measure can have up to $2^{|\Omega|-1}$ extreme points in the worst case, enumerating all extreme points is still very hard due to the mere amount of extreme points. Furthermore, for many computations one actually does not need the extreme points, because necessity measures are quite simply structured. For example, the lower expectation of a random variable X w.r.t. a necessity measure ν can be simply computed as a Choquet integral or with the help of the Moebius inverse (cf., e.g., [Troffaes and Hable, 2014, p. 332f]). Beyond computational aspects, note also that though necessity measures are mathematically special cases of coherent lower probabilities, their interpretation can be quite different: For coherent lower probabilities under the epistemic interpretation (cf., [Couso and Dubois, 2014]), one assumes that there is a precise probability measure that is unknown and the credal set describes all precise probability measures that cannot be excluded as candidates for the underlying unknown true probability measure. Such an interpretation motivates emphasis on the extreme points of the credal set as the extreme possible true measures. However, in the context of possibility theory, a necessity measure does not need to allow for such an interpretation, see Couso and Dubois [2014]. Thus, one has to be careful when combining non-additive measures stemming from different theories.

In the meanwhile, also results for the extreme points of the credal set of so-called probability boxes (p-boxes, for short) are available. The class of p-boxes is a class of coherent lower probabilities whose credal sets are generated as the set of all probability measures whose distribution function F lies between some lower distribution function $\underline{F}$ and some upper distribution function $\overline{F}$. This class is a superclass of the class of necessity measures and is usually interpreted in an epistemic fashion, i.e., one assumes that there is a true measure with cumulative distribution function F lying in the interval $[\underline{F}, \overline{F}]$. Probability boxes have a wide range of application (see [Ferson et al., 2002]) including also applications in classical statistics: For example, if one quantifies the statistical sampling uncertainty of the empirical distribution function of a sample of a random variable with the help of a nonparametric Kolmogorov-Smirnov type confidence band ([Kolmogorov, 1941, Smirnov, 1939, Feller, 1948, Miller, 1956]) then one has essentially a p-box. For the case of p-boxes on a finite space Ω, in Montes and Destercke [2017] sharp bounds for the maximal number of extreme points and an enumeration procedure for computing the extreme points are given.

Part C: Analysis of non-standard data structures

In multivariate statistics, there is a broad literature (cf., e.g., [Mosler, 2013]) aiming at generalizing the concept of a quantile and especially the median as a measure of location to the multivariate situation. One basic tool for establishing such generalizations is the notion of *data depth* and the notion of *outlyingness*. A depth-function measures how deep a certain data point lies w.r.t. a data cloud or w.r.t. a probability distribution. Dually, an outlyingness-function measures the outlyingness of data points. The concept of data depth as initially proposed[19] by Tukey [1975] as a graphical tool for visualizing bivariate data, initiated a broad line of research establishing further notions of data dept like simplicial depth ([Liu et al., 1990]), Oja depth ([Oja, 1983]), zonoid depth [Koshevoy and Mosler, 1997] or L^p depth ([Zuo and Serfling, 2000]) to cite just a few. A more recent proposal for data depth is the Monge-Kantorovich depth ([Chernozhukov et al., 2017]) based on the Monge-Kantorovich theory of measure transportation, where one uses an optimal transportation map between a distribution of interest on $\mathbb{R}^d$ and a reference distribution on the d-dimensional unit ball. In contrast to the most other data depth notions, Monge-Kantorovich depth is also able to capture non-convex aspects of the data. Data depth has a broad range of applications in multivariate data analysis, like for example in description of multivariate distributions (e.g., [Liu et al., 1999, Serfling, 2004, Wang and Serfling, 2005]), outlier detection (e.g., [Serfling, 2006, Zhang, 2002]), multivariate density estimation (e.g., [Fraiman et al., 1997]), depth-based classification or clustering (e.g., [Ruts and Rousseeuw, 1996, Christmann, 2002, Jörnsten, 2004]), rank and sign tests (e.g., [Brown and Hettmansperger, 1989, Hettmansperger and Oja, 1994], or linear regression (e.g., [Rousseeuw and Hubert, 1999]).

Most treatments of data depth are basically devoted to the Euclidean d - space as a linear space and put focus on affine invariance[20] as a desirable property[21]. This property seems natural especially if one has a geometrical understanding of $\mathbb{R}^d$ (cf. also [Chaudhuri,

[19]An idea very similar to Tukey's half-space depth was already used in [Hodges, 1955] for a bivariate sign test.

[20]This means that the depth-function is invariant under a simultaneous affine transformation of both the considered data point and the data cloud/probability distribution, cf., [Mosler, 2013, p. 3].

[21]An axiomatic approach to data depth can be found in Liu et al. [1990] and Zuo and Serfling [2000], who propose four properties that could generally be considered desirable for a statistical depth function: a) affine invariance, b) maximality at the center, c) linear monotonicity relative to the deepest point and d) vanishing at infinity.

1996]). However, in some situations of application like for example in multidimensional poverty/inequality analysis (cf., e.g., Alkire et al. [2015]) or in the analysis of multidimensional psychological latent constructs like for example "autoritarism" (cf., e.g., Duckitt et al. [2010]), a geometrical understanding is not suggesting itself. An order theoretic underpinning looking at the point-wise ordering, where for example a person could be termed as less poor than another person if she is less poor w.r.t. all considered dimensions of poverty, seems to be more natural in such cases.

Contribution 4) G. Schollmeyer. Lower quantiles for complete lattices. Technical Report 207, Department of Statistics, LMU Munich, 2017a. URL `http://www.statistik.uni-muenchen.de/forschung/technical_reports/index.html`.

> **Short summary:** Contribution 4 develops a quantile concept for the case of data or random variables with values in a complete lattice. The concept turns out to be related to the notion of data depth known from multivariate statistics. In the contribution we analyze how different properties, like for example uniqueness or representation invariance, which is a concept very similar to the notion of affine invariance in multivariate data depth, can be achieved for the case of lattice-valued data.

Contribution 4 develops a generalization of the univariate quantile concept for random variables/samples with values in a **complete lattice** $(V, \leq)$. The contribution analyzes, to which extent one can generalize the notion of a quantile and which intuitively desirable properties like *i*) representation invariance, *ii*) "non-arbitraryness", *iii*) generality, *iv*) uniqueness, *v*) "simplicity", *vi*) sharpness and *vii*) richness of the obtained quantile system can still be reached in this generalized context. It turns out that an asymmetric approach leads to an acceptable result: For a real-valued random variable, a lower quantile with level $\alpha \in (0, 1]$ can be characterized as that real number $z \in \mathbb{R}$ that is minimal in the set

$$M_\alpha := \{x \in \mathbb{R} \mid P(X \leq x) \geq \alpha\},$$

where P is the underlying image measure of the underlying random variable X or the empirical measure associated to the considered data sample, respectively. The above characterization straightforwardly generalizes to the case of random variables with values in a partially ordered set (poset). However, in the general situation of poset-valued random variables, the set M_α can contain more than one minimal element (or also no minimal element), such that the property *iv*) uniqueness (and also property *v*) "simplicity") comes into trouble. Because of reasons of symmetry, it is not always possible to reasonably choose in a non-arbitrary way one element from the set of minimal elements of M_α to reach uniqueness. If the underlying space is not only a poset but a complete lattice, then instead of choosing from the set M_α one can alternatively simply take the infimum $\bigwedge M_\alpha$ as a candidate for a quantile. This candidate has already very promising properties, it is unique, unambiguously defined and furthermore the elements $\bigwedge M_\alpha$ are always comparable for different levels α, a fact that contributes to the property *v*) "simplicity". Depending on the complexity of $(V, \leq)$, the system $Q := \{\bigwedge M_\alpha \mid \alpha \in [0, 1]\}$ of quantiles can be arbitrarily sparse. Additionally, property *vi*), sharpness, demanding that every quantile $q \in Q$ should be minimal in the set $\{x \in V \mid P(X \leq x) \geq P(X \leq q)\}$, is generally not satisfied. Fortunately, at least property vi) can be rescued by a further construction: Since the family $\mathcal{K} := \{q \in V \mid q \text{ minimal in } \{x \in V \mid P(X \leq x) \geq P(X \leq q)\}\}$ is a kernel system[22] (cf., [Kwuida and Schmidt, 2011]), one can construct the modified quantile system

$$Q^* := \{\bigvee\{z \in \mathcal{K}, z \leq q\} \mid q \in Q\}$$

that now satisfies also property vi). While the quantile system Q^* allows for a qualitative kind of data analysis (the elements of Q^* are actually elements of the underlying ordered set $(V, \leq)$), the mapping

$$\lambda : V \longrightarrow [0,1] : x \mapsto P(X \leq \bigwedge\{q \in Q^* \mid q \geq x\})$$

(called **level-function** in the contribution) can be used for a quantitative kind of data analysis. Here, λ can be understood as a generalization of the univariate distribution function $F : \mathbb{R} \longrightarrow \mathbb{R} : x \mapsto P(X \leq x)$. Opposed to the usual distribution function, the function λ still has a nice pivot type property in the general situation of random variables with values in an only partially ordered set: Under weak conditions one has $\forall \alpha \in \mathrm{Im}(X) : P(\lambda \circ X \leq \alpha) = \alpha$. In many situations, the function λ could be understood as a kind of outlyingness-function: If for example $(V, \leq)$ is the set of all convex sets of $\mathbb{R}^d$, ordered by set inclusion and if we treat points x in $\mathbb{R}^d$ as singletons $\{x\}$, then the function λ is a sort of outlyingness function that is dual to (a transformation of) Tukey's half-space depth ([Tukey, 1975]).

Contribution 5) G. Schollmeyer. Application of lower quantiles for complete lattices to ranking data: Analyzing outlyingness of preference orderings. Technical Report 208, Department of Statistics, LMU Munich, 2017b. URL `http://www.statistik.uni-muenchen.de/forschung/technical_reports/index.html`.

Short summary: In Contribution 5 we apply the lattice-valued quantile concept of Contribution 4 to the analysis of ranking data. We analyze a data set of the wisdom of the crowd phenomena for ranking data where 26 university students had to rank the former 44 US presidents according to their presumed order of presidency. The quantile concept allows for a notion of outlyingness of the supplied rankings. We analyze commonalities of selected outlying strata of rankings, ranging from analyzing only the most central rankings to the analysis of the whole set of rankings. We also indicate how one can use the quantile concept to construct an aggregation function in the context of Social Choice Theory. In a final step we additionally sketch how to utilize the quantile concept for developing tests of model fit for statistical models of ranking data like the Insertion Sorting Rank model (Biernacki and Jacques [2013]) or the Thurstonian model ([Thurstone, 1927], cf. also [Lee et al., 2014]).

The proposed quantile concept of Contribution 4 is applied to the analysis of ranking data in Contribution 5. Ranking data appear for example in Social Choice Theory where a set of n

[22] In this context, a kernel system is a subset of V that contains the least element of V and that is furthermore closed under arbitrary suprema.

voters rank a set $C = \{C_1, \dots, C_q\}$ of q alternatives (for example candidates running for an office). One can see this as dealing with n data points in the space of all binary relations on the candidate set C. The set of all binary relations on C, ordered by set inclusion, builds a complete lattice and thus we can fully exploit the quantile concept for samples with values in a complete lattice developed in Contribution 4. The order theoretic notion of outlyingness can then be used for example to cluster the rankings given by the n voters by increasing outlyingness strata. One can analyze for a given outlyingness threshold c the commonalities of the orderings with an outlyingness lower than c by looking for example at the intersection of all such orderings. In Contribution 5 we analyze a data set that is related to the wisdom of the crowd phenomena in the context of ranking data, see [Lee et al., 2014]. There, 26 students were asked to rank the former 44 US-presidents according to their presumed order of presidency.

Beyond the descriptive analysis of ranking data, the contribution also shortly sketches, how characteristics of the obtained commonalities of the strata can be used to analyze the appropriateness of statistical models for ranking data. Concretely, one can quantify for the obtained partial orderings, how close they are to interval orders[23]. To do so, one can simply count the number of counterexamples to the condition $x < y \;\&\; z < w \Longrightarrow x < w$ or $z < y$ that characterizes an interval order (cf., [Fishburn, 1970]). For the ranking data set of the 44 former US-presidents it seems that for example for a Thurstonian model (cf., [Thurstone, 1927]) or for the Insertion Sort Rank model (ISR, cf., [Biernacki and Jacques, 2013]), the actually obtained orderings are far more close to an interval order as one would expect under the assumed statistical models.

In the context of Social Choice Theory it is also possible to use the order theoretic outlyingness concept developed in Contribution 4 to construct a straightforward aggregation function for n-tuples of orderings on some space C of alternatives, called profiles in the context of Social Choice Theory: To aggregate the n orderings $R_1, \dots, R_n$ given by n voters about different alternatives (for example candidates in an election) into one "consensus" ordering that represents the groups preferences, one can simply choose that profile(s) of the group that is (are) the most central one(s) in the group to represent the groups preferences. (If there are more than one most central ordering, then one can either choose arbitrarily one of them or alternatively, one can apply a further aggregation rule to the most central orderings.) With this procedure, which is called **commonality sharing** and that is further analyzed in Contribution 6, one would get a kind of aggregation function that could be seen as a completely order theoretic generalization of the univariate median[24]. This generalization of the median actually differs from another generalization of the median to ranking data, namely the Kemeny rule (cf., [Kemeny, 1959]) that can be understood as another generalization of the univariate median in the sense that it is defined as the minimizer of a $\mathcal{L}^1$-norm.

[23] An interval order is a strict partial order $(V, <)$ satisfying $\forall x, y, z, w \in V: \; x < y \;\&\; z < w \Longrightarrow x < w$ or $z < y$. If V is countable, then $(V, <)$ is an interval order iff there exist real-valued functions l and u on V such that $\forall x \in V: u(x) > l(x)$ and $x < y \iff u(x) < l(y)$, see Fishburn [1970]. This is the reason for the name interval order.

[24] Note that in the simple univariate case, a median is exactly a value that is most central w.r.t. the data points or the underlying probability distribution.

Contribution 6) C. Jansen, G. Schollmeyer, T. Augustin. A probabilistic evaluation framework for preference aggregation reflecting group homogeneity. *Revised version under review in Mathematical Social Sciences*, original version available as a technical report under `https://epub.ub.uni-muenchen.de/29269/`, the revised version can be found under `http://gschollmeyer.userweb.mwn.de/prob_evaluation_revision.pdf`.

Short summary: In Contribution 6 a probabilistic framework for evaluating the quality of different aggregation functions of Social Choice Theory is developed. The analysis is done in dependence on a notion of group specific homogeneity, for which a minimal axiomatization, as well as a concrete homogeneity measure satisfying the axioms are proposed. In a simulation study, selected aggregation functions, including the commonality sharing rule based on Contributions 4 and 5 are analyzed.

The aggregation procedure described above, called "commonality sharing" in Contribution 5, is analyzed together with other classical aggregation functions known from Social Choice Theory in Contribution 6. The analysis is based on a statistical framework that incorporates information about the homogeneity of the group involved in the group decision. One supposes that the homogeneity of groups can vary very much from group to group. Answering the question of what is a good aggregation function could thus also be very dependent on the groups homogeneity. The contribution is therefore concerned firstly with formalizing a notion of homogeneity. In our contribution, we give a minimal set of conditions, a reasonable **measure of homogeneity** should satisfy (namely i) sensitivity for consensus, ii) anonymity and iii) monotonicity w.r.t. a switch of voters to the majority). We also give a concrete measure of homogeneity that satisfies these conditions and also discuss other measures known from Social Choice Theory (e.g., Kendall's and Smith's W that does not satisfy our requirements). For a statistical analysis of the performance of aggregation functions under different scenarios of homogeneity, actually the complete distribution of the randomly observed profiles for a group is needed. Since from an applicational point of view, because of the high dimensionality of the profile space, the specification of a reasonable underlying distribution seems to be very hard to achieve, we go here an indirect way: Instead of specifying the distribution of the profiles, one only specifies (for example data-based or expert-based) the distribution of the homogeneity, quantified by the homogeneity measure from above. Because of its one-dimensionality, the specification of the distribution of the homogeneity measure is far easier than the specification of the whole profile distribution. To statistically analyze the performance of different aggregation functions, in a second step one has to specify a **"quality measure"** that quantifies how well a consensus order supplied by an analyzed aggregation function represents the groups preferences.

Given the specified distribution of the homogeneity measure one can then analyze the distribution of the quality measure. Since different distributions of the profile can lead to the same distribution of the homogeneity measure, the distribution of the quality measure is not uniquely identified and there is a whole set of possible distributions of the quality measure compatible with the specification of the homogeneity measure. To analyze the quality measure one thus has at least two different possibilities: One can a) either analyze the distribution of

the quality measure under the whole set of all profile distributions that are compatible with the specified distribution of the homogeneity measure, or one can b) try to single out one specific distribution of the quality measure compatible with the specification of the distribution of the homogeneity measure, for example that distribution which maximizes entropy. While a) is thought of in the spirit of imprecise probabilities where one gets a whole set of distributions of the quality measure, from which one can compute for example upper and lower expectations, the approach b) can be understood as some objective Bayes type analysis, where one gets one precise distribution of the quality measure that can be used to compare the performance of different aggregation functions.
In a short simulation study, we analyzed six common aggregation functions including Kemeny's rule, as well as the commonality sharing procedure from Contribution 5. It turned out that, as expected, if one uses an $\mathcal{L}^1$-type quality measure, then Kemeny's rule performs better than the other aggregation rules because it actually is defined as the optimizer of this quality measure. Also for other quality measures used in the study the general picture does not change much. (But Kemeny's rule is not every time the "best", for very heterogeneous groups, a dictator-ship type aggregation function performs a little bit better w.r.t. the chosen quality measure.) Compared to the other rules, the commonality sharing rule behaves very similar to Kemeny's rule, which is not so surprising given that both Kemeny's rule, as well as the commonality sharing rule are different generalizations of the median to ranking data.

Remarks and perspectives Part C

As already noted, if in Contribution 4 one takes the lattice co $(\mathbb{R}^d)$ of all convex compact sets of $\mathbb{R}^d$ ordered by set inclusion and if one identifies data points $x \in \mathbb{R}^d$ with singletons $\{x\}$, then the level-function λ can be seen as an outlyingness-function that is dual to (a transformation of) Tukey's half-space depth. However, since we are now thinking in the space co $(\mathbb{R}^d)$, we can also deal with imprecisely observed data, namely with convex data. Thus, the method is slightly more general than Tukey's half-space depth, here. In particular, one can also use it (and this was actually the motivation for the development of the lattice-valued quantile concept, compare the remarks and perspectives part A) to get a descriptive method for analyzing the distribution of set-valued estimators (for example the estimators for the identification regions of Contributions 1 and 2) by firstly drawing B bootstrap samples that lead to B bootstrapped set-valued estimates from which one can then compute a $\gamma \cdot 100\%$ lower quantile to get an insight into the spread of the set-valued estimator. Of course, there are far more non-conventional applications of the quantile concept thinkable, since one only needs the structure of a complete lattice: For example in the context of symbolic data analysis (cf., e.g., [Bock and Diday, 2012]), if one is interested for instance in the constitution of the households in a country w.r.t. age, then one has to deal with distribution-function-valued data. Since the space of all distribution functions with the ordering $F \leq G \iff \forall x \in \mathbb{R} : \; F(x) \leq G(x)$ builds a lattice,[25] one can use the quantile approach to get a rough descriptive insight into the distribution of the age distributions in the households.

Furthermore, the quantile concept can also be applied in the context of formal concept analysis (see, e.g., [Ganter and Wille, 2012]): In **formal concept analysis** one starts with a so-called **formal context** (G, M, I), where G is a set of objects, M is a set of attributes and

[25]Generally, this lattice is not complete, but this is only a technical subtlety that will not lead to any problem, here.

$I \subseteq G \times M$ is a binary relation between G and M with the interpretation $(g, m) \in I \iff$ object g has attribute m. In the context of data analysis, G is typically the set of observed data points and M models the multivariate attributes of the data points. For example in a social survey, $G = \{g_1, \ldots, g_m\}$ is a set of m respondents that were given a set $M = \{m_1, \ldots, m_n\}$ of n statements and gIm then means that respondent g agrees with the statement m. In formal concept analysis one then builds up a hierarchy of so-called **formal concepts** that are pairs (A, B) where A is a set of objects and B is a set of attributes with the property that

1. Every object $g \in A$ has every attribute $m \in B$ (i.e.: $\forall g \in A \forall m \in B : gIm$).
2. There is no further object $g \in G \backslash A$ that has also all attributes of B (i.e.: $\forall g \in G : (\forall m \in B : gIm) \Longrightarrow g \in A$).
3. There is no further attribute $m \in M \backslash B$ that is also shared by all objects $g \in A$ (i.e.: $\forall m \in M : (\forall g \in A : gIm) \Longrightarrow m \in B$).

A formal concept can be understood as the mathematical formalization of a concept that says which objects belong to the concept and which attributes characterize the concept. The set A of a formal concept (A, B) is called **extent** and the set B is called **intent** of the concept. The set of all formal concepts is naturally equipped with the **subconcept relation**[26]

$$(A, B) \leq (C, D) \iff A \subseteq C \,\& B \supseteq D,$$

which means that the formal concept (A, B) is a subconcept of the concept (C, D) (i.e.: $(A, B) \leq (C, D)$) iff it is more specific in the sense that, compared to the concept (C, D), less objects belong to the concept (A, B) which share more common attributes. The basic theorem on concept lattices ([Ganter and Wille, 2012, p. 20]) then states that the set of all formal concepts equipped with the subconcept relation builds a complete lattice. This complete lattice, called concept lattice, can then be used together with the quantile approach to get a notion of outlyingness of formal concepts, which would translate in our example to an outlyingness for response patterns the respondents gave to the different statements of M. (To every respondent one can naturally relate the most specific formal concept that has this respondent in its extent and the set of all statements the respondent agreed to in its intent.) This notion of outlyingness in some sense captures the extremeness of opinions with respect to the analyzed population of respondents. Note also that the basic theorem on concept lattices furthermore states that every complete lattice can be represented as a lattice of formal concepts. Thus, if we think in terms of concept lattices, we do not lose any of the generality of the lattice-valued quantile approach. Actually, also the generalization of Tukey's half-space depth from above can be thought in this language, and can also be implemented in terms of formal concept analysis by looking at an appropriate formal context (G, M, I). To do so, one has to take G as the set of all points in $\mathbb{R}^d$ that where present in the data set, either as precise point valued observations or as convex-set-valued observations. The set M has to be taken as the set of all half-spaces spanned by d points that are either precisely observed points or the extreme points of observed convex sets. Then, one has to define $(g, m) \in I \iff$ point g (or convex set g) lies (completely) in the half-space m. Finally, one can compute the quantiles in the corresponding concept lattice. (To do so, one does not have to compute the whole concept lattice.) Of course, for data in $\mathbb{R}^2$ one could do the computation far more efficiently by adopting

[26] Actually, for formal concepts (A, B) and (C, D) one automatically has $A \subseteq C \iff B \supseteq D$. Thus, one could equivalently define $(A, B) \leq (C, D) \iff A \subseteq C$ or $(A, B) \leq (C, D) \iff B \supseteq D$.

the $\mathcal{O}(n \log n)$-algorithm given in Rousseeuw and Ruts [1996] to the case of convex-set-valued data.

Part D: Relational data analysis: Stochastic dominance for partially ordered sets

Stochastic (first order) dominance is a partial ordering between random variables that captures a notion of one variable being stochastically clearly "higher" in outcome than another random variable. It has many applications in a variety of disciplines like for example welfare economics (cf., e.g., [Arndt et al., 2012, 2015]), decision theory (cf., e.g., [Levy, 2015]), portfolio analysis (cf., e.g., [Kuosmanen, 2004]), nonparametric item response theory (cf., e.g., [Scheiblechner, 2007]), medicine (cf., e.g., [Leshno and Levy, 2004]), toxicology (cf., e.g., [Davidov and Peddada, 2013]) or psychology (cf., e.g., [Levy and Levy, 2002]) to name just a few. For real-valued random variables, stochastic dominance principles, which in its roots can be traced back to the work of J. Bernoulli ([Bernoulli, 1713], cf. [Schneeweiß, 1967, p. 38]) are a broadly researched topic with well developed theory and many applications (e.g., already the bibliography [Bawa, 1982] contains more than 400 references, cf., also [Mosler and Scarsini, 1993] or [Levy, 1992]). For multivariate first order stochastic dominance, though theoretically also well established (cf., e.g., [Lehmann, 1955, Levhari et al., 1975, Kamae et al., 1977] for characterizations of stochastic dominance), empirical applications of multivariate stochastic dominance like for example Arndt et al. [2015] seem to be rare, which is maybe due to the computational hardness of checking dominance in the multivariate situation.

G. Schollmeyer, C. Jansen, and T. Augustin. Detecting stochastic dominance for poset-valued random variables as an example of linear programming on closure systems. Technical Report 209, Department of Statistics, LMU Munich, 2017. URL `http://www.statistik.uni-muenchen.de/forschung/technical_reports/index.html`.

Short summary: Contribution 7 gives a method for detecting first order stochastic dominance for random variables with values in a partially ordered set (poset). The method makes substantial use of linear programming and is furthermore generalized to linear optimization on closure systems. Beyond checking for stochastic dominance, also proposals for statistical inference and regularization based on Vapnik-Chervonenkis theory are given. The generality of the developed methods are illustrated on data examples in the context of multivariate inequality analysis, item impact and differential item functioning in item response theory, analyzing distributional differences in spatial statistics and guided regularization in the context of cognitive diagnosis models.

In Contribution 7, **first order stochastic dominance** for random variables $X, Y : (\Omega, \mathcal{F}) \longrightarrow (V, \mathcal{F}', \leq)$ with values in a **partially ordered set** (poset)[27] $(V, \leq)$ is analyzed. There, the focus is not on analytically given random variables, instead, a method for detecting stochastic dominance for random variables that are given by empirical samples and their associated empirical laws is established. (Of course, statistical inference is also treated in the contribution.)

[27] This includes especially the multivariate case of $\mathbb{R}^d$ where the natural order $x \leq y \iff \forall i \in \{1, \dots, d\} : x_i \leq y_i$ is used. Note also that every finite poset $(V, \leq)$ can mathematically be represented as a multivariate case where the dimension equals the order dimension of $(V, \leq)$, cf. [Dushnik and Miller, 1941, Trotter, 2001].

Since for poset-valued random variables the values are generally not totally ordered, first order stochastic dominance is not simply characterized by the distribution function, anymore. One elegant characterization in the poset-valued situation is the upset-characterization

$$X \leq_{SD} Y \iff \forall A \in \mathcal{U}((V,\leq)) : \hat{P}(X \in A) \leq \hat{P}(Y \in A), \tag{1}$$

where $\mathcal{U}((V,\leq))$ is the set of all upsets of $(V,\leq)$. A set $A \subseteq V$ is called an **upset** if for arbitrary $a,b \in V$ we have $a \in A \quad \& \quad b \geq a \Longrightarrow b \in A$. Intuitively, the notion of an upset A can be understood as formalizing a reasonable concretion of the notion of "big values" by saying for an upset A that a value x is *big* if it is in A and stating that x is *not big* if it is not in A: Every such set A formalizing a reasonable concretion of the term *big* should satisfy the property that if a value $a \in V$ is termed *big*, and if $b \geq a$, then also b should be termed *big*. The statement of stochastic dominance $X \leq_{SD} Y$ then could be verbalized as "The probability of X taking big values is smaller than or equal to the probability of Y taking big values, independent of the concretion of the term *big*. Since the family of all upsets is usually very large, explicitly checking (1) is practically not possible and thus, the upset characterization is commonly not used for checking stochastic dominance.

A linear programming approach for detecting stochastic dominance

In the contribution we still use the upset characterization[28] for detecting stochastic dominance. The reason is that in the field of order and lattice theory, the family of upsets is actually a well understood closure system[29]. It turns out that checking first order stochastic dominance in fact can efficiently be done by solving a linear program of the form

$$\max_{A \text{ upset}} \hat{P}(X \in A) - \hat{P}(Y \in A). \tag{2}$$

and checking, if the maximal value of (2) is lower than or equal to zero.[30] Here, one needs not to explicitly look at every upset A, instead, upsets A can efficiently be represented by its indicator functions. The condition of being an upset can be nicely incorporated by inequality constraints since the indicator functions of upsets are exactly the isotone $\{0,1\}$-valued functions. The main simplifying fact is here that the integrality constraints (i.e. the demand that the decision variables are integer, here concretely, binary) on the indicator functions can actually be dropped by only demanding that all involved decision variables are in the real-valued interval $[0,1]$. So, here one has to solve not a binary, but only a linear program. Concretely, one has to solve a linear program with $|V|$ decision variables and $|\lessdot| + 2|V|$ inequality-constraints, where $\lessdot$ is the covering relation associated to $\leq$ and where for the poset V it is enough to look at that values of V that were actually observed in the sample. This means that the problem is of course manageable in realistic situations of application. In the contribution, we also analyze a dataset in the context of inequality analysis where $n = 1515$, $|V| = 922$ and $|\lessdot| = 3122$.

[28] Other techniques of checking dominance can be found e.g. in [Mosler and Scarsini, 1991, Østerdal, 2010] where a characterization of stochastic dominance in terms of a mass transportation problem is utilized.

[29] A closure system $\mathcal{S}$ on a space Ω is a family $S \subseteq 2^{\Omega}$ of subsets of Ω that contains Ω itself and is closed under arbitrary intersections.

[30] If the actually observed maximal value of (2) is zero, then we know $X \leq_{SD} Y$, but beyond this, the minimal value of (2), i.e., the optimal value of (2) obtained by replacing maximization with minimization, gives furthermore some sort of a measure of the extent of dominance which additionally can be used for statistical testing. This is an advantage compared to dominance checking approaches based on a mass transportation characterization, cf., Sections 3.3 and 3.2.2 of the contribution.

A generalization: Linear programming on closure systems

The reason for the ease of translating the problem of detecting stochastic dominance into a linear program, especially the fact that one can drop the integrality constraints, lies mainly in the order theoretic fact that the family of all upsets is not only a closure system, it is furthermore also closed under arbitrary unions. A natural question is then: "Are similar techniques also applicable for closure systems that are not closed under arbitrary unions?" It turns out that also general closure systems can be described by inequality constraints and thus the optimization of a linear function on a general closure system can still be done via binary programming, but with the main difference that the integrality constraints of the involved binary programs generally cannot be dropped or can only partially be dropped. To illustrate the usefulness of being able to solve a binary program on a closure system, we briefly sketch two situations of statistical data analysis (discussed in Section 4.1 and 4.2 and applied in Section 6.4 and 6.2 of the contribution) where closure systems somehow naturally arise:

Exemplary closure systems relevant for statistical applications

Spatial statistics: The closure system of all convex sets of $\mathbb{R}^2$ (cf., Sections 4.1 and 6.4 of the contribution) could be interesting in the context of spatial statistics. For example in ecology, one would possibly be interested in differences in the spatial distribution of different species, for example female and male pacific cods in the Bering Sea, analyzed in Syrjala [1996]. For the statistical analysis one can use for example generalizations of the Kolmogorov-Smirnov test or the Cramér von Mises test as discussed in Syrjala [1996]. Another way of statistical analysis would be to generalize[31] the Kolmogorov-Smirnov test slightly differently by looking at every convex area of the Bering Sea, computing the difference in the proportions of male and female cods and finally analyzing the maximal and minimal difference over all convex sets. This would be exactly solving a binary program on the closure system of all convex sets[32] of $\mathbb{R}^2$. While the maximal difference of proportions of female and male cods provides a quantity that can be used for a statistical test, the convex set where the maximimum/minimum of the difference in proportions is actually attained can give further descriptively interesting insights into the data situation. To solve the binary program one has to efficiently describe the underlying closure system. A direct description by explicitly enumerating all convex sets is very hard due to the high number of convex sets: For n points $(x_1, \ldots, x_n)$ of $\mathbb{R}^2$ the closure system $H := \{A \cap \{x_1, \ldots, x_n\} \mid A \subseteq \mathbb{R}^2, A \text{ convex } \}$ can have 2^n sets in the worst case. However, it is possible to describe a closure system by the help of so-called formal implications[33] known from formal concept analysis: A **formal implication** is a pair (A, B) of subsets of the underlying space (in our case $\{x_1, \ldots, x_n\}$), usually denoted by $A \longrightarrow B$, with

[31] Compared to other generalizations, like, e.g., Peacock [1983], Fasano and Franceschini [1987] or Justel et al. [1997], the generalization described here is completely independent of the choice of the used coordinate system. Of course, there are other, computationally much more manageable consistent and affine invariant possibilities of generalization, e.g., by applying consistent, affine invariant data depth functions, like, e.g., Tukey's half-space depth. (concerning consistency, cf., [Koshevoy, 2002, Hassairi and Regaieg, 2007, Cuesta-Albertos and Nieto-Reyes, 2008]).

[32] Of course, the closure system of all convex sets is infinite, but it is actually enough to look at the finite set of points where one actually did observe cods.

[33] Formal implications also play an important role in related fields like e.g. in mining association rules ([Agrawal et al., 1993, Piatetsky-Shapiro, 1991, Lakhal and Stumme, 2005]). There, some sort of stochastic generalizations of formal implications, called association rules, are analyzed.

the interpretation as the statement "All sets of a family of sets that contain all elements of A always also contain all elements of B. A closure system can then be described by the set $\mathfrak{I}$ of all formal implications that are actually true for the closure system. To describe a closure system by formal implications it is actually not necessary to enumerate all valid implications, but only a so-called **implication base**, which is a minimal subset of the set of all valid implications that generates the set of all valid implications in the sense that every closure system in which all implications of the base are valid is such that also all other implications from $\mathfrak{I}$ are valid. The computation of an implication base is generally computationally expensive (cf., [Kuznetsov, 2004, Ryssel et al., 2014]), but for the special case of convex sets, because of Carathéodory's theorem about convex hulls[34], one can explicitly give an implication base as

$$\{\{a,b,c\} \longrightarrow \mathrm{co}\,(\{a,b,c\}) \mid a,b,c \in \{x_1,\ldots,x_n\}, \mathrm{co}\,(\{a,b,c\}) \supsetneq \{a,b,c\}\},$$

where $\mathrm{co}\,(\{a,b,c\})$ denotes the intersection of the convex hull of the set $\{a,b,c\}$ and the set $\{x_1,\ldots,x_n\}$. The demand of the validity of the formal implications can be incorporated into the binary program via inequality constraints. All in all, this means that one can solve the optimization problem by solving a binary program with n decision variables and $\mathcal{O}(n^3)$ inequality constraints. Unfortunately, generally none of the integrality constraints can be dropped, so the binary program would be very hard to solve already for medium sizes of n.

Item response theory: A further natural example of a closure system can be found in dichotomous item response theory[35] (IRT): Let M be a set of persons and let G be a set of items in an IRT data set. Let furthermore be $I \subseteq G \times M$ the relation that describes the responses of the persons to the items via $(g,m) \in I \iff$ person g solved item m. Then, a natural closure system to consider is the closure system of all formal concept extents of the formal context (G,M,I), i.e., the system

$$H := \{S_Y := \{g \in G \mid \forall m \in Y : (g,m) \in I\} \mid Y \subseteq M\}.$$

of sets S_Y of that persons that solved at least all items from a given set Y. This closure system is often also very large[36]. However, it appears that again it can efficiently be described with methods of formal concept analysis. If one is interested in the analysis of item impact or differential item functioning (DIF)[37] for example with respect to gender, one can analyze in a completely non-parametric manner differences in the proportions of female and male persons who solved at least all items from a set Y of items and then look at that set Y for which this difference is maximal/minimal. This is again a binary programming problem on a closure system that can also be seen as a generalization of the classical Kolmogorov-Smirnov test to a very high dimensional multivariate situation. The example given in Section 6.2 of the contribution consists of a set of 45 items which corresponds to a Kolmogorov-Smirnov type test in the 45-dimensional space $\{0,1\}^{45}$ which can be solved in an acceptable amount of time,

[34] Carathéodory's theorem states that if a point $x \in \mathbb{R}^d$ lies in the convex hull of a set P of points, then there exists a subset $Q \subseteq P$ of at most $d+1$ points such that x lies also in the convex hull of Q.

[35] In item response theory one is concerned with a sound way of inferring the values of presumed underlying latent traits like "intelligence" from manifest observed data like the answers of respondents to questions of a test battery. For an introduction to item response theory, see, e.g., [Baker, 2001].

[36] For example the data set analyzed in the contribution has about 8.900.000.000 elements.

[37] One speaks of item impact of an item m, if for different groups of persons the probability of solving this item differs. In a similar way, an item m shows differential item functioning, if the probability of solving the item is different for different groups of persons that are comparable in the sense that they have the same underlying abilities. For more details on differential item functioning, see, e.g., [Osterlind and Everson, 2009].

which is due to the fact that in this situation, for the binary program, fortunately, much of the integrality constraints can be dropped, see Section 4.2 of the contribution. Furthermore, the "effective dimension" is actually not 45 here, which makes the computational, and also the statistical complexity smaller. In terms of Vapnik-Cherovonenkis theory (see below), the V.C.-dimension is only 22, here. If one is not only interested in differences between response probabilities, but also in differential item functioning, which is present if not the unconditional response probabilities, but the response probabilities conditional on the person abilities differ, one can make use of reweighting techniques. In a first step one can reweight the subsamples of female and male persons in such a way that the abilities (measured by item scores) are (approximately) identically distributed. In a second step, one can analyze the joint distribution of the item scores and the response patterns. To do so, one can effectively use the method of conceptual scaling known from formal concept analysis. (For a more detailed explanation, see Section 2.2. and Section 6.2 of the contribution.) Since one uses the item scores for accounting for the different distributions of the latent abilities in the male and the female group, the resulting test for differential item functioning can be seen as some multivariate version of the Mantel Haenszel procedure (see [Holland et al., 1988], cf., also, e.g., Strobl et al. [2015], Tutz and Schauberger [2015], Tutz and Berger [2016] for other more modern tests for differential item functioning.)

Inference: permutation tests, conservative distribution free Vapnik-Chervonenkis bounds and regularization

The second part of the contribution is concerned with the analysis of the distribution of the maximal value of the linear program (2) (or the corresponding binary programs for the general case of linear optimization on a closure system) that can be also used as a reasonable test statistic for a statistical test. For stochastic dominance in the univariate case, the maximal value of (2) is exactly the test statistic D^+ of the two-sample Kolmogorov-Smirnov test. In contrast to this situation where the Kolmogorov-Smirnov test statistic is distribution free for the class of continuously distributed random variables, the situation changes for multivariate generalizations. Generally, a concise analysis of the distribution of the maximal value of the program (2) seems to be difficult here, because the distribution of the test statistic is dependent on both the exact structure of the involved closure system (which is usually only implicitly given) and the interplay between the closure system and the underlying true law P.[38]value a Thus, an alternative would be here to rely on permutation tests to do inference. However, it is still possible to get analytical conservative bounds for inference based on concepts of Vapnik-Chervonenkis theory[39], see Section 5.2 of the contribution. Beyond giving conservative bounds, Vapnik-Chervonenkis theory also allows for a kind of "regularization" of the test statistic by reducing the Vapnik-Chervonenkis dimension of the underlying closure system by appropriately reducing the size of the closure system. The possibility of regularization is discussed in Section 5.3 of the contribution and illustrated exemplarily for two examples (namely in the context of inequality analysis based on stochastic dominance (Section 5.3.1) and in the

[38] There actually exists some literature about the asymptotic distribution of the optimal value of a random linear program (e.g., [Babbar, 1955, Sengupta et al., 1963, Prèkopa, 1966]) but this literature seemingly does not apply here, because in our case, under the null hypothesis, the random objective function is symmetrically distributed around the zero vector such that the assumption of a unique optimal basis for the asymptotic linear program (cf. [Prèkopa, 1966, Theorem 5]) is not satisfied.

[39] For an introduction to Vapnik-Chervonenkis theory, see, e.g., [Devroye et al., 2013].

context of cognitive diagnosis modeling that is related to nonparametric item response theory (especially knowledge space theory) and also to formal concept analysis (Section 5.3.2)). For the application of regularization techniques (called **"taming"** in the contribution) one needs to compute the Vapnik-Chervonenkis dimension of the underlying closure system. Thus, characterizing and computing the Vapnik-Chervonenkis dimension of exemplary closure systems, which is also of interest of its own, is also treated in Sections 5.2.1 and 5.2.2 of the contribution. Actually, there is a close connection between the description of closure systems via formal implications and the notion of shatterable sets from Vapnik-Chervonenkis theory that is genuinely involved in the definition of the Vapnik-Chervonenkis dimension of a family of sets, see Section 5.2.1 of the contribution. Surprisingly, there seems to be not much research that directly connects formal concept analysis and Vapnik-Chervonenkis theory, the only works in this direction, the authors of the contribution are aware of, are the papers Anthony et al. [1990a,b], Chornomaz [2015], Albano and Chornomaz [2017], Albano [2017a,b], Makhalova and Kuznetsov [2017].

Applications of relational data analysis

In Section 6 of the contribution we also apply the developed techniques of relational data analysis to different data examples and discuss the flexibility of relational data analysis to deal with deficient or non-standard data. Section 6.1 is concerned with detecting first order stochastic dominance in the context of multivariate inequality analysis. We use the German General Social Survey (ALLBUS, [GESIS - Leibniz - Institut fur Sozialwissenschaften, 2015]) and analyze differences between the male and female respondents with respect to three dimensions of multivariate inequality analysis[40]: Income, education and health. All three dimensions have specific data aspects: For example income is not observed precisely, as is often the case. For those respondents who answered the direct question about income one has precise values. For those who did not answer the direct question, a categorized follow up question about the income given in intervals $[a, b)$ was additionally asked. Thus, one has some precise values, some interval values and some missing values for respondents who completely refused to answer the income questions. The relational method for checking dominance can easily deal with such kind of data[41] because one only has to specify, when a reported income is higher than another. This could be done for example by treating all incomes as intervals[42] and say for example that an interval $[a, b)$ is greater than an interval $[c, d)$ if $d \leq a$. Additionally, one has the flexibility to impose further modeling assumptions. For example, one could assume that the coarsening behavior is unknown but identical in both analyzed subgroups. Then, for analyzing stochastic dominance one can furthermore conclude that reported intervals $[a, b)$ and $[c, d)$ with identical end-points (i.e. $a = c$ & $b = d$) are comparable (i.e. $[a, b) \leq [c, d)$ & $[a, b) \geq [c, d)$). Of course, it is also possible to replace the intervals by their mid-points in an ad hoc manner, but this is not so much in the spirit of the proposed relational method of data analysis. Note further

[40] If one is interested for example in poverty in the sense of Sens capability approach (cf., [Sen, 1999]), then of course one would have to analyze far more "dimensions of poverty", in the contribution we only exemplarily analyze dimensions of individual capability, cf., also [Arndt et al., 2006] for a feasibility study of the operationalization of the capability approach for the Socio-Economic Panel (SOEP).

[41] Of course, the completely missing values are problematic because they are incomparable to any other reported income. However, this is actually adequate because it reflects the fact that from a missing value, without imposing some modeling assumptions, one has of course really no information about the income. The only way to deal with this issue without statistical modeling seems to be here to shift the focus from the population of actual interest to the population of the responders and to stay silent about the non-responders.

[42] A precise income of c Euro can be treated as the interval $[c, c]$.

that also more sophisticated ideas of statistical modeling are applicable, here: If one is able to model the distribution of the true values within the intervals[43], one can (at least in principle) incorporate this distribution in the analysis. This can also help in treating item non-responders (i.e. persons that did only answer some of the questions about the three dimensions). For the concrete computation one may have to discretize this distribution if it is modeled continuously in the first place. By incorporating different thinkable coarsening pattern scenarios one could get a stylized form of a sensitivity analysis.

The dimension education of the ALLBUS survey is a totally ordered 9-level scale implementing the International Standard Classification of Education (ISCED) 2011 (see [UNESCO Institute for Statistics (UIS), 2012]) for Germany. However, the ISCED 2011 also allows for other forms of implementation that would yield only a partially ordered measurement of scale. For example, for two different highest educational achievements one can use expert-based value judgments about which of them is "higher" w.r.t. multivariate poverty or if they are incomparable. Another way would be to say that one bases the dimension education on the whole educational path and say that a person A is more poor than another person B w.r.t. the dimension *Education* only if both persons followed the same educational path but person A stopped earlier with a lower highest educational achievement than person B. This partial ordering would lead to a less decisive analysis, but it has the potential to reveal, how much a more classical analysis would dependent on the choice of a totally ordered scale for the dimension education.

Finally, the dimension health is the reported answer to the question about how the respondents would describe their health status in general. For the answer, both a 5-category and a 6-category response scale were used in a split ballot design to test for a possible impact of different response scales. The method of stochastic dominance can deal with such issues of split designs in different ways: Firstly one can do a separate analysis for every split, which would be especially appropriate if systematic differences between the response scales are present. Results of the analysis that are valid in both splits can then be seen as valid, independent of the choice of the scale. Secondly, if both response scales adequately operationalize the same construct, one can analyze all splits jointly by matching both scales to each other based on their respective empirical distribution functions. This is possible because one knows that because the splitting was randomly assigned, the measured construct has the same distribution in every split.[44]

The analysis of stochastic dominance in our data example shows a nearly strict dominance: The female subpopulation is nearly strictly dominated by the male subpopulation w.r.t. the joint analysis of the three dimensions income health and education and the results are highly statistically significant. This illustrates that "weak forms" of relational data analysis are still able to provide non-trivial substance matter conclusions that are very reliable due to the "weak conceptualizations" of involved terms like that of inequality.

Beyond stochastic dominance, also for general linear programming on closure systems we

[43]For example one could assume that the distribution of the unobserved values within the intervals is independent of the presence of coarsening.

[44]Of course, due to possibly different measurement error in different splits, the actual measurements can differ in their distribution. But if the measurement errors are independent of the construct and from each other, this will only produce a "smearing" of the measurements, which can lead to cases where stochastic dominance w.r.t. the actual construct is actually present, but it is not present anymore for the measurements. A transition of non-stochastic dominance w.r.t. the construct into stochastic dominance w.r.t. the measurements can fortunately not happen, which shows the power of relational modeling.

give application examples in the contribution: In Section 6.2 we use an IRT data set[45] and apply a high-dimensional Kolmogorov-Smirnov type test for joint item impact and differential item functioning w.r.t. gender. It turns out that also for such a nonparametric procedure that takes the whole joint distribution of all 45 item answers into account, one has still enough power to detect systematic differences between male and female respondents. The analysis of the involved statistical complexity shows that the Vapnik-Chervonenkis dimension of the involved closure system is 22, which is less then half of the "formal" dimension of 45, indicating that data sets are sometimes much more structured than one may think. This, together with the possibility of regularization that also provides theoretical bounds in terms of Vapnik-Chervonenkis theory, could help in making the "curse of dimensionality" (Bellman, 1957) more into a "blessing" (Breiman, 2001). Additionally, in Section 5.3.2 of the contribution, in the context of **cognitive diagnosis models** (see, e.g., [Rupp et al., 2010]), we discuss, how one can incorporate external knowledge about the underlying closure system to efficiently regularize it in a more guided manner. The discussed possibility of **guided regularization** is illustrated in Section 6.3 with a subsample from the Trends in International Mathematics and Science Study 2007 (TIMSS, [Mullis et al., 2005]). Finally, we also illustrate the usage of the closure system of convex sets of $\mathbb{R}^2$ in **spatial statistics**. For a subsample of a data set of dead and alive oak trees in a secondary wood in Urkiola Natural Park analyzed in Laskurain [2008], we test for differences in the spatial distribution between dead and alive oak trees.

Contribution 8) G. Schollmeyer, C. Jansen, and T. Augustin. A simple descriptive method for multidimensional item response theory based on stochastic dominance. Technical Report 210, Department of Statistics, LMU Munich, 2017b. URL `http://www.statistik.uni-muenchen.de/forschung/technical_reports/index.html`.

> **Short summary:** Contribution 8 uses concepts of (multivariate first order) stochastic dominance to establish a descriptive approach for a relational analysis of person ability and item difficulty in the context of item response theory. It is shown that the proposed descriptive approach avoids a paradox of multidimensional item response theory that was firstly described in Hooker et al. [2009].

Contribution 8 develops a simple descriptive method for data analysis in the context of item response theory. The applied technique consists in a more or less purely relational data analysis that relies on ideas of multivariate first order stochastic dominance. A useful feature of stochastic dominance in this situation is that *i*) the data need only be of a partially ordered scale of measurement and *ii*) stochastic dominance provides the flexibility of suspending judgment by neither saying "random variable X is stochastically greater than random variable Y" nor saying "Y is stochastically greater than X". In item response theory one is interested in the measurement of latent traits, for example the latent ability of persons or the latent difficulty of a test item in a psychometric test. In a pre-philosophical manner, one can argue that the notion "ability" always actually already asks for being complemented to an

[45] The data set is a subsample from the general knowledge quiz Studentenpisa conducted online by the German weekly news magazine SPIEGEL (see [SPIEGEL Online, 2009]).

"ability wherefore?" and thus, the concept *ability* seems to be a genuinely relational concept that could appropriately be analyzed with a relational underpinning. Despite of this aspect, classical models of item response theory do model the "ability" of persons and the "difficulty" of items per se, either via real-valued parameters in the unidimensional Rasch model or with real-valued vectors in multidimensional generalizations[46]. While the classical unidimensional Rasch model is to some extent guided by measurement[47] theoretic considerations[48], for non-Rasch type multidimensional IRT models, measurement theoretic argumentations seem to be not very widely used in classical psychometric literature. However one judges the usefulness of questions concerning the need of measurement theoretic scrutinizations, in the context of multidimensional item response theory, within many multidimensional models (including multidimensional Rasch-type models, see [Jordan, 2013]), one is confronted with a paradox[49] firstly described in Hooker et al. [2009]: If multidimensional IRT models are used in entrance tests, it can be advantageous for applicants to deliberately answer some questions wrongly: It can happen that by applying for example a multiple-hurdle rule ([Segall, 2010]) after estimating multidimensional abilities, that a person is not approved though she solved all items another person that was approved solved, plus some more. In measurement theoretic terms one can see this as a violation of the homomorphism property w.r.t. the relation

$$\text{“Person } A \text{ is at least as able as person } B \text{ , if she has solved every item solved by person } B\text{.”} \qquad (R)$$

In contrast to many other thinkable meaningful relations concerning the presumed latent structure, this relation can be clearly located within the empirical relational structure. To deal with this seemingly paradoxical situation, one can follow different routes, for example, in Hooker et al. [2009], one proposal was to do a constrained estimation of the ability parameters that respects the empirical relation (R) from above.

Instead of "repairing" classical multidimensional IRT models, Contribution 8 goes a purely descriptive, relational way: By utilizing ideas of stochastic dominance, we establish empirically supported partial orderings on the set of items and on the set of persons that model the difficulties of items and the abilities of persons in a relational manner: If the difficulties of different items in a psychometric test are known, then one can naturally define a person A to be "more able" than a person B if for every item that was solved by person B there bijectively exists another item solved by Person A which is as least as difficult as the item solved by Person B. In a dual manner, if the abilities of the persons are given, then one can introduce a difficulty-relation between items. Since beforehand one neither knows the difficulties of items nor the abilities of persons, one has to simultaneously analyze pairs of ability- and difficulty-relations that support each other. It turns out that from the hierarchy of such empirically mutually supportive pairs builds one can easily compute the smallest and the greatest element: In particular, the smallest element (i.e., the weakest pair of relations) is related to so-called simple formal implications[50] from formal concept analysis and the greatest element provides some stronger stochastic type ordering of abilities and difficulties that can

[46] For a fundamental scrutinization of the quantitative nature of latent traits, see [Michell, 2008].

[47] The term measurement will be used in the sequel in the sense of representational measurement theory.

[48] If one is willing to understand the notion of the empirical relational structure in such a broad sense that one is prepared to relate probabilities to the empirical relational structure, then with a great deal of goodwill, one can see the Rasch model as an instantiation of an *additive conjoint measurement* situation, see [Perline et al., 1979], for a contrary argumentation, see, e.g., Kyngdon [2008] or Heene [2013].

[49] Another point of view would actually be to see this paradox not as a paradox at all and to say that everything is in best order.

[50] A formal implication $A \longrightarrow B$ is called simple if A is a singleton.

be understood as a type of a stochastic generalization of the simple formal implications of formal concept analysis. All empirically mutually supportive pairs respect the relation (R) and thus avoid the paradox. The developed descriptive method has some conceptual similarities to descriptive approaches of item tree analysis (ITA, see [Van Leeuwe, 1974]) that are applied for example in the context of knowledge space theory (cf., Doignon and Falmagne [2012]), and which can be understood as another variant of a stochastic generalization of notions of formal concept analysis. (Actually, there is a neat relation between formal concept analysis and knowledge space theory, see [Rusch and Wille, 1996]). The Basic Local Independence Model (BLIM) from knowledge space theory is thus also analyzed in the contribution. It shows up that for quasi-ordinal knowledge spaces and maximum likelihood estimation under fixed careless-error and lucky-guess-probabilities, the paradox is also avoided.

Remarks and perspectives Part D

Contribution 8 made substantial use of the notion of stochastic dominance. Thus, for the computation of empirically mutually supportive pairs the linear programming approach of Contribution 7 can be used. To compute the ability relation between n persons one then has to solve n^2 linear programs, which can be very hard if n is very high. However, one can make use of the structural properties of the ability relation. For example, if person A has solved more items than person B then one already knows that person B cannot be more able than person A. (The ability relation is a homomorphic image of the item score relation.) Furthermore, since the ability relation is a preorder and thus transitive, if for example person A is more able than person B and person B is more able than person C, then one already knows that person A is also more able then person C and one does not have to check dominance for the pair (A, C) explicitly. Also, if person A is more able than person B but not more able than person C, then person B cannot be more able than person C. The transitivity of the ability relation and the difficulty relation is naturally given due to the construction of these relations. This is a main difference to descriptive techniques used in item tree analyis (ITA) like the classical ITA algorithm or the inductive ITA algorithm that are sometimes used in knowledge space theory (cf., e.g., [Schrepp, 1999, 2002]). There, instead of stochastic dominance, ideas of statistical preference (cf., [De Schuymer et al., 2003b,a]) are used and one says that an item A is more difficult than an item B if up to some amount of exceptions all persons who solved item A did also solve item B. This relation, because of the involved exceptions, is not transitive and one therefore has to somehow “repair” the intransitivity (see the algorithmic descriptions given in [Schrepp, 2006]).

A naturally arising question relevant for both contributions is the question of generalization to higher scales of measurement: Can one generalize the stochastic dominance checking approach of Contribution 7 to data with a higher scale of measurement and can one introduce descriptive ability and difficulty notions in the context of Contribution 8 that have more than a partially ordinal character? It turns out, that for both contributions, reasonable generalizations can be achieved:

Concerning Contribution 7, first order stochastic dominance can be alternatively characterized in a decision theoretic manner as $X \leq_{SD} Y$ iff $\mathbb{E}(u \circ X) \leq \mathbb{E}(u \circ Y)$ for every bounded, non-decreasing Borel-measurable utility function $u : (V, \leq) \longrightarrow \mathbb{R}$.[51] If the underlying space $(V, \leq)$ is equipped with a richer structure, then one can get a reasonable, more decisive notion of

[51]For this characterization one has to assume, that $(V, \leq)$, equipped with an appropriate topology builds a partially ordered Polish space, cf., e.g. [Mosler and Scarsini, 1993].

stochastic dominance by demanding $\mathbb{E}(u \circ X) \leq \mathbb{E}(u \circ Y)$ only for that utility functions that also respect the underlying additional structure of V. One example would be the case where V is a p-dimensional structure, where every dimension has a cardinal scale of measurement but different dimensions cannot be compared. Then one naturally has the partially ordered structure $(V, \leq)$ where $x \leq y \iff \forall i \in \{1, \ldots, p\} : x_i \leq y_i$. But additionally, one has a natural relation $R \subseteq \leq \times \leq$ on pairs of comparable elements of V that can express for a pair $x < y$ “how much bigger“ y is, compared to x, in relation to other pairs $x' < y'$. Concretely, one can naturally define $(x', y')R(x, y)$ iff the extent to which y is bigger than x is clearly higher than the extent to which y' is bigger than x', which is the case if the difference between x and y in every dimension is lower than the difference between x' and y'. Thus

$$R = \{((x', y'), (x, y)) \mid \forall i \in \{1, \ldots, p\} : y'_i - x'_i \leq y_i - x_i\}. \tag{3}$$

and the notion of stochastic dominance can be modified to $X \leq_{SD} Y$ iff $\mathbb{E}(u \circ X) \leq \mathbb{E}(u \circ Y)$ for all non-decreasing, bounded, Borel-measurable $u : V \longrightarrow \mathbb{R}$ that are additionally R-homomorphisms, meaning that $\forall((x', y'), (x, y)) \in R : u(y') - u(x') \leq u(y) - u(x)$ holds.[52] To compute the modified version of stochastic dominance one can directly work with the utility functions and implement the homomorphism conditions via linear constraints. A more detailed description of the implementation via linear programming is given in a more general context of decision theory under severe uncertainty in Jansen, Schollmeyer, and Augustin [2017b].
For Contribution 8 one can also introduce some partial cardinal notion of item difficulty and person ability. For example, for three items A, B and C one can say that item A “plus” item B are more difficult than item C if for every person p who solved items A and B there bijectively exists another person q that is not more able than person p and solved item C. This could be also generalized to more than two items that are compared in “sum” to the “sum” of other items. Of course, the computation of all such relations would be very hard, especially the linear program of Contribution 7 seems to be not straightforwardly adoptable to this generalized situation. Instead, a formulation as a mass transportation or a network flow problem seems to be more promising, here.

[52] Note that this modification is conceptually already far different from classical first order stochastic dominance. For example, in the case $p = 1$ it essentially translates to $X \leq_{SD} Y \iff \mathbb{E}(X) \leq \mathbb{E}(Y)$, which is a complete order on all random variables (with existing expectations). In particular, if for example Y takes very low values with some probability, then w.r.t. $\leq_{SD}$, it can be still higher than an arbitrary X, provided that with some positive probability Y takes also high enough values that could compensate for the low values. This property is very different to classical first order stochastic dominance where one cannot offset low values with high values. Another modification in the spirit of second order stochastic dominance would be to restrict the relation R of (3) to quadrupels $((x', y'), (x, y))$ with $x \leq x'$ and $y \leq y'$ to model a notion of decreasing returns to scale. This seems to be more or less equivalent to the scaled convex order, cf., [Koshevoy and Mosler, 2007].

Brief overall conclusion

As the more technical/detailed perspectives were already discussed in the corresponding paragraphs for every contributing part, I would like to finish this overview with a more general (and of course a little overstated) conclusion.

One of the background motivations behind this thesis was some wish in trying to contribute a little bit to the development of methods for data analysis that are intended to provide some flexibility for some sort of "careful data analysis" that tries to avoid a "rigorous attitude" of statistical modeling "*...which is write down a convenient model based on mathematical convenience and now act as if thats exactly the truth and everything follows from that.*" [Wassermann, 2016]. As we have seen in the discussion of ill-posedness in the context of partial identification, also within the methodology of partial identification, which is often attributed as some cautious type of analysis, it seems to be not enough to simply model situations of partially identified models by using for example random set theory and then to stick back to such a "rigorous type modeling" by again treating the new statistical model that now only lives in a more complex space, as if it is exactly true.

Generally, statistical modeling seems to always come along with some form of schematization, where one already knows that the model is not completely adequate. This subtly shapes statistical modeling and in fact foils any type of "rigorous modeling" in such situations. At the same time, attempts of "weak" or "cautious" modeling are well-advised to not squander trust by using attributes like "reliable" only as a label.

To put the above reservations more into a perspective or a wish-list, I would finally like to agree with Freedmans point of view in admitting that

> "*...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]

Further (cited) publications by the author

C. Jansen, T. Augustin, and G. Schollmeyer. Decision theory meets linear optimization beyond computation. In A. Antonucci, L. Cholvy, and O. Papini, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 14th European Conference, ECSQARU 2017, Lugano, Switzerland, July 10–14, 2017, Proceedings*, pages 329–339. Springer, 2017a.

C. Jansen, G. Schollmeyer, and T. Augustin. Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences. In A. Antonucci, G. Corani, I. Couso, and S. Destercke, editors, *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, volume 62 of *Proceedings of Machine Learning Research*, pages 181–192. PMLR, 2017b.

J. Plass, M. Cattaneo, T. Augustin, G. Schollmeyer, and C. Heumann. Reliable categorical regression analysis for non-randomly coarsened observations: An analysis with German labour market data. Technical Report 206, Department of Statistics, LMU Munich, 2017a. URL `http://www.statistik.uni-muenchen.de/forschung/technical_reports/index.html`.

J. Plass, M. Cattaneo, G. Schollmeyer, and T. Augustin. On the testability of coarsening assumptions: A hypothesis test for subgroup independence. *International Journal of Approximate Reasoning*, 2017b. doi: http://dx.optdoi.org/10.1016/j.ijar.2017.07.014. (to appear).

G. Schollmeyer and T. Augustin. On sharp identification regions for regression under interval data. In F. Cozman, T. Denœux, S. Destercke, and T. Seidenfeld, editors, *ISIPTA'13: Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications*, pages 285–294. SIPTA, 2013.

Further cited work

T. Abe and Y. Funaki. The non-emptiness of the core of a partition function form game. *International Journal of Game Theory*, 46(3):715–736, 2017.

C. H. Achen. *Interpreting and Using Regression*. Sage, 1982.

R. Agrawal, T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, SIGMOD '93, pages 207–216. ACM, 1993.

A. Albano. The implication logic of (n,k)-extremal lattices. In K. Bertet, D. Borchmann, P. Cellier, and S. Ferré, editors, *Formal Concept Analysis: 14th International Conference, ICFCA 2017, Rennes, France, June 13-16, 2017, Proceedings*, pages 39–55. Springer, 2017a.

A. Albano. *Polynomial growth of concept lattices, canonical bases and generators: extremal set theory in formal concept analysis*. PhD thesis, Technische Universität Dresden, 2017b. URL `http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-226980`. accessed 19.08.2017.

A. Albano and B. Chornomaz. Why concept lattices are large: extremal theory for generators, concepts, and VC-dimension. *International Journal of General Systems*, pages 1–18, 2017.

S. Alkire, J. Foster, S. Seth, M. E. Santos, J. M. Roche, and P. Ballon. *Multidimensional Poverty Measurement and Analysis*. Oxford University Press, 2015.

C. M. Allen, S. D. Griffith, S. Shiffman, and D. F. Heitjan. Proximity and gravity: modeling heaped self-reports. *Statistics in Medicine*, 36(20):3200–3215, 2017.

M. Anthony, N. Biggs, and J. Shawe-Taylor. The learnability of formal concepts. In *Proceedings of the Third Annual Workshop on Computational Learning Theory*, COLT '90, pages 246–257. Morgan Kaufmann Publishers Inc., 1990a.

M. Anthony, N. Biggs, and J. Shawe-Taylor. *Learnability and Formal Concept Analysis*. University of London. Royal Holloway and Bedford New College. Department of Computer Science, 1990b.

A. Antonucci, C. P. de Campos, D. Huber, and M. Zaffalon. Approximate credal network updating by linear programming with applications to decision making. *International Journal of Approximate Reasoning*, 58:25 – 38, 2015.

C. Arndt, S. Dann, R. Kleimann, H. Strotmann, and J. Volkert. *Das Konzept der Verwirklichungschancen (A. Sen): empirische Operationalisierung im Rahmen der Armuts- und Reichtumsmessung, Machbarkeitsstudie, Endbericht an das Bundesministerium für Arbeit und Soziales*. 2006. URL `http://nbn-resolving.de/urn:nbn:de:0168-ssoar-265292`. accessed 20.08.2017.

C. Arndt, R. Distante, M. A. Hussain, L. P. Østerdal, P. L. Huong, and M. Ibraimo. Ordinal welfare comparisons with multiple discrete indicators: A first order dominance approach and application to child poverty. *World Development*, 40(11):2290–2301, 2012.

C. Arndt, N. Siersbæk, and L. P. Østerdal. Multidimensional first-order dominance comparisons of population wellbeing. WIDER Working Paper 2015/122, 2015. URL `http://hdl.handle.net/10419/129471`. accessed 20.08.2017.

W. Arulampalam, V. Corradi, and D. Gutknecht. Modeling heaped duration data: An application to neonatal mortality. *Journal of Econometrics*, 200(2):363 – 377, 2017.

T. Augustin. *Optimale Tests bei Intervallwahrscheinlichkeit*. Vandenhoeck und Ruprecht, 1998. In German, with an English summary on pages 247-249.

T. Augustin, F. P. A. Coolen, G. de Cooman, and M. C. M. Troffaes, editors. *Introduction to Imprecise Probabilities*. Wiley, 2014.

D. Avis. A revised implementation of the reverse search vertex enumeration algorithm. In G. Kalai and G. M. Ziegler, editors, *Polytopes — Combinatorics and Computation*, pages 177–198. Birkhäuser, 2000.

D. Avis and K. Fukuda. Reverse search for enumeration. *Discrete Applied Mathematics*, 65 (1-3):21–46, 1996.

M. M. Babbar. Distributions of solutions of a set of linear equations (with an application to linear programming). *Journal of the American Statistical Association*, 50(271):854–869, 1955.

F. B. Baker. *The Basics of Item Response Theory*. ERIC, 2001.

V. S. Bawa. Research bibliography stochastic dominance: A research bibliography. *Management Science*, 28(6):698–712, 1982.

R. Bellman. *Dynamic Programming*. Princeton University Press, 1957.

A. Beresteanu and F. Molinari. Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814, 2008.

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

R. Berk, L. Brown, A. Buja, K. Zhang, and L. Zhao. Valid post-selection inference. *The Annals of Statistics*, 41(2):802–837, 2013.

J. Bernoulli. *Wahrscheinlichkeitsrechnung (Ars conjectandi)*. 1713. translated and published by R. Haussner.

C. Biernacki and J. Jacques. A generative model for rank data based on insertion sort algorithm. *Computational Statistics & Data Analysis*, 58:162 – 176, 2013.

G. Birkhoff. *Lattice Theory*. American Mathematical Society, 1940.

H.-H. Bock and E. Diday. *Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data*. Springer, 2012.

O. N. Bondareva. Some applications of linear programming methods to the theory of cooperative games. *Problemy kibernetiki*, 10:119–139, 1963. [in Russian]. English translatio in Selected Russian Papers in Game Theory 1959-1965. Princeton University Press, 1968.

L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

L. Breiman et al. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231, 2001.

A. Bronevich and T. Augustin. Approximation of coherent lower probabilities by 2-monotone measures. In T. Augustin, F. P. A. Coolen, S. Moral, and M. C. M. Troffaes, editors, *ISIPTA'09: Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 61–70. SIPTA, 2009.

B. Brown and T. P. Hettmansperger. An affine invariant bivariate version of the sign test. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 117–125, 1989.

E. Carrizosa, J. Gordillo, and F. Plastria. Classification problems with imprecise data through separating hyperplanes. Technical report, Optimization Online, 2007. URL `http://www.optimization-online.org/DB_HTML/2007/09/1781.html`. accessed 20.08.2017.

M. Cattaneo and A. Wiencierz. Likelihood-based imprecise regression. *International Journal of Approximate Reasoning*, 53(8):1137–1154, 2012.

M. Černỳ and M. Rada. On the possibilistic approach to linear regression with rounded or interval-censored data. *Measurement Science Review*, 11(2):34–40, 2011.

P. Chaudhuri. On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872, 1996.

V. Chernozhukov, H. Hong, and E. Tamer. Inference on parameter sets in econometric models, 2004. Working Paper, MIT. URL `www.stanford.edu/~doubleh/papers/victor.pdf`, accessed 20.08.2017.

V. Chernozhukov, H. Hong, and E. Tamer. Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284, 2007.

V. Chernozhukov, S. Lee, and A. M. Rosen. Intersection bounds: estimation and inference. *Econometrica*, 81(2):667–737, 2013.

V. Chernozhukov, A. Galichon, M. Hallin, and M. Henry. Monge–Kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 45(1):223–256, 2017.

Y. Chiba and T. J. VanderWeele. A simple method for principal strata effects when the outcome has been truncated due to death. *American Journal of Epidemiology*, 173(7):745–751, 2011.

B. Chornomaz. Counting extremal lattices. working paper or preprint, 2015. URL `https://hal.archives-ouvertes.fr/hal-01175633`.

A. Christmann. Classification based on the support vector machine and on regression depth. In Y. Dodge, editor, *Statistical Data Analysis Based on the L1-Norm and Related Methods*, pages 341–352, Basel, 2002. Birkhäuser.

F. Ciliberto and E. Tamer. Market structure and multiple equilibria in airline markets. *Econometrica*, 77(6):1791–1828, 2009.

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.

F. P. A. Coolen and K. J. Yan. Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference*, 126(1):25–54, 2004.

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.

I. Couso and D. Dubois. Maximum likelihood under incomplete information: Toward a comparison of criteria. In M. B. Ferraro, P. Giordani, B. Vantaggi, M. Gagolewski, M. Ángeles Gil, P. Grzegorzewski, and O. Hryniewicz, editors, *Soft Methods for Data Science*, pages 141–148. Springer, 2017.

I. Couso and L. Sánchez. Mark-recapture techniques in statistical tests for imprecise data. *International Journal of Approximate Reasoning*, 52(2):240–260, 2011.

J. Cuesta-Albertos and A. Nieto-Reyes. The Tukey and the random Tukey depths characterize discrete distributions. *Journal of Multivariate Analysis*, 99(10):2304 – 2311, 2008.

O. Davidov and S. Peddada. Testing for the multivariate stochastic order among ordered experimental groups with application to dose–response studies. *Biometrics*, 69(4):982–990, 2013.

A. C. Davison and D. V. Hinkley. *Bootstrap Methods and their Application.* Cambridge University press, 1997.

G. De Cooman and D. Aeyels. On the coherence of supremum preserving upper previsions. In *Proceedings of PMU 96 (Sixth International Conference on Information Processing and Management of Uncertainty in Knowledge Based Systems, Granada, Spain, July 1-5*, pages 1405–1410, 1996.

B. De Schuymer, H. De Meyer, and B. De Baets. A fuzzy approach to stochastic dominance of random variables. In T. Bilgiç, B. D. Baets, and O. Kaynak, editors, *Tenth International Fuzzy Systems Association World Congress*, pages 253–260. Springer, 2003a.

B. De Schuymer, H. De Meyer, B. De Baets, and S. Jenei. On the cycle-transitivity of the dice model. *Theory and Decision*, 54(3):261–285, 2003b.

D. Denneberg. *Non-additive Measure and Integral.* Kluwer Academic Publishers, 1994.

T. Denœux, M.-H. Masson, and P.-A. Hébert. Nonparametric rank-based statistics and significance tests for fuzzy data. *Fuzzy Sets and Systems*, 153(1):1–28, 2005.

J. Derks and J. Kuipers. On the number of extreme points of the core of a transferable utility game. In P. Borm and H. Peters, editors, *Chapters in Game Theory: In Honor of Stef Tijs*, pages 83–97. Springer, 2002.

J. Derks, H. Haller, and H. Peters. The selectope for cooperative games. *International Journal of Game Theory*, 29:23–38, 2000.

S. Destercke. Independence and 2-monotonicity: Nice to have, hard to keep. In W. Liu, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 11th European Conference, ECSQARU 2011, Belfast, UK, June 29–July 1, 2011. Proceedings*, pages 263–274. Springer, 2011.

L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer, 2013.

T.-N. Do and F. Poulet. Kernel-based algorithms and visualization for interval data mining. In D. A. Zighed, S. Tsumoto, Z. W. Ras, and H. Hacid, editors, *Mining Complex Data*, pages 75–91. Springer, 2009.

J.-P. Doignon and J.-C. Falmagne. *Knowledge Spaces*. Springer, 2012.

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

J. Duckitt, B. Bizumic, S. W. Krauss, and E. Heled. A tripartite approach to right-wing authoritarianism: The authoritarianism-conservatism-traditionalism model. *Political Psychology*, 31(5):685–715, 2010.

B. Dushnik and E. W. Miller. Partially ordered sets. *American Journal of Mathematics*, 63 (3):600–610, 1941.

N. Eid, J. Maltby, and O. Talavera. Income rounding and loan performance in the peer-to-peer market. 2016. Available at SSRN: `https://ssrn.com/abstract=2848372` accessed 17.08.2017.

J. Fan. Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, pages 196–216, 1993.

G. Fasano and A. Franceschini. A multidimensional version of the Kolmogorov-Smirnov test. *Monthly Notices of the Royal Astronomical Society*, 225(1):155–170, 1987.

W. Feller. On the Kolmogorov-Smirnov limit theorems for empirical distributions. *The Annals of Mathematical Statistics*, pages 177–189, 1948.

S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. *Sandia National Laboratories*, 2002. URL `http://citg.home.tudelft.nl/fileadmin/Faculteit/CiTG/Over_de_faculteit/Afdelingen/Afdeling_Waterbouwkunde/sectie_waterbouwkunde/people/personal/gelder/publications/citations/doc/citatie147.pdf`.

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

P. C. Fishburn. Intransitive indifference with unequal indifference intervals. *Journal of Mathematical Psychology*, 7(1):144–149, 1970.

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

D. A. Freedman. A rejoinder on models, metaphors, and fables. *Journal of Educational Statistics*, 12(2):206–223, 1987.

D. A. Freedman. Some issues in the foundation of statistics. In B. C. van Fraassen, editor, *Topics in the Foundation of Statistics*, pages 19–39. Springer, 1997.

R. Frisch. *Statistical Confluence Analysis by Means of Complete Regression Systems*, volume 5. Universitetets Økonomiske Instituut, 1934.

K. Fukuda and A. Prodon. Double description method revisited. In M. Deza, R. Euler, and I. Manoussakis, editors, *Combinatorics and Computer Science: 8th Franco-Japanese and 4th Franco-Chinese Conference Brest, France, July 3–5, 1995 Selected Papers*, pages 91–111. Springer, 1996.

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

M. Gentry and T. Li. Identification in auctions with selective entry. *Econometrica*, 82(1): 315–344, 2014.

GESIS - Leibniz - Institut fur Sozialwissenschaften. Allbus compact (2015): Allgemeine Bevölkerungsumfrage der Sozialwissenschaften, 2015. URL `https://www.gesis.org/allbus/inhalte-suche/studienprofile-1980-bis-2016/2014`. GESIS Datenarchiv, Köln. ZA5241 Datenfile Version 1.1.0.

P. B. Gilbert, R. J. Bosch, and M. G. Hudgens. Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics*, 59(3):531–541, 2003.

D. B. Gillies. Solutions to general non-zero-sum games. In *Contributions to the Theory of Games*, volume 4, pages 47–85. Princeton University Press, 1959.

P. Groeneboom, G. Jongbloed, and J. A. Wellner. Estimation of a convex function: characterizations and asymptotic theory. *Annals of Statistics*, pages 1653–1698, 2001.

Z. Gu, E. Rothberg, and R. Bixby. Gurobi optimizer reference manual. 2012. URL `http://www.gurobi.com`. accessed 19.08.2017.

A. Hassairi and O. Regaieg. On the Tukey depth of an atomic measure. *Statistical Methodology*, 4(2):244 – 249, 2007.

M. Heene. Additive conjoint measurement and the resistance toward falsifiability in psychology. *Frontiers in Psychology*, 4:246, 2013. ISSN 1664-1078. doi: 10.3389/fpsyg.2013.00246.

D. F. Heitjan and D. B. Rubin. Ignorability and coarse data. *The Annals of Statistics*, pages 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)*, pages 235–249, 1994.

J. L. Hodges. A bivariate sign test. *The Annals of Mathematical Statistics*, 26(3):523–527, 1955.

P. W. Holland, D. T. Thayer, H. Wainer, and H. Braun. Differential item performance and the Mantel-Haenszel procedure. In *Test Validity*, pages 129–145, 1988.

G. Hooker, M. Finkelman, and A. Schwartzman. Paradoxical results in multidimensional item response theory. *Psychometrika*, 74(3):419–442, 2009.

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

M. Jaeger. On testing the missing at random assumption. *Machine Learning*, 4212:671–678, 2006.

P. Jordan. *Paradoxien in quantitativen Modellen der Individualdiagnostik*. PhD thesis, Universität Hamburg, 2013. URL `http://ediss.sub.uni-hamburg.de/volltexte/2013/6198`. accessed 20.08.2017.

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

A. Justel, D. Peña, and R. Zamar. A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics & Probability Letters*, 35(3):251–259, 1997.

T. Kamae, U. Krengel, and G. L. O'Brien. Stochastic inequalities on partially ordered spaces. *The Annals of Probability*, 5(6):899–912, 1977.

J. G. Kemeny. Mathematics without numbers. *Quantity and Quality*, 27:577–591, 1959.

A. Kolmogorov. Confidence limits for an unknown distribution function. *The Annals of Mathematical Statistics*, 12(4):461–463, 1941.

G. Koshevoy and K. Mosler. Zonoid trimming for multivariate distributions. *The Annals of Statistics*, 25(5):1998–2017, 1997.

G. A. Koshevoy. The Tukey depth characterizes the atomic measure. *Journal of Multivariate Analysis*, 83(2):360 – 364, 2002.

G. A. Koshevoy and K. Mosler. Multivariate Lorenz dominance based on zonoids. *Advances in Statistical Analysis*, 91(1):57–76, 2007.

T. Kroupa. Geometry of possibility measures on finite sets. *International Journal of Approximate Reasoning*, 48(1):237–245, 2008.

H. Küchenhoff, T. Augustin, and A. Kunz. Partially identified prevalence estimation under misclassification using the kappa coefficient. *International Journal of Approximate Reasoning*, 53(8):1168 – 1182, 2012.

T. Kuosmanen. Efficient diversification according to stochastic dominance criteria. *Management Science*, 50(10):1390–1406, 2004.

S. Kuznetsov. On the intractability of computing the Duquenne-Guigues base. *Journal of Universal Computer Science*, 10(8):927–933, 2004.

L. Kwuida and S. E. Schmidt. Valuations and closure operators on finite lattices. *Discrete Applied Mathematics*, 159(10):990–1001, 2011.

A. Kyngdon. The Rasch model from the perspective of the representational theory of measurement. *Theory & Psychology*, 18(1):89–109, 2008.

L. Lakhal and G. Stumme. Efficient mining of association rules based on formal concept analysis. *Formal concept analysis*, 3626:180–195, 2005.

N. Laskurain. *Dinámica espacio-temporal de un bosque secundario en el Parque Natural de Urkiola*. PhD thesis, Universidad del Pais Vasco/Euskal Herriko Unibertsitatea, 2008.

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

E. L. Lehmann. Ordered families of distributions. *The Annals of Mathematical Statistics*, 26 (3):399–419, 1955.

J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 2017. accepted for publication.

M. Leshno and H. Levy. Stochastic dominance and medical decision making. *Health Care Management Science*, 7(3):207–215, 2004.

D. Levhari, J. Paroush, and B. Peleg. Efficiency analysis for multivariate distributions. *The Review of Economic Studies*, 42(1):87–91, 1975.

H. Levy. Stochastic dominance and expected utility: survey and analysis. *Management Science*, 38(4):555–593, 1992.

H. Levy. *Stochastic Dominance: Investment Decision Making under Uncertainty*. Springer, 2015.

H. Levy and M. Levy. Experimental test of the prospect theory value function: A stochastic dominance approach. *Organizational Behavior and Human Decision Processes*, 89(2):1058 – 1081, 2002.

J. C. Lindsey and L. M. Ryan. Methods for interval-censored data. *Statistics in Medicine*, 17 (2):219–238, 1998.

R. Y. Liu, J. M. Parelius, K. Singh, et al. 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. Y. Liu et al. On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1):405–414, 1990.

T. Makhalova and S. O. Kuznetsov. On overfitting of classifiers making a lattice. In K. Bertet, D. Borchmann, P. Cellier, and S. Ferré, editors, *Formal Concept Analysis: 14th International Conference, ICFCA 2017*, pages 184–197. Springer, 2017.

C. F. Manski. Anatomy of the selection problem. *The Journal of Human Resources*, 24(3): 343–360, 1989.

C. F. Manski. Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323, 1990.

C. F. Manski. Partial identification with missing data: concepts and findings. *International Journal of Approximate Reasoning*, 39(2-3):151–165, 2005.

C. F. Manski and E. Tamer. Inference on regressions with interval data on a regressor or outcome. *Econometrica*, 70(2):519–546, 2002.

J. Marschak and W. H. Andrews. Random simultaneous equations and the theory of production. *Econometrica*, 12(3/4):143–205, 1944.

I. K. McDonough and D. L. Millimet. Missing data, imputation, and endogeneity. *Journal of Econometrics*, 199(2):141 – 155, 2017.

J. Michell. Is psychometrics pathological science? *Measurement*, 6(1-2):7–24, 2008.

L. H. Miller. Table of percentage points of Kolmogorov statistics. *Journal of the American Statistical Association*, 51(273):111–121, 1956.

E. Miranda, I. Couso, and P. Gil. Extreme points of credal sets generated by 2-alternating capacities. *International Journal of Approximate Reasoning*, 33(1):95–115, 2003.

F. Molinari. Partial identification of probability distributions with misclassified data. *Journal of Econometrics*, 144(1):81–117, 2008.

I. Montes and S. Destercke. On extreme points of p-boxes and belief functions. *Annals of Mathematics and Artificial Intelligence*, 2017. doi: 10.1007/s10472-017-9562-x.

K. Mosler. Depth statistics. In C. Becker, R. Fried, and S. Kuhnt, editors, *Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather*, pages 17–34. Springer, 2013.

K. Mosler and M. Scarsini. Some Theory of Stochastic Dominance. In K. Mosler and M. Scarsini, editors, *Stochastic orders and decision under risk*, pages 203–212. Institute of Mathematical Statistics, 1991.

K. Mosler and M. Scarsini. *Stochastic Orders and Applications: A Classified Bibliography.* Lecture Notes in Economics and Mathematical Systems. Springer, 1993.

T. S. Motzkin, H. Raiffa, G. L. Thompson, and R. M. Thrall. The double description method. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games II*, pages 51–73. Princeton University Press, 1953.

I. Mullis, M. Martin, G. Ruddock, C. O'Sullivan, A. Arora, and E. Erberber. *TIMSS 2007 assessment frameworks.* Boston College, 2005.

E. A. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1): 141–142, 1964.

H. Oja. Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, 1 (6):327–332, 1983.

L. P. Østerdal. The mass transfer approach to multivariate discrete first order stochastic dominance: Direct proof and implications. *Journal of Mathematical Economics*, 46(6):1222–1228, 2010.

S. J. Osterlind and H. T. Everson. *Differential Item Functioning*. Sage Publications, 2009.

J. Peacock. Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices of the Royal Astronomical Society*, 202(3):615–627, 1983.

S. Peng. *Three essays in econometrics*. PhD thesis, Cornell University Library, 2017. URL `https://ecommons.cornell.edu/handle/1813/51674`. accessed 20.08.2017.

R. Perline, B. D. Wright, and H. Wainer. The Rasch model as additive conjoint measurement. *Applied Psychological Measurement*, 3(2):237–255, 1979.

G. Piatetsky-Shapiro. Discovery, analysis and presentation of strong rules. *Knowledge Discovery in Databases*, pages 229–248, 1991.

J. Plass, A. Omar, and T. Augustin. Towards a cautious modelling of missing data in small area estimation. In A. Antonucci, G. Corani, I. Couso, and S. Destercke, editors, *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, volume 62 of *Proceedings of Machine Learning Research*, pages 253–264. PMLR, 2017.

M. Ponomareva and E. Tamer. Misspecification in moment inequality models: back to moment equalities? *The Econometrics Journal*, 14(2):186–203, 2011.

A. Prèkopa. On the probability distribution of the optimum of a random linear program. *SIAM Journal on Control*, 4(1):211–222, 1966.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2014. URL `http://www.R-project.org`.

O. Reiersl. Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica*, 9(1):1–24, 1941.

G. Rohwer and U. Pötter. *Grundzüge der sozialwissenschaftlichen Statistik*. Juventa, 2001.

J. P. Romano and A. M. Shaikh. Inference for the identified set in partially identified econometric models. *Econometrica*, 78(1):169–211, 2010.

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

P. J. Rousseeuw and I. Ruts. Algorithm AS 307: Bivariate location depth. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 45(4):516–526, 1996.

A. Rupp, J. Templin, and R. Henson. *Diagnostic Measurement: Theory, Methods, and Applications*. Guilford, 2010.

A. Rusch and R. Wille. Knowledge spaces and formal concept analysis. In H.-H. Bock and W. Polasek, editors, *Data Analysis and Information Systems: Statistical and Conceptual Approaches Proceedings of the 19th Annual Conference of the Gesellschaft für Klassifikation e.V. University of Basel*, pages 427–436. Springer, 1996.

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

U. Ryssel, F. Distel, and D. Borchmann. Fast algorithms for implication bases and attribute exploration using proper premises. *Annals of Mathematics and Artificial Intelligence*, 70 (1-2):25–53, 2014.

H. Scheiblechner. A unified nonparametric IRT model for d-dimensional psychological test data (d-ISOP). *Psychometrika*, 72(1):43, 2007.

H. Schneeweiß. *Entscheidungskriterien bei Risiko*. Springer, 1967.

H. Schneeweiß, J. Komlos, and A. S. Ahmad. Symmetric and asymmetric rounding: a review and some new results. *Advances in Statistical Analysis*, 94(3):247–271, 2010.

M. Schomaker, S. Hogger, L. F. Johnson, C. J. Hoffmann, T. Bärnighausen, and C. Heumann. Simultaneous treatment of missing data and measurement error in HIV research using multiple overimputation. *Epidemiology*, 26(5):628, 2015.

M. Schrepp. Extracting knowledge structures from observed data. *British Journal of Mathematical and Statistical Psychology*, 52(2):213–224, 1999.

M. Schrepp. Explorative analysis of empirical data by boolean analysis of questionnaires. *Zeitschrift für Psychologie*, 210(2):99–109, 2002.

M. Schrepp. Ita 2.0: a program for classical and inductive item tree analysis. *Journal of Statistical Software*, 16(10):1–14, 2006.

D. O. Segall. Principles of multidimensional adaptive testing. In W. J. van der Linden and C. A. Glas, editors, *Elements of Adaptive Testing*, pages 57–75. Springer, 2010.

A. Sen. *Development as Freedom*. Oxford University Press, 1999.

J. K. Sengupta, G. Tintner, and B. Morrison. Stochastic linear programming with applications to economic models. *Economica*, 30(119):262–276, 1963.

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.

G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

L. S. Shapley. Cores of convex games. *International Journal of Game Theory*, 1(1):11–26, 1971.

N. V. Smirnov. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Moscow University Mathematics Bulletin*, 2(2), 1939.

M. Spadoni and L. Stefanini. Computing the variance of interval and fuzzy data. *Fuzzy Sets and Systems*, 165(1):24–36, 2011.

SPIEGEL Online. Studentenpisa - Alle fragen, alle Antworten, 2009. URL `http://www.spiegel.de/unispiegel/studium/0,1518,620101,00.html`. In German. accessed 18.08.2017.

L. Steyer. Adapting support vector machines to generalised interval data: Implementation of a suitable kernel for convex sets and its comparison to a minimax approach. Master Thesis, Department of Statistics, LMU Munich, 2017.

J. Stoye. Partial identification and robust treatment choice: an application to young offenders. *Journal of Statistical Theory and Practice*, 3(1):239–254, 2009.

C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1):307, 2008.

C. Strobl, J. Kopf, and A. Zeileis. Rasch trees: A new method for detecting differential item functioning in the rasch model. *Psychometrika*, 80(2):289–316, 2015.

M. Studený and V. Kratochvíl. Linear core-based criterion for testing extreme exact games. In A. Antonucci, G. Corani, I. Couso, and S. Destercke, editors, *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, volume 62 of *Proceedings of Machine Learning Research*, pages 313–324. PMLR, 2017.

M. Studený and T. Kroupa. Core-based criterion for extreme supermodular functions. *Discrete Applied Mathematics*, 206:122 – 151, 2016.

S. E. Syrjala. A statistical test for a difference between the spatial distributions of two populations. *Ecology*, 77(1):75–80, 1996.

L. L. Thurstone. A law of comparative judgment. *Psychological Review*, 34(4):273, 1927.

N. Torelli and U. Trivellato. Modelling inaccuracies in job-search duration data. *Journal of Econometrics*, 59(1):187–211, 1993.

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

M. C. M. Troffaes and R. Hable. Computation. In T. Augustin, F. P. A. Coolen, G. de Cooman, and M. C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 329–337. Wiley, 2014.

W. T. Trotter. *Combinatorics and Partially Ordered Sets: Dimension Theory*. JHU Press, 2001.

J. W. Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians*, volume 2, pages 523–531, 1975.

G. Tutz and M. Berger. Item-focussed trees for the identification of items in differential item functioning. *Psychometrika*, 81(3):727–750, 2016. ISSN 1860-0980.

G. Tutz and G. Schauberger. A penalty approach to differential item functioning in Rasch models. *Psychometrika*, 80(1):21–43, 2015.

UNESCO Institute for Statistics (UIS). *International standard classification of education: ISCED 2011.* UIS, Montreal, Quebec, 2012.

L. V. Utkin and T. Augustin. Powerful algorithms for decision making under partial prior information and general ambiguity attitudes. In F. Cozman, R. Nau, and T. Seidenfeld, editors, *ISIPTA '05: Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications*, pages 349–358. SIPTA, 2005.

L. V. Utkin and T. Augustin. Decision making under incomplete data using the imprecise Dirichlet model. *International Journal of Approximate Reasoning*, 44(3):322–338, 2007.

L. V. Utkin and A. I. Chekh. A new robust model of one-class classification by interval-valued training data using the triangular kernel. *Neural Networks*, 69:99–110, 2015.

L. V. Utkin and F. P. A. Coolen. Interval-valued regression and classification models in the framework of machine learning. In F. Coolen, G. de Cooman, T. Fetz, and M. Oberguggenberger, editors, *ISIPTA'11: Proceedings of the Seventh International Symposium on Imprecise Probability: Theories and Applications*, pages 371–380. SIPTA, 2011.

J. F. Van Leeuwe. Item tree analysis. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden*, 1974.

S. Vansteelandt, E. Goetghebeur, M. G. Kenward, and G. Molenberghs. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, 16(3):953–979, 2006.

V. Vapnik. *Estimation of Dependences Based on Empirical Data*. Springer, 2006.

A. Wallner. Extreme points of coherent probabilities in finite spaces. *International Journal of Approximate Reasoning*, 44(3):339 – 357, 2007.

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

L. Wassermann. Inference for high dimensional regression, 2016. URL `https://bfi.uchicago.edu/events/machine-learning-what%E2%80%99s-it-economics`. accessed 17.08.2017. Talk given at the conference "Machine Learning: What's in it for Economics?" (23.09.2016).

G. S. Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372, 1964.

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.

K. Yokote. Weighted values and the core in NTU games. *International Journal of Game Theory*, 46(3):631–654, 2017.

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

Z. Zhang. Profile likelihood and incomplete data. *International Statistical Review*, 78(1): 102–116, 2010.

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

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. .5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

Schollmeyer, Georg

Name, Vorname

München, 10.11.2017

Ort, Datum

Unterschrift Doktorand/in

Formular 3.2